HUMAN ENGINEERING GUIDE TO EQUIPMENT DESIGN

(Revised Edition)

Sponsored by

Joint Army-Navy-Air Force Steering Committee

Edited by

Harold P. Van Cott, Ph. D., and Robert G. Kinkade, Ph. D.

American Institutes for Research
Washington, D.C.

1972

Library of Congress Catalog Card Number 72-600054.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402 - Price $8

AUTHORS, REVIEWERS, AND CONTRIBUTORS

JAMES W. ALTMAN, Ph.D. President, Synectics Corporation. Coauthor, chapter 12. *Designing for Maintainability.*

CHARLES A. BAKER, Division Chief, National Highway Traffic Safety Administration. Coauthor, chapter 3. *Visual Presentation of Information.*

JOHN W. BLACK, Ph. D. Professor of Speech, The Ohio State University. Reviewer, chapter 5. *Speech Communication.*

GEORGE E. BRIGGS, Ph. D. Professor of Psychology, The Ohio State University. Reviewer, chapter 7, *Data Entry Devices and Procedures.*

DONALD E. BROADBENT, Sc. D. Medical Research Council, Cambridge, England. Reviewer, chapter 4. *Auditory and Other Sensory Forms of Information Presentation.*

GLENN L. BRYAN, Ph. D. Director, Psychological Sciences Division, Office of Naval Research. Coauthor, chapter 13. *Training System Design.*

LEE S. CALDWELL, Ph. D. Research Psychologist, U.S. Army Medical Research Lab. Reviewer, chapter 11. *Engineering Anthropology.*

ARMAND N. CHAMBERS, Ph. D. Applied Sciences Associates, Inc. Contributor, chapter 4. *Auditory and Other Sensory Forms of Information Presentation.*

ALPHONSE CHAPANIS, Ph. D. Professor of Psychology, The Johns Hopkins University. Coauthor, chapter 8. *Design of Controls;* and chapter 15. *Human Engineering Tests and Evaluations.*

DONALD W. CONOVER, Ph. D. Vice President of Research, Man Factors, Inc. Contributor, chapter 9. *Design of Individual Workplaces.*

BILLY M. CRAWFORD, Systems Effectiveness Branch, Aerospace Medical Research Laboratory. Coauthor, chapter 12. *Designing for Maintainability.*

BRUCE H. DEATHERAGE, Ph.D. Applied Research Laboratories, The University of Texas at Austin. Author, chapter 4. *Auditory and Other Sensory Forms of Information Presentation.*

E. RALPH DUSEK, Ph. D. Chief, Behavioral Science Division, U.S. Army Institute for Environmental Medicine. Reviewer, chapter 11. *Engineering Anthropology.*

GORDON A. ECKSTRAND, Ph. D. Chief, Training Research Division, Wright-Patterson AFB. Reviewer, chapter 13. *Training System Design.*

GEORGE FROST, Human Engineering Division, Aerospace Medical Research Laboratory, Wright-Patterson AFB. Author, chapter 6. *Man-Machine Dynamics.*

ROBERT GLASER, Ph. D. Director, Learning Research and Development Center, University of Pittsburgh. Reviewer, chapter 13, *Training System Design.*

WALTER F. GRETHER, Ph.D. Senior Scientist, Aerospace Medical Research Lab., Wright-Patterson AFB. Coauthor, chapter 3. *Visual Presentation of Information.*

H. T. E. HERTZBERG, Research Physical Anthropologist, Aerospace Medical Research Lab., Wright-Patterson AFB. Author, chapter 11. *Engineering Anthropology.*

JERRY S. KIDD, Ph. D. Professor, School of Library and Information Sciences, University of Maryland. Coauthor, chapter 1. *System and Human Engineering Analyses.*

ROBERT G. KINKADE, Ph. D. Executive Scientist, American Institutes for Research. Coauthor, chapter 8. *Design of Controls;* Reviewer, chapter 12. *Designing for Maintainability;* and Coauthor, chapter 14. *Training Device Design.*

E. K. H. KROEMER, Ph. D.-Ing. Human Engineering Division, Aerospace Medical Research Lab., Wright-Patterson AFB. Contributor, chapter 11. *Engineering Anthropology.*

KARL D. KRYTER, Ph. D. Director, Sensory Science Research Center, Stanford Research Institute. Author, chapter 5. *Speech Communication.*

MARTIN I. KURKE, Ph. D. Bureau of Narcotics and Dangerous Drugs, U.S. Department of Justice. Reviewer, chapter 15. *Human Engineering Tests and Evaluations.*

NORMAN H. MACKWORTH, Ph. D. Research Associate, Psychiatric Department of the Medical School, Stanford University. Reviewer, chapter 3. *Visual Presentation of Information.*

MAURICE P. RANC, Jr. Manager, Product Support, AAI Corporation. Contributor, chapter 9. *Design of Individual Workplaces;* Reviewer, chapter 14. *Training Device Design.*

JAMES J. REGAN, Ph. D. Chief Psychologist, Naval Training Device Center. Coauthor, chapter 13. *Training System Design.*

ROBERT SEIBEL, Ph. D. Associate Professor, The Pennsylvania State University. Author, chapter 7. *Data Entry Devices and Procedures.*

JOHN W. SENDERS, Department of Psychology, Brandeis University Contributor, chapter 6. *Man-Machine Dynamics.*

H. WALLACE SINAIKO, Ph. D. Institute for Defense Analyses. Reviewer, chapter 1. *System and Human Engineering Analyses;* chapter 2. *Man as a System Component.*

HOWARD W. STOUDT, Ph. D. Assistant Professor of Physical Anthropology, Harvard School of Public Health. Reviewer, chapter 9. *Design of Individual Workplaces.*

ROBERT M. THOMSON, Ph.D. Reviewer, chapter 8. *Design of Controls;* Author, chapter 10. *Design of Multi-Man-Machine Work Areas.*

GILBERT C. TOLHURST, Ph.D. Professor of Speech and Psychology, University of Massachusetts. Reviewer, chapter 4. *Auditory and Other Sensory Forms of Information.*

HAROLD P. VAN COTT, Ph. D. Director, Institute for Research in Organizational Behavior, American Institutes for Research. Coauthor, chapter 1. *System and Human Engineering Analyses;* and chapter 2. *Man as a System Component;* Reviewer, chapter 10. *Design of Multi-Man-Machine Work Areas;* Coauthor, chapter 15. *Human Engineering Tests and Evaluations.*

MELVIN J. WARRICK, Ph. D. Associate Director, Human Engineering Division, Aerospace Medical Research Laboratory, Wright-Patterson AFB. Coauthor, chapter 2. *Man as a System Component.*

GEORGE R. WHEATON, Research Scientist, Institute for Research in Psychobiology, American Institutes for Research. Coauthor, chapter 14. *Training Device Design.*

ROBERT M. WHITE, Head, Anthropology Lab., U.S. Army Natick Labs. Reviewer, chapter 11. *Engineering Anthropology.*

In addition, acknowledgement is made of the efforts of the following persons for their contributions:

Dr. Randall M. Chambers, Mr. Renato Contini, Dr. Duane F. Kasten, Dr. James J. Keenan, Dr. Charles R. Kelley, Dr. Austin Kibler, Dr. John Lyman, Dr. Ross A. McFarland, Miss Diane E. Olson, Dr. Howard Page, Dr. H. M. Parsons, Dr. Richard W. Pew, Dr. Robert G. Pfefferkorn, Dr. Warren H. Teichner and Dr. Wesley E. Woodson.

FOREWORD

The first edition of *Human Engineering Guide to Equipment Design*, edited by Morgan et al. and published by McGraw-Hill (1963), was prepared under the sponsorship of a Joint Army-Navy-Air Force Steering Committee. Its primary purpose as is the purpose of this revised edition, was "to provide a guide in human engineering which the designer can use in the same manner as handbooks in other areas to assist in solving design problems as they arise . . .". Because preparation of the 1963 *Guide* took almost 10 years, and because the fields of human-factors engineering and military technology were expanding rapidly, it soon became evident that plans for an updated and expanded version of the *Guide* should be initiated.

In November of 1963, the Assistant Secretary of the Navy for Research and Development requested the Office of Naval Research to undertake a restudy of the human factors affecting military design and to develop a comprehensive specification or handbook for guidance in applying human engineering techniques to new equipment and system designs. Accordingly, a Joint Army-Navy-Air Force Steering Committee (JSSC) was again constituted with representatives from each service: Drs. Marshall J. Farr and James W. Miller, both then of the Engineering Psychology Branch, Office of Naval Research; Dr. Walter F. Grether, Aerospace Medical Research Laboratory, Wright-Patterson AFB, Ohio; Maj. Robert J. Lacey, then of the Headquarters, Air Force Systems Command, USAF; Jacob L. Barber, Behavioral Sciences Division, Army Research Office; and Dr. William F. Wokoun, Jr., then with the Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland. Dr. Melvin J. Warrick, Air Force Aerospace Medical Research Laboratory, later replaced Maj. Lacey; Dr. Leon T. Katchmar, Army Human Engineering Laboratories, Aberdeen Proving Ground, Maryland, replaced Dr. Wokoun; and Gerald S. Malecki, Engineering Psychology Programs, ONR, replaced Dr. Miller.

In 1965 the JSSC prepared a statement of requirements for a revision of the 1963 *Guide*, and in the same year the American Institutes for Research (AIR), Silver Spring, Maryland, was awarded the contract. The monitoring office for the contract, N00014-70-C-0365 (formerly Nonr-5021 (00)), was the Engineering Psychology Branch, Psychological Sciences Division, Office of Naval Research, Washington, D.C. However, all three military services contributed to the funding of the contract. Essentially, AIR accepted responsibility for managing the preparation of new and/or revised chapters, and for editing and integrating them into a publishable manuscript. The selection of the chapter authors and reviewers was made jointly by the JSSC and AIR.

Many individuals, in addition to those specifically referenced in the chapters, have helped the JSSC and AIR in reviewing the material. Their anonymous help is most gratefully acknowledged.

Although work on this *Guide* was begun in 1965, substantive changes which incorporated current data were still being made in some chapters as late as Spring, 1970.

Users of this *Guide* are invited to comment on its usefulness and to suggest improvements. Comments and suggestions may be directed to any member of the JSSC, or to Dr. Martin A. Tolcott, Director, or Gerald S. Malecki, Assistant Director, Engineering Psychology Programs, Office of Naval Research, Arlington, Virginia 22217.

The Joint-Services Steering Committee
MARSHALL J. FARR, *Chairman*

JOINT ARMY-NAVY-AIR FORCE STEERING COMMITTEE

Jacob L. Barber, Jr.
Office of the Chief of Research and Development
Department of the Army
(Formerly with the Army Research Office)

Marshall J. Farr
Office of Naval Research

Walter F. Grether
Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base

Leon T. Katchmar
U.S. Army Human Engineering Laboratory
Aberdeen Proving Ground

James W. Miller
Department of Commerce
(Formerly with the Office of Naval Research)

Melvin J. Warrick
Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base

William F. Wokoun, Jr.
MUZAK Corporation
(Formerly with U.S. Army Human Engineering Laboratory)

PREFACE

This book was written primarily for those who must make human engineering decisions about equipment design and use. The literature on human engineering, accumulating as rapidly as it has over the past 25 years, is much too voluminous and diverse for any one individual to review, integrate, and translate into practical recommendations. By providing in a single volume authoritative data, principles, and design practices, and by presenting a comprehensive bibliography, it is hoped this *Guide* will be a useful reference that will assist not only in identifying and resolving problems, but also in determining what is important at various design stages. It will not replace the human engineering specialist nor do his work for him. No book can substitute for systematic analysis or eliminate the need for research when data are needed to resolve unique and new problems.

In achieving a level of presentation useful to a group with diverse backgrounds, one risks either under- or oversimplification, particularly in areas that lie between system engineering and human engineering, and between human engineering and behavioral science. The first two chapters are intended to partially bridge these gaps for a user who is uncertain of his knowledge of system analysis or of human performance concepts. They summarize the logic of system analysis and present some of the more fundamental concepts of human behavior. Only in this way, however, does the *Guide* resemble a typical text; otherwise the intent is to provide information directly relevant to human engineering design problems.

The reader familiar with the earlier *Human Engineering Guide to Equipment Design* (McGraw-Hill, 1963) will find points of similarity and difference in this new *Guide*. Old chapters have been revised in the light of new research data and of changes in technology. To these have been added such topics as human performance concepts, data entry devices, training concepts and training-equipment design, and human engineering tests and evaluations. The index has been expanded to facilitate more rapid access, and the reference list has been enlarged to provide more comprehensive support and clarification of the material contained in the *Guide*. The format of the chapters has also been changed. A summary and table of contents appear at the beginning of each one. Even with these extras, the coverage of the *Guide* is a compromise with length. Some important areas commonly associated with human factors are neglected either on the premise that they are covered more completely elsewhere, or that they are somewhat tangential to human engineering as a system planning, design, and evaluation process. The scope of research and development in the area of environmental support and control was too great for inclusion in this *Guide*. Therefore, the reader is referred to the several excellent sources in this area, namely: *Bioastronautics Data Book* (P. Webb, Ed. NASA SP3006, 1964); *Compendium of Human Responses to the Aerospace Environment* (E. M. Roth, Ed. NASA CR 1205, 1969); and *A Bibliographical Sourcebook of Compressed Air, Diving and Submarine Medicine* (E. C. Hoff, Ed., Government Printing Office, 1948–1966).

Rather than make recommendations for specific systems, we have assumed that much of human engineering data can be applied to an array of systems. We recognize, however, that both data and recommendations are subject to interpretation and modification based on the operational and environmental properties of the particular systems. This process of judicious translation necessarily is left to the user.

We wish to acknowledge the use of certain text, tables, charts, graphs, and illustrations which appeared in the earlier *Human Engineering Guide to Equipment Design.* Since that work was performed under contract for the U.S. Government, it is in the public domain and specific citations are not required.

Many people enthusiastically contributed their skill to the preparation of this book. It is hoped this acknowledgment will, in some small way, serve as thanks for their well-spent time. Special acknowledgment goes to the Technical Editor, Mrs. Shirley A. Kennon, who served a coordinative function, provided editorial review of each chapter, and worked with the authors on format and style matters. The Editors are indebted to Dr. Oscar S. Adams, Manager, Advanced Studies Division, Lockheed-Georgia Co., who reviewed many chapters of the *Guide* for their practical value. Special thanks are also due to the Administrative Assistants, Mrs. Nancy S. Brown and Mrs. Janet G. Strasel; to the Graphic Arts Specialist, Mrs. Annie I. Heffner, who was responsible for the many figures and illustrations in the text; and to Mrs. Doris G. Donohue, Mrs. Lily G. Griner, and Mrs. Sara Stein for their accurate typing and proofing, and countless other tasks. Grateful acknowledgment is also expressed to Mr. Irving Rudin and Mrs. Sara Curry of the Naval Research Laboratory for helpful comments and suggestions relative to the editing and final review of the manuscript.

The Editors are also indebted to their professional colleagues at AIR-Washington, who provided valuable technical suggestions, advice, and support at various stages of the book's preparations. In particular we acknowledge the support of Dr. Edwin A. Fleishman, Director of AIR-Washington, and Dr. Angelo Mirabella, Principal Investigator for this project at AIR-Washington; and the assistance of William J. Baker, Erwin W. Bedarf, Ronald P. Carver, Armand N. Chambers, Charles A. Darby, Jr., Alfred J. Farina, Jr., Herbert L. Friedman, George H. Johnson, Arthur L. Korotkin, Marjorie J. Krebs, George C. Theologus, and George R. Wheaton who provided additional technical reviews of selected chapters.

Finally, the Editors would be remiss not to express sincere appreciation to the members of the Joint Army-Navy-Air-Force Steering Committee who expertly guided this work from its inception. Their contribution is described in the Foreword.

HAROLD P. VAN COTT
ROBERT G. KINKADE
Editors

CONTENTS

Chapter 1

System and Human Engineering Analyses

Jerry S. Kidd

University of Maryland
College Park, Md.

Harold P. Van Cott

American Institutes for Research
Washington, D.C.

During the design and development of a system the human engineer not only represents man as a user but provides information about him as a system component. As part of total system analysis, human engineering analysis consists of methods whereby decisions can be made concerning the design of the system and particularly the safety, effectiveness, role, and integration of man in the system. This chapter summarizes these methods as applied to the definition of system requirements and constraints, to the descriptions of functions and event sequences, and to the allocation of component processes.

CONTENTS

This chapter was reviewed by H. Wallace Sinaiko.

1. System and Human Engineering Analyses

1.1 Background and Introduction

Engineering is a process of planning, designing, fabricating, and testing things: a truck, rifle, radio, or space vehicle. These products can be viewed as falling along a continuum of originality, novelty, or uniqueness: a continuum that moves from the familiar or traditional to something without precedent. (See Figure 1-1.)

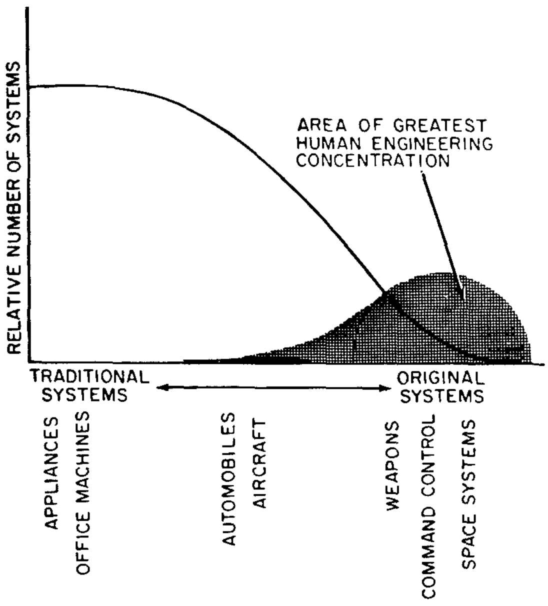


FIGURE 1-1. Hypothetical relationship between the frequency of existing systems and the originality of the systems.

The products generated by the engineering process affect human welfare in many ways. People may benefit directly from the product. They may be the users, operators, or maintainers of the product. The human engineer plays a particularly important role in product and system design because he influences the selection among design alternatives as they relate to people.

In one of his dual roles, the human engineer represents the potential user in regard to ease of operating the equipment, his safety, comfort, etc. In the other role, he evaluates man as a system component and his contribution to the total system.

In traditional systems, engineering is evolutionary, each successive system being derived from a previous one. Knowledge gained from past developments reduces trial and error, and the design-development cycle is shortened.

In new, original systems, the engineering process, particularly in its early stages, is highly conceptual. The engineer cannot rely on rules of thumb. New insights, approaches, techniques, and data are required and analytic models must be developed and exploited before undertaking detailed design. As a consequence the development cycle is lengthened.

How much emphasis the human engineer places on his alternative roles varies with the system. If it is a traditional one, an appliance, a keyboard, or a car, his emphasis tends to be on the welfare of the user—the housewife's safety in using a stove or toaster, the comfort of the typist in using a keyboard, the ease of operation by the driver of the car. In the case of a new, unique system, the emphasis shifts to a concern for the performance demanded of man by the system—his accuracy, speed, and consistency over time. In the first instance the approach is primarily man oriented; in the second it is primarily system oriented.

The goal of the human engineer working with the design of a traditional system is largely one of improvement or optimization. In the design of a new car model, the human engineer may be interested in improving ride quality or road handling ability. Perhaps only a minor modification in seat suspension or steering dynamics will represent a significant step toward improved consumer acceptance. In the case of a new system, such as SEALAB (a submersible facility for human operators to perform oceanographic research), developed by the Navy, the human engineer is less concerned with achieving an optimum man-machine interface than with the design of an interface which is operable and supports the performance required of man by the system. In the latter case, it is often necessary to make compromises to achieve a workable system.

In traditional systems, which have become embedded in a culture, such compromises would not be acceptable because the latitude for innovation is considerably narrowed. Here any dramatic change might substantially disrupt the vested interest in the economics or acceptance of the system. An example is the design of the typewriter keyboard, which has become so prevalent that the introduction of a new, superior keyboard would be unlikely not only because of the investment in the present keyboard, but also because of the number of trained operators capable of using it.

Historically the major concentration of technical and scientific effort by human engineering has been on military systems, systems which fall in the original area of the traditional-original continuum. For this reason many of the analytic methods described later in this chapter and in other parts of the book are oriented toward new, original systems derived from military research and development. Because many military products have been adopted or adapted for commercial use, one can distinguish not only between the nature of engineering on traditional and original systems, but also between the engineering of military and commercial systems. Most definitions of human engineering fail to make such distinctions either in concepts or in practice even though such distinctions are useful in defining goals, criteria, emphasis, approach, and methods. (See Table 1–1.)

The human engineer's approach will vary depending on whether his concern is with a traditional or an original system. While methods vary less in emphasis than in kind, it might be said that in the case of a new system the human engineer is concerned with conceptually synthesizing a total system, with determining the role that human performance will play in such a system, and then with designing the environment and man-machine interface to make that performance possible. Mission, function, and task analysis methods can be of use in the human engineering of new systems because they allow the development of a conceptual framework within which human performance and performance constraints can be studied, and they are the best way of understanding the complete system.

TABLE 1–1. HUMAN ENGINEERING OF TRADITIONAL AND ORIGINAL SYSTEMS

Goals	Traditional systems improvement	Original systems operability
Criteria	Comfort Safety Consumer acceptance	Reliability. Accuracy. Efficiency.
Emphasis	User's welfare	Human component performance.
Approach	Man-oriented	System-oriented.
Method	Analytic-experimental	Synthetic-analytic.

In the case of traditional systems, where the role of man has largely been defined by past systems, there is little emphasis on mission and functions analysis, and task analysis is used less to construct the performance required of man than to analyze his performance to determine how it might be improved through equipment redesign.

1.2 System Analysis

System analysis, also sometimes called system engineering, can be described as having the following general purposes:

1. Scheduling. In the development of a complex system, system analysis is necessary to identify all of the requirements and the logical and sequential order in which they must be accomplished. It is, therefore, an important input to design, production, and test schedules as well as to system management in defining the resources, the skills and money needed to produce a fully operational system.

2. Identifying Limiting Factors. System analysis enables the designer to determine the factors that potentially limit or constrain the performance of a system. These may include environ-

mental limitations, hardware factors, information acquisition, flow factors, personnel problems, and costs.

3. Establishing System Performance Criteria. System analysis provides the criteria which must be met by the several interrelated functional elements of a system. These criteria thus became standards both for design as well as for test and evaluation.

4. Identifying and Explicating Design Options. Through explicit comparison of expected performance with criteria, system analysis enables the designer to decide better the utilization of men and machines. Each stage of a conceptual sequence of operations can be the instance for a check of design decisions. Insofar as functional requirements suggest or compel the survey of options, system analysis is often the instigator of invention.

5. Evaluation of Systems. System analysis is a prerequisite to the test and evaluation of systems. Performance measures of the system and its subsystems are needed to find out how a system can be expected to perform under actual operating conditions or whether one system can be expected to be better than another. These measures can be selected intelligently only from the criteria of performance furnished in a system analysis.

System analysis includes the following steps or phases:

1. The explication of system requirements and constraints.
2. The description of system functions.
3. Detailed descriptions of operational event sequences (including environmental conditions).
4. Detailed descriptions of component processes.

While these steps are interdependent, they are not always followed in a rigid sequence. Some parts of the total system may already have been designed and tested; and time constraints may change the order of approach. Nevertheless, they form a conceptual structure helpful to understanding the relationships of human engineering to the total system engineering process.

1.3 System Requirements and Constraints

System requirements delineate what the system must be able to do: its objectives. System constraints are limits within which accomplishment must take place. Requirements include the mission or purpose of the system as a whole, and the operational characteristics or performance requirements which detail the specific goals, objectives, and standards of the system's mission. Constraints include the environmental, resource, cost, and time limits imposed on system design by the state of the art, by nature, or by the procuring activity (see for example, Air Force Systems Command (AFSC) Manual 80-3, USAF, 1966).

Mission objectives are stated in general terms to avoid their revision in cases where performance requirements are altered in response to technological, budget, or other system constraints. Examples of mission objectives are: (a) injecting a space vehicle into orbit, or (b) interception of manned bombers, or (c) strategic bombing of targets, or (d) airborne surveillance, or (e) the management of resources.

Performance requirements spell out the system's mission requirements. Operational performance requirements are applicable to operations and equipment that are directly concerned with active performance of the mission, while support requirements are those pertaining to events and equipment concerned with preparing and maintaining the operational capability of the system. Examples of operational performance requirements might include: (a) the vehicle must be capable of sustaining a crew of three men, or (b) the vehicle must have an effective operating range of 250 miles, or (c) the system must produce permanent scientific photographic recordings of earth geography, or (d) the vehicle must attain a planned earth orbit at an altitude of 250 nautical miles, or (e) the system must update an operational plan within eight hours (each instance referring, of course, to a different system).

A significant aspect of system requirements analysis for the human engineer is the detailed analysis of the mission of the system over time (Kearns and Ritchie, 1961). Figure 1–2 illustrates how a total space mission is broken into phases before examining the critical require-

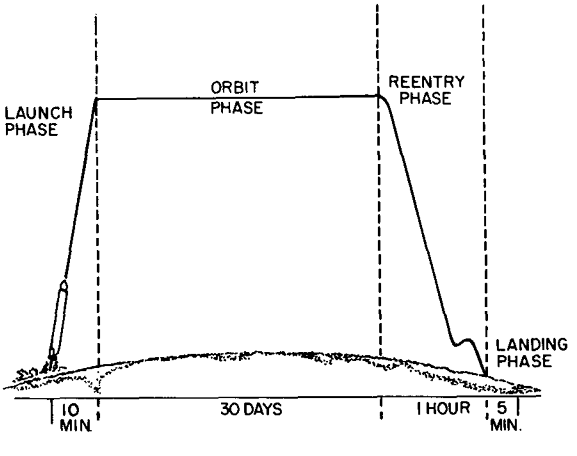


FIGURE 1–2. An example of a mission description (after Kearns and Ritchie, 1961).

ments of each phase. This analysis by phase is useful in determining what the environment will be during each phase, the supporting systems that will be needed, and the limiting conditions that will be imposed upon equipment and men.

Having broken the mission into its major phases, each one is then examined to determine its composition. Figure 1–3 shows how the orbital phase is broken down into the subphases of rendezvous and docking.

The analysis of system requirements must also take into consideration other system constraints. For example: (a) the vehicle must be capable of operating under all weather conditions, or (b) no "7-level or higher" ground support skill will be required to maintain the system at organizational levels, or (c) total training equipment for air and ground crews must be limited to 4.5 million dollars, or (d) development must be completed by December, 1975, or (e) mean allowable down time for central control processor is one hour per week.

The purpose of analyzing system requirements and constraints is to identify the specific functions the system must perform. This, in turn, leads to a determination of the kinds of human and instrument capabilities required to satisfy the functional requirements. From a human engineering standpoint, this means identifying the personnel functions and also the specific human performance, personnel, and training requirements of the system.

1.4 The Analysis of System Functions

A system function may be viewed as a broadly defined operation or activity which contributes to the system's mission or goal. It is usually the primary reason for including a particular subsystem (equipment, or crew member) in the design system. These functional operations or activities may include: (a) detecting signals, (b) measuring information, (c) comparing one measurement with another, (d) processing information, and (e) acting upon decisions to produce a desired condition or result on the system or the environment. The identification, analysis, and synthesis of system functions is the sequence

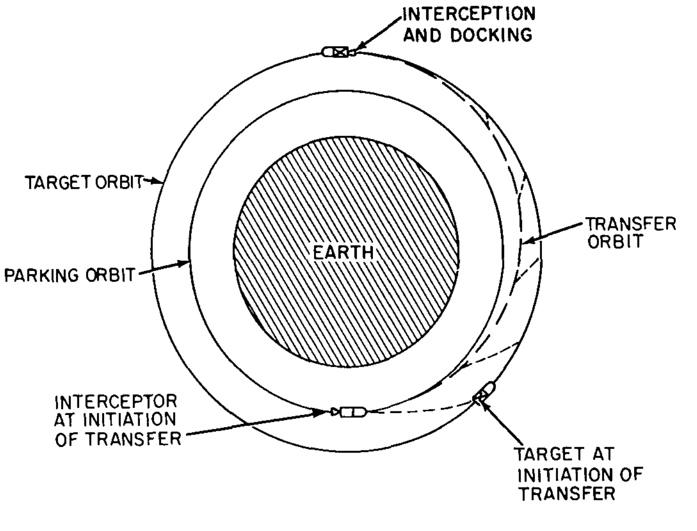


FIGURE 1–3. Orbital transfer phase (after Ryken et al., 1963).

which translates system requirements and constraints into an organized program for design implementation. Design requirements are the human or instrumented capabilities, or combinations of these, that may be used to accomplish the system functions. They identify the processes which convert available inputs into required outputs, as in guiding a missile, transmitting air traffic control orders, arming a warhead, etc. (see for example, AFSC Manual 80-3).

Initially, only gross system functions are identified, and not the personnel or equipment that will accomplish the functions. System requirements determine system functions, and the function itself determines what equipment and human performance is necessary to carry out that function. For example, the mission requirements of the space vehicle referred to earlier might include: (a) injecting the manned vehicle into a planned orbit from a ground launch, (b) rendezvousing with a target satellite, (c) leaving the orbit, (d) re-entering the earth's atmosphere, and (e) landing on the surface of the earth at a designated base. Each major mission segment imposes specific performance requirements which must be satisfied; for example, "safe mechanical attachment to the target satellite" is a requirement for the rendezvous phase. From this information, system engineers identify the specific functions the system must perform in order to fulfill the stated requirement. One function would be that of "stabilizing the attitude of the vehicle." Since only the function is identified at this point, function analysis must be performed to determine the specific capabilities needed; e.g., can this function best be accomplished by means of an automatic attitude stabilization system, or by a human operator.

The purpose of function analysis is to determine how each function can be performed in the system and to consider the feasible alternative combinations that will lead to successful completion of the mission. This is the step in which the design approach is determined. For example, if navigation has been identified as a system function, a human capability might be "manual dead reckoning," an equipment capability might be "an inertial system," and a man-equipment capability might be "radar dead reckoning." Regardless of the specific procedures involved in analyzing system functions, three major steps will normally be involved: (a) examining each system function to determine the kinds of capabilities needed to meet the system performance requirements, (b) exploring possible combinations of man-equipment capabilities through trade-off studies and time-line analyses, (c) determining which design approach will maximize overall system cost effectiveness.

When gross functions are restructured into lower level subfunctions, the kinds of subsystems, equipment, or man-equipment combinations that will satisfy the specific functional requirements are often apparent. However, each capability must be analyzed in terms of interface requirements, and the probable effect of each alternative should be evaluated with respect to other aspects of system performance. In some cases, the availability of existing equipment will dictate the most realistic combination from a cost-effectiveness standpoint.

1.5 Operator-Centered Analyses

1.5.1 Task Analyses

Previous practical experience with system design has indicated the value of specifying the performance characteristics of the tasks to be accomplished by human operators. Such specification covers the psychological aspects of the indication to be observed (stimulus and channel), the action required (response behavior, including decision making), the skills and knowledges required for task performance, and probable characteristic human errors and equipment malfunctions (Miller, 1953). The gathering and organization of such information is called task analysis.

The objective or purpose of task analysis is to provide the basic "building blocks" for the rest of the human engineering analysis. The development of this information through task analysis can be divided sequentially into two major parts: (a) subtask derivation and (b) skill and knowledge analysis. In subtask derivation, information pertinent to the entire subtask is obtained, including the location at which it is performed, and its relationship to existing tasks. To find out which skills and knowledge are required involves an examination of the various steps or parts of the subtasks. This analysis results in a statement of the psychological requirements of the tasks, the kinds of discriminations that must be made, the decision-making, motor and other skilled responses required.

The kinds of information that subtask derivation may include are:

1. A statement of the task, as derived from the personnel function. The statement should contain an action verb and indicate the purpose of the task in terms of a system goal or subgoal, an example would be the words "actuate power switch."

2. The category of task should be noted, i.e., whether it is an operator, maintenance, or support task.

3. The location in which the task is performed. This separates tasks performed in such separate places as the maintenance shop, supply area, manufacturer's plant, on board ship, in flight, on the flight line, etc. (See Folley et al., 1960.)

The task may be peculiar to one or more segments of a work cycle or mission as identified by mission analysis. The segment in which it falls should be named.

The task may be discontinuous and readily divisible into discrete subtasks frequently encompassing procedures such as checking. Or it may be a continuous control action that affects a continuously varying stimulus, as in steering an automobile. Each task should be identified as one of these two types.

The frequency of task occurrence should be estimated, if the task is periodic. If performance of the task depends on the occurrence of malfunctions, human errors, or contingencies, the probability of task performance should be estimated. This probability estimate may be revised as operational evidence becomes available. One possible format for the analysis is shown in Figure 1-4.

As human functions are identified, they are entered in Row 1 of the format. These functions represent major types of human activity such as maintenance, operation, or command. Row 2 is used to enter the task within the function. The maintenance function, for example, can be broken down into the following tasks: checking, adjusting, servicing, troubleshooting, repair.

Column 3 is used to enter the subtask, such as checking power supplies, servicing punch clutch mechanism, or trouble-shooting a particular component. The action stimulus, or event, that instigates performance of the task is entered in

Function: (1)	Operate aircraft power plant and system controls
Task: (2)	Control jet engine operation

Subtask (3)	Action Stimulus (4)	Required Action (5)	Feedback (6)	Task Classification (7)	Potential Errors (8)	Time (9) Allow-able (9a)	Time (9) Neces-sary (9b)	Work Station (10)	Skill Level (11)
3.1 Adjust engine r.p.m.	4.1 Engine r.p.m. on tachometer	5.1 Depress throttle control downward	6.1 Increase in indicated tachometer r.p.m.	7.1 Operator task, aircraft commander	8.1 a. Misread tachometer b. Fail to adjust throttle to proper r.p.m.	9a.1 10 sec.	9b.1 7 sec.	10.1 aircraft commander's seat	11.1 Low

FIGURE 1-4. Format for task allocation and analysis.

Column 4. This stimulus may be an out-of-tolerance display indication, a requirement of periodic inspection, a command, or a failure, and so on. The required action or response is indicated in Column 5. This may be a control movement, a voice communication, a command, or other human response. In Column 6 is entered the feedback or indication of the adequacy of the response, such as "power indicator light is ON."

The task classification, Column 7, is a code assigned to each task to indicate its similarity to other tasks. Beginning with the first task, a separate code number is assigned to a task if it is unlike any previous task in terms of the entries in Columns 4, 5, and 6. If it is similar to a previous task with respect to all three columns, it is given the same code number as that task, plus an A suffix. If it is similar with respect to only one column, the same code number with a C suffix should be used. This coding scheme provides a basis on which to sort and classify tasks according to their similarity.

Column 8 is used to enter potential sources of human error, abnormal conditions, and equipment malfunctions. Column 9 is divided into two sections, Allowable and Necessary Time. Allowable time is the time within which a task must be performed and is entered in Column 9a. Necessary time is the time required for a man to perform the task and is entered in Column 9b. Allowable times are obtained from functional-analysis data of the type found in time profiles. Necessary time data may be gotten from several sources. If off-the-shelf equipment is being used, the time data can be obtained from operational records or from empirical tests. If only static mockups of the type used in design engineering inspections are available, acceptable time estimates can be made for many operations. Real-time simulation introduces the dynamics of task performance and produces more precise time estimates.

The geographic location or work station at which the task is performed is entered in Column 10. Finally, skill levels are entered in Column 11. These skill level entries are obtained by having a minimum of three trained job analysts rate tasks on a five-point scale according to their difficulty. The separate judgments are then aggregated, and the mean difficulty level is used as a measure of the skill level required to perform the task.

Job allocation is accomplished when the subtasks have been assigned to specified personnel positions. This is achieved by successive sorting of the data in Columns 7, 10, and 11. First, the entries in Column 10 are sorted into groups of tasks performed at common work stations. This eliminates the assignment to one individual of tasks performed at several work stations. Next, the data of Column 7 are arranged into groups of tasks of a similar nature that are to be performed at a given work station. Finally, the entries in Column 11 are sorted into groups of tasks with a common skill level requirement.

1.5.2 Extension of Task Analysis

In its early forms, task analysis tended to be relatively static and descriptive. However, extensions toward the dynamic and prescriptive were inherent in the original formulations. For example, it soon became apparent that the basic technique permitted an assessment of the workload imposed on the operator over time.

A kind of time-line analysis can be made for critical functions to determine whether time constraints would necessitate a reallocation of functions. Analysis of a sequential list of system functions against a time base can reveal: (a) periods of peak personnel and equipment workloads, (b) situations involving a conflict in demands upon personnel or equipment, and (c) additional requirements not previously apparent from function analysis.

Such analysis permits an evaluation of the number and durations of functional activities assigned to each element of the system. It provides an overview of what system personnel are expected to accomplish during various time increments for each major functional area of the system. Severe overloading should be eliminated by adding personnel to the work station, by dividing work stations, or by other means. An underloaded position may be combined with another underloaded position with the same skill level on similar tasks. A given capability may be used to cut across or combine two or more separate functions, since some functions must be repeated at different points in the function sequence. Also, where separate functions are very similar to one another, it may be possible to combine them, resulting in the repetition of a single function so as to increase system effectiveness and reduce personnel requirements.

Another, related extension of task analysis concentrates on predicting the occurrence of operator errors and identifying the locus of other contingencies. A contingency is any nonroutine situation with which the system may have to deal in performing required functions. The number and variety of contingencies for a given system may be very large. However, engineering data on system requirements, equipment and human performance requirements, environmental factors, and other mission data should provide a basis for anticipating possible nonroutine situations. The majority of these are likely to occur within the following categories:

1. Malfunctions, such as accidents, failures, etc., of either personnel or equipment;
2. Extreme climatic or environmental conditions, such as snowstorms, dust, hurricanes, acceleration, humidity, vibration or radiation.
3. Enemy activities, such as detection, interception, use of weapons, etc.

After the specific contingencies pertinent to the system have been determined, the design alternatives for resolving items must be selected. Contingency analysis may indicate new functional requirements or may necessitate a revision of previously established functions. For example, in the case of a potential malfunction of a critical nature, the system must be able to: (a) detect the malfunction, (b) determine its location and extent, (c) evaluate its effect on the system and its consequent effect on the mission, (d) determine the corrective action to be taken, (e) carry out the corrective action, and (f) retest the system after the corrective action has been completed. These newly identified functions may then generate equipment design requirements.

Characteristic equipment malfunctions and/or human errors should also be identified. In the development stage, information of this type may be obtained by examining maintenance records and observing the operation of similar equipment. Some data may be derived from contingency analyses or from examination of component characteristics. All these data should be reviewed and updated when subsystem and system tests provide a more valid basis for their derivation.

Because specific error information is difficult to obtain during the development stage, it may be difficult or impractical to require specific estimates of the probability of error occurrences. The following general scale may be used, however, to supplement the categorization and description of the error.

1. Error *highly likely* to occur unless specific avoidance technique is utilized (probability greater than 0.5).
2. Error *likely* to occur unless specific avoidance technique is utilized (probability greater than 0 but less than 0.5).
3. Error *unlikely* to occur even if specific avoidance technique is not utilized (probability of 0 or very close to it).

The quantification of human reliability for specific tasks has been attempted. If the designer desires to quantify human error performance more precisely than the categorical scheme described above he should consult Payne and Altman (1962) and Swain (1964) for a description of existing data and techniques.

The system consequences of each equipment malfunction and human error should be described. An estimate of the seriousness of these consequences may be made along the following scale:

1. Hazard to personnel and/or equipment.
2. Degradation of system performance.
3. Degradation of subsystem performance.
4. Degradation of component performance.
5. Little effect on system performance.

The estimate of consequence seriousness or extent may be judged during the development phase by analyzing block and circuit diagrams relevant to the malfunctioning part of the data chain in which human error was made.

When these two analyses, the error analysis and the system consequences analysis, are considered together, the result is called an "error made and effects" analysis. The designer must eliminate the possibility of an operator action that would have a catastrophic effect on personnel or equipment safety even if the event has a low probability of occurrence. On the other hand, the designer should not be concerned with eliminating those human errors or stylistic differences in performance that have little or no effect on other personnel, on equipment, or on system performance. There are several alternatives in eliminating or minimizing the probability of a human error occurring. The safest approach is to design the system so that the operator cannot commit the error. Alternate approaches that can be considered are: warning and caution labeling, procedure development, and training emphasis. Specific aspects of system development to minimize human error are discussed throughout the *Guide*. The elimination of error-potential situations, especially those causing hazards to personnel and/or equipment, should be a prime consideration of the system designer.

1.5.3 Sequential Analysis

More sophisticated extensions of task analysis go well beyond the time-line concept introduced above while retaining the basic attribute of a time-ordered sequence. There is a wide variety of applications of these techniques, such as describing production processes or clerical operations, or data processing, or information-handling systems, and many other kinds of activities in large systems, small subsystems, or even in one-man operations. This method is useful for developing a picture of system operation; for indicating the functional relationships between system elements; for tracing the flow of materials, information, etc.; for showing the physical or sequential distribution of operations; or for identifying the inputs and outputs of particular subsystems.

A chart can be made presenting in a systematic, easy-to-reference fashion the information that is derived from sequential analysis. It can depict the information-decision-action pattern for each function performed by the operator. An example of such a chart is shown in Figure 1–5 (Coakley and Fucigna, 1955) which presents one section of a decision analysis of a radar-tracking system.

Figure 1–5 shows that part of the chart for the pre-acquisition phase. Each function performed during a phase is represented by a different symbol: a triangle for information, a circle for decision, and a square for action. The chart then shows, in exact sequence, the inputs (information), decisions, and outputs (actions) of man. Note that man receives information from the equipment, the target, and learned operating procedures. Man then makes a decision and performs two actions. These actions in turn bring more information.

The most advanced extensions of the basic notions of task analysis are exemplified by the refinements of Operational Sequence Diagrams (Kurke, 1961). The advantages of this approach derive from the capability of the diagrams to make explicit the consequences of alternative design configurations. Furthermore, the approach achieves a unification of the concepts of event, information, decision, and action which are otherwise difficult to handle.

The basic symbology is illustrated in Figure 1–6, and an actual application covering the comparison of two design alternatives is shown in Figure 1–7. Kurke also suggests that the diagram procedure can be combined with symbolic logic and numerical probability estimates to yield quantitative predictions of system failure (or the inverse, system reliability).

1.6 The Use of Analytic Techniques

Although the conceptual procedures presented above were illustrated with specific examples, the uses of these procedures are possibly not yet clear in a practical sense. In order to achieve this level of understanding, it is necessary to review the process of system development and indicate explicitly where it is possible to interject human engineering influence.

In the classical pattern of industrial human engineering, the first instance for human engineering participation may come from a Request for Proposal or some similar instrument from a governmental agency. At this time, the probable response of the engineering enterprise is to form a "proposal team" in which the human engineer is included.

As now widely practiced, the proposal preparation process is, in effect, a preliminary design activity. Many crucial engineering decisions are made at this first stage, never to be controverted. In other words, if the human engineer does not provide inputs at this stage, his total influence on the final version of the equipment or system is likely to be marginal.

In some cases, specific design decisions may be allocated to the human engineer. Workspace

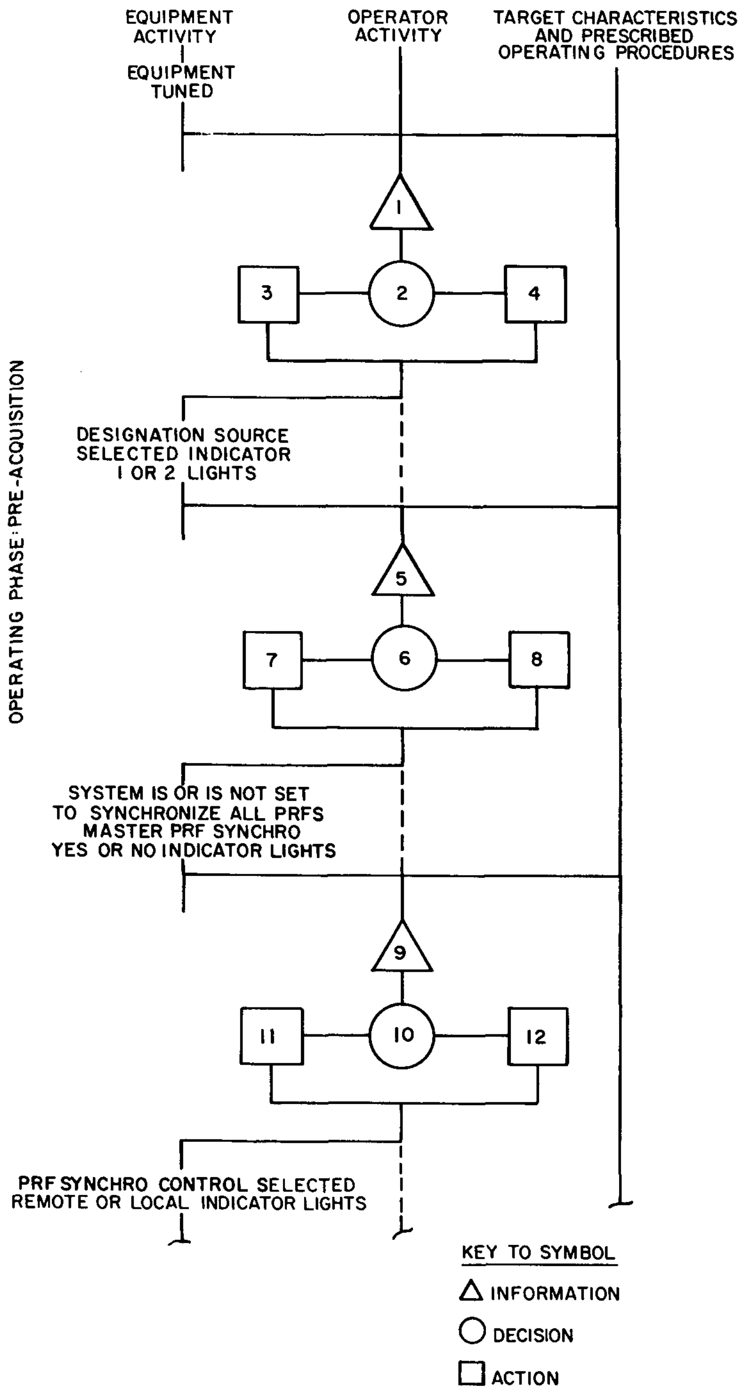


FIGURE 1-5. Decision, action, information diagram for radar-tracking system (Coakley and Fucigna, 1955).

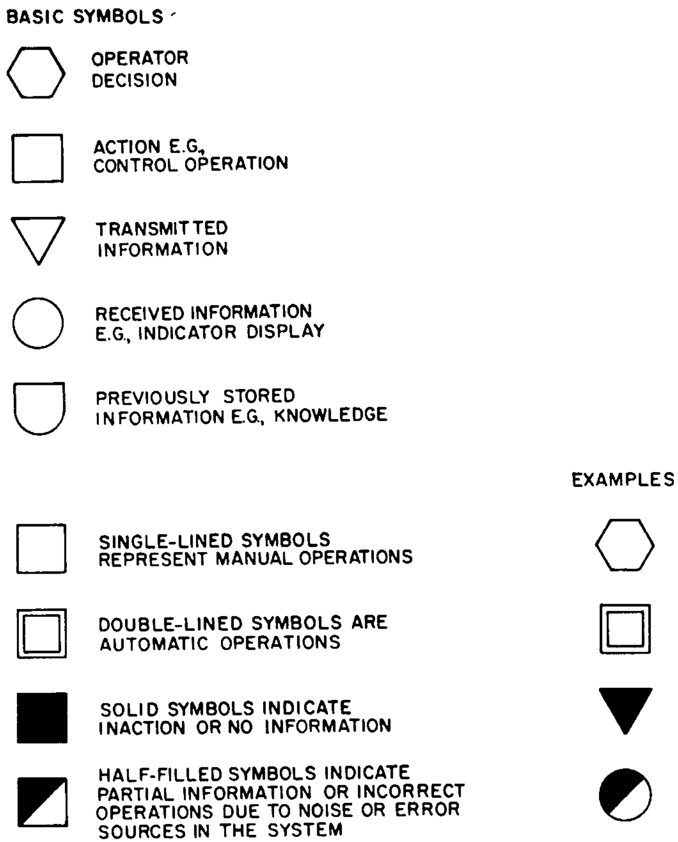


FIGURE 1–6. Symbols used in operator sequence diagrams (Folley et al., 1960).

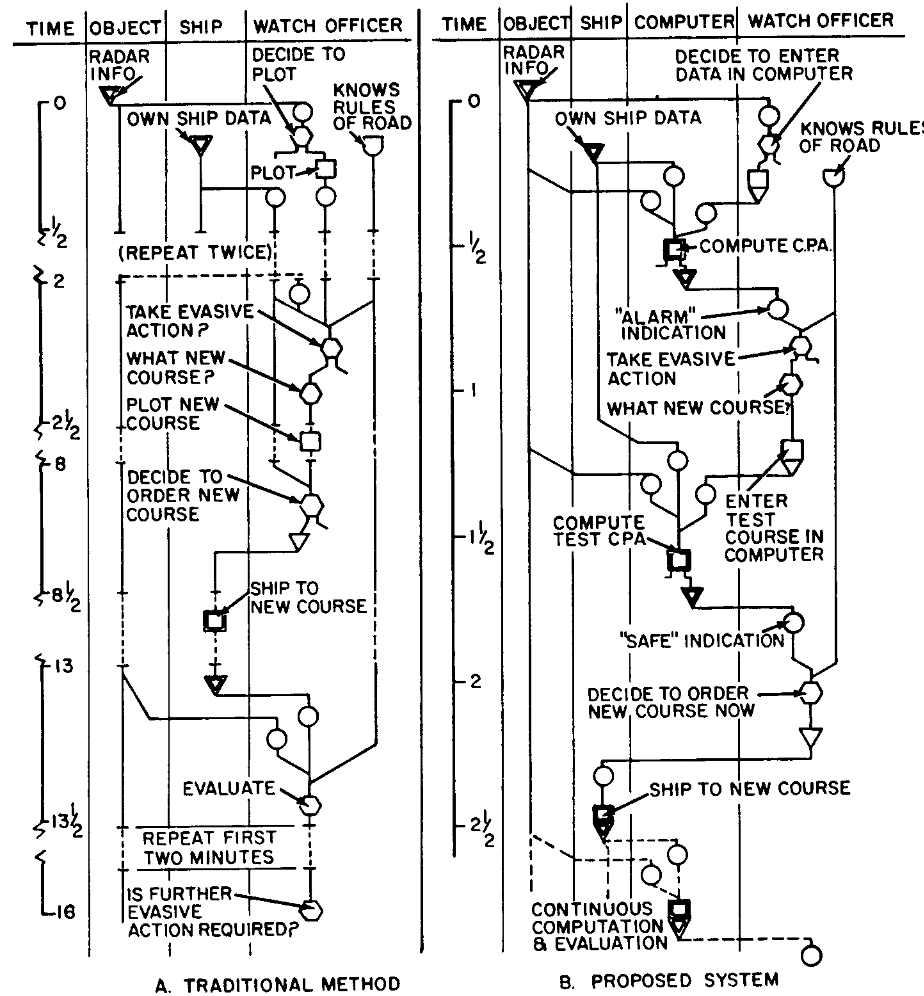


FIGURE 1–7. Detail analysis of alternate collision-avoidance systems (Kurke, 1961).

arrangements and control/display panel layouts, for example, are often assigned directly to the human engineer. Other members of the team will then have the task of working out ways to implement the human engineering design.

In most cases, however, the human engineer will function in an advisory capacity, volunteering information or responding to specific requests for information which will influence design decisions. Thus, one way of looking at the human engineer's job is that he is primarily a collector, organizer, and provider of information in a problem-solving/decision-making situation. Many of the kinds of information which are likely to be needed can be found in the chapters of this *Guide*, other texts, and the professional literature of the field. However, the human engineer is well advised not to limit his attention to these sources but to be alert to other sources, particularly those which bear on the operational contexts in which the equipment or system is to be used and on the engineering characteristics of the equipment itself.

An example of the increasingly vigorous tendency to incorporate contextual information into the human engineering part of the design process, along with the more traditional categories of anthropometrics, psychophysiological capacities and limitations, perceptual-cognitive abilities, and so forth, comes from recent work on intelligence and communications systems (Tonik, 1967). The inquiries into context are generally labeled as studies of "user needs." Ideally, some member of the design team is familiar in detail with the operational environment and the problems of prospective users or operators of the system (or equipment) being designed. More pointedly, some member of the design team makes it his business to know what the user needs or expects the system to do for him—and speaks for the user at the instance of making design decisions. The human engineering member of the design team is a logical candidate for this duty because of his training and experience in relating requirements to potential solutions.

Once the design process is complete and pilot or prototype model fabrication is initiated, the primary job of the human engineer in industry is to monitor the process. It is often the case that an excellent "blueprint" must be modified to accommodate various production procedures and constraints. The human engineer should act to insure that features desirable from a human factor standpoint are not needlessly comprised to facilitate production objectives.

It is in the test and evaluation phase, however, where a critical resurgence of human engineering participation and responsibility is likely to occur. Most acceptance test procedures are standardized and are conducted against physical criteria (e.g., weight, space, power, heat resistance, etc.). However, operational tests will (or should) utilize the human element. Insofar as they do so, the controls needed, if valid conclusions are to be drawn, are similar to those needed in the conduct of a psychological experiment. Similarly, the measures of system performance will be analogous in their mathematical/statistical character to measures of human behavior. Fortunately, there is a growing tendency to conduct system and equipment tests according to the dictates of good experimental design.

Given the general review of the equipment development process as annotated above, it is possible to consider a more specific work aid for the human engineer. This is presented in the form of a set of questions, in more or less serial order, which the human engineer is likely to be called upon to deal with in the process of the design, fabrication, test and evaluation of a system.

1. Preliminary design phase

   (*a*) Why is this system being sought?
   What mission will the system be expected to fulfill?
   More particularly, what is this new system expected to do that existing systems are not doing or are not doing well enough?

   (*b*) How is the system to fulfill its mission?
   What are the stages of mission execution?
   What functions must be accomplished by the system at each stage?

   (*c*) In what environments must the system function?
   What particular hazards will obtain?
   What stresses or demands are likely to be placed on the system?

   (*d*) Who will benefit by system operation?
   Who will use the system?

What kinds and numbers of operator and/or maintenance personnel are available?

(*e*) What are the major technological options?

What alternative configurations are feasible?

What particular resource or class of resource is most crucial to prospective system effectiveness?

2. Advanced design phase

(*a*) What functions should be assigned to human operator and support personnel?

What conditions will impose peak task loads on the operator or operators?

What conditions (e.g., long periods of inactivity) will tend to degrade operator performance?

What pattern of decision—action will occur at crucial mission stages?

(*b*) What information is required by operators (and/or support personnel) in order to fulfill their functions?

What is the probable pattern of channels and of flow rate for this information?

In what form (i.e., code, mode, format) will the information be most useful to the operator?

(*c*) How many humans are needed to man and support the system under normal and peak load conditions?

What special skills, capabilities or attributes are needed for effective operator performance?

What special training, if any, will be required?

Is such training feasible?

What resources will be required to implement the training?

(*d*) How should the assigned functions be distributed among operator and support personnel?

How should the work stations be arranged?

What instrumentation is required at each work station?

How should this instrumentation be laid out?

(*e*) What specific devices, tools, or controls are most appropriate to the pattern of task actions that will be imposed on operator and support personnel?

What kinds of aids, guides, indicators, locks, interlocks, cover plates, etc. would be useful to facilitate correct actions and prevent operator errors?

What means are available to allow quick recovery or to maintain the safety and integrity of the system in the event of operator error or failure?

3. Mock-up to prototype fabrication phase

(*a*) What options are available for eliminating, combining, or simplifying any of the instrumentation?

(*b*) What will be the effect, if any, on human performance, safety, or morale of any proposed changes in configuration or instrumentation?

(*c*) What safeguards, if any, are required to insure adherence to the design plan and functional requirements of the system?

What quality control procedures are required to insure the validity of human factors considerations in the final product?

4. Test and evaluation phase

(*a*) By what means can test and evaluation be made as realistic as possible in terms of the ultimate operator and support personnel and in terms of the operational conditions?

(*b*) What criteria of system and operator performance are logical in terms of the mission and functions assigned?

What measurement procedures will yield data which are valid with respect to such criteria?

What test instrumentation is required?

(*c*) What form of test design will yield unequivocal answers to questions of the effectiveness, operability and maintainability of the system?

What is the most economical way of implementing the test design required?

1.7 The Values of Human Engineering in System Analysis

The payoff in human engineering during system analysis is determined by concrete operating dividends which manifest themselves in the performance of the final system. Some typical criteria, which can be used as yardsticks for assessing the effectiveness of the human engineering effort, are described below (see for example, AFSC Manual 80-3).

1. Improved Performance. Just as the "right man" in the "right job" with the "right tools" is likely to be a productive industrial worker, so the man with the "right assignment," the "right tasks," the "right equipment," and the "right environment" is likely to be an efficient equipment operator or maintenance technician.

2. Reduced Training Costs. A man needs less training to operate a device when both the device and the operational procedures are properly engineered for human use. If he fits the device and the device fits him, and if the task procedures are optimum, he reaches the required standard or proficiency with less expenditure of time, money, and effort.

3. Improved Manpower Utilization. Tasks and tools are optimally human engineered when they minimize the need for special skills or high aptitudes in the human operator. More of the manpower pool can then be trained to do the tasks, thereby improving manpower utilization. A measure of the human engineering contribution to system design is the percentage of the available manpower that can perform the required tasks. In the face of limited manpower resources, it is essential that our human resources be used with utmost effectiveness.

4. Fewer Losses From Accidents and Misuse. Poor equipment design causes many of the accidents commonly attributed to human error. Poorly designed equipment tends to be used incorrectly or not at all. Man-hours and equipment wasted are a drain on the economy that can be at least partially reduced by better application of human engineering principles.

5. Increased Economy of Production and Maintenance. Simplification of design often results in devices that are not only easier to operate, but simpler to manufacture and maintain. The application of the principles of human engineering will frequently lead to more economical production and maintenance, in addition to increased operator effectiveness.

6. Improved User Acceptance. Although keeping its personnel comfortable and happy may not be the primary goal of a system, it is essential that a pilot or other system operator be given freedom to "do his job" rather than be forced to fight his own equipment. A good system designed for ease of operation and maintenance, as well as for the safety and protection of its personnel, inspires confidence and increases efficiency. While some frustrations are inevitable, one goal of human engineering is to ensure that frustrations caused by troublesome or hazardous equipment are reduced to a minimum.

References

Coakley, J. D., & Fucigna, J. T. Human engineering recommendations for the instrumentation of radar. Report No. AN/FPS-16 (Xn-2), 1955, Radio Corporation of America, Moorestown, N. J.

Folley, J. D., Jr., Altman, J. W., Glaser, R., Preston, H. O., & Weislogel, R. L., (Eds.) Human factors methods for system design. Report No. AIR-B90-60-FR-225, 1960, The American Institutes for Research, Pittsburgh.

Kearns, J. H., & Ritchie, M. L. Cockpit control-display subsystems engineering. ASD Technical Report No. 61-545, October 1961, Wright-Patterson AFB, Ohio.

Kurke, M. I. Operational Sequence Diagrams in systems design. *Human Factors*, 1961, **3**, 66-73.

Miller, R. B. A method for man-machine task analysis. Report No. WADC-TR-53-136, June 1953, Wright Air Development Center, WPAFB, Ohio. (Also available as Report No. AIR-A41-53-FR-51, 1953, The American Institutes for Research, Pittsburgh.)

Payne, D., & Altman, J. W. An index of electronic equipment operability: Report of development. Report No. AIR-C43-1/62-FR, January 1962. The American Institutes for Research, Pittsburgh.

Ryken, J. M., Emerson, J. E., Onega, G. T., & Bilz, J. L. Study of requirements for the simulation of rendezvous and docking of space vehicles. AMRL-TDR-63-100, Wright-Patterson Air Force Base, Ohio, October, 1963 AD 425-499.

Swain, A. D. Some problems in the measurement of human performance in man-machine systems. *Human Factors*, 1964, **6**, 687-700.

Tonik, A. B. (Ed.) Information retrieval: The user's viewpoint; an aid to design. From the Fourth Annual National Colloquium on Information Retrieval, 1967. International Information, Inc., Philadelphia.

United States Air Force. AFSC Manual, Handbook of instructions for aerospace personnel subsystem design. Report No. AFSCM 80-3, Transmittal No. 14, HIAPSD, July 1966, Andrews AFB, Washington, D. C.

Selected Readings

Chapanis, A. On some relations between human engineering operations research and system engineering. In D. P. Eckman (Ed.), *Systems: research and design.* New York:John Wiley and Sons, Inc., 1961.

Dickman, S., & Katzenberger, F. X. Systems Engineering. *Sperry Rand Engineering Review*, 1968, **21**(1), 2-8.

Ellis, D. O., & Ludwig, F. J. *System philosophy.* Englewood Cliff, N.J.:Prentice-Hall, 1962.

Wilson, W. E. *Concepts of engineering system design.* New York:McGraw-Hill, 1965.

Chapter 2

Man as a System Component

Harold P. Van Cott

American Institutes for Research
Washington, D.C.

Melvin J. Warrick

Aerospace Medical Research Laboratory
Wright-Patterson Air Force Base, Ohio

This chapter introduces some scientific and technical fundamentals of human performance that relate to equipment design. Man is viewed as an information processing system with component sensing, processing, memory, and response subsystems. A discussion of these subsystems and their parameters provides a background for subsequent chapters. The focus is on human performance and performance constraints as they interact with equipment design and the environment in which man works.

CONTENTS

This chapter is based, in part, on Chapter 1 of the previous *Guide*. It was reviewed by H. Wallace Sinaiko.

2. Man as a System Component

2.1 A General Model of Human Performance

Using information processing as a model, four human subsystems can be identified: (a) sensing, (b) information processing, (c) memory or storage, and (d) responding. The interrelationships among these functional subsystems are shown in Figure 2–1.

The human sensing subsystem detects and encodes energies from the physical environment. The information processing subsystem acts upon stored and sensed inputs, discriminates signals from noise, recognizes information patterns, makes decisions, and selects appropriate responses from among available options. It also controls regulatory and reflexive processes for survival and adaptation. The memory subsystem provides long-term and short-term storage of encoded information. The response subsystem transduces processed information into such actions as postural adjustments of the body and limbs, search and scan movements of the eyes, and production of speech.

2.2 Sensing Subsystem

2.2.1 Operational Characteristics

Man's sensing system consists of receptors responsive to specific types and ranges of physical energy. Energy that activates a receptor or sensing organ is a *stimulus.* Some sense organs respond to external stimuli (e.g., the eye to luminous energy) while others respond to internal

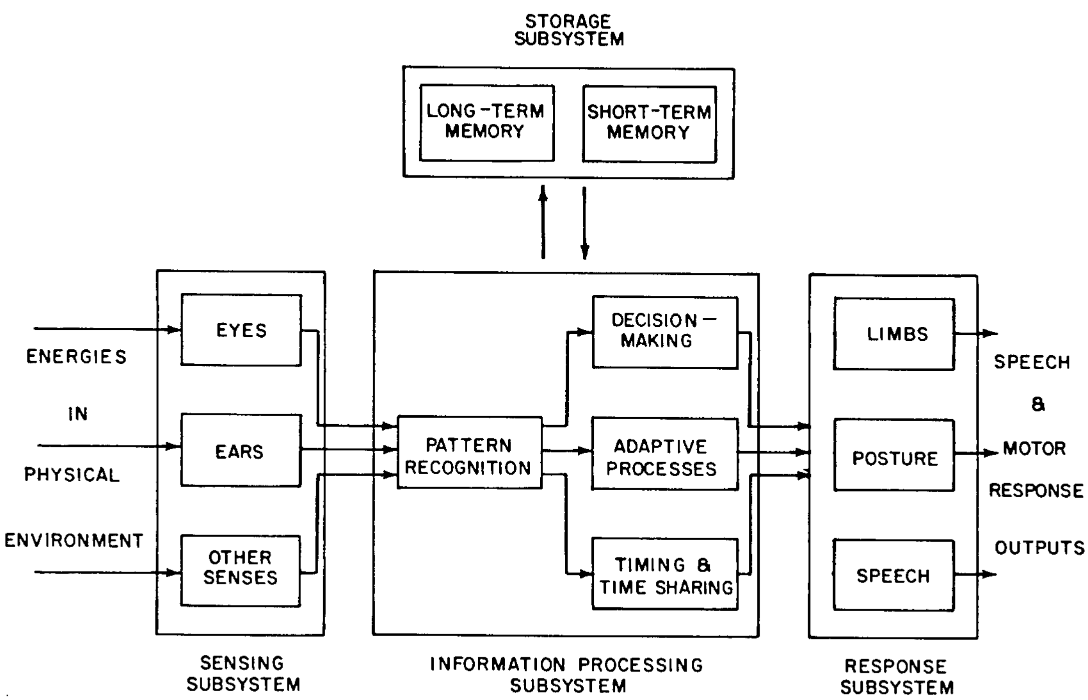


FIGURE 2–1. The human information processing system (after Welford, 1960; McCormick, 1964; and Pew, 1965).

TABLE 2–1. MAN'S SENSES AND THE ENERGIES THAT STIMULATE THEM

Sensation	Sense organ	Stimulation	Origin
Sight	Eye	Some electromagnetic waves.	External.
Hearing	Ear	Some amplitude and frequency variations of pressure in surrounding media.	External.
Rotation	Semicircular canals	Change of fluid pressures in inner ear.	Internal.
	Muscle receptors	Muscle stretching	Internal.
Falling and rectilinear movement.	Otoliths	Position changes of small, bony bodies in inner ear.	Internal.
Taste	Specialized cells in tongue and mouth.	Chemical substances	External on contact.
Smell	Specialized cells in mucous membrane at top of nasal cavity.	Vaporized chemical substances.	External.
Touch	Skin	Surface deformation	On contact.
Pressure	Skin and underlying tissue	Surface deformation	On contact.
Temperature	Skin and underlying tissue	Temperature changes of surrounding media or objects, friction, and some chemicals.	External on contact.
Pain	Unknown, but thought to be free nerve endings.	Intense pressure, heat, cold, shock, and some chemicals.	External on contact.
Position and movement (kinesthesis).	Muscle nerve endings	Muscle stretching	Internal.
	Tendon nerve endings	Muscle contraction	Internal.
	Joints	Unknown	Internal.
Mechanical vibration.	No specific organ	Amplitude and frequency variations of pressure.	External on contact.

Adapted from Mowbray and Gebhard (1958).

stimuli (e.g., kinesthetic receptors to limb movements). Table 2–1 lists the human sense organs, the types of energy that stimulate them, their corresponding sensations, and the origin of stimulation. Further information about vision, audition, and other sensing subsystems and their design implications may be found in Chapters 3, 4, and 5 of this *Guide.*

2.2.2 Design Implications

In the design of a man-machine system, the human senses may be considered candidates for directly detecting and measuring conditions about the performance of the system or about events in its environment. Human sensing may, of course, be augmented by electromechanical or optical sensors that detect energies outside the human limits. Properly processed prior to display, these energies can also be recognized, interpreted, and utilized by man as a source of information.

For the designer to take advantage of a human sensor's sensitivity, resolving capability, or information transmission capacity, he must do so indirectly by changing or selecting conditions external to it: the energy or information source, its location, intensity and coding; the rate and manner of information presentation; etc. Specific design implications and principles are described in Chapters 3, 4, and 5.

2.2.3 Correlation with the Physical Stimulus

As the magnitude of a physical stimulus changes along a given dimension, the magnitude of the corresponding sensation also changes. The relationship, however, is not perfect. Human sensing subsystems are typically nonlinear; equal increments in stimulus magnitude do not produce equal increments in sensation over the entire range of stimuli. For example, doubling sound energy does not double loudness. As a matter of fact about a tenfold increase is required for a sound to seem twice as loud. In general the perceived magnitude of a stimulus is proportional to some power of the physical magnitude. This power (exponent) is different for different stimuli. Although it is less than one for loudness, there are sensations such as shock, temperature, heaviness, and others, for which the exponent is greater than one.

While sensation along a given stimulus dimension varies primarily with that dimension, it is affected by changes along other dimensions of the same stimulus. For example, loudness of a tone varies with its frequency as well as with its amplitude. This is why compensation circuits are used in auditory equipment. Similarly, the apparent brightness of a light varies with the color as well as with the radiant energy of the source.

2.2.4 Sensitivity Range

The largest and smallest amounts of energy to which a particular sense is responsive define its upper and lower limits. Table 2–2 lists the practical upper limits and the lowest thresholds for various sense organs, giving the smallest detectable amount of energy and the largest amount permissible before pain or permanent damage occurs. Stimuli near either of these limits lead to unreliable sensing and should be avoided as signal sources.

The lower sensitivity threshold is not really a specific, precise value, but, rather, the average, mean or median, of a statistical distribution of energy values. Typically the threshold is defined

TABLE 2–2. STIMULATION-INTENSITY RANGES OF MAN'S SENSES

Sensation	Smallest detectable (threshold)	Largest tolerable or practical
Sight	10^{-6} mL.	10^{4} mL.
Hearing	2×10^{-4}dynes/cm^2.	$<10^{3}$ dynes/cm^2.
Mechanical vibration	25×10^{-5} mm average amplitude at the fingertip (Maximum sensitivity 200 Hz).	Varies with size and location of stimulator. Pain likely 40 dB above threshold.
Touch (pressure)	Fingertips, 0.04 to 1.1 erg (One erg approx. kinetic energy of 1 mg dropped 1 cm.) "Pressure," 3 gm/mm^2.	Unknown.
Smell	Very sensitive for some substances, e.g., 2×10^{-7} mg/m^3 of vanillin.	Unknown.
Taste	Very sensitive for some substances, e.g., 4×10^{-7} molar concentration of quinine sulfate.	Unknown.
Temperature	15×10^{-5} gm-cal/cm^2/sec. for 3 sec. exposure of 200 cm^2 skin.	22×10^{-2} gm-cal/cm^2/sec. for 3 sec. exposure of 200 cm^2 skin.
Position and movement	0.2–0.7 deg. at 10 deg./min. for joint movement.	Unknown.
Acceleration	0.02 g for linear acceleration 0.08 g for linear deceleration 0.12 deg./sec^2 rotational acceleration for oculogyral illusion (apparent motion or displacement of viewed object).	5 to 8 g positive; 3 to 4 g negative. Disorientation, confusion, vertigo, blackout, or redout.

Based in part on Mowbray and Gebhard (1958).

as that stimulus value which, under ideal conditions, is detected approximately 50% or in some cases 75%, of the time (Figure 2–2). In practical applications, much higher values are required.

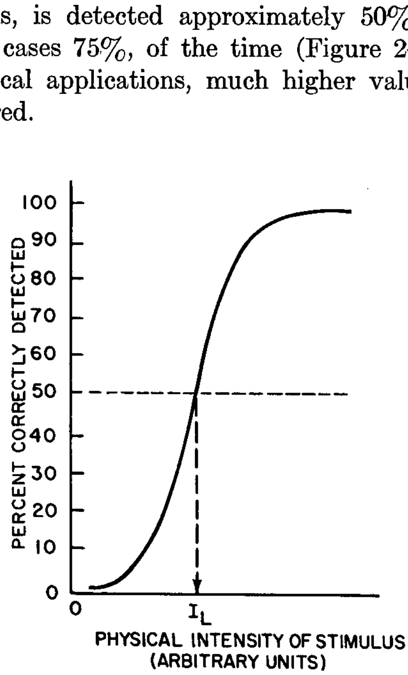


FIGURE 2–2. Determining the lower threshold (I_L) from a cumulative distribution of percent of times stimuli are detected.

Although one tends to think about upper and lower limits only in reference to stimulus amplitude or intensity, sensitivity ranges may be determined for any physically measurable stimulus attribute. Table 2–3 lists the frequency sensitivity ranges for several stimuli. Similar sensitivity-range data or threshold values are available for size, movement, point and extended light sources, and many other stimulus attributes. Methods for determining threshold values are discussed in Section 2.2.7.

2.2.5 Relative Discrimination Sensitivity (ΔI)

In many operator and maintenance tasks it is necessary to determine whether the intensity, frequency, etc., of Stimulus A is equal to, less than, or greater than Stimulus B. Such tasks, called *relative discrimination* tasks, involve man's ability to detect a small difference between two stimuli, or a change (Δ) in one stimulus. The least change in a stimulus or the least difference between two stimuli that can be detected is called the Just Noticeable Difference (JND). JNDs can be determined for physical stimuli and even for subjective qualities—comfort, pleasantness, pain, etc. Generally speaking, ΔI is a JND in intensity; Δf is a JND in frequency; $\Delta\lambda$ is a JND in wavelengths, and so on. The term ΔI is used here to describe the parameter (JND) as well as to refer to intensity alone.

It has been found that, over the typical useful range of some stimuli, the ΔI corresponding to a JND bears a constant relationship to the reference stimulus magnitude I, i.e., $\Delta I/I=C$. In other words, as the base or reference stimulus increases in magnitude, increasing increments in stimulus magnitude are required to be detected. This relationship is illustrated in Figure 2–3. For example, the intensity of a signal light would have to be increased more for the change in brightness to be noticed if the light were in a brightly illuminated room than if it were in a dimly illuminated one.

For brightness and loudness the ratio $\Delta I/I$ corresponding to a JND remains reasonably constant over much of the useable range of stimulus intensities (Stevens, 1951). Unfortunately this is not generally the case for other stimulus dimensions.

TABLE 2–3. FREQUENCY-SENSITIVITY RANGES OF THE SENSES

Stimulus	Lower Limit	Upper Limit
Color (hue)	300 nm (300 × 10^{-9} m.)	800 nm.
Interrupted white light	Unlimited	50 interruptions/sec. at moderate intensities and duty cycle of 0.5.
Pure tones	20 Hz	20,000 Hz.
Mechanical vibration	Unlimited	10,000 Hz at high intensities.

Adapted from Mowbray and Gebhard (1958).

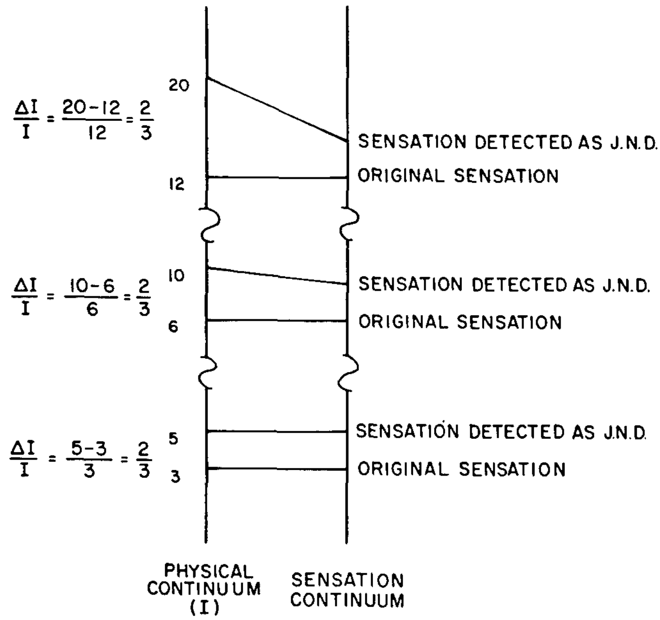


FIGURE 2-3. The relationship between I, ΔI, and Just Noticeably Different (JND) sensation (after Guilford, 1954). From *Psychometric Methods* (2nd ed.) by J. P. Guilford. Copyright 1954 by McGraw-Hill Book Co. Reprinted with permission of McGraw-Hill Book Co.

Tables 2-4 and 2-5 present the number of ΔI's and Δf's within the sensitivity ranges of several sense organs. It will be noted that sense organs differ in sensitivity and that the relative sensitivity of a particular sense organ is different for different stimulus dimensions. Although valuable as a guide in determining or evaluating the characteristics of a proposed display, it must be remembered that tables such as 2-4 and 2-5 are typically based on ideal and static conditions and hence are, for most practical purposes, too "optimistic."

TABLE 2-4. RELATIVE DISCRIMINATION OF PHYSICAL INTENSITIES

Sensation	Number discriminable
Brightness	570 discriminable intensities, white light.
Loudness	325 discriminable intensities, 2,000 Hz.
Vibration	15 discriminable amplitudes in chest region using broad contact vibrator with 0.05-0.5 mm amplitude limits.

After Mowbray and Gebhard (1958).

2.2.6 Absolute Judgments

Identification of a stimulus or its magnitudes must often be made on an absolute rather than a relative basis. This is true when it is either impractical or undesirable to make a reference stimulus available. The pilot who estimates his altitude visually, the infantryman who judges the speed of a tank, the spotter who estimates an

TABLE 2-5. RELATIVE DISCRIMINATION OF FREQUENCY

Sensation	Number discriminable
Hues	128 discriminable hues at medium intensities.
White light	375 discriminable interruption rates between 1-45 interruptions/sec. at moderate intensities and duty cycle of 0.5.
Pure tones	1,800 discriminable tones between 20 Hz and 20,000 Hz at 60-dB loudness.
Interrupted white noise	460 discriminable interruption rates between 1-45 interruptions/sec. at moderate intensities and duty cycle of 0.5.
Mechanical vibration	180 discriminable frequencies between 1 and 320 Hz.

After Mowbray and Gebhard (1958).

artillery location from its sound, are all making "absolute judgments."

One method used to obtain a measure of man's ability to make absolute judgments is to present him with a set of stimuli successively. As each stimulus is presented the observer assigns an absolute rating to the designated stimulus attribute: its specific color, brightness, size, etc. The observer's ratings may be numerical or descriptive. If the task is to rate the brightness of lights, lights are presented in random order and the observer assigns a numerical rating to the brightness of each light. The average and the variability of the ratings are then computed.

Tables 2–6 and 2–7 compare some of the senses in terms of the ability to make absolute discriminations of intensity and frequency, respectively. It will be noted that the number of absolute discriminations that can be made is much smaller than the number of discriminable differences shown in Tables 2–4 and 2–5. This has important implications for coding stimuli for displays. For example, man can discriminate up to 180 differences in vibratory frequency; however, as a means of presenting information on an absolute basis, only 4 to 6 different frequencies should be used. (See also Section 2.3.2 and Chapter 3.)

TABLE 2–6. ABSOLUTE IDENTIFICATION OF INTENSITY

Sensation	Number identifiable
Brightness	3 to 5 discriminable intensities with white light of 0.1–50 millilamberts.
Loudness	3 to 5 discriminable intensities with pure tones.
Vibration	3 to 5 discriminable amplitudes.

After Mowbray and Gebhard (1958).

TABLE 2–7. ABSOLUTE IDENTIFICATION OF FREQUENCY

Sensation	Number identifiable
Hues	12 or 13 discriminable hues.
Interrupted white light	5 or 6 discriminable interruption rates.
Pure tones	4 or 5 discriminable tones.

After Mowbray and Gebhard (1958).

2.2.7 Psychophysical Methods

Data such as those presented in the immediately preceding sections are obtained by the psychophysical methods. In essence, the human is used as a meter to obtain measures of the experienced absolute or relative magnitude of a physical stimulus. Table 2–8 summarizes the characteristics of these psychophysical methods, the statistical indices employed, and the problems to which they are most applicable.

In using any psychophysical method, several precautions must be taken. First, because there is inherent variability in human judgments and reports, repeated measurements on the same subject must be taken. The same stimulus or signal magnitudes must be presented repeatedly, singly, or in combination with other stimuli depending on the method used, to obtain threshold values and the sampling error. Secondly, a sample of observers representative of the population from which the typical operator would be selected must be used. Third, the psychophysical measurement situation must be representative of the task conditions for which the threshold is being determined. If the problem is to determine the faintest target that can be detected on a radar scope, the psychophysical measurement situation must be analogous to the real scope-monitoring task with respect to such parameters as the task environment and duration, and the precision and speed requirements for detection. Detailed treatments of the psychophysical methods and their application may be found in Stevens (1951) and Chapanis (1959).

An alternate to some of the classical psychophysical methods is "Signal Detection Theory" (Tanner and Swets, 1954). Applications of this theory to human engineering problems are discussed in Pew (1965), Cooper (1965), and Chapter 4 of this *Guide*.

2.2.8 Search and Vigilance

If an operator is kept active or is a member of a team, his detection performance will be better than if he is monitoring in an impoverished environment (Poulton, 1966). If the operator has been trained with frequencies of signal occurrence different from those in the operational task, his performance will be poor. If the operator reports only signals that he is very positive of,

TABLE 2–8. METHODS OF PSYCHOPHYSICS

Method	Procedure	Statistical index	Problems to which most applicable
Adjustment (average error)	Observer adjusts signal intensity until it is subjectively equal to, or in some relation to, a criterion.	Average of settings (average error of settings measures operator precision).	Absolute thresholds; equality; equal ratio or interval.
Minimal change (limits)	Signal is varied up and down in magnitude. Observer reports when it meets criterion.	Average value of signal at transition point of observer's judgment.	All thresholds; equality.
Constant stimuli	Comparison stimuli are paired at random with a fixed standard stimulus. Observer reports whether each comparison stimulus is greater or less than standard.	JND equals stimulus distance between 50 and 75 percent points on psychometric function.	All thresholds; equality; equal ratio or interval.
Paired comparison	Stimuli (need not be physical energies) are presented in pairs, exhausting all combinations of stimuli taken two at a time. Observer reports which of each pair is greater in respect to given attribute or criterion.	Proportion of judgments calling one stimulus greater than another. Proportions may be translated into scale values via the assumption of a normal distribution of judgments.	Order on a scale (e.g., of comfort, desirability, etc.). Equal interval.
Rating scale (absolute judgment).	Each of a set of stimuli is given an "absolute" rating or index in terms of some selected attribute. Rating may be numerical or descriptive.	Average or median rating assigned by observers.	Order on a scale. Also used to determine useable codes (e.g., color coding).

After S. S. Stevens (1951). From *Handbook of Experimental Psychology*, by S. S. Stevens (Ed.). Copyright 1951 by John Wiley and Sons. Reprinted with permission of John Wiley and Sons, Inc.

his frequency of correct detections may be reduced. However, as the criterion for reporting is relaxed, the number of incorrect as well as correct detections increases. Other conditions that affect signal detectability are presented in Table 2–9. Conditions that affect the detectability of visual targets are described in Chapter 3, and those that affect auditory vigilance are found in Chapter 4.

2.2.9 Reliability of Sensing

Normally, the human sensing system is reliable, consistent, and precise. The mean-time-between-failures of the human system is much longer than for most electromechanical sensing systems. With respect to consistency, errors may not be attributable to the human sensing system itself, but rather may stem from such non-human sources as the equipment, information flow, task design, and the operating environment. If, for example, display tubes could be made that did not blur or distort character shapes, or did not contain noise or flicker, human sensing would make a more consistent contribution to total system reliability. Principles of man-machine interface design stated in other parts of this *Guide* point the way to achieving an acceptable degree of human sensing reliability.

2.2.10 Design Implications of Sensing-System Parameters

The way in which equipment is designed to present information to the human component must be related to the parameters of the human sensing system. These parameters vary from individual to individual and from sense organ to sense organ. Therefore, it is necessary in many cases to conduct experiments with human subjects to determine functional engineering specifications. Such subjects should be repre-

TABLE 2–9. TASK CONDITIONS AFFECTING SIGNAL DETECTABILITY DURING PROLONGED VIGILANCE

Improved probability of detection		
Simultaneous presentation of signals to dual channels		Buckner & McGrath (1963), Gruber (1964).
Men monitoring display in pairs; members of pairs permitted to speak with one another; 10 minutes rest each 30 minutes of work; random schedule inspection by supervisor.		Bergum & Lehr (1962).
Introduction of artificial signals during vigilance period to which a response is required.		Garvey, Taylor & Newlin (1959), Faulkner (1962).
Introduction of knowledge of results of artificial signals		Baker (1960).
Artificial signals identical to real signals		Wilkinson (1964).
Decreased probability of correct detections		
Introduction of artificial signals for which a response is not required.		Colquhoun (1961).
Excessive or impoverished task load on operator		Poulton (1960).
Introduction of a secondary display monitoring task		Jerison (1963), O'Hanlon & Schmidt (1964), Ware, Baker & Sheldon (1964), Wiener (1964).
Operator reports only signals of which he is sure		Broadbent & Gregory (1963).
Change in probability of detection with time		
A short pretest period followed by infrequently appearing signals during vigilance.	High initial probability of detection, falling off rapidly.	Colquhuon & Baddeley (1964).
Few pretest signals before vigilance period.	Reduces decrement in probability of detection with time.	Colquhuon & Baddeley (1964).
Prolonged continuous vigilance	Decreases probability of correct signal detection.	Mackworth & Taylor (1963).

sentative of the education, training, and physical parameters of the operator population to be assigned to the system. Table 2–10 lists eight key sensing parameters, some of their more important implications for engineering design, and the classes of equipment that will be affected.

Even though a design meets or exceeds the thresholds for detection or for differential sensitivity, it may not be adequate for sensing under adverse operating conditions. A designer may assume, for example, that having attained threshold visibility for the symbols on a radar scope, further increases would be a luxury. While this assumption may be valid under ideal conditions, it is not likely to be in an operational environment where stress, monotony, or boredom are added, and the operator has been assigned extra tasks. It follows that, prior to acceptance, designs should be tested under conditions as nearly like the operational environment and operating workload situation as possible—adhering to work-rest cycles and the other factors that could degrade performance. See Chapter 15 for further information on tests and evaluation.

2.3 Information Processing Subsystem

2.3.1 Amount of Information Transmitted (I_t)

In considering man as an information processing system, two parameters are of interest: the amount of information that man can transmit and the rate with which he can transmit it.

TABLE 2–10. IMPLICATIONS OF SENSING SUBSYSTEMS PARAMETERS ON EQUIPMENT DESIGN

Parameter	Implications of parameter for equipment design	Equipment affected
Detection sensitivity (lower threshold).	Defines minimal intensity and frequency of signals that can be detected by a sense organ.	Alarms, voice, and visual displays.
Detection sensitivity (upper limit).	Defines limit on intensity and frequency beyond which sensitivity is lost and/or damage may occur to sense organ	Alarms, ambient illumination, protective equipment (e.g., goggles, ear protectors), noise suppression.
Differential sensitivity (difference threshold).	Defines intensity or frequency by which: (a) signal A must be increased or decreased for the change to be detected, (b) signals A and B must differ to be detected.	Scope resolution, scale, and pointer design.
Sensitivity range (upper limit minus lower threshold).	Defines maximum "bandwidth" of a physical energy that can be used for signal presentation & display purposes.	Voice communications equipment (headsets, speakers); visual displays (e.g., sonar, radar, photogrammetry, etc.).
Information transmission capacity.	Determines maximum number and type of codes possible within a stimulus dimension. Determines maximum rate of information presentation. Determines maximum rate of operator decision-making.	Map, display board, and scope symbology; coded warning signals; information up date rates; desirability of control dynamics to aid operator response; amount of information presented.
Speed	Determines maximum rate of information presentation, operator response speed, and system response.	Determines information presentation & update rate.
Reliability	Affects overall design, utility, and cost of system.	All man-machine interfaces.
Variability	Information presentation parameter values must be selected on basis of performance of "typical" operators.	All man-machine interfaces.

Amount of information is frequently expressed in terms of "bits"—the logarithm to the base two of the number of equally likely alternatives. The number of bits is equal to the number of two-choice discriminations required to specify a particular event from alternative ones. Unless information is properly distributed between the man and the other parts of a system, the operator may either be overloaded with information that he cannot process, or be unable to process that which is presented rapidly enough.

The effectiveness of a communication system, or of a coding system, can be assessed in terms of the relationship between input and output. The greater the correlation between input and output, the better the system. If the input and the corresponding output can be determined, it is also possible to determine how much of the input is actually transmitted to the output,how much is lost, and how much of the output is due to noise introduced into the channel during transmission.

The amount of information transmitted through a "human channel" can be calculated from a two-way data matrix consisting of S stimulus categories and R response categories. The cells of the matrix (S stimulus by R response categories) contain the frequencies with which a particular stimulus produces a particular response. By adding the contents of the cells by rows and by columns, the number of times that each stimulus was presented (N_k), and the number of times that each response category was used (N_j) can be determined. From these sums and the cell frequencies, the following "probabilities" can be computed (Garner and Hake, 1951; see also Attneave, 1959; Miller, 1956; and Garner, 1962).

$p(j, k) = N_{jk}/N =$ The probability of the joint occurrence of a particular stimulus k and a particular response j. (2–1)

$p(j) = N_j/N =$ The probability of occurrence of each response j. (2–2)

$p(k) = N_k/N =$ The probability of occurrence of each stimulus k. (2–3)

$p_k(j) = N_{jk}/N_k$

$= p(j, k)/p(k) =$ The conditional probability of response j, given stimulus k. (2–4)

$p_j(k) = N_{jk}/N_j$

$= p(j, k)/p(j) =$ The conditional probability of stimulus k having occurred, given response j. (2–5)

From these values, the amount of information transmitted (I_t) may be determined. Accordingly,

$$I_t = I_r - E_r, \quad (2\text{–}6)$$

where I_r is a measure of response uncertainty or information, and

$$I_r = -\sum_{j=1}^{R} p(j) \, log_2 p(j) \quad (2\text{–}7)$$

and E_r is a measure of noise generated within the human channel, and

$$E_r = -\sum_{k=1}^{S} p(k) \sum_{j=1}^{R} p_k(j) \, log_2 p_k(j). \quad (2\text{–}8)$$

In these computations it is assumed that the information, I_r, in the responses is the limiting factor. If, for example, only two response categories were ever used, the amount of information transmitted could never exceed one bit, irrespective of the number of stimulus categories used. If stimulus information is the limiting factor, "k" and "j" in the preceding equations are simply interchanged.

The measures of information transmission are useful for a wide variety of human engineering problems, ranging from selecting an optimum number of coded shapes for knobs, or colors for an array of warning signals, to constructing rating scales for evaluating equipment.

2.3.2 Capacity and Coding for Input

One of the questions most frequently asked of a human engineer is: if a stimulus dimension is to be used to code and display information, into how many distinct steps or levels may it be divided so that each level or step conveys a distinct meaning to an observer?

A similar problem arises in determining the number of categories to be used in rating scales for evaluating equipment. Here the question is: how many judgment categories are actually used and needed by the raters? For example, if an observer uses only the categories "fast" and "slow" to describe the speed of an aircraft, his response will not convey very much information. But can he reliably use more categories; if so, how many? Fundamentally, these problems all resolve into the basic question of how much information is useable to and transmitted by the observer.

A number of investigators have measured the amount of information transmitted by both unidimensional and multidimensional stimuli. Table 2–11 lists the transmission limits in bits for several unidimensional stimuli, frequency, intensity, etc. Table 2–11 also shows the number of stimulus categories that could be used to code information from any given stimulus dimension. For example, only seven different discrete pitches can be used for coding.

Several facts of interest are revealed by Table 2–11. First, the channel capacity of vision is higher than that of other senses. Second, within a given sense, different stimulus dimensions are associated with different capacities for transmission. The greater the range between the upper and lower detection thresholds for a given sensory dimension, the greater the capacity of the sense organ to transmit information along that dimension. Third, the number of absolute judgments that can be made along any one dimension varies considerably from sense to sense and among stimulus dimensions. For example, while the eye can reliably identify only five different brightnesses, it can identify as many as 13 colors. In practical terms, this means that color codes can convey more information than brightness codes.

The data in Table 2–11 may appear to contradict common sense. We know, for example, that

TABLE 2–11. THE CHANNEL CAPACITY OF SENSES FOR DIFFERENT UNIDIMENSIONAL STIMULI

Sense	Stimulus dimension	Channel capacity (bits)	Discriminable categories	Investigator
Vision	Dot position (in space)	3.25	10	Hake & Garner (1951).
	Dot position (in space)	3.2	10	Coonan & Klemmer (in Miller, 1956).
	Size of squares	2.2	5	Eriksen & Hake (1955).
	Dominant wavelength	3.1	9	Eriksen & Hake (1955).
	Luminance	2.3	5	Eriksen & Hake (1955).
	Area	2.6	6	Pollack (in Miller, 1956).
	Line length	2.6–3.0	7–8	Pollack (in Miller, 1956).
	Direction of line inclination.	2.8–3.3	7–11	Pollack (in Miller, 1956).
	Line curvature	1.6–2.2	4–5	Pollack (in Miller, 1956).
Taste	Salt concentrations	1.9	4	Beebe-Center et al. (1955).
Audition	Intensity	2.3	5	Garner (1953).
	Pitch	2.5	7	Pollack (1952, 1953).
Vibration (on chest)	Intensity	2.0	4	Geldard (in Miller, 1956).
	Duration	2.3	5	Geldard (in Miller, 1956).
	Location	2.8	7	Geldard (in Miller, 1956).
Electrical shock (skin).	Intensity	1.7	3	Hawker (1960).
	Durations	1.8	3	Hawker & Warn (1961).

we can identify thousands of words, many makes and models of automobiles, hundreds of faces, and so on. The explanation lies in the multitude of stimulus dimensions used (Miller 1956). Most objects or everyday experiences differ from one another in many ways: spoken words differ in waveform and duration; objects, faces, and automobiles differ in size, color, and shape, and in dozens of other ways. Each dimension conveys some information about the object that is not conveyed by the others. When, for example, an observer judges multidimensional stimuli, the number of absolute discriminations he can make increases (see Table 2–12). In practice, this may mean that the information-carrying capacity of radar or map symbols, of coding schemes for controls, etc., can all be made to transmit more information if multidimensional stimuli are used. Thus, color combined with shape would increase the information contained in a map symbol; shape plus size could accomplish a similar effect in coding control knobs. Although the possibilities of multidimensional coding are attractive, it should be remembered that such codes do not necessarily increase the speed or rate with which information is transmitted through the operator. One should not increase the number of stimulus categories simply because techniques are available to do so. It is frequently of more practical value to use multidimensional stimuli to present redundant information thus increasing the reliability of the information transmitted, rather than the amount.

TABLE 2–12. THE CHANNEL CAPACITY OF SENSES FOR MULTIDIMENSIONAL STIMULI

Stimulus dimension	Channel capacity (bits)	Discrim-inable categories	Investigator
Size, brightness, and hue (varied together).	4.1*	18	Eriksen (1954).
Frequency, intensity, rate of interruption, on-time fraction, total duration, and spatial location.	7.2	150	Pollack & Ficks (1954).
Colors of equal luminance	3.6	13	Halsey & Chapanis (1954).
Loudness and pitch	3.1	9	Pollack (1953).
Position of points in a square (no grid).	4.6	24	Klemmer & Frick (1953).

* Note: The capacity of each dimension separately was approximately 2.7 bits.

2.3.3 Transmission Delays and Reaction Times

In many systems rate of information transmission or response time affects system performance; consequently, the system's effectiveness may be critically dependent on the lags in the human component. Some of the factors that affect human reaction time are:

1. The sense used,
2. The characteristics of the input signal,
3. Signal rate,
4. Whether or not anticipatory information is provided,
5. The response requirements of the task, and
6. Individual differences in age, sex, training, experience, and instructions.

In general, reaction time is shortest with simple, conspicuous, signals, and increases as the number of signals to be attended to increases or as signal intensity decreases.

2.3.4 Upper Limit of Information-Transmission Rate

Several attempts have been made to determine the upper limit of the rate of human information transmission. Unfortunately, there is no single, simple upper limit; rather, it varies from task to task. This is not to say, however, that it is undesirable to determine specific limits for those tasks where the ceilings compromise system performance or man-machine compatibility. Pierce and Karlin (1957), for example, were faced with the practical problem of determining human transmission rate for purposes of comparing it with the 50,000 and 50,000,000 bits/sec. capacities of telephone and television. The task selected was reading, and it was concluded that the upper rate for that task is about 43 bits/sec.

2.3.5 Information Conservation and Information Compression

The discussion of man as an information processor has, so far, dealt primarily with his ability to *conserve* the information contained in the input. However, many tasks require man to act as more than a simple information transmission channel. An operator may combine information from a variety of sources; he may add information from memory or storage to the input information; he may respond selectively to only certain input stimuli. In other words, he is compressing information, rather than conserving it. While much more must be learned about human information compression, there appears to be a linear relationship between the amount of information compression required by a task and amount of time required by an operator to accomplish compression (Posner, 1964).

2.3.6 Pattern Recognition

Compared with electro-optical scanners, man has an unusual ability to recognize patterns. His ability to recognize terrain features, photographic details, etc. cannot be duplicated by modern technology. This ability to recognize patterns supplements man's visual acuity.

A number of conditions affect pattern recognition (Garner, 1962). For example:

1. Pattern discrimination becomes poorer as the number of stimuli to be discriminated increases.
2. As patterns increase in redundancy, they are discriminated one from another more accurately, but more slowly.
3. Patterns that are symmetrical are more accurately identified than patterns that are random or asymmetrical.
4. Patterns that are easily associated with other familiar patterns are more readily identified than novel patterns.

2.3.7 Learning

For many years it was assumed that learning is unique to living organisms. However, in recent years computer scientists have developed systems exhibiting adaptive and learning characteristics. Gibson (1965) defines an *adaptive* system as one whose parameters change to counteract degradation in performance brought about by a change in the system environment. He defines a *learning* system as one that optimizes performance based on the recognition and use of familiar features in a present situation. Several concepts important to the understanding of learning systems, whether they be human or mechanical, are briefly summarized here and covered in detail in Chapter 13.

Any system that learns must recognize familiar patterns in a situation, generalize from one situation to another, and make selective responses that optimize performance.

The process of recognizing familiar features or patterns in a current situation involves matching selected features of the present situation with features that have been experienced in the past. This requires some form of storage of past features. Gibson (1965) suggests two mechanisms by which stored information might be used to recognize patterns. The first is *template* matching; features of sensed objects or events are compared with and matched against previously learned standards or prototype templates. The second mechanism is *property-listing*, in which elements or features of the present situation are compared with a stored list of elements and features. Since there is never a complete identity between a past situation and a current one, matching by either template or property-listing must also involve an ability to generalize from one situation to another. Psychologists call this stimulus generalization.

One fundamental concept in learning is the necessity for information that will inform the learner that a given response was or was not successful. This is sometimes called knowledge-of-results, feedback, or reinforcement. The selective reinforcement of correct responses by feedback or knowledge-of-results strengthens successful or optimizing responses and weakens others.

A graph of the changes in performance of the learner is called a learning curve. Such a curve consists of some measure of performance (speed, accuracy, errors, hits, etc.) plotted against some measure of the amount of practice or training (e.g., trials, days). Learning curves differ as a function of a variety of system and environmental parameters: e.g., with the amount, type, and timing of reinforcement; with the difficulty or discriminability of stimulus conditions; with motivation; and with the distribution in time of stimuli to be reacted to, etc. In comparing training equipments or procedures on the basis of learning curves these and many other factors must, of course, be considered.

All human learning is subject to the effects of transfer of training—that is, the influence of past learning on present learning. When both the situation and the actions involved in the present situation are similar to those of the past, the rate of learning is accelerated. This effect is called *positive* transfer. For example, using a simulator for training assumes that what is learned in the simulator (e.g., piloting an aircraft or making tactical decisions) will carry over to the operational situation being simulated. Furthermore, it is assumed that this carryover will be helpful, not detrimental.

Negative transfer occurs when past learning interferes with new learning. This can occur when the stimuli in the new situation are similar to the stimuli in the old situation, but different responses are required. An example of this type of negative transfer occurred when a particular control in an aircraft operated in one direction for early models, but was reversed in subsequent models. The stimuli were the same, but the required response was different. Negative transfer can also occur when the stimuli in the new situation are different from the stimuli in the old situation, but the responses are the same. The more different the stimuli the more difficult the response generalization. If the two stimuli are quite similar, it is relatively easy to learn to make the same response to either of them, but where the stimuli are different, it is difficult to associate the same response with both stimuli.

Man's capability to learn provides a number of advantages, often permitting equipment designs to be simplified and enabling systems to achieve a generality of functioning impossible without man. At the same time, man's learning capability imposes certain requirements on design. First, a task initially rejected because it is difficult, may, with practice, become one that can be performed easily. This means that tasks must be pretested. Second, particularly if the system is complex, a means for training on the system tasks must be provided by a separate and frequently costly item of equipment, (e.g., a training device or simulator). Third, the designer must consider possible negative transfer effects. By standardizing tasks across systems (e.g., using common display and control configurations on different models of aircraft, vehicles, etc.), and by designing tasks so that they are consistent with previously learned behaviors, the probability of negative transfer effects is minimized.

2.3.8 Decision-Making

A decade ago, one basic task of man in systems was the continuous control of one or more of the parameters of the system's performance. With recent widespread application of computers many tasks are now more nearly discrete than continuous, and are concerned with decisions about system performance and about the environments in which the man-machine system is operating. These decisions are frequently characterized by risk or uncertainty and sometimes by outcomes having serious consequences in terms of human, economic, or social values.

In the design of decision-making systems, one problem is central: how should the system be designed to assist the decision-maker in making effective decisions? This problem raises a number of theoretical and practical questions: What are the criteria for an effective decision? What characteristics of decision-making situations should be considered in evaluating equipment and task design alternatives? How do men behave in making decisions? Answers to these questions have implications for solving practical design problems such as the type and amount of information to be sensed, processed, and displayed; the method by which man-machine decision-making should be conducted; or the strategy that should be used in making decisions.

Decision-making may be performed under certainty, under risk, or under uncertainty. Decision-making under certainty occurs when decision alternatives are known and when each leads invariably to a specific outcome. A soldier makes a decision under certainty when he selects a route that involves the shortest travel time. Decision under risk exists when the alternatives are known but each alternative leads to a set of possible outcomes, each outcome having a known probability of occurrence. Decision-making under uncertainty is that in which the probabilities of specific outcomes or the variety of possible outcomes, are unknown. This latter type of decision-making is characteristic of many tactical and strategic decision-making situations.

The first theories of decision-making were developed by economic theorists. Their "*economic man*" was assumed to be completely rational, objective, and fully informed; and the effectiveness of his decisions could be assessed in terms of maximizing utility. While this theory has been applied in various forms to problems of consumer behavior, a theory of riskless choice that assumes a rational and fully informed decision-maker has limited application in human engineering—the nearest analog appears to be troubleshooting decisions.

Some theories of decision-making under risk suggest that an individual behaves in a manner that will maximize *expected utility.* Theories of this type assume that the alternative choices can be ranked in order of desirability, and that expected utility is behaviorally meaningful in terms of the performance of the decision-maker. Experiments on decision-making under risk show that subjects have different probability (risk) preferences and typically do not act in accord with the objective probabilities. This means that a theory of decision-making based on maximization of utility fails to account fully for the facts of human decision-making behavior for all types of situations.

Research on risky decisions has led to a distinction between objective and subjective probabilities. There is, unfortunately, no simple, linear relationship between the real odds and the odds as perceived by the subjects. An extension of this idea is referred to as *Subjectively Expected Utility,* (SEU). Risks will be taken only if the pay-off is judged to be sufficiently high. If the value to the subject of a "dollar to be gained" is multiplied by his estimate of the probability of success, one obtains the SEU for the given situation. This concept is helpful in accounting for decision-making behavior in a variety of situations where payoffs can be manipulated.

In many real systems, decisions occur sequentially and the information available for later decisions is contingent upon the nature and consequences of earlier decisions. Unfortunately, present formal theories fail to provide a precise descriptive model of how man will behave when confronted with such a dynamic decision-making situation. In such situations the concepts of utility and expected value appear to have less meaning than the concept of man as a statistical hypothesis tester. For example, in command or administrative systems, the decision-maker is less concerned with the consequence and payoff of a single event than with the outcome of a series of sequential events. Several experiments demonstrate that, as an intuitive statistician, man is fairly good in testing the statistical properties of sequences of events. He can, for example, estimate the mean of a series of numbers from a small sample and attach confidence to his judgment.

2.3.9 Design Implications

Pew (1965) indicates that the most fruitful result of recent research on dynamic decision-making has been to question seriously the desirability of forcing a human operator to integrate all the data that relate to a decision and, in addition, to make the final choice of action. This situation is found in many command systems in which an operator must obtain, assimilate, and store data to make choices. It would appear better to relieve the operator of the evaluative function and to assign that role to an independent decision maker:

> For example, if an operator receives the datum that defensive radar has identified an enemy aircraft vectored toward North America, and from experience can assess the likelihood ratio based on this datum between two hypotheses—an enemy attack and friendly visit by enemy—then a computer can utilize this information in combination with similar evaluative reports from other specialists to arrive at the most reasonable conclusion about the true state of the world. (Pew, 1965)

Many of the parameters in decision-making tasks are difficult to identify and measure. However, theories and approaches have been formulated that provide useful guidance for design and suggest parameters of system design that do affect human decision processes.

2.4 Memory Storage Subsystem

While attempts have been made to estimate the capacity of the human memory, such estimates are of more academic than practical interest. Knowing, for example, that man can store the equivalent of 10^8 to 10^{15} bits of information (Geyer and Johnson, 1957) is of less practical interest than knowing the design implications of the more moderate demands that system tasks make on memory, the stability of storage over time, how information should be coded for storage in human memory, how information can best be retrieved from memory, etc.

2.4.1 Theories of Storage

Theories of human memory (Broadbent, 1957; Melton, 1964) suggest that there are two types of memory storage. One type called short-term memory deals with events that have just recently occurred, i.e., within seconds or minutes. Air traffic controllers who quickly store and retrieve information about aircraft in a traffic pattern, or battle staff analysts using rapidly updated displays to estimate characteristics of an order of battle, rely on short-term memory.

The second type of memory is called long-term memory. It involves the integration and recall of information acquired over longer periods of experience, practice, and training. This is typical of the storage and recall of emergency procedures, routes, or operational plans. While a clear functional or physiological distinction between short-term and long-term memory is not presently possible, there are differences in behavior that substantiate their existence.

Long-term memory has many implications for training and for training-equipment design. For example, how should one present information for relatively permanent storage and precise recall later? Short-term memory, on the other hand, is involved in receiving transitory inputs from an operating system, temporarily storing the information, and almost immediately recalling it in order to control equipment or to make operational decisions. Examples of tasks involving short-term memory are: (a) on-the-spot diagnoses of equipment malfunctions, (b) predicting tracks and future positions from an air traffic display, or (c) being told a string of digits, a phone number, or a number to be entered into a calculator that is used in a matter of seconds.

2.4.2 Properties of Short-Term Memory—Serialized Input

When a series of letters, digits, or words is presented to an operator for immediate recall some kind of trace is created in the nervous system. The trace may be like a non-localized coded pulse train rather than an actual change in an element of the nervous system. Whatever the mechanism, the important fact is that the trace decays very rapidly. Even so, short-term recall can be improved by proper task and equipment design. For example, the probability of correct recall increases as the number of items is decreased. As the time between the presentation of an item and its recall is decreased, the probability of correct recall increases. If both the time interval and the number of items are kept constant but the rate of information presentation is increased, the probability of correct recall first increases and then falls off. When the number of items remains constant and the time interval between items is decreased the probability of correct recall increases. When the meaningfulness or familiarity of the items is increased, the probability of correct recall increases. These and similar findings are important to the designer in deciding how much information can be presented to an operator and how to encode it. If, for example, a pilot is told a number to enter into a computer, it should not exceed five to seven digits even under the best of conditions.

2.4.3 Recoding Input Information

Miller (1956) points out that when information is stored by a human, whatever it is that is stored may not be in the same form as the input, nor even in the form of bits; rather the input may be recoded into "chunks." A simple example would be the recoding of the binary number 1100100 to the decimal number 100. Speed and accuracy of recall will increase when the input can readily be recoded into chunks. Thus, the upper limit on the number of input items that can be stored and correctly recalled—about seven digits, letters, or numbers—can be extended if the items can be recoded into smaller chunks. One would not, for example, normally design a display to present numbers in binary form when they could be presented in the shorter and more familiar decimal form.

2.5 Response Subsystem

There is a useful analogy between the human's response subsystem and an electromechanical servomechanism. Common to both is feedback. Feedback permits responses to be made accurately despite unexpected or varying loads and deflections imposed on the output by external forces. Like a servomechanism, the human response subsystem is subject to three kinds of errors: transient, steady-state, and transient oscillatory. The steady-state error is large when output friction is involved, transient errors are large when inertia must be overcome. Oscillatory behavior on starting or stopping a movement occurs if there is output friction, particularly when the required corrections are sensitive.

2.5.1 Fatigue, Boredom, and Learning

Fatigue is a reduction in a muscle group's ability to contract. Accompanying physiological changes are such symptoms as tension, tremor, and pain that adversely affect performance and warn of the fatigued condition. Redesigning or reallocating the task, modifying the working environment, and changing the work-rest cycle are means of preventing or reducing this kind of fatigue and its effects.

Boredom is both difficult to measure and to counteract. Like fatigue, it may degrade performance and increase its variability. Proper design of tasks and of equipment helps postpone the onset of boredom and reduce its severity and duration. In a prolonged vigilance task, for example, inserting false signals or displaying to the operator a record of his performance or productivity may be employed to maintain interest.

Performance improves with learning; response speed tends to increase and response variability tends to decrease. Failure to properly take into account the capability of the operator to learn can result in either under- or over-design of operational and training equipment. To assume that an operator can adapt to virtually any complex situation can result in equipment and tasks so complicated that no feasible amount of practice or training can assure reliable performance. On the other hand, failure to take into account man's ability to learn can lead to automating functions which the human could perform satisfactorily, reliably, and more economically in terms of total system development, production, and operations costs. Occasionally a dilemma arises. What starts out as a novel, interesting, even demanding task, once mastered, can become routine and boring. Again, however knowledge of results helps to sustain interest.

2.5.2 Information in Motor Response

For most practical purposes, the rate with which information can be transmitted by the response system is assumed to range from less than 3 bits per sec. to a maximum of about 9 bits per sec. depending on the task, the organization of the information, and the readout elements involved. For verbal responses a maximum rate of approximately 7.9 bits per sec. has been suggested, while for keypressing the maximum rate is about 2.8 bits per sec. (Alluisi, Muller, and Fitts, 1957; Fitts, 1954). Such rates, however, are obtained only with considerable practice; they cannot be expected in operational situations, nor for prolonged periods.

2.5.3 Control Dynamics

In many contemporary systems man is a critical element in the control system. In steering or guidance, for example, he must continuously detect and correct errors. To assess the stability and control characteristics of the total man-machine system, the designer needs a describing or transfer function, a manipulatable mathematical equation, for the man just as he needs one for the other components of the system. See Chapter 6.

2.6 Human Constraints on Design

Although man is an extremely versatile functional component, his use in a system imposes problems and constraints on the designer. These problems and constraints result from the physical and behavioral variations among men, the structural and functional limitations of the body, the requirements for man's safety and comfort, and the need to maintain his physiological functions.

2.6.1 Human Variability

Human beings differ in strength, weight, bodily dimensions, and even in body composition. They differ with respect to sensory abilities; speed, accuracy, and timing of responses; ability to learn and adapt; capacity to withstand fatigue and stress; and in many other ways. Even the same man's performance capabilities vary from moment to moment, from one time of the day to another, and from day to day. All these variations, at least the possibility of such variations, must be considered in designing equipment, environments, and tasks.

Structural differences among people, e.g., arm length, hip breadth, hand width, etc., have obvious relevance to design. Similarly, differences in visual acuity, ability to track moving targets, memory, speed of learning, etc., interact with equipment and task design. In some instances variations in human characteristics tend to be related. There are, for example, correlations among body weight, height, and strength. In other instances, different characteristics are virtually unrelated. Neither sensory ability nor memory appears to vary appreciably with physical size or strength. Correlations among man's physical characteristics are considered in the design and sizing of personal equipment and work areas. However, the correlations, if any, between man's physical dimensions and his abilities are not generally considered in system design.

2.6.2 Describing Variability

The "normal" distribution. Frequency plots of human structural and performance characteristics often approximate, or are assumed to approximate, the normal, bell-shaped, Gaussian distribution. (See Figure 2–4.)

The most common measure of the central tendency of such a distribution is the arithmetic mean (M), i.e., $\frac{1}{N}\sum_{1}^{N} X$, where X is each measurement, and N is the number of cases. The most precise measurement of the variability of a normal distribution is the standard deviation σ; i.e., $\sqrt{\sum_{1}^{N} x^2 \Big/ N}$, where x is the deviation of each measurement from the mean (i.e., $x = X - M$). In other words σ (sigma) is the root-mean-square (r.m.s.) deviation from the mean. Sixty-eight percent of the measures in a normal distribution lie between 1σ below the mean and 1σ above the mean. Although both M and σ can be computed when the distribution is not normal, it is difficult to interpret them precisely. Detailed discussions of measures of central tendency and variability, and of their interpretation, can be found in an elementary statistics text.

Cumulative distributions. To obtain design values, particularly those based on anthropometric dimensions, cumulative frequency distributions are commonly used. A cumulative distribution shows the number (or proportion) of people whose measurements fall at and below given values.

A cumulative distribution does not require that the basic measurements be normally distributed. However, if the cumulative distribution

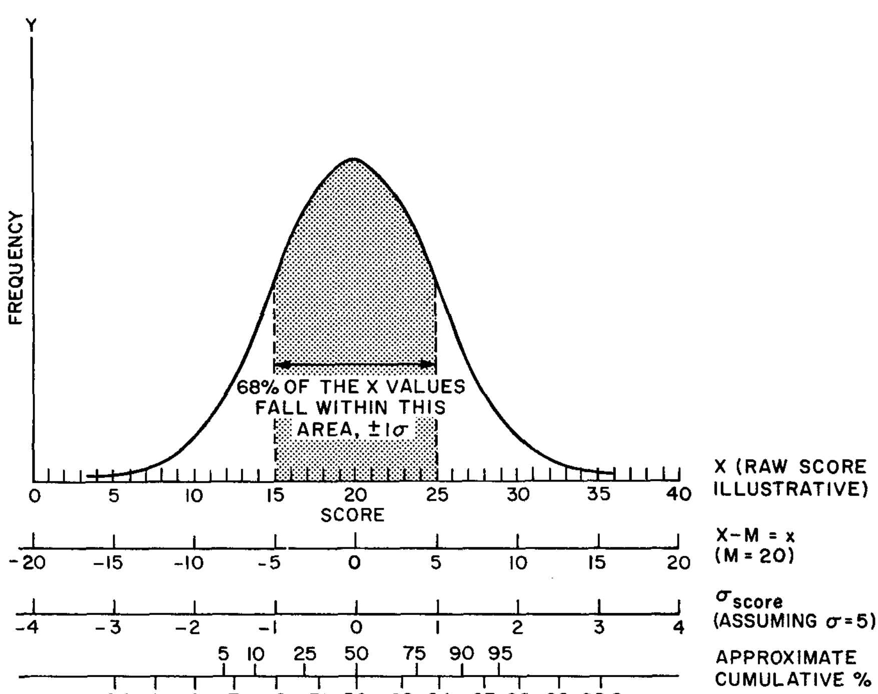


FIGURE 2–4. Normal bell-shaped distribution. Many sets of measurements of human characteristics approximate this distribution. Here X is a measure of a characteristic, and Y is the relative frequency of different values of X.

is based on normally distributed measurements, half the measurements will fall at or below the mean. From a cumulative distribution one can also readily obtain "percentile values"—the percent of people at or below (or above) a particular measurement.

In designing to accommodate man's physical dimensions, it is common practice to use the 5th and 95th percentiles of the population of users. Although cumulative distributions are most frequently used to present anthropometric information, such distributions may also be used to present other kinds of data (see, for example, Figure 2–2).

Other distributions. Not all human characteristics or measures of human performance are normally distributed. Some distributions are too peaked, some too flat; others are skewed so that the peak of the curve is off-center. In these cases either the arithmetic mean (M) or the standard deviation (σ) or both can be misleading. If the distribution is skewed, for example, the mean does not correspond to the 50th percentile.

In some cases the distribution of human performance measures does not even approximate a normal curve. Measures of social behavior, for example, often yield so-called J-curves. To illustrate, if one notes the arrival times of people required to be at work by 8:00, the number of people arriving during successive time intervals may increase more and more rapidly to a maximum at about 7:55, and then fall off abruptly. A J-curve is also obtained when one plots the frequencies of various degrees of compliance with regulations, warnings, stop signs, etc.

In a few instances, human performance measures are distributed rectangularly. For example, the frequency with which automobile accidents occur along a stretch of turnpike has been found to be approximately rectangular, the frequency of accidents being approximately the same at all points in the region of observation.

Another useful theoretical distribution is the Poisson. It is typically used in describing or analyzing the frequencies with which a relatively rare event occurs during an interval of time or space. For example, when the number of people having 0, 1, 2, 3 etc. accidents in a certain time period (e.g., a year) is plotted, a Poisson-like distribution is typically obtained.

Although the shapes of these various distributions may affect equipment design only indirectly, they are of considerable importance in systems analyses. One fundamental aspect of such analyses is illustrated in the next section.

2.6.3 The Variability of Errors

In the measurement of man-machine system performance, two general types of error are identified: variable and constant (Figure 2–5). Variable errors, as the term implies, differ from one measurement to the next. Constant errors are those that are consistent in both direction and amount. Rifle shots that are tightly clustered to the right of the bull's eye reflect a constant error but little variable error.

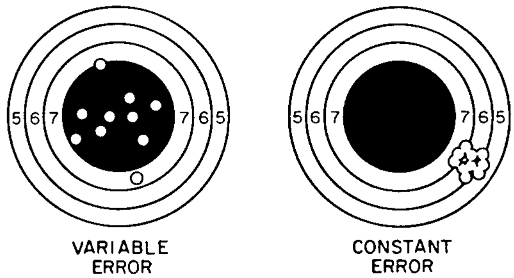


FIGURE 2–5. Variable and constant errors in the dispersion of hits on a target.

Variable errors are often attributable to a combination of unpredictable factors. For this reason they are sometimes called random errors. One source of variable error is the human. Fortunately, human variable errors tend to decrease with practice.

A constant error is usually attributable to a single systematic source, either in the man or in the system with which he is working. A misaligned rifle sight would, for example, result in a constant error. Were constant errors not confounded with, and obscured by, variable errors, they would be relatively easy to identify and eliminate, or at least reduce, through design or corrective training. Unfortunately the two types of error usually occur together and it is necessary to collect and analyze considerable data to reliably quantify a constant error and identify its source. If the distribution of the total error is normal, then the mean of a sample of the errors is an estimate of the constant error and the standard deviation (σ) is an estimate of the variable error.

Constant errors from two or more sources can be added algebraically or vectorially. Variable errors add according to the squares of their standard deviations. Thus, the accumulation of two variable errors is

$$\sigma^2_{a+b} = \sigma^2_a + \sigma^2_b \qquad (2\text{–}9)$$

where σ_{a+b} is the variable error of the system, σ_a is the variable error of the first source and σ_b is that of the second. This equation must be modified to include a correlational term when the errors are related such that those from one source vary with those of the other source.

In man-machine systems, when one variable error is appreciably larger than the others, the source of the larger error should be attacked first. For example, if the σ of the error of plotting range is 100 yd. and that for reading from the plot is 300 yd., a reduction of 50 yd. in the σ of the *plotting* error, the smaller error source, will have a negligible effect on the total system error. (Using the equation above, $100^2+300^2=316^2$, and $50^2+300^2=304^2$, an improvement of only 12 yd.) In contrast, a reduction of 50 yd. in the σ of the *reading* error, the larger error source, decreases system error by almost the full 50 yd. (Using the above equation, $100^2+250^2=269^2$, an improvement of 47 yd.) Cutting the larger error in half, to 150, reduces system error to 180 yd.; but cutting the smaller error in half only decreases system error to 304 yd.

In "go-no-go" situations, when errors are either "all or none," in or out of tolerance, etc., the probability of no error is computed by simply multiplying the separate probabilities of no error, assuming that they are not correlated.

$$1-P_{eT} = (1-P_{e1})(1-P_{e2}) \qquad (2\text{–}10)$$

where P_{eT} is the probability of an error in one or both sources, P_{e1} is the probability of error in one source (man or machine), and P_{e2} is the probability of error in the second source. The same formula can be expanded to represent several sources of error arranged in a series.

2.6.4 Anthropometric Constraints

As a physical component of systems, man has weight and occupies space. These physical dimensions impose constraints on the design of work areas, seats, desks, console tops, and on the size of doorways, escape hatches, and maintenance accesses. See Chapter 10.

Man's limits—the force he can apply, the speed and types of movements he can make in pushing, pulling, and turning objects with his hands and feet—must be considered in designing controls, tools, and other work objects. Detailed discussions will be found in Chapters 8 and 11.

As a functional, responding component, man also imposes constraints on design in terms of his requirements for seeing, reaching, and moving. These constraints affect the arrangement of controls and displays, their distance from the operator, and their configuration and location with respect to the movement demands of tasks. Such constraints and requirements are described in Chapters 3, 9, 10, and 11.

2.6.5 Requirements for Survival and Physiological Functioning

Man imposes the requirement that the environment in which he is to function be one that insures survival and supports normal physiological processes. In normal terrestrial environments, requirements for adequate ventilation, temperature and noise control, the control of toxic products, provisions for food and for waste elimination, etc., although important, are not unique or difficult to meet. However, at high altitude, under water, in space, or in other exotic and military environments, design for life support becomes extremely critical. Although such life-support considerations obviously affect system design, they are not discussed in detail in this *Guide* except as they have direct implications for performance, and for the design of operator equipment.

References

Alluisi, E. A., Muller, P. F., Jr., & Fitts, P. M. An information analysis of verbal and motor responses in a forced-paced serial task. *Journal of Experimental Psychology*, 1957, **53**, 153–158.

Attneave, F. *Applications of information theory to psychology*. New York:Holt, Rinehard and Winston, 1959.

Baker, C. H. Observing behavior in a vigilance task. *Science,* 1960, **132,** 674–675.

Beebe-Center, J. G., Rogers, M. S., & O'Connell, D. N. Transmission of information about sucrose and saline solutions through the sense of taste. *Journal of Psychology,* 1955, **39,** 157–160.

Bergum, B. O., & Lehr, D. J. Vigilance performance as a function of task and environmental variables. Report No. 11, 1962, U.S. Army Air Defense Human Research Unit, Fort Bliss, Tex.

Broadbent, D. E. A mechanical model for human attention and immediate memory. *Psychological Review,* 1957, **64,** 205–215.

Broadbent, D. E., & Gregory, M. Division of attention and the decision theory of signal detection. *Proceedings of the Royal Society* (London), 1963, **158**(b), 222–231.

Buckner, D. N., & McGrath, J. J. (Eds.), *Vigilance: A Symposium.* New York:McGraw, Hill, 1963.

Chapanis, A. *Research techniques in human engineering.* Baltimore:The Johns Hopkins Press, 1959.

Chapanis, A. Engineering psychology. In P. R. Farnsworth (Ed.), *Annual Review of Psychology,* **14,** Palo Alto:Annual Reviews, Inc., 1963.

Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology.* New York: Wiley & Sons, 1949.

Colquhoun, W. P. The effect of "unwanted" signals on performance in a vigilance task. *Ergonomics,* 1961, **4,** 41–51.

Colquhoun, W. P., & Baddeley, A. D. Role of pretest expectancy in vigilance decrement. *Journal of Experimental Psychology,* 1964, **68,** 156–160.

Cooper, G. R. Decision theory. In R. E. Machol (Ed.), *System engineering handbook.* New York: McGraw-Hill Book Co., 1965.

Eriksen, C. W. Multidimensional stimulus differences and accuracy of discrimination. Technical Report No. WADC 54–165, 1954, Wright Air Development Center, WPAFB, Ohio.

Eriksen, C. W., & Hake, H. W. Absolute judgments as a function of the stimulus range and the number of stimulus and response categories. *Journal of Experimental Psychology,* 1955, **49,** 323–332.

Faulkner, T. W. Variability of performance in a vigilance task. *Journal of Applied Psychology,* 1962, **46,** 325–328.

Fitts, P. M. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology,* 1954, **47,** 381–391.

Garner, W. R. An informational analysis of absolute judgments of loudness. *Journal of Experimental Psychology,* 1953, **46,** 373–380.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York:John Wiley & Sons, 1962.

Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychological Review,* 1951, **58,** 446–459.

Garvey, W, D., Taylor, F. V., & Newlin, E. P. The use of "artificial signals" to enhance monitoring performance. Report No. 5269, 1959, U.S. Naval Research Lab., Washington, D.C.

Geyer, B. H., & Johnson, C. W. Memory in man and machines. *General Electric Review,* 1957, **60,** 29–33.

Gibson, J. E. Adaptive and learning control systems. In R. E. Machol (Ed.), *System Engineering Handbook.* New York:McGraw-Hill, 1965.

Guilford, J. P. *Psychometric methods.* (2nd ed.) New York:McGraw-Hill, 1954.

Gruber, A. Sensory alternation and performance in a vigilance task. *Human Factors,* 1964, **6,** 3–12.

Hake, H. W., & Garner, W. R. The effect of presenting various numbers of discrete steps on scale reading accuracy. *Journal of Experimental Psychology,* 1951, **42,** 358–366.

Halsey, R. M., & Chapanis, A. Chromaticity-confusion contours in a complex viewing situation. *Journal of the Optical Society of America*, 1954, **44**, 442–454.

Hawker, G. R. Cutaneous communication: Absolute identification of electrical intensity level. *Journal of Psychology*, 1960, **49**, 203–212.

Hawker, G. R., & Warn, J. S. Maximum It for absolute identification of cutaneous electrical intensity level. *Journal of Psychology*, 1961, **49**, 279–288.

Jerison, H. J. On the decrement function in human vigilance. In D. N. Buckner & J. J. McGrath (Eds.), *Vigilance: A Symposium*. New York:McGraw-Hill, 1963.

Klemmer, E. T., & Frick, F. C. Assimilation of information from dot and matrix patterns. *Journal of Experimental Psychology*, 1953, **45**, 15–19.

Mackworth, J. F., & Taylor, M. M. The d′ measure of signal detectability in vigilance-like situations. *Canadian Journal of Psychology*, 1963, **17**, 302–325.

McCormick, E. J. *Human factors engineering* (2nd ed.). New York:McGraw-Hill, 1964.

Melton, A. W. (Ed.) Categories of human learning. In A. W. Melton (Ed.), *Symposium on the psychology of human learning, University of Michigan*, **1962**. New York: Academic Press, 1964.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 1956, **63**:

Mowbray, G. H., & Gebhard, J. W. Man's senses as informational channels. Report No. 936, May 1958, Applied Physics Lab., Johns Hopkins University, Silver Spring, Maryland. (Reprinted in H. W. Sinaiko, (Ed.) *Selected papers on human factors in the design and use of control systems*. New York:Dover Publications, Inc., 1961.)

O'Hanlon, J., Jr., & Schmidt, E. A. The effect on the level of vigilance of an adjacent secondary vigilance task. Report No. 750–4, 1964, Human Factors Research, Inc., Los Angeles.

Pew, R. E. Human information processing concepts for system engineers. In R. E. Machol (Ed.) *System engineering handbook*, New York: McGraw-Hill, 1965.

Pierce, J. R., & Karlin, J. E. Reading rates and the information rate of a human channel. *The Bell System Technical Journal*, 1957, **36**, 497–516.

Pollack, I. The information of elementary auditory displays, I. *Journal of the Acoustical Society of America*, 1952, **24**, 745–749.

Pollack, I. The information of elementary auditory displays, II. *Journal of the Acoustical Society of America*, 1953, **25**, 765–769.

Pollack, I., & Ficks, L. Information of elementary multi-dimensional auditory displays. *Journal of the Acoustical Society of America*, 1954, **26**, 155–158.

Posner, M. I. Information reduction in the analysis of sequential tasks. *Psychological Review*, 1964, **71**(b), 491–504.

Poulton, E. C. The optimal perceptual load in a paced auditory inspection task. *British Journal of Psychology*, 1960, **51**, 127–139.

Poulton, E. C. Engineering psychology. In P. R. Farnsworth (Ed.), *Annual Review of Psychology*, 1966, **17**.

Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: John Wiley & Sons, 1951.

Tanner, W. P., & Swets, J. A. A decision-making theory of visual detection. *Psychological Review*, 1954, **61**, 401–409.

Ware, J. R., Baker, R. A., & Sheldon, R. W. Effect of increasing signal load on detection performance in a vigilance task. *Perceptual and Motor Skills*, 1964, **18**, 105–106.

Welford, A. T. The measurement of sensory-motor performance: Survey and appraisal of twelve years progress. *Ergonomics*, 1960, **3**, 189–230.

Wiener, E. L. Multiple channel monitoring. *Ergonomics*, 1964, **7**, 453–460.

Wilkinson, R. T. Artificial "signals" as an aid to an inspection task. *Ergonomics*, 1964, **7**, 63–72.

Chapter 3

Visual Presentation of Information

Walter F. Grether

Aerospace Medical Research Laboratory
Department of the Air Force
Dayton, Ohio

Charles A. Baker

National Highway Traffic Safety Administration
Washington, D.C.

How visual displays are adapted to human use is considered in this chapter. The purposes of the displays and the conditions and methods of their presentation are discussed first, followed by a section defining measurement terms, including visual acuity in object detection and identification. A delineation of proper workplace illumination is presented next, then sections covering such subjects as visual coding techniques through use of color and shape, design recommendations for warning coding techniques through use of color and shape, design recommendations for warning and signal devices, and for mechanical indicators. Electronic displays, printed information, and projection of images complete the range of material discussed.

CONTENTS

This chapter is based in part on Chapter 2 of the previous *Guide*. It was reviewed by Norman H. Mackworth. Sections 3.2, 3.3, 3.5, 3.6, 3.8, and 3.9 were written by Dr. Grether, Sections 3.1, 3.4, and 3.7 were written by Mr. Baker.

3. Visual Presentation of Information

3.1 Visual Presentation of Information

Because an operator often needs more visual information than his unaided senses can gather and present to him, other sources of information, such as compasses, radarscopes, speedometers, etc., have been created as adjuncts to the human senses. This chapter is a guide to the design and use of these devices.

3.1.1 General Display Considerations

There is much more to designing a good visual display than merely making it visible. The displayed information must be understandable so that correct decisions or control actions can be made without unacceptable delay. This means that a display must be designed to suit the particular conditions under which it will be used.

Conditions of use. To provide the proper display, the designer should consider at least the following conditions:

1. Viewing Distance. The maximum expected viewing distance should influence the size of details shown on a display. The usual reading distance for printed materials is about 16 in. Many indicators are designed for reading at arm's length, to permit the operator to reach switches or adjust knobs on indicators. This limit is generally set at 28 in., and is used in determining the recommended dimensions for scale markings and numerals. For other reading distances, these dimensions must be adjusted up or down in direct proportion to the distances.

2. Illumination. The size of display details should be suited to the lowest expected illumination level. The color of the illumination also should be considered. Occasionally red illumination is used to preserve the operator's dark adaptation, but red light limits the use of some types of color coding on displays. Warning signal lights, cathode-ray tubes, and other displays which emit rather than reflect light are usually hindered rather than helped by other lighting in the work station. When using general illumination for displays, careful design is required to minimize glare sources on the cover glass.

3. Angle of View. The preferred viewing angle is usually 90° to the plane of the display. On large display panels, or where more than one operator views the same display, there might be a considerable deviation from the 90° angle of view. This situation can give rise to excessive parallax distortion or cause parts of the display to be hidden, unless such offset viewing has been compensated for in design.

4. Presence of Other Displays. Usually, an operator divides his attention among several displays. Inconsistencies in the manner of presentation among the displays can confuse or slow up the operator. However, displays should not look too much alike because the operator might read the wrong one. Obviously, each display must be well identified.

5. Compatibility With Related Controls. The design of a display may be affected by the controls associated with it. Ideally, displays and controls should be designed and located so that the untrained operator will select the correct control and operate it as expected. That is to say, the display should imply the form of the operator's response.

Method of use. The way in which the operator will use the information presented to him is an important consideration in the design of displays. An analysis should be made, therefore, of the type of action the operator will be expected to take while receiving or after he receives information from the display.

3.1.2 The Function of the Display

Display function determines not only the information to be shown but the methods of presentation. Various purposes for displays are:

1. Continuous system control.
2. System status monitoring.
3. Briefing.
4. Search and identification.
5. Decision-making.

Displays for continuous system control. Continuous system control involving human inputs has been referred to as "tracking" by life scientists and as "man-machine dynamics" by engineers and physical scientists. The studies conducted on continuous system control have emphasized closed-loop systems in which the effects of the control output on the systems are fed back to the displays. The block diagram in Figure 3–1 depicts such a system.

Frequently, these displays present continuous functions which require the operator to make continuous inputs (tracking). The steering wheel inputs to an automobile, ship, or submarine, and the attitude control inputs to an aircraft are examples of continuous control closed-loop systems. In driving a car, the visual feedback loop is primarily the real world as seen through the windshield and rear-view mirror. Supplemental information, such as speed, is displayed by instruments. In all-weather flight or submarine operation, most parameters of the real world are sampled, sensed, and displayed by instruments to provide man with the information he needs for system control. The following factors should be considered in selecting displays for continuous tracking tasks.

1. The direction of motion of the displayed parameter should be compatible with the motion relationship of (a) the system parameter and (b) the control.

2. The information updating frequency and accuracy of the displayed parameter should be compatible with system control requirements.

3. The scale factor or resolution of the displayed parameter must exceed control accuracy requirements by at least a factor of two.

4. In higher order control systems, derivative terms or computed predicted values should be used to "quicken" the displayed parameter to unburden the operator from making complex decisions in predicting system dynamics.

Displays for system status monitoring. There may be many system parameters having a low frequency of change; yet, the operator must be made aware of changes when they do occur. The engine heat indication on a car, or step load changes in a power station are examples of low-frequency information inputs which can be critical to system operation. Man, being a poor monitor, reacts unreliably to low-frequency events; thus the display of such information requires special attention. In selecting displays for system status monitoring, consider that:

1. Changes critical to safety or effectiveness require a place of high priority in the operator's central field of view. Stimuli demanding high attention should be used for these indications.

2. When different sources of information are displayed, the arrangements or combinations should be in functional or sequential groups. Such techniques as pointer alignment allow for rapid and accurate monitoring.

3. In certain applications displays should provide information on recommended operator procedures as well as the status of the system. This is particularly applicable to checkout procedures.

4. When many subsystems require monitoring, it may be necessary to use machine filtering of information to avoid overloading the operator or to save panel display space. "Auto-scanning" systems can be programmed to display only those information sources that are changing or fall outside some tolerance threshold.

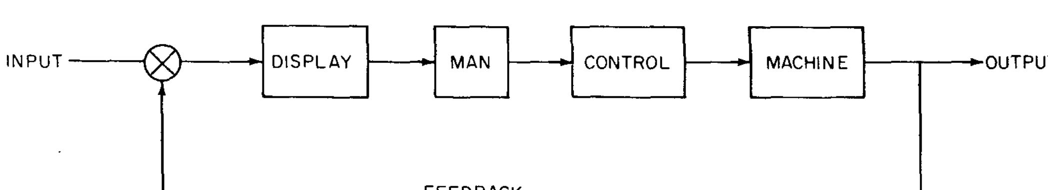


FIGURE 3–1. Typical block diagram of a closed-loop system. Such systems show continuous functions which require tracking by the operator.

Displays for briefing. Displays for briefing purposes provide for long-term storage of information. A critical factor in presentation is a format that enhances memory and recall. There are many applications for such displays: briefing of pilots in flight planning, training of maintenance personnel to recall checkout sequences, and briefing operators who must find target complexes in unfamiliar terrain. While many principles of learning apply to long-term memory, the basic one is to insure that the data content can be verbalized as much as possible. To achieve this end, the following factors must be considered:

1. In briefing for target forms or patterns, the salient distinguishing characteristics should be associated with the verbal names of familiar objects having a similar appearance. For example, if a target complex looks like a horseshoe, this feature should be emphasized on the display. (See Attneave, 1954 and 1957; Foster, 1964.)
2. In using graphic or pictorial codes, select symbols which have high population stereotypes, i.e., those shapes which nearly everyone agrees convey the information symbolized. (See Howell and Fuchs, 1961; and Mackworth, 1966.)
3. When certain operations require that a number of tasks be performed in sequence, one memory aid is the use of acronyms or words whose letters cue the operator to the correct order. For example, the words BEST WAY cue the following procedure: *B*allistic correction, *E*levation error, *S*lew correction, *T*ime correction, *W*ind insert, *A*ttitude correction, and *Y*aw correction. Also, if a sequence of tasks is to be performed on a complex panel, it can be more easily recalled if the pattern of the operations can be described.

Displays for search and identification. Search and identification displays are frequently applied in reconnaissance and surveillance tasks. These are often photographic, electro-optical, or radar displays which provide point-to-point correspondence with the sensed features of the real world. However, when the reconnaissance deals with locating and identifying sources emitting electromagnetic, infrared, or other forms of energy, then symbolic, graphic, or numeric displays may be used. Man's perceptual or pattern recognition skills are ideally suited for processing information from pictorial reconnaissance displays because: (a) he is able to perform in adverse signal to noise conditions, (b) he is relatively unaffected by the changing aspect angles of the sensors, and (c) he uses inference based on the context in which the target signal is found. Nevertheless, when man is required to perform target search and identification in real time, the amount of information to be processed per unit time can exceed his capabilities. The following factors need to be considered in designing pictorial display systems for search and identification:

1. If a target must be identified by its shape characteristics alone, it must have adequate contrast, resolution, and angular size.
2. If a target is to be identified by its size, intensity, or color characteristics, these must be properly scaled.
3. If a target is to be located on a display that is cluttered with other forms of similar size, shape, or color, long search times must be allowed for identification. Display enhancement and filter techniques will improve search in cluttered fields.
4. If target signals are near visual threshold in luminance, the dark adaptation of the operator's eye must be maintained at proper levels.
5. A denotable target position, such as range and bearing, or X–Y coordinates, must be easily derived by the operator.

Displays for decision-making. When displays are used for decision-making, the operator will decide the meaning of information presented to him instead of performing a preplanned response to a specified signal. Types of tasks which involve interpretation of information for decision-making include (a) trouble-shooting malfunctioning equipment, (b) estimating enemy intentions, and (c) determining, from sensed information (such as noisy sonar displays), the presence or absence of targets.

3.1.3 The Combination and Integration of Displays

Frequently, more than one indication can be presented by a single display, offering economies of panel space and eye movement as well as the advantage of simplification. The number that can profitably be combined is limited, however.

1. If indications are compressed within a small area, particularly if they are superimposed, identification of the desired information can be more difficult.

2. Some techniques of instrument combination involving mirrors or optical projection (e.g., stereoscopic displays) restrict the eye position of the operator, reduce his overall mobility, and require artificial lighting—even in daylight.

Combining related kinds of information. The above-listed disadvantages notwithstanding, there have been some desirable and successful combinations. In the indicator shown in Figure 3–2, two radio-compass needles are superimposed over a magnetically slaved, rotating card. The zero position on the card always points toward North. The fixed marker at the top of the card indicates aircraft heading. Bearing to a station can be read directly under the radio-compass needle tuned to that station. From the angles formed by the two radio-compass needles, the pilot can estimate his ground position and course in relation to the two ground stations. Thus, this one display shows magnetic direction, aircraft heading, and bearings to two radio stations.

Displaying derivatives. In higher order systems, it is frequently possible to display derivative terms which can assist the operator in system control. An example of such a display is depicted in Figure 3–3, a plan-position display for instrument landing of a helicopter. The cross hairs represent the desired landing location, the aircraft symbol represents the analog position over the ground, and the circle represents derivative terms of position error. The pilot's task is to maneuver the helicopter over the landing site. As this is being accomplished, the cross hairs will move to the aircraft symbol. To accomplish this task, the pilot puts in a control response to cause an attitude change in the vehicle (pitch and roll), which, in turn, results in an X and Y translation acceleration, which is a derivative of vehicle velocity. In time, this translational velocity becomes a position change. It is evident that the pilot's direct control of aircraft attitude is several time integrations removed from his displayed error, i.e., position error.

The pilot's task can be greatly simplified if some, or all, of these derivative terms are pre-

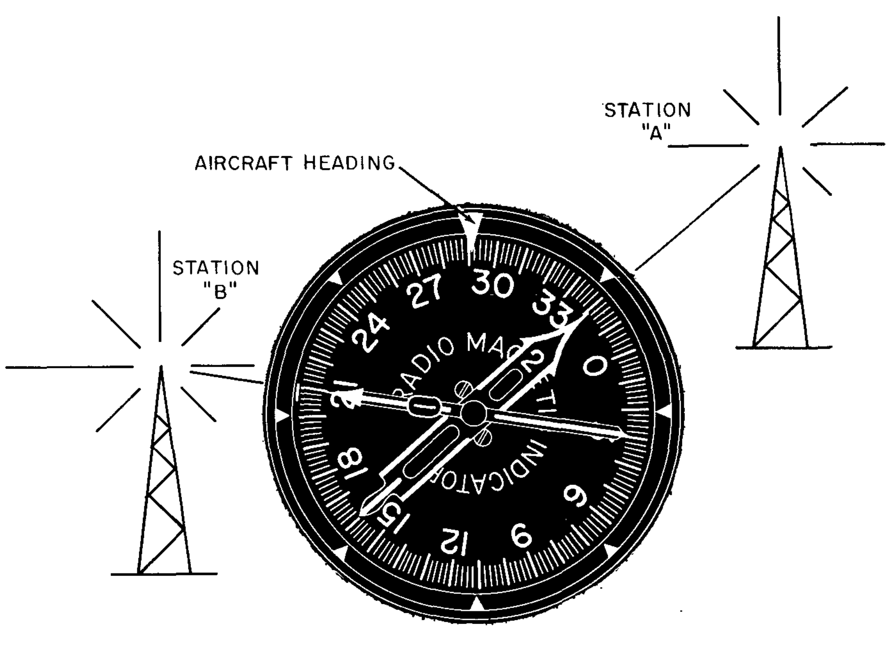


FIGURE 3–2. An example of a display which combines related information. Aircraft headings and magnetic bearings to each of two radio stations are presented simultaneously.

sented to him. In Figure 3–3, the circle can be used to present an analog of such derivative information as the vehicle's X and Y velocities, the vehicle's X and Y accelerations, or the vehicle's attitude rates. The position error is always evident to the operator, and the desired correction is to maintain the position of the "quickened" circle over the cross hairs. The operator is provided "quickened" information about his future position by the location of the circle. By using these position derivatives, the display responds to control inputs with shorter lags than those with which the vehicle changes in position. This display technique improves operator performance. However, the gain factors of these derivative terms must be optimized for the dynamics of each system involved.

Summing derivatives and position error. In some cases it is possible to present a computed error value as an optimum combination of several separate system variables. The aircraft Zero Reader shown in Figure 3–4 embodies such a combination. In this indicator, the vertical line presents a combination of positions relative to the localizer path, relative heading (first derivative), and angle of bank (second derivative); and the horizontal line presents a combination of positions relative to the glide path and aircraft pitch angle (first derivative).

By keeping the two lines of the Zero Reader centered (zeroed) the pilot flies the aircraft so as to make an asymptotic approach to the desired localizer and glide path. The localizer line can be centered by controlling the angle of bank, and

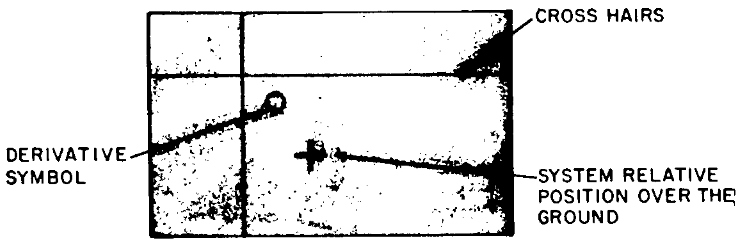


FIGURE 3–3. Derivative symbol in two axes showing where system derivatives are vectoring the vehicle. Cross hairs indicate desired position over the ground in two axes, X and Y.

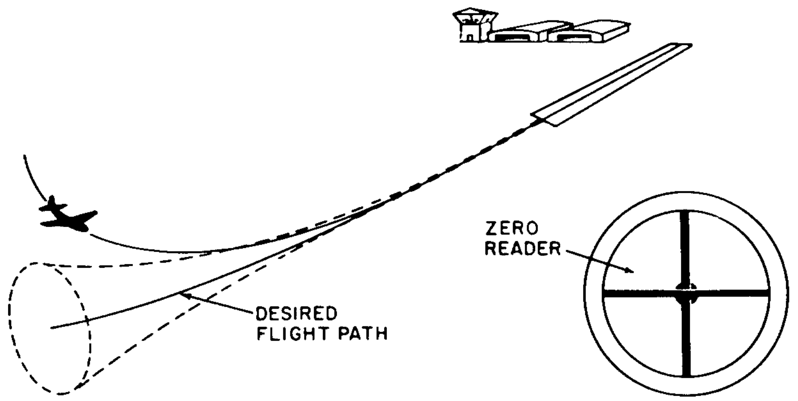


FIGURE 3–4. Position error summed with derivative terms. Although the aircraft is not on the beam center, the display indicates that a proper (pre-computed) approach to the beam center is being accomplished.

the glide-path line can be centered by controlling the pitch angle. Both of these are rather easily controlled, and involve minimum time lags which simplify the pilot's task when making an instrument landing.

In summary, a combination of system variables is particularly suitable where the first and higher-order derivatives or their dynamic equivalents can be combined usefully with the basic value the operator is trying to reach or maintain. The number of derivatives to be combined, and the ratios to be used in the combination, must be determined for each application because different types of equipment, or different kinds of operations, might require different ratios.

A disadvantage of this kind of combination is that individual values are lost unless they also are presented as separate indications. If he keeps the lines zeroed, the operator will achieve the result he desires, but he will not be shown the details of accomplishment. In the case of the Zero Reader, he will not know the exact flight path, which might be required for clearance of ground obstacles. A predictor display which presents the computed time course behavior of the system can provide useful flight path trends.

Predictor display. Figure 3–5 shows a predictor display concept in which the cross hairs remain fixed and the moving element represents the system. The future time course behavior of the system in X and Y coordinates is coded by concentric circles. The largest circle is in real time and therefore is position error. The terminal circle (a dot) represents where the system will be at some arbitrary time in the future if the present control inputs are programmed to return to trim position. Intermediate circles represent graduated intervals of time and indicate predicted system position which is computed faster than real time. Many applications for this display technique have considerable merit. (See Kelley, 1960 and 1968.)

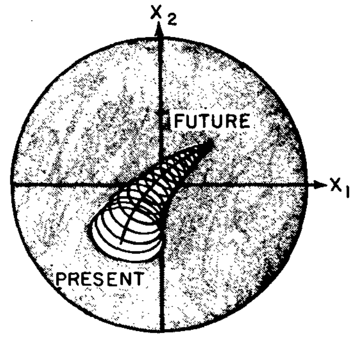


FIGURE 3–5. Predictor display with graduated time intervals. Predicted position in future time is represented by all but the largest circle, which represents present position (Kelley, 1960).

3.2 Visual Detection, Identification, and Estimation

Essential to a visual information system are data on human abilities to detect and identify targets. The term "target" is used to mean any object, pattern, or marking that operators must see in order to perform successfully. In many cases, the designer can do little about the visibility of targets, for example, those seen in optical fire-control and bomb-sight systems. In other cases, however, target visibility can be enhanced by means of proper design.

3.2.1 Definition of Terms and Units of Measurement

Important terminology and units of measurement in target visibility are:

Visual angle. This is the angle subtended at the eye by the viewed object (see Figure 3–6). Usually, this is given in minutes of arc, as computed by the following formula:

$$\text{Visual angle (min.)} = \frac{(57.3)(60)L}{D} \qquad (3\text{–}1)$$

where L = the size of the object measured perpendicular to the line of sight, and D = the distance from the eye to the object. The 57.3 and 60 in the formula are constants for angles less than 600 min.

Illumination. (Also called illuminance.) This is a measure of the areal density of light reaching an object or surface. The common unit of measurement is the foot-candle (ft.–c.). A foot-candle is the density of light falling on the inner surface of a sphere of one foot radius when a point source of light with an intensity of one international candle (c) is placed at the sphere's center.

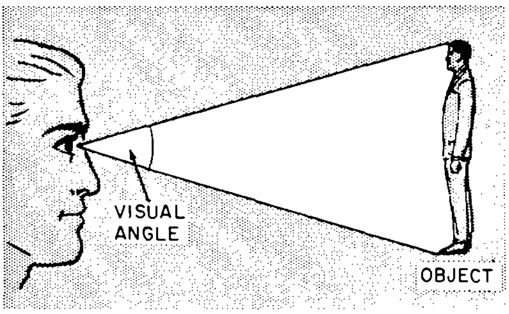


FIGURE 3-6. The meaning of visual angle.

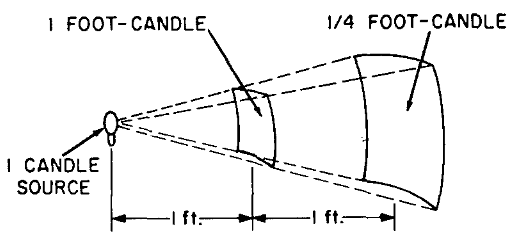


FIGURE 3-7. Illustration of inverse square law of illumination.

The illumination diminishes with distance according to the inverse square law (Figure 3-7) as follows:

$$\text{Illumination ft.-c.} = \frac{\text{Intensity of source (c)}}{\text{Distance (ft.)}^2} \qquad (3\text{-}2)$$

The illumination also diminishes according to the cosine of the angle between the surface and the direction of incoming light.

Luminance. For most display purposes, this is the important measurement. It is the amount of light per unit area reflected from or emitted by a surface. Although this measurement is frequently called brightness, strictly speaking, brightness is the resulting subjective sensation and is influenced by contrast, adaptation, and other factors besides the physical energy in the stimulus. Luminance is commonly expressed by a variety of units for which conversion factors are given in Table 3-1. The preferred units of luminance are defined below:

1. Lambert (L). A unit of luminance equal to that of a perfectly diffusing and reflecting surface illuminated by a standard candle at a distance of one centimeter (cm.).
2. Millilambert (mL.). One thousandth of a Lambert, and very nearly equal to a foot-Lambert (ft.-L.).
3. Foot-Lambert (ft.-L.). A unit of luminance equal to that of a perfectly diffusing and reflecting surface illuminated by one foot-candle. The approximate luminance values for a variety of commonly experienced conditions is shown in Figure 3-8.

Reflectance. (Also called albedo.) This is the relation between illumination reaching a surface and the resulting luminance. A perfectly diffusing and reflecting surface would be one that absorbs no light and scatters the illumination in the manner of a perfect mat surface. Such a surface would have a reflectance of 100%. If illuminated by 1 ft.-c., it would have a luminance of 1 ft.-L. for all viewing angles. In actual practice, the maximum reflectance achievable

TABLE 3-1. CONVERSION FACTORS FOR LUMINANCE UNITS

Units	Foot-lamberts	Lamberts	Milli-lamberts	Candles per square inch	Candles per square foot	Candles per square centimeter
ft.-L	-------------	1.076×10^{-3}	1.076	2.21×10^{-3}	3.18×10^{-1}	3.43×10^{-4}
L	9.29×10^{2}	-------------	1.0×10^{3}	2.054	2.96×10^{2}	3.18×10^{-1}
mL	9.29×10^{-1}	1.0×10^{-3}	-------------	2.054×10^{-3}	2.957×10^{-1}	3.183×10^{4}
c/in²	4.52×10^{2}	4.87×10^{-1}	4.87×10^{2}	-------------	1.44×10^{2}	1.55×10^{-1}
c/ft²	3.14	3.38×10^{-3}	3.38	6.94×10^{-3}	-------------	1.076×10^{-3}
c/cm²	2.92×10^{3}	3.14	3.14×10^{3}	6.45	9.29×10^{2}	-------------

Note: Value in units in left-hand column times conversion factor equals value in units shown at top of column.

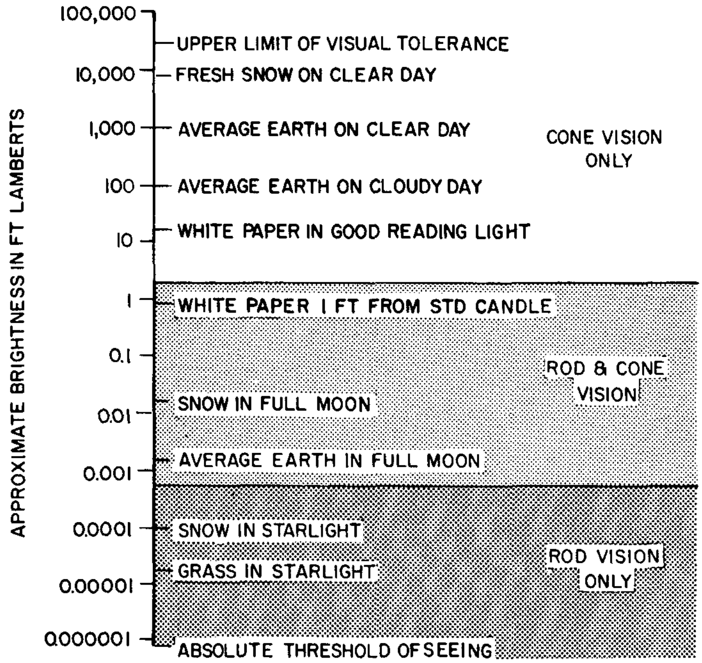


FIGURE 3–8. Examples of various levels of luminance.

for a nearly perfectly diffusing (mat white) surface is about 95%. Reflectance is defined by the following formula:

$$\text{Reflectance } (\%) = 100 \times \frac{\text{Luminance (ft.-L.)}}{\text{Illumination (ft.-c.)}} \qquad (3\text{–}3)$$

Contrast. This is a measure of luminance difference, usually between that of a target (L_t) and its background (L_b) as computed by the formula:

$$\text{Contrast } (\%) = 100 \times \frac{L_b - L_t}{L_b} \qquad (3\text{–}4)$$

Contrast can vary from 100% (positive) to zero for targets darker than their backgrounds, and from zero to infinity (negative) for targets brighter than their backgrounds. Reflectance can be substituted for luminance in the above formula for computing contrast.

For targets such as the first two shown in Figure 3–9, with equal light and dark areas, the luminance of the white areas should be treated as background, when applying the above formula. Sometimes the luminance of the lighter area, rather than that of the background, is used as the denominator in computing contrast. When this is done, contrast cannot exceed 100%. However, this method of computing contrast provides a restricted range and is not recommended.

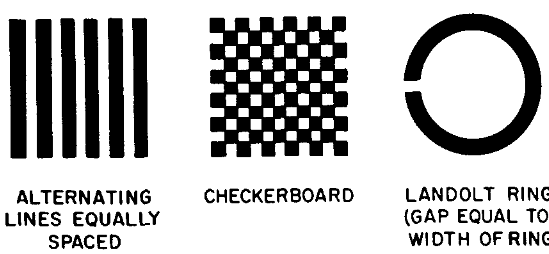


FIGURE 3–9. Common types of test targets for minimum separable acuity.

3.2.2 Factors Affecting Visual Acuity

The smallest detail that the eye is capable of resolving at a specified distance is referred to as minimum visual acuity. Visual acuity is expressed by the visual angle, in minutes of arc, or as the reciprocal of this angle.

The probability of detecting a target depends on its size, when other conditions such as knowledge about its location, the amount of light on the target, its luminance contrast, its shape, or the amount of time the viewer has to search for it are held constant.

Usually, visual acuity is determined for a 50% threshold, a target of the size that has a 50% probability of being reported. In some cases, the designer might want higher probability values, such as 95% or 99% thresholds, so that targets are almost always detected. In order to determine the size of target that will be detected almost all of the time, multiply 50% thresholds by two. Figure 3–10 shows that doubling the size of a 50% threshold target assures nearly 100% probability of detection.

Measuring Visual Acuity

There are several kinds of visual-acuity measurements: minimum separable, minimum perceptible, stereoscopic, and vernier acuity.

Minimum separable acuity. Also called gap resolution, this is the measurement of the smallest space the eye can detect between the parts of a target. Three kinds of target commonly used for measuring gap resolution are shown in Figure 3–9, above.

The curve in Figure 3–11 (Moon and Spencer, 1944) shows the smallest Landolt ring gap the eye can detect with different background luminances. This curve shows that as the amount of light is increased the eye can detect smaller and smaller gaps. Acuity increases as the amount of light increases. The normal human eye can detect a gap of about 1 min. of visual angle at ordinary, indoor-light levels with targets that have a high luminance contrast.

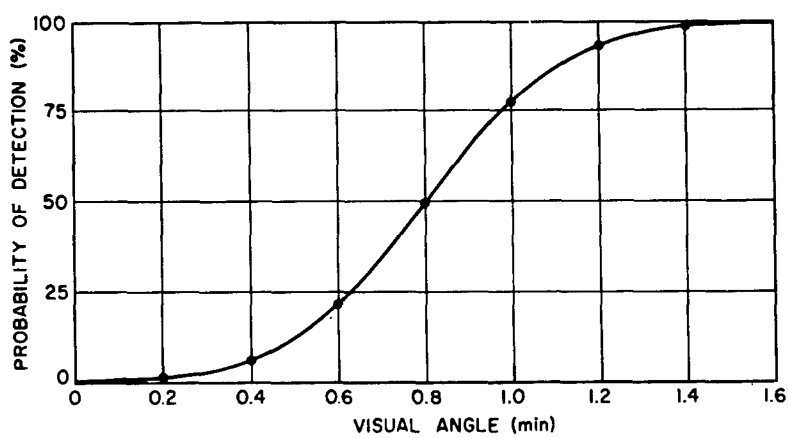


FIGURE 3–10. The cumulative probability of detection as a function of visual angle (Blackwell, 1946).

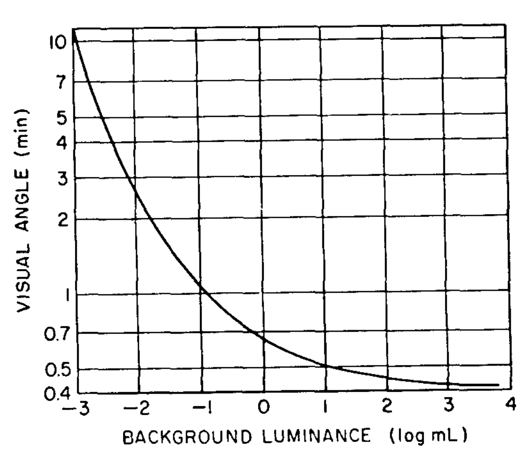


FIGURE 3–11. Minimum separable visual acuity as a function of background luminance (Moon & Spencer, 1944).

The acuity curve of Figure 3–11 is valid only for targets that are darker than their backgrounds. For white targets on black backgrounds or for letters and numbers cut out of a panel and lighted from behind, the curve appears different. As the luminance of white-on-black or cut-out targets increases, the acuity of the eye increases, at first, just as it does for black targets. But for luminances above about 10 mL., acuity for white-on-black targets drops off quickly because the white parts blur. The technical visual term for this phenomenon is irradiation.

With less contrast, there is lower acuity. For example, it is harder to see black on gray than it is to see black on white. This is illustrated by the curves of Figure 3–12 (Cobb and Moss, 1928), which show some measurements made with the Landolt ring seen earlier in Figure 3–9 at three different background luminances.

Minimum perceptible acuity. Also called spot detection, this is the measurement of the eye's ability to detect the smallest possible target, a spot that can be lighter or darker than its background. Spot detection depends both on luminance level and contrast, as is shown in Figure 3–13. The curves labeled 1 to 100% in this illustration are for targets that are either brighter or darker than their backgrounds. The curves for contrast above 100% are for targets brighter than their backgrounds (this is the only way of getting contrast above 100%). The thresholds in this graph are for a 99% probability of detection—that is, almost certain detection. While the data in Figure 3–13 are for simple circular targets, other simple compact targets, such as squares, hexagons, or octagons, would give comparable threshold values. As targets deviate from simple compact shapes, however, the detection thresholds will be increased (for such threshold data, see Lamar et al., 1947, and Kristofferson, 1957).

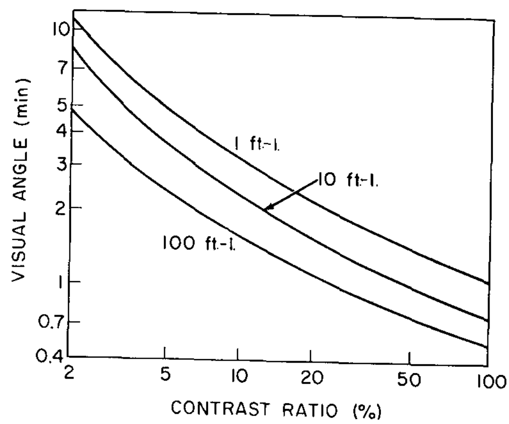


FIGURE 3–12. Minimum separable acuity as a function of background luminance (Cobb & Moss, 1928).

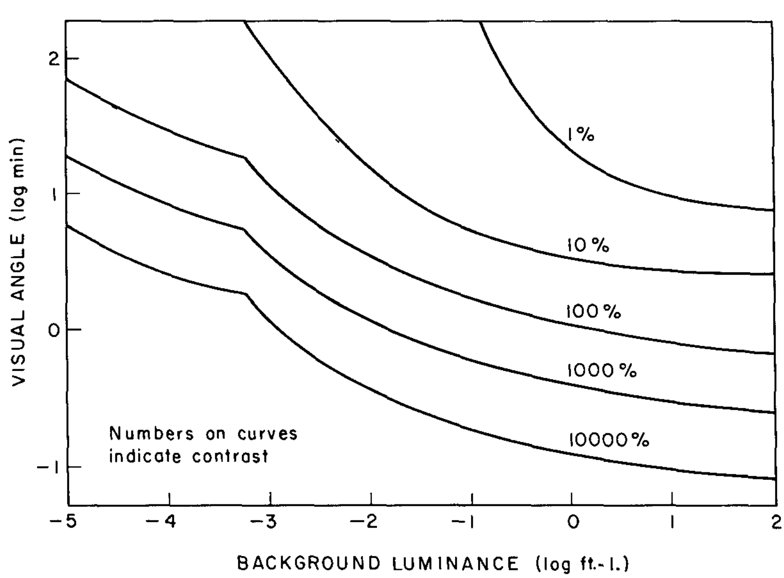


FIGURE 3–13. Minimum perceptibility, or spot detection, for circular targets as a function of contrast, and background luminance (Blackwell, 1946).

There is no lower detectable-size limit for spots that are brighter than their backgrounds. The eye can detect a bright spot, no matter how small it is, as long as it is bright enough. For instance the star Mira can be seen even though it subtends an angle of only 0.056 sec.

Lines and squares are also visible against bright backgrounds even when they are very small. For example, against a bright sky, the normal human eye can see a wire 1° long 75% of the time, even if the wire is only 0.43 sec. visual angle wide. Similarly, a dark square is visible against a bright sky 75% of the time even if the square subtends a visual angle of only 14 sec. (Hecht et al., 1947).

Stereoscopic acuity. Since the left and right eye are about 2.2 in. apart, the brain combines the slightly different images—or pictures—formed in the two eyes into a single image that has depth. The two pictures seen differ most for objects near the eyes and least for far-away objects. To look at something close, the eyes converge until the lines of sight meet on the subject. As the gaze is shifted to something farther away, the eyes diverge until, for very distant targets, the lines of sight are nearly parallel.

The threshold of stereoscopic acuity is the difference between the parallactic angles of two targets that are at just-noticeably-different distances from each other (see Figure 3–14.) The curve of Figure 3–15 (Berry et al., 1950) shows how stereoscopic acuity increases (visual angle decreases) as the amount of light increases.

Vernier acuity. Vernier acuity is measured by using two lines which are separated. The threshold of vernier acuity is the smallest lateral displacement of one line from the other that can be detected. This kind of acuity is used in reading some instruments and also optical range finders in tanks and ships. The curve in Figure 3–16 (Berry et al., 1950) shows that exceptionally small details can be resolved by the human eye and that vernier acuity is enhanced with increasing amounts of light.

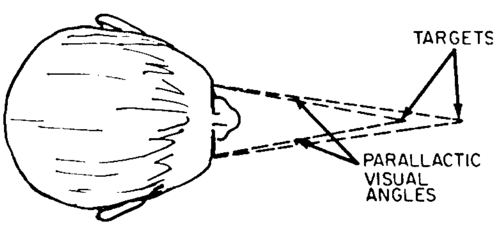


FIGURE 3–14. Explanation of stereoscopic acuity.

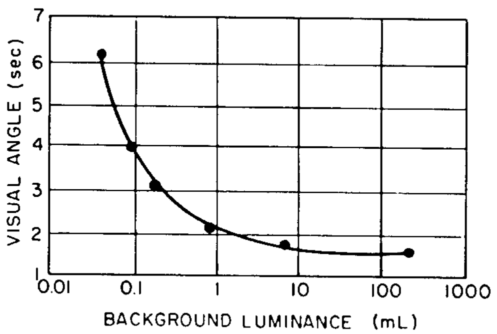


FIGURE 3–15. Stereoscopic acuity as a function of background luminance (Berry et al., 1950).

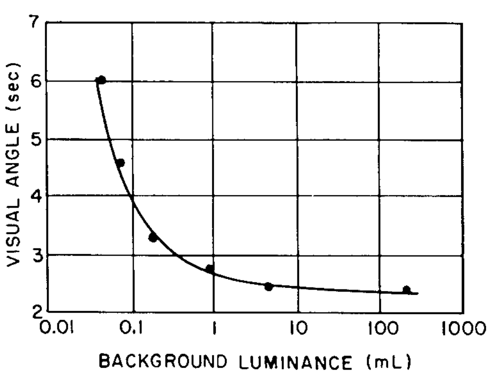


FIGURE 3–16. Vernier acuity as a function of background luminance (Berry et al., 1950).

Effect of Adaptation Level

Visual acuity is best when the eyes have *not* been exposed to high levels of luminance. For example, if the captain of a ship has been out on the flying bridge scanning the sunlit ocean and then steps into the pilot house, he is momentarily "blinded." If it is important for him to be able to see certain displays immediately, the letters, numbers, or markings should be large enough or bright enough to be read in this "semiblind" condition. If visual displays have to be read by people who have just been exposed to the open sea, clouds, desert, or snow, the luminance level of displays should be at least 0.01 as bright as the pre-exposure field (Brown, 1953).

With the eye adapted to the luminance level of the target and its surrounding area, visual acuity is maximized. But it is reduced when the target and immediate surround are at a lower brightness than the greater surround. Acuity also is slightly reduced if the surround is considerably darker than the target.

The curve in Figure 3–17 (Ireland et al., 1967) shows the effect of luminance for surrounds both darker and lighter than the target and its immediate background. These data are for a dark Landolt ring acuity target on a lighter background, with a gap of 1.93 min. visual angle. The curve shows how the threshold contrast increases as the background luminance is increased over that of the surround. Reducing the background luminance below that of the surround luminance, on the other hand, has very little effect. In fact, Lythgoe (1932) found acuity to be slightly improved when the surround luminance was about one-half the luminance of the task area. It is clear from such data as these that targets should not be in a shadow or near a large area of much higher brightness if visual acuity is to be at a maximum.

Effects of colored lighting. Colored illumination can be obtained by placing a filter in front of the light source or simulated by placing a filter in front of the observer's eyes; both methods have the same effect for the viewer. In many applications, colored filters will reduce visual acuity, because the total amount of light reaching the eye is reduced. Colored illumination is recommended only under certain conditions because it distorts the colors of objects. Red filtering is helpful to people who must keep their eyes ready to work in darkness: pilots, tank drivers, x-ray technicians, and photographers. In addition, selective color filtering is recommended for rooms in which operators must watch low brightness displays such as radar scopes.

Dynamic visual acuity. All of the preceding measures of visual acuity are for *static* conditions. In other situations, either the observer or the target is moving. "Dynamic visual acuity" is generally defined in terms of the smallest detail that can be detected when the target is moving. Angular movement of the target decreases the threshold of visual acuity.

Some threshold data for dynamic visual acuity from Burg (1966) are shown in Figure 3–18. These data are for minimum separable acuity and target motion through 180°. These and other data (see also Miller and Ludvigh, 1962) indicate that the loss of acuity increases rapidly as the rate of motion exceeds 60° per sec. Such thresholds for dynamic visual acuity could be considerably higher with a shorter viewing time or target travel distance.

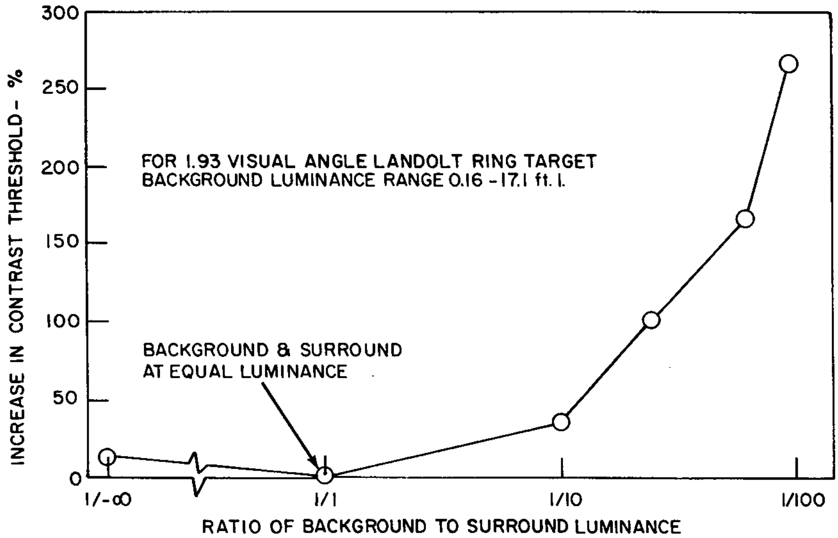


FIGURE 3–17. Effect on visual acuity of surround luminance (Ireland et al., 1967).

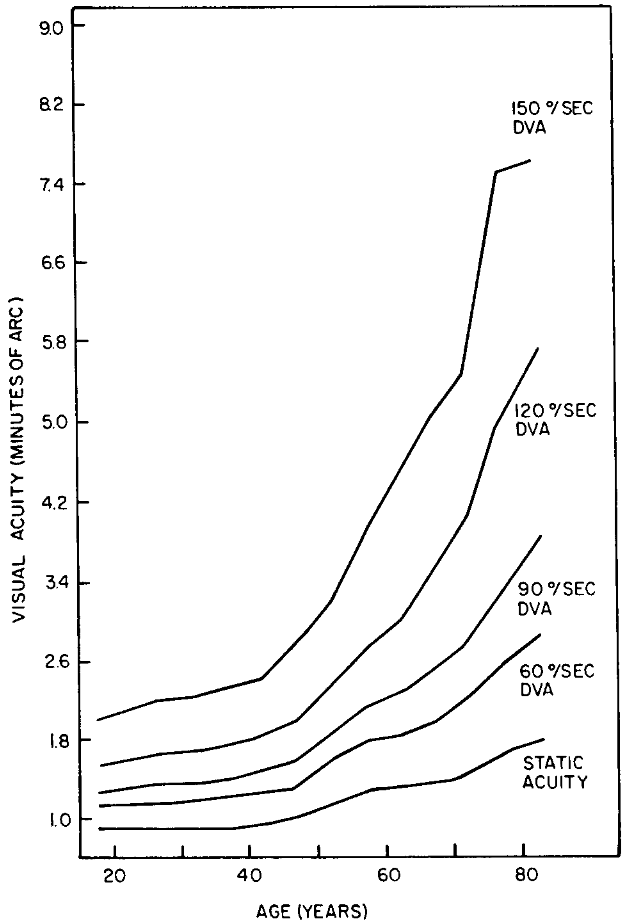


FIGURE 3–18. The variation of static and dynamic visual acuity with age (After Burg, 1966).

3.2.3 Searching for Targets

Having to search for targets is more complicated than being able to observe them with prior knowledge of their location. Some parts of the eye are more sensitive than other parts, and successful searching depends on how the searcher moves his eyes. Light reaching the eye is picked up by the retina, which is composed of two sets of light-sensitive cells, the rods and cones. The cones function best under daylight conditions; the rods function best under twilight or night conditions. Color discrimination is made by the cones only, but light-dark discrimination involves both rods and the cones. Resolution or visual acuity is much poorer for the rods.

The eye sees differently from diverse angles as well as under varying ambient light conditions because of a small area in the center of the retina called the fovea which contains all cones and no rods. Since cones are insensitive to twilight or night conditions, foveal vision is not effective in seeking out dim targets after dark. But, as the angle of view from the fovea is increased, the concentration of rods becomes denser, and night vision is enhanced.

Acuity in peripheral vision. For daylight conditions, visual acuity diminishes rapidly with distance from the fovea as shown by data from Blackwell and Moldauer (1958) and Taylor (1961) plotted in Figure 3–19. This means that for a near-threshold target to be seen, the eyes must be fixated within an angle as small as 1°. At progressively greater peripheral angles, the target size must be increased to see the critical details. Using the data of Taylor, for example, a target at double the threshold size (relative

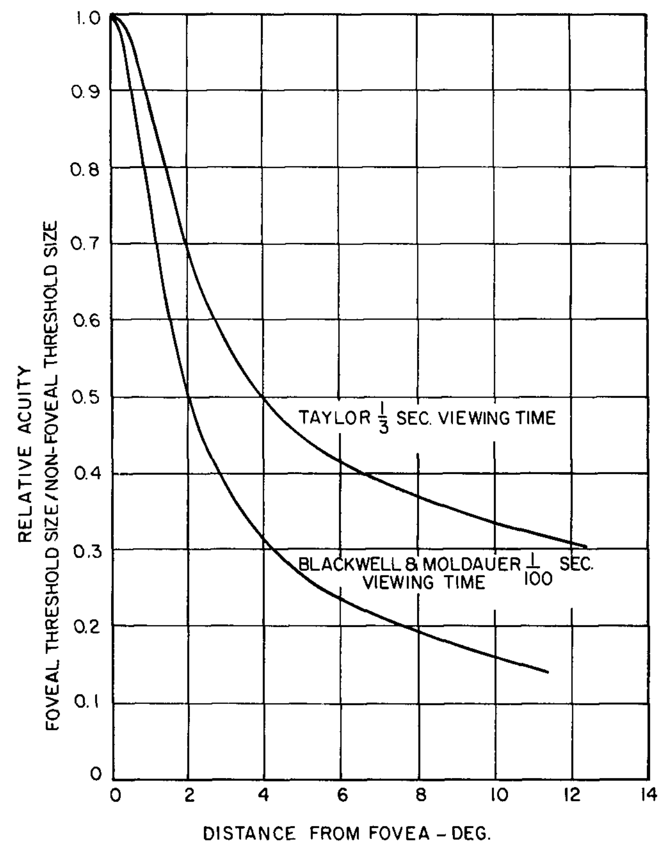


FIGURE 3–19. Change of visual acuity with distance from the fovea at high luminance, 76 ft.-L. Data are for circular targets which varied in size (Blackwell & Moldauer, 1958; and Taylor, 1961).

acuity 0.5) should be visible at an angular displacement of about 4° or less from the fovea. These data illustrate the loss in resolution or visual acuity when the target is perceived by sensors outside the fovea under daylight conditions. To be perceived by peripheral vision, the target must be several times larger than it has to be when seen by foveal vision.

Peripheral acuity at low light levels is illustrated by the data of Mandelbaum and Sloan (1947) shown in Figure 3–20. There is an intermediate luminance range, around 0.002 mL., where the best acuity is about 4° to 8° from the fovea, although this acuity is only about 0.1 of the acuity in foveal vision in daylight. At the lowest light levels at which just about anything can be seen (0.00001 to 0.000001 mL.), peripheral vision is unaffected by the target angular distance from the fovea. At this low luminance level, acuity is poor by comparison with peripheral vision in daylight.

Eye fixations during search. In visual search the observer successively fixates his eyes on different points in the area being covered. The spot on which he fixates is imaged on the fovea where visual acuity is greatest. The normal rate of such fixations during search is about 3 per sec. (White and Ford, 1960), but may become slower as the observer makes longer fixations on objects of interest. Typically, the fixations are concentrated in certain areas. If something of potential interest is found, numerous small fixation shifts will be made around this specific area. At the same time, other parts of the total search area will be overlooked.

Empty field myopia. Eyes searching an empty sky or homogeneous field will see an object somewhat out of focus when it comes within the

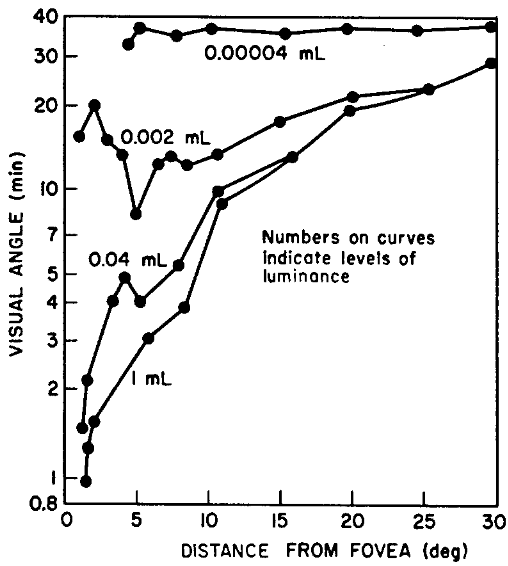


FIGURE 3–20. Change of required target size with distance from fovea for targets of varying luminance (Mandelbaum & Sloan, 1947).

visual range. This condition is called empty field myopia because the eyes normally accommodate for near rather than distant vision. The distance at which individuals will accommodate when searching an empty field is one to two meters (Whiteside, 1957).

Empty field myopia is not a significant handicap in typical military search situations. Usually there will be some objects in the visual field, such as vehicle structure, clouds, or a distant horizon, on which the eyes can accommodate. Lack of feedback on where the observer has searched (repeated searching of the same areas) is also a major factor in searching an empty field.

3.2.4 The Detection of Colored Targets

Acuity is increased by luminance contrast. Making the target and background two different colors does not increase acuity if there is high luminance contrast. If luminance contrast is low, color contrast improves visual acuity appreciably.

Detecting surface color. Although no one color will be most visible against all types of terrain (desert, snow, sea, or jungle), a bright orange target is seen best at great distances against most backgrounds. Also, because fluorescent materials enhance luminance contrast, they are recommended for survival equipment, identification panels, and other targets where detectability is desired.

Detecting colored lights. There is a detection problem when color-coded lights are at a great distance. Signal color recognition depends on eye illumination from the light. This is proportional to the intensity of the light and inversely proportional to the square of the distance. It is also affected by the luminance of the background and the color of the light. The curves of Figure 3–21 (Hill, 1947) show in millicandles the threshold illumination for color recognition of point source signal lights when viewed against neutral backgrounds of various luminances. These are 90% thresholds. The curves show that red, green, and white are nearly equal in terms of threshold intensity for color identification. For yellow the threshold is somewhat higher because of a tendency to confuse yellow with white.

3.2.5 Target Visibility Against Patterned Backgrounds

Visual detection is difficult when the target appears against a mottled or patterned background, especially when other objects in the field resemble the target in size, color, shape, or luminance. Targets should then be much larger than one would predict from threshold acuity data. Observing the following rules will make targets more visible against non-uniform backgrounds:

1. Choose a color and luminance that contrast most with the colors in the background.
2. Pick white or bright colors for targets on dark backgrounds, and vice versa.
3. Use a fluorescent color for targets against dark backgrounds.
4. Use as large an area of solid color as possible. Do not use stripes or checks; patterns like these are not visible at great distances—they fuse and reduce contrast.
5. Divide target color into two contrasting colored areas. One or the other of these two colors will contrast with most backgrounds: white and red, bright yellow and black, bright yellow and blue, and bright green and red.

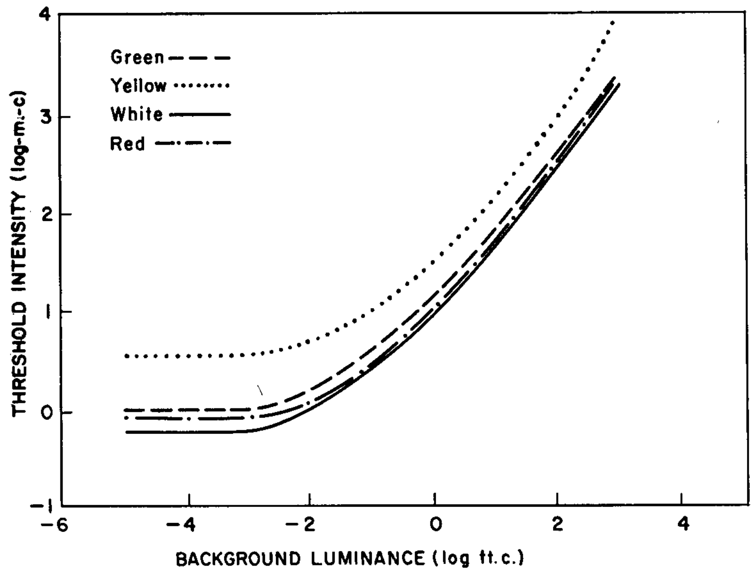


FIGURE 3–21. Intensity thresholds of point-source signal lights as a function of the background luminance for signal-light color. Figures are for 90% thresholds (Hill, 1947).

3.2.6 Prediction of the Sighting Distance for Military Targets

Researchers have tried to predict the distance at which specified military targets should be visible to human observers. A series of nomographs has been developed by Duntley (1948). These nomographs are also available in Middleton (1958). These threshold data for predicting performance in the field indicate expected performance under a variety of environmental conditions. Such threshold data are applicable only when the observer is stationary; when the observer knows the approximate target location; when the target has a simple shape and uniform brightness; and when the target background is of uniform brightness. Unfortunately, most military target conditions do not conform to these optimal conditions and the threshold data must be adjusted to predict actual performance. Typically an adjustment of a factor of two is satisfactory but the designer always assumes some risk when using threshold data to predict field performance.

Air-to-ground target recognition. Military observers in aircraft and helicopters must cope with angular motion, cluttered backgrounds, limited search time, and obstructed vision in some directions. There is a 10° to 15° forward and downward viewing limitation in most fixed-wing aircraft. Terrain features and haze add further restrictions. If the observer is also the pilot, other duties compete with search activities.

A field study of ground target recognition from helicopters was conducted by Moler (1962). While a few targets were acquired at the maximum possible sighting range, a large proportion were never reported at all. Higher speeds and greater restrictions on downward visibility would make observers in fixed-wing aircraft less successful than Moler's observers in helicopters. (See Erickson, 1965; Davies, 1965; and Snyder, 1964.)

3.2.7 Recognition of Targets in Imaging Displays

Photo-interpreters, or observers of radar or infrared displays, must search for targets in a picture of terrain obtained by direct photography or other types of sensors. These pictures may be displayed as photographs, transparencies, projected images, images on cathode ray tubes, or telescopic images. Observers must recognize distinctive target patterns or reference landmarks appearing against highly patterned backgrounds. Factors that affect visibility include target size, amount of clutter in the background, contrast, display area, display movement, and prior knowledge of target location.

Target size. The human eye is capable of identifying the letters of the alphabet if these letters subtend a visual angle of at least 5 min. of arc. This ability defines 20–20 vision as measured by the Snellen eye chart. Letters are highly discriminable; target patterns of the type involved in image interpretation generally are not. Figure 3–22 (Steedman and Baker, 1960) shows the relation of target size to accuracy and identification speed for image interpretation targets. From the curve, it is evident that when the visual angle subtended by the largest dimension of the target is smaller than 12 min., an increase occurs in relative search-to-identification time and in errors of identification. These data indicate that targets should subtend, as a minimum, 12 min. of arc to insure reasonably accurate recognition. With low contrast, or difficult target patterns, the minimum size for recognition should be increased by a factor of two or three.

Display clutter. How quickly targets will be recognized in complex displays is affected by the amount of clutter there. As the number of irrelevant targets increases, there is a nearly proportional increase in the required search time (Boynton et al., 1957, and Baker et al. 1960). Thus, with only a short search time available, the number of missed targets will increase proportionally with increases in clutter.

Resolution. Increasing the display resolution, up to the limit of the observer's visual acuity, will generally aid target recognition (Bennett et al., 1963). If such increased resolution is obtained by an increase in magnification or scale factor, however, there will usually be compensating disadvantages from increased image motion, reduced viewing time, and reduced area of ground coverage (Simon and Craig, 1965). Thus, resolution must be balanced against these other factors.

Contrast and color. Closely related to resolution are the contrast relations on the image. Target recognition will be facilitated if the total contrast range (difference between lightest and darkest areas) is high; if the distinguishable shades of gray (gray-scale) are at a maximum; and in colored images, if the distinguishable colors are at a maximum. The contrast direction (positive or negative) should be that most familiar to the observer. For example, photographic

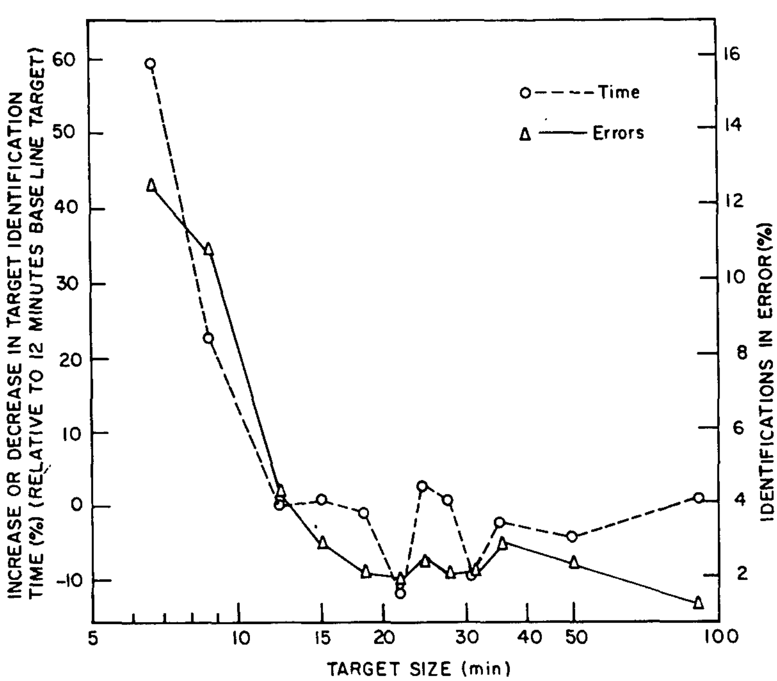


FIGURE 3–22. Effect of target size on target search and identification (after Steedman & Baker, 1960).

maps should be positives rather than negatives. For radar images it is customary to show reflective targets as bright areas on the display, and low-return surfaces as dark areas. A study by Van Ausdall and Self (1964) using side-looking radar, indicates only a slight advantage for the conventional contrast direction for experienced radar observers.

Image motion and viewing time. Where imaging displays are used to observe the ground in real time from aircraft and space vehicles, motion of the image restricts viewing time and, thereby, the number of targets that will be recognized. Boynton and Bush (1957), using complex static displays, found that recognition improved with viewing time up to (and probably beyond) 24 sec. Representative data for moving displays from side-looking radar obtained by Van Ausdall and Self (1964) indicate that some new targets were still being recognized after 2 min. of viewing time. About 50% of the recognitions, however, occurred in the first 30 sec. after the targets came into view. For imaging displays of large area, or containing fine detail, still longer viewing times will be required.

Display size. Factors of space, weight, and cost usually dictate the choice of a minimum acceptable rather than an optimal size for imaging displays. It is unlikely, therefore, that the display would ever be too large. To determine the minimum required display size for identification of a target on the basis of its form alone, the formula given below may be used:

$$D = \frac{1.54LRM}{T}, \qquad (3\text{–}5)$$

where D = Display size in the axis that R is being displayed (in.),
L = Display viewing distance (in.),
R = range on ground being displayed (statute mi.),
M = minimum target visual angle for detection (min. of arc), and
T = greatest target dimension (ft.).

In applying the above formula, standard values may be assumed for two of the parameters. Viewing distance (L) will normally be at least 16 in. unless lenses are used to relieve eyestrain. Displays for use by pilots on aircraft instrument panels will normally be about 28 in. from the eyes. For the minimum target visual angle (M) a value of 12 min. may be used if no better information is available. For displays with inherently poor resolution or low contrast, this value may be doubled or tripled. Where other cues aid in target identification, such as distinctive land masses or a limited target set, display size may be reduced substantially.

For some display situations, image motion and viewing time must also be considered in the choice of display size. The available viewing time for a point on the ground, based upon display size, scale factors, and vehicle speed can be computed by the formula below:

$$T = \frac{0.049D}{SV}, \qquad (3\text{–}6)$$

where T = viewing time (sec.),
S = scale factor $\dfrac{\text{distance on display}}{\text{distance on ground}}$,
D = max. display dimension (in.), and
V = vehicle velocity (knots).

In applying the above formula, it should be recognized that only targets passing through the maximum dimension of the display would be visible for the computed time.

Restriction of search area. Observers having no prior knowledge of target locations on mapping displays miss a large proportion of the real targets, and report a large number of false targets. In actual use, given geographic coordinates of a specific target, or, target position related to a prominent landmark, observers will recognize targets more quickly and accurately (McKechnie, 1967).

Use of multiple observers. A team of two or three observers will often give better target recognition performance than a single observer will give. (See Bolin, Sadacca, and Martinek, 1965.) The gain in performance by using more than one observer is not large nor consistent, however. In terms of overall target recognition performance (maximum real and minimum false targets reported), the team should give best results when its members work cooperatively rather than independently.

3.2.8 Optical Aids to Target Detection

A variety of optical aids can be used to improve human visual capability in searching for targets. The most common of these are: magnification, stereoscopic viewing, and filters.

Magnification. Targets can be detected or spotted at greater distances than those at which they can be identified. A ship at sea can be seen long before its type can be determined. To recognize something, the observer has to be able to see details that are visible at a distance.

Telescopes, binoculars, and magnifiers improve target detection and identification, although the degree of improvement is never equal to the magnification ratio—a 5-power telescope does not mean that one can see five times as far or five times as well because magnification also does the following: (a) lowers the brightness of the target, (b) lowers the contrast between the target and its background, (c) makes the image of the target fuzzier, (d) increases the image movement of a moving target, (e) makes the image of a steady target seem to move if there is any unsteadiness or vibration in the instrument, (f) cuts down the size of the field of view.

The degree of these effects depends on the number of glass parts in the device, how accurately the lenses and prisms are aligned, the quality of the optical surfaces, whether the glass surfaces are coated, and the stability of the instrument mounting. Another factor subtracting from effectiveness of optical devices is the quality of the air (haze, dust, etc.).

Two general questions important in choosing a telescopic device are: What are its design specifications, and, In what way (on a steady or moving platform, in daylight, or at night) will it be used? In answer to the first question, magnification, exit-pupil size, diameter of the field, and light transmission are some of the most important parameters. These factors might conflict, however; any optical device has to be a compromise. For example, a large exit-pupil size makes the device much easier to use, but exit-pupil size is inversely proportional to magnification. Again, increasing the diameter of the field permits increasing the exit-pupil size, but this means increasing the size of the optical parts, leading to a bigger and heavier device.

To answer the second question (in what way will the device be used?)—using good binoculars and telescopes is more effective at night than is supposed. The increase in magnification—or size of the image—more than compensates for the loss of light. For daylight use, if the observer knows where the target is, or if he has to search a relatively small area, moderate-to-high magnification will be helpful. If he has to search a large area, if the target is moving fast, if he is moving, or if he has only a short time to look, magnification will be more hindrance than help. In most cases, an observer will do best by using his unaided eye or low power magnification for initial scanning. Then, when he has spotted a target or located a suspicious area, magnification aids can help him to identify the target.

Design Recommendations

1. Binocular devices are better than monocular ones, especially for night use. The advantage of binocular devices is small in daylight when brightness levels are high.
2. For binoculars in aircraft or moving ground vehicles, use hand-held or mounted binoculars of 3 to 4 power. Powers higher than this are not recommended because of the vehicle's motion and the small field of view.
3. For daytime use on a steady platform (the ground or a ship in normal seas), magnifications up to 6 power are recommended for hand-held binoculars, and magnifications up to 20 power for mounted binoculars.
4. For *night* use on a steady platform, magnifications up to 10 power are recommended for hand-held binoculars, and magnifications up to 20 power for mounted binoculars.
5. An exit-pupil size of 6 mm gives the best performance—other things being equal.
6. The lines of reticles should be thin enough not to cut out targets but thick enough to see easily. In any event, their size should not exceed 2 min. visual angle.
7. In devices that will be used at night, reticles should be illuminated with deep-red light (above 600 mμ) and should have an adjustable intensity control.
8. Rubber eye guards should be put around the eyepiece to help the observer rest his eye at the right place and block out ambient illumination.

9. Binocular devices should provide for interpupillary distance adjustment between 2.2 and 3 in.

Stereoscopic viewing. Stereoscopic aids for better judgment of distance have been used to advantage in aerial photograph analysis and interpretation. Adding a third dimension to radar displays has not proven to be as successful. Maintaining fusion of the two images is difficult in continued stereoscopic viewing. Also, the images presented to the two eyes must have sharp detail and correct geometrical relationships for depth. For these reasons, stereoscopic viewing is not a generally useful optical aid for visual displays (Liebowitz and Sulzer, 1965).

Use of filters. In many cases, filters, such as sun glasses, have a detrimental effect on vision by reducing the amount of light reaching the eyes, and in distorting color relationships. On the other hand, filters are beneficial to vision in certain situations: (a) sunglasses with neutral filters of about 15% transmission are recommended for reducing the light outdoors; (b) polarizing filters are beneficial for reducing specular glare from water, highways, and other reflecting surfaces and, for use as sunglasses; (c) red goggles are beneficial for preserving dark adaptation while in a light environment; and (d) selective color filters are useful for viewing radars in lighted rooms.

3.2.9 Estimating Size, Distance, and Speed

The human eye has extraordinary capacities for seeing small details and faint amounts of light and for detecting small differences between objects. It is poor, however, at estimating absolute values. For example, under ideal conditions, the eye can see a difference in the brightness of two areas if they differ by as little as 1%, but even experienced photographers have difficulty estimating the amount of light in a room within 100% of the true value.

This characteristic of the human eye is important in size, distance, speed, or acceleration estimations when absolute or numerical terms are required. Thus, the size of an unfamiliar target cannot be estimated accurately unless its distance is known. If the observer estimates distance the estimate distortion will distort the corresponding size estimate even further.

Distances to targets are usually underestimated when they are seen across smooth water or snow, particularly when other objects that provide distance cues in the field of view are lacking. If the distance (or size) of some other object is known, the distance to an unfamiliar target usually can be estimated with some accuracy. If there are no other objects in view—as when a target is high in the sky or across a large body of smooth water—estimates of distance are usually too short. It is almost impossible to estimate the distance of a target seen against a clear sky unless it is close or its size is known.

Estimates of the speed of moving objects—such as aircraft, birds, or automobiles—are poor and are probably related to estimates of distance as well as target size. This explains why small objects (insects and birds) appear to move faster than big ones (aircraft). Little is known about human ability to estimate speed changes (acceleration) except that it is inaccurate and unreliable.

3.3 Workplace Illumination

Efficient use of vision in operating equipment requires an illumination level that can vary from a low one of a specified color, such as that on the bridge of a ship or in an aircraft cockpit, to a high level of white illumination as found in offices and machine shops.

Important considerations for adequate workplace lighting can be listed as follows:

1. The distribution of light.
2. Luminance contrasts of the viewed objects and their details.
3. The quality and color of the illuminants and workplace surfaces.
4. The intensity of illumination.

3.3.1 Light Distribution

The design engineer is concerned with two primary sources of light: natural sunlight, provided directly or indirectly, and artificial light. Three methods of distributing artificial light over the task area may be considered. (See Figure 3–23.)

1. Direct Light. In this method, rays from the light source fall directly on the task area.

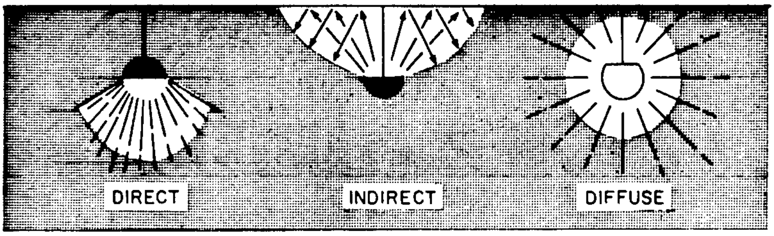


FIGURE 3–23. Light distribution methods.

Typically a light bulb with an opaque bowl inverted *over* it is used.

2. Indirect Light. Most of the light rays are reflected from the walls and ceiling before they reach the task area. This is achieved by placing an opaque bowl *under* the light.

3. Diffused Light. In this method, the light source can be enclosed in a translucent bowl so that the light is emitted from a larger surface area.

There are many ways of combining these methods to produce different illumination results. Which one to use depends on the visual task.

Direct lighting gives the most light at the working surface where 90%–100% of the output of the luminaire (the entire lighting unit—socket, bulb, and directing or diffusing element) is directed downward. The most prominent faults of direct lighting are contrasts, shadows, and glare.

Indirect lighting offers general, even illumination without shadow or glare thus causing less visual fatigue. When the source is shielded, 90%–100% of the light is directed upward toward ceiling and upper walls.

Diffuse lighting requires less electrical power than indirect lighting but does cause some glare and shadows. However, the use of fluorescent units with baffles reduces the problem.

Combinations of direct or indirect light with diffuse lighting can provide modifications if needed, e.g., a fluorescent desk lamp combines the downward-directed characteristics of direct lighting with diffusing characteristics of fluorescent tubes.

Reducing glare. When a relatively bright light source or its reflected image appears in the visual field, the resulting glare causes decreased visibility and visual discomfort. Within the visual field, direct glare comes from a light source, and specular glare comes from a reflecting bright surface. (See Figure 3–24.)

Direct glare can be reduced by:

1. Avoiding bright light sources within 60° of the center of the visual field.
2. Using shields, hoods, and visors to keep the direct light source out of the viewer's eyes.
3. Using indirect lighting.
4. Using several low-intensity sources rather than one high one.

Specular glare can be reduced by:

1. Using diffuse light.
2. Using dull, mat surfaces (flat paints, desk blotters) rather than polished surfaces.
3. Arranging direct light sources so that the viewing angle to the work area is not equal to the angle of incidence from the source.

Surround brightness. The surround luminance should be 10% less than that of the visual task area. But there should be some illumination falling on the surrounding areas. The immediate-surround surfaces should have reflectance factors no higher than the targets in the central visual field when both areas are equally illuminated.

3.3.2 Surface Reflectance

Reflectance results when light beams incident on a surface are redirected. There are several kinds of reflectance (Figure 3–25): (a) diffuse light is reflected if the surface is composed of rough, irregular particles; (b) specular, incident light on a polished surface, such as a mirror, will be reflected at an angle equal to the angle of incidence; and (c) compound reflectors are surfaces having both specular and diffuse reflectance characteristics.

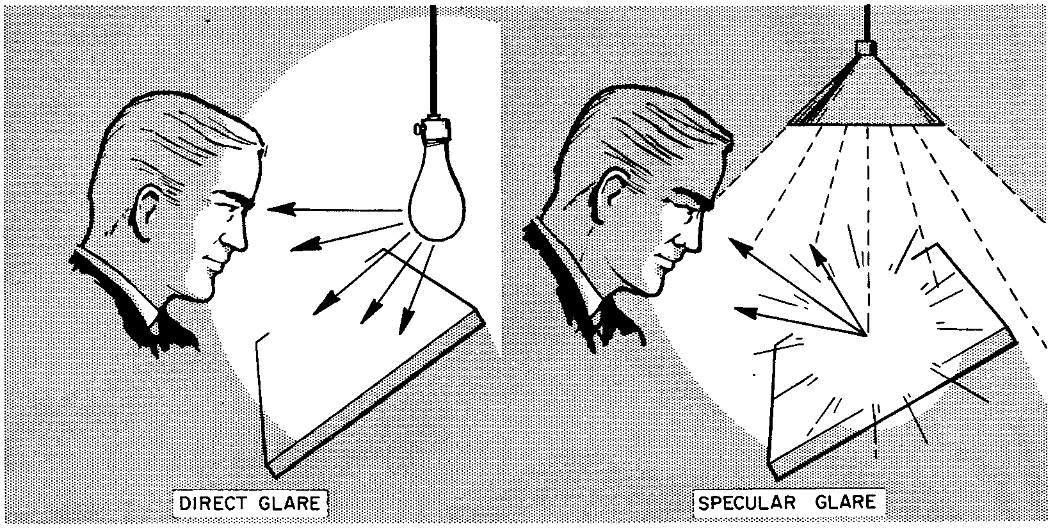


FIGURE 3–24. Sources of visual glare.

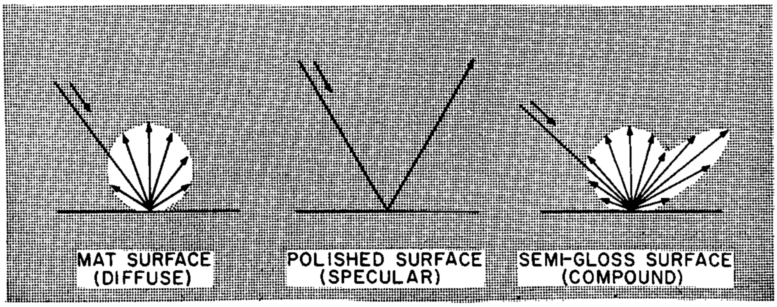


FIGURE 3–25. Light reflection from various surfaces.

Selective reflectance. An object's color results from selective reflectance and absorption of particular wavelengths of incident light. Something appears to be red because the longer wavelengths (red) are reflected, and the shorter ones (blue) are absorbed. The highest reflectance occurs with the color white. Selective and spectral reflectance are dictated by the percentage of reflected light at arbitrary wavelength steps.

Reflectance of surround. The reflectance of ceilings, walls, floors, furnishings, or machinery contributes significantly to the general illumination level of a room. A less-intense light source is required in a workplace with highly reflecting surfaces. To reduce specular reflections, large surfaces such as walls and desk tops should not be polished. Figure 3–26 illustrates general recommendations for surface reflectances in offices, study rooms, machine shops, power stations, etc. (See also IES Lighting Handbook, 1966.)

3.3.3 Shadow and Surface Color

Shadows can be avoided by using diffused light, and light colors on all surfaces where interreflections will increase the amount of light in obscured areas. Dark shades of gray, green, blue, red, and brown (see Table 3–2) should not be used on large surfaces; pastels and light gray are recommended.

For general room lighting, the designer should strive for an evenness of distribution without shadows and glare. The levels required can be determined from Table 3–3.

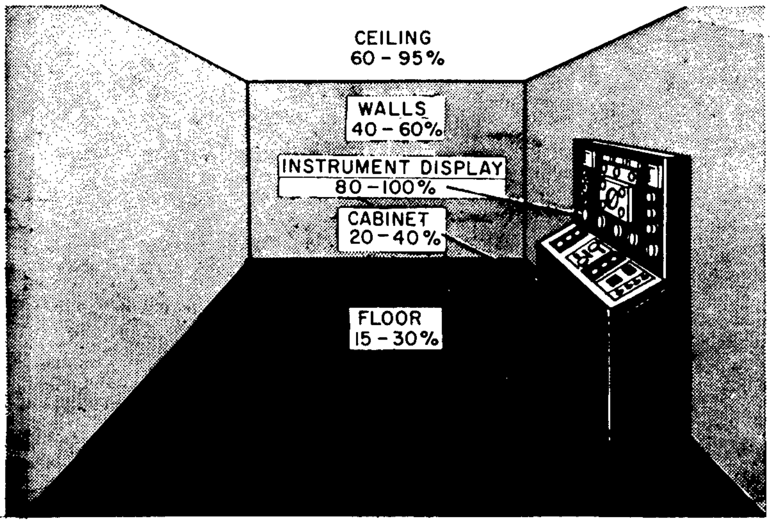


FIGURE 3–26. General recommendations for workplace reflectances.

TABLE 3–2. APPROXIMATE REFLECTANCE FACTORS FOR VARIOUS SURFACE COLORS

Color	Reflectance	Color	Reflectance
White	85		
Light:		Dark:	
Cream	75	Gray	30
Gray	75	Red	13
Yellow	75	Brown	10
Buff	70	Blue	8
Green	65	Green	7
Blue	55		
Medium:		Wood Finish:	
Yellow	65	Maple	42
Buff	63	Satinwood	34
Gray	55	English Oak	17
Green	52	Walnut	16
Blue	35	Mahogany	12

Woodson & Conover (1964).

TABLE 3–3. GENERAL ILLUMINATION LEVELS AND TYPES OF ILLUMINATION FOR DIFFERENT TASK CONDITIONS AND TYPES OF TASKS

Task condition	Type of task or area	Illuminance level (Ft.-c)	Type of illumination
Small detail, low contrast, prolonged periods, high speed, extreme accuracy.	Sewing, inspecting dark materials, etc.	100	General plus supplementary, e.g., desk lamp.
Small detail, fair contrast, speed not essential.	Machining, detail drafting, watch repairing, inspecting medium materials, etc.	50–100	General plus supplementary.
Normal detail, prolonged periods.	Reading, parts assembly; general office and laboratory work	20–50	General, e.g., overhead ceiling fixture.
Normal detail, no prolonged periods.	Washrooms, power plants, waiting rooms, kitchens	10–20	General, e.g., random natural or artificial light
Good contrast, fairly large objects.	Recreational facilities	5–10	General.
Large objects	Restaurants, stairways, bulk-supply warehouses.	2–5	General.

By mounting bench and desk lamps near the working area, supplementary lighting can be obtained. Flexible-neck lamps are recommended because they can be adjusted to avoid glare and shadows.

3.3.4 Reflection Hazards

Any reflection from a windshield or window that reaches the operator's eyes reduces his ability to see out of the windshield or window. This hazard can be avoided by:

1. Use of glare shields placed in the light path so as to block light otherwise reflected from the windshield. (See Figure 3–27.)
2. Lights and lighted surfaces in such positions that reflections will not reach the eye.
3. Areas with dark, mat surfaces.
4. A minimum illumination level consistent with adequate visibility of the indicators.

3.3.5 Indicator and Panel Lighting

Illumination requirements. In some situations, minimum illumination permitting adequate indicator reading must be maintained both for dark adaptation and to avoid objectionable reflections. The curve in Figure 3–28 (Chalmers et al., 1950) shows the relationship between the luminance of indicator marks and efficiency in reading them. Other situations call for high illumination in the workspace. Exposure to lightning flashes, flares, searchlights, or atomic flashes can result in temporary loss of visual adaptation, as shown by the data of Brown (1964) in Figure 3–29. If such exposures can be anticipated, the loss of capability to read vital instruments and control panels will be minimized if high illumination is provided in the workspace.

Table 3–4 provides summary recommendations for lighting systems, color, and luminance levels for a variety of situations. The various lighting techniques are discussed below.

While uniform light distribution is difficult to achieve with known lighting techniques, fixtures located for approximate equality are helpful. Otherwise, some indicators will not be legible while others will be too bright. A ratio of 7:1 between the brightest and dimmest is the maximum tolerable range. In addition, direct or reflected light shining in the eyes of the operator must be avoided.

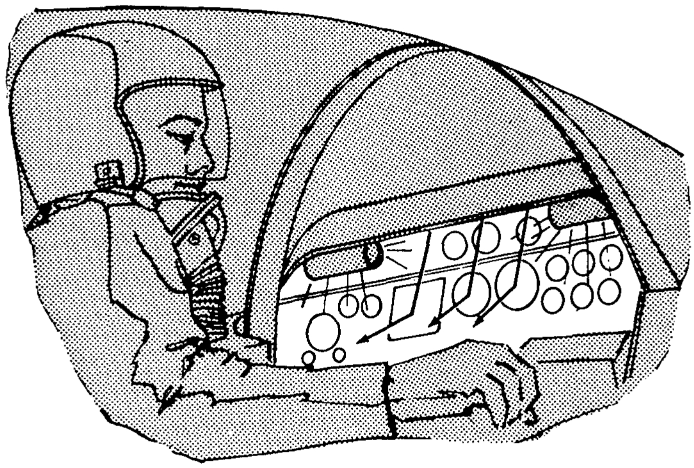


FIGURE 3–27. Glare shield and downward direction of floodlights for avoidance of windshield reflections.

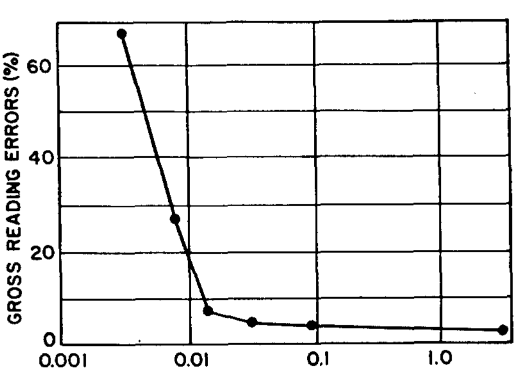


FIGURE 3–28. Instrument reading errors as a function of lighting (after Chalmers et al., 1950).

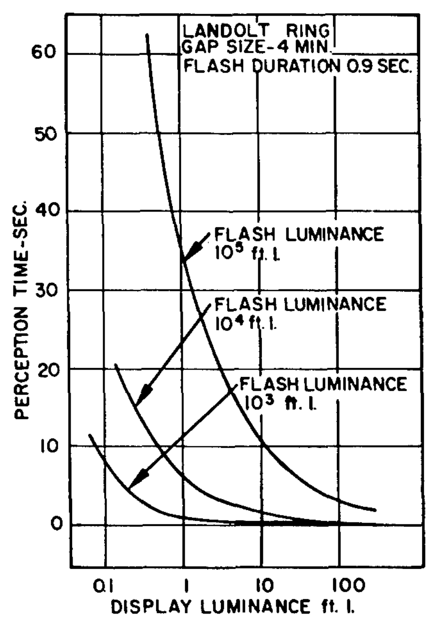


FIGURE 3–29. Effect of high flash luminance on perception time (after Brown, 1964).

TABLE 3–4. RECOMMENDATIONS FOR INDICATOR, PANEL, AND CHART LIGHTING

Condition of use	Recommendations		
	Lighting technique	Luminance of markings (ft.-l)	Brightness adjustment
Indicator reading, dark adaptation necessary.	Red flood, integral or both, with operator choice.	0.02–0.1	Continuous throughout range.
Indicator reading, dark adaptation not necessary but desirable.	Red or low-color-temperature white flood, integral, or both, with operator choice.	0.02–1.0	Continuous throughout range.
Indicator reading, dark adaptation not necessary.	White flood	1–20	Fixed or continuous.
Reading of legends on control consoles, dark adaptation necessary.	Red integral lighting red flood, or both, with operator choice.	0.02–0.1	Continuous throughout range.
Reading of legends on control consoles, dark adaptation not necessary.	White flood	1–20	Fixed or continuous.
Possible exposure to bright flashes.	White flood	10–20	Fixed.
Very high altitude, daylight restricted by cockpit design.	White flood	10–20	Fixed.
Chart reading, dark adaptation necessary.	Red or white flood with operator choice.	0.1–1.0 (on white portions of chart).	Continuous throughout range.
Chart reading, dark adaptation not necessary.	White flood	5–20	Fixed or continuous.

3.3.6 Selection of Lighting Method

There are two general techniques of instrument lighting: flood lighting and integral lighting. These are illustrated in Figure 3–30.

Flood Lighting

With this technique, light is provided by luminaires not integral to the indicator or panel. Because the light rays have a fairly large angle of incidence, the light source can be located above the indicator allowing specularly reflected light to go downward, away from the windshield or the operator's eyes.

The advantages of this technique are:

1. Uniform light distribution.
2. Illuminated decals, knobs, switches, and indicators.
3. Illuminated space between indicators which aids distance perception.
4. A minimum number of luminaires is required, and they can be made easily accessible for replacement of bulbs.
5. Luminaires do not obscure the edges of indicators as they may in integral lighting.

The disadvantages of flood lighting are:

1. Considerable light is scattered to other areas.
2. It is often difficult to position the luminaires without obstructing vision, or otherwise cluttering workspace.
3. Shadows are cast by indicator bezels as the angle of incidence is reduced.

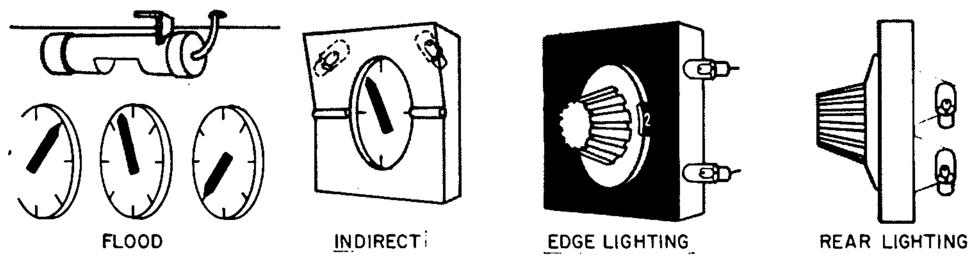


FIGURE 3–30. General methods of indicator and panel lighting.

Integral Lighting

Integral lighting that is tailored to or built into individual instruments or panels has these advantages:

1. Minimum light scatter to other areas.
2. Lighting tailored to each indicator or panel.
3. Concealed light fixtures do not obstruct the workspace.

Among the disadvantages are:

1. The surfaces between indicators and between illuminated panel markings are not lighted, thereby causing instrument faces and panel markings to appear to be floating in space.
2. For most systems it is difficult to obtain uniform illumination on all parts of an instrument, and on different indicators or markings on the same panel.

The more common types of integral lighting are the following:

Indirect lighting. Light is provided around the rim of the indicator by reflection from a light shield or by transmission through plastic. Usually the light sources are at the top so that light which escapes will go downward. With this method it is difficult to get uniform light distribution and avoid shadow areas. The fixtures are likely to occlude edges of instruments when oblique reading is necessary.

Rear lighting. Light is transmitted from the rear through translucent markings in an otherwise opaque covering over transparent plastic. Difficulties arise in lighting of pointers and other moving parts, and in making the lamps accessible for replacement.

Wedge lighting. For lighting of small and medium sized instruments, such as used in aircraft, light can be applied through wedge-shaped cover glasses. The thick edge of the wedge, and the source lamps, are placed at the top, so escaping light goes downward. Usually a reverse wedge glass is placed over the light conducting wedge, directing stray light downward.

Edge lighting. Light conducted by internal reflections through flat plates of transparent plastic escapes through apertures in the otherwise black coating over the plastic. Lamps inserted through polished holes in the plastic are distributed to give fairly uniform light. As normally prepared, the markings on the plastic are white under reflected light but can be any other light color. The illumination provided by the edge lighting can also be any desired color. Although instrument dials can be illuminated with edge lighting, there is difficulty in lighting instrument pointers. This method is better suited to the lighting of legends and other markings on control panels. (See MIL-P-7788D.)

Electroluminescent lighting. This is a relatively new type of integral lighting in which a laminated conducting plate glows when an electrical potential is applied, a technique particularly suitable for lighting of legends on panels of all types. As further improvements are made in this technique, it should provide an excellent solution to many workspace lighting applications. Major advantages are:

1. Uniform brightness and color of panel markings.
2. Increasing variety of colors which can be useful for coding.

For some electroluminescent applications, it is desirable to use low-reflectance coatings, which appear dark under reflected light. By this method markings can be made invisible until illuminated. Low-reflectance coatings are also beneficial for use in high ambient illumination conditions. The reduced surface reflections result in higher contrast between the illuminated markings and the background. (See Peterson and Smith, 1965.)

3.4 Visual Coding

On the highway or in the marketplace, man uses visual codes which warn him of danger or which convey information. Spots of color, numbers, letters, lines, arrows, lights, as well as color codes for wires and resistors are used for these purposes. Most codes are symbolic and must be learned, but once mastered, they can be an effective way of getting information. On the other hand, poorly designed codes can cause confusion and accidents. Although this section deals only with visual coding, the designer should consider codes giving information to the operator's sense of hearing or of touch, especially when he must direct his visual attention to many other information sources.

3.4.1 Visual Coding Methods

Information can be visually coded by color, size, luminance, shape, inclination, flash rate, and stereo depth. Table 3–5 compares and summarizes the advantages and disadvantages of these codes.

TABLE 3–5. COMPARISON OF CODING METHODS

Code	Number of code steps*		Evaluation	Comment
	Maximum	Recommended		
Color:				
Lights	10	3	Good	Location time short. Little space required. Good for qualitative coding. Larger alphabets can be achieved by combining saturation and brightness with the color code. Ambient illumination not critical factor.
Surfaces	50	9	Good	Same as above except ambient illumination must be controlled. Has broad application.
Shapes:				
Numerals & letters	Unlimited		--------	Location time longer than for color or pictorial shapes. Requires good resolution. Useful for quantitative and qualitative coding. Certain symbols easily confused.
Geometric	15	5	Fair	Memory required to decode. Requires good resolution.
Pictorial	30	10	Good	Allows direct association for decoding. Requires good resolution. Good for qualitative coding only.
Magnitude:				
Area	6	3	Fair	Requires large symbol space. Location time good.
Length	6	3	Fair	Requires large symbol space. Good for limited applications.
Brightness	4	2	Poor	Interferes with other signals. Ambient illumination must be controlled.
Visual number	6	4	Fair	Requires large symbol space. Limited application.
Frequency	4	2	Poor	Distracting. Has merit when attention is demanded.
Stereo-depth	4	2	Poor	Limits population of users. Highly limited application difficult to instrument.
Inclination	24	12	Good	Good for limited application. Recommended for quantitative code only.
Compound codes	Unlimited		Good	Provides for large alphabets for complex information. Allows compounding of qualitative and quantitative codes.

* The maximum number assumes a high training and use level of the code. Also a 5% error in decoding must be expected. The recommended number assumes operational conditions and a need for high accuracy.

Color Coding

The available number of colors for coding depends on whether colored lights or reflected colors are used. In general, more saturated colors can be obtained with lights than with surface pigments or reflected colors. But more coding steps are available with surface pigments under certain conditions. Surface pigments may lead to serious distortion if used under certain kinds of illumination. For example, color coding on maps and charts might be completely lost if they are used in rooms lighted with red lights. However, colored light codes are not affected by the color of the general illumination.

Colored lights for coding. The number of easily identifiable spectral colors depends on luminance, size (in visual angle), and color of the lights. The ten spectral colors in Figure 3–31 can be identified with a 2% error after a relatively short training period when the lights have a luminance of 1 mL or more (Chapanis and Halsey, 1956). By using high-purity interference filters, it is possible to approximate pure spectral colors. The angular substance of the color source should not be less than 15 min. of arc for highly accurate color recognition (Bishop and Crook, 1960).

Combining color coding of lights with other stimulus attributes of light provides a code with a larger number of identifiable categories. Bishop and Crook (1960) combined purity (saturation) and luminance with ten colored light sources (similar to those of Chapanis and Halsey, above). Highly trained observers were able to identify 60 such combinations (purity levels of 30% to 70% and luminances of 1, 10, and 100 ft.-L) with reasonable accuracy (5% error). However, when practice levels were not maintained, the observers lost some of their skill in identifying the stimuli. The studies by Bishop and Crook indicate that 30 compound stimuli could be easily learned and used as reliable operation codes.

Air and marine navigators view color signal lights at great distances. The small visual angles subtended by these sources can be considered as point sources of light. To be able to recognize a signal light color depends upon (a) the intensity of the light source, (b) the luminance of the background, and (c) the particular colors observed. (See Figure 3–21.) Yellow signal lights require the greatest intensity. Red and green signal lights are rarely interchanged. On the other hand, the yellow light was frequently confused with white. While these signals can be seen at lower intensities than shown on the graph, the colors cannot be identified correctly.

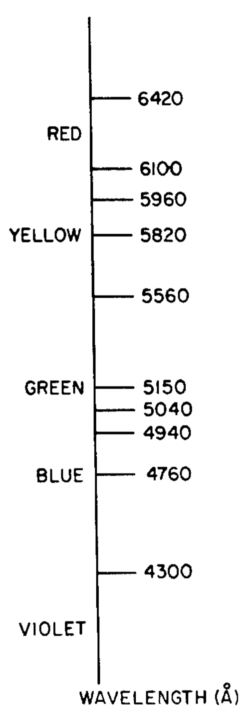


FIGURE 3–31. These ten spectral colors can be identified accurately with little training (Chapanis & Halsey, 1956).

TABLE 3–6. NINE EQUALLY DISCRIMINABLE SURFACE COLORS

Hues	Code number	Munsell book number	Excitation purity	Dominant wavelength
1	1.5	3R	37.2	629
2	3	9R	65.8	596
3	5.5	9YR	81.8	582
4	8.5	1GY	76.0	571
5	11.5	3G	27.5	538
6	15	7BG	35.0	491
7	18	9B	56.5	481
8	20.5	9PB	52.7	460
9	24	3RP	36.5	510

Conover (1959).

Surface colors for coding. Using 25 maximally saturated Munsell colors, Conover (1959) derived and validated an equal discriminability scale of surface colors. His studies indicate that a normal observer can be easily trained to identify nine maximally saturated surface colors with almost error-free performance. Table 3–6 recommends Munsell hues which are maximally discriminable. When fewer than nine colors are needed, the hues (1 through 9) should be selected to have the largest numerical separations. For example, if three hues are needed, select 1, 5, and 9 from Table 3–6. As indicated earlier, surface colors should subtend a visual angle of at least 15 min. of arc for accurate recognition.

If Munsell colors are selected so that they vary in chroma (saturation) and value lightness from black to white, 50 colors can be identified with high accuracy (Hanes and Rhoades, 1959). However, a high training level must be maintained to make accurate identification of these compound color codes, and illumination conditions must be held constant.

Color coding and color blindness. About 6% of healthy, adult males have markedly reduced sensitivity to colors. Most of these so-called color-blind people can correctly identify some colors—they merely do not see as many colors as color-sighted people do. Actually, only 0.003% of males are completely color-blind, i.e., are unable to see any color except black and white.

If colored light codes are to be presented to color-blind people, there are only three colors that should be used. These are aviation red, green, and blue as defined by the Army-Navy Aeronautical Specification AN-C-56. These colors must meet the exact requirements of the specifications because there are many other green, red, and blue hues that color-blind people cannot identify correctly. They can be used only at moderate distances. At great distances, blue is often confused with green. White or yellow should not be added to the code because color-blind persons confuse red with yellow and green with white.

The nine surface colors listed in Table 3–7 are ideal for coding because both color-sighted and color-blind people can recognize them relatively easily. The numbers refer to those in Federal Specification TT–C–595, except the one for blue (10B 7/6), which is a Munsell standard specification. Note that the color code, as contrasted to the colors in Table 3–6, combines saturation and lightness.

Shape Coding

The number of identifiable shapes in letter and number form is unlimited. For shapes in geometric form, however, learning and retention are limited to around 15 unless special training and continued practice are undertaken.

TABLE 3–7. SURFACE COLORS FOR COLOR-NORMAL AND COLOR-BLIND PEOPLE

Color	Spec. No.*	Color	Spec. No.*
Red	1110	Gray	1625
Orange	1210	Buff	1745
Yellow	1310	White	1755
Blue	10B 7/6	Black	1770
Purple	2715		

* From Fed. Spec. II-C–595 except for blue, which is from Munsell (1942).

Numbers and letters. Use of letter and number codes is restricted by space requirements and ease of learning associated meanings. Well-lighted letters and numbers having sharp outlines and high contrast can be identified accurately if they are greater than 10 min. of visual angle.

Although many letters and numbers can be used for coding, some are easily confused with others because of their shapes. For example, 9's are often confused with 0's and 8's. It is possible, however, to design these numbers to improve their discriminability.

The graph of Figure 3–32 (Green et al., 1953) shows the search time required to locate a specific numeral as a function of the density of numerals on the display. Two numeral orientations were used. In one case, the numerals were upright with respect to the observer. In the other case, numerals were randomly oriented. When a color code with half the numbers blue and half yellow was used, subjects found the specific number in about half the time if they were told the color. Subsequent studies by Green and Anderson (1956) verified that search time is approximately proportional to number of target color symbols. Targets of other colors on the display caused only slight increases in search time. These findings indicate that the target is more quickly found if (a) orientation is held constant, (b) the density of targets is decreased, and (c) color codes are added to assist search.

Numerical labels. Components are sometimes given coded three-digit numbers to help technicians locate a particular component. Adjacent switches therefore may have similar numbers. Troubles arise when these number codes are used as identifying symbols in passing verbal orders. Three-digit numbers are often transposed in memory, especially the second and third numerals. Serious accidents have resulted in power stations when the wrong circuit was switched on. It is better to use a number for the first and third character but a letter for the middle character. This avoids the possibility of switch #259 being pulled instead of #295.

Geometric and pictorial shapes. If geometric and pictorial shapes are needed for coding, choose shapes that are compatible and have association with objects to be coded. As a secondary criterion, design the symbols to be highly

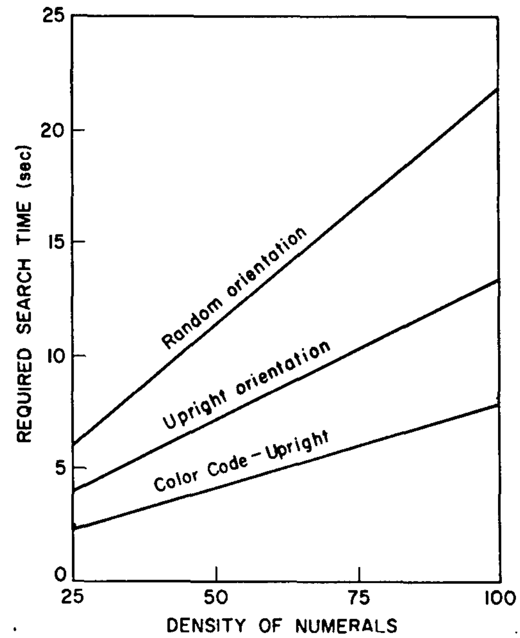


FIGURE 3–32. Search time to find a specific number as a function of numeral density for two orientations and color coding (Green et al., 1953).

discriminable. For example, the shapes shown in Figure 3–33 (Sleight, 1952) are easily identified and rarely confused. They are given in the order of the time required to identify each from a complex field of shapes. If the maximum dimension is at least 12 min. of visual angle, if the outline of the shape is sharp, and if the contrast is high, these forms can be identified correctly nearly 100% of the time.

Comparison of shape and color codes. Pictorial shapes depicting the real-world objects they represent, such as aircraft, ship, and missile silhouettes, can be easily learned, remembered, and used. For an extensive study on pictorial shapes for coding see Howell and Fuchs (1961). The data presented in Figure 3–34 indicate that pictorial shapes of military objects are superior to highly discriminable geometric forms. However, if highly similar symbols are used, search time and errors increase. When the task is a visual search for targets in a cluttered field of view, color codes have an excellent application because search times for these are smaller than for any of the three shape codes. However, learning the meaning of color codes requires more effort. Shapes combined with color permit the advantages of each to be used, as shown in Figure 3–34.

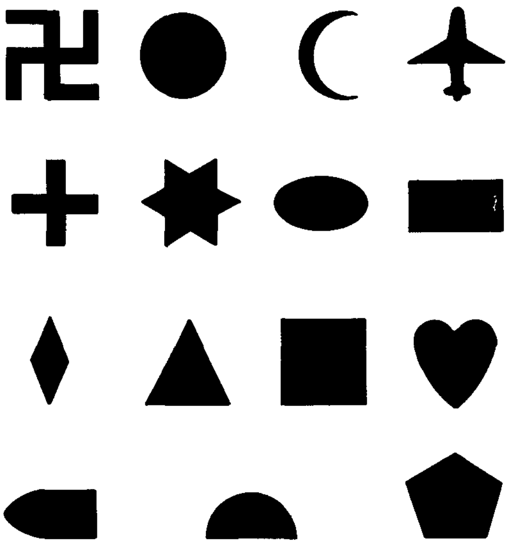
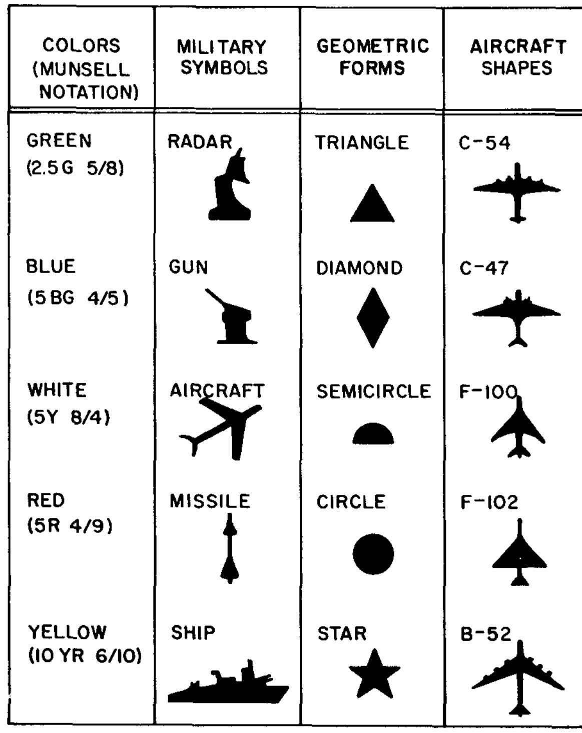


FIGURE 3–33. Fifteen highly discriminable shapes. They are presented (from top to bottom) in the order of ease of identification (Sleight, 1952).



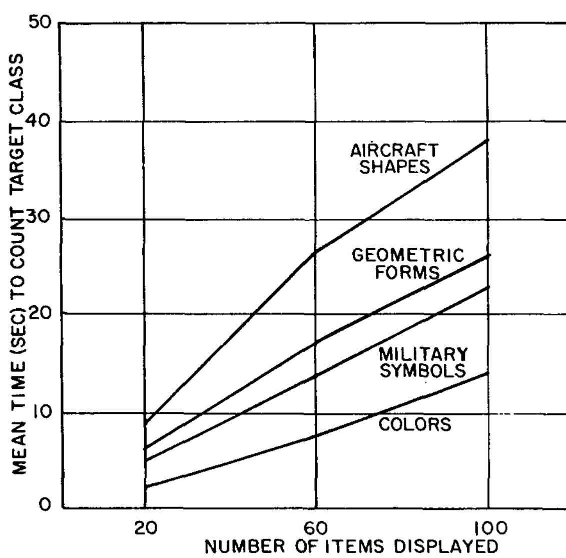


FIGURE 3–34. Average counting time as a function of display density, comparing color coding with the three shape codes (Smith and Thomas, 1964).

Shape variation. Codes can be generated by systematically varying some parameter of a basic geometric form, for example, using a rectangle and changing the height to width ratio. Muller et al., (1955) studied the absolute recognition of ellipses with axis ratios ranging from 0.0 (a line) to 1.00 (a circle). The axis ratios for maximally discriminable eight-symbol and five-symbol alphabets are presented in Figure 3–35.

The eight-symbol alphabet is the maximum size that can be identified accurately (i.e., less than 5% errors) after several hours of practice. The five-symbol alphabet can be identified with almost no error (less than 1%). To attain these accuracies the major axis should subtend a visual angle of about 30 min. of arc or more. Orientation of the symbol is not important.

This type of coding actually combines variations of shape, area, or size and would be particularly suitable for CRT display where dynamic variation in shape and area can be used as combined codes.

Magnitude Coding

Information can be coded by correlating symbol magnitude (area or linear extent), luminance, frequency, etc. with some actual characteristic of the target. The most obvious correlation is the area of the symbol with the size of the target; but the area of the symbol could represent the range of the target. Line length could represent velocity of the target.

In making up a code, the first step is to pick the upper and lower limits of the scale, then de-

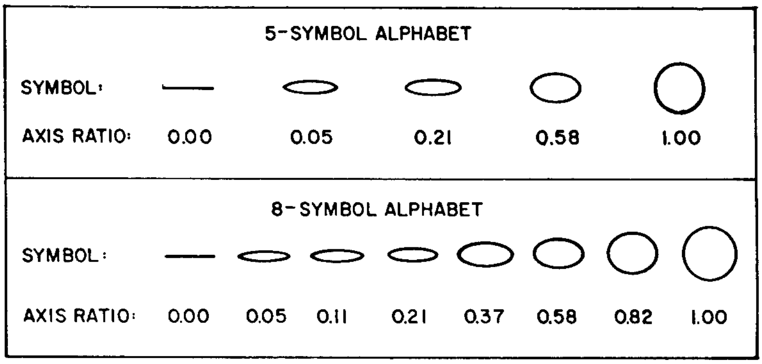


FIGURE 3–35. Five- and eight-symbol alphabets designed for maximum discriminability. The five-symbol alphabet can be used with less than 1% error (Muller et al., 1955).

cide how many steps there are to be. To space or scale the steps to get the same amount of accuracy all along the scale, apply the following general rule: in scaling a code, values are less likely to be confused if they are equally spaced on a logarithmic scale (i.e., have a constant ratio). Thus, five sizes should be 0.01, 0.032, 0.10, 0.32, and 1.0 in.^{2}, with each area being 3.2 times the next smaller one.

Area coding. Data on accuracy in reading symbols that are coded with various numbers of area steps are presented in Figure 3–36. This figure shows that as the number of areas exceeds five, there is a significant increase in operator errors in decoding. The data are based on near optimum, i.e., equal ratio, scaling of areas. Muller et al. (1955) have refined the equal ratio scaling to allow for the end point, or anchor, effects. Figure 3–37 provides three-, four-, and five-symbol area alphabets found to be maximally discriminable. Actually, there is only a slight deviation from the constant-ratio rule.

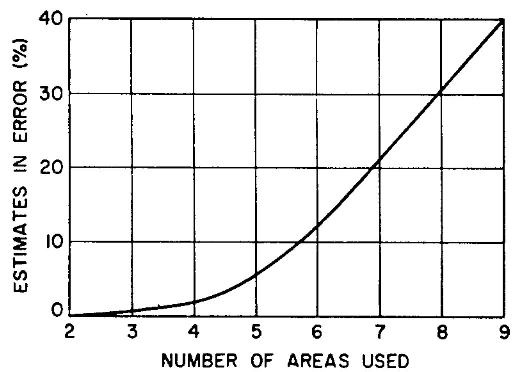


FIGURE 3–36. Accuracy of absolute identification of area magnitude as a function of the number of areas used. For each alphabet size, equal ratio scaling of areas was used (Baker and Grether, 1954).

Using these five areas will result in slightly less than 5% errors in decoding. The four-area code will result in less than 2% errors, and the three-area code less than 1% error. These data are valid if the ratio of smallest to largest area is at least six and when the smallest area subtends a visual angle of at least 5 min. of arc. For normal viewing distances (30 in.) these particular symbols ranged in diameter from 0.05 to 0.30 in.

Line length coding. The data on area in Figure 3–36 are generally applicable to codes using line length. The line lengths should be equally spaced on a logarithmic scale to provide near maximum decoding accuracy. Combining line length codes with other codes is discussed below under inclination coding.

Visual number coding. A target can be coded by correlating some dimension of information with the number of dots comprising the target signal. For example, a one-dot signal would represent a target value that is different from a two- or three-dot signal. In 1924, Oberly did a study exposing dots for less then 1/10 of a sec. so that observers were unable to count but had to estimate the number of dots exposed. (See Figure 3–38.) He found that errors are negligible for identification of signals coded by five dots or less; above six dots, errors rise rapidly. These

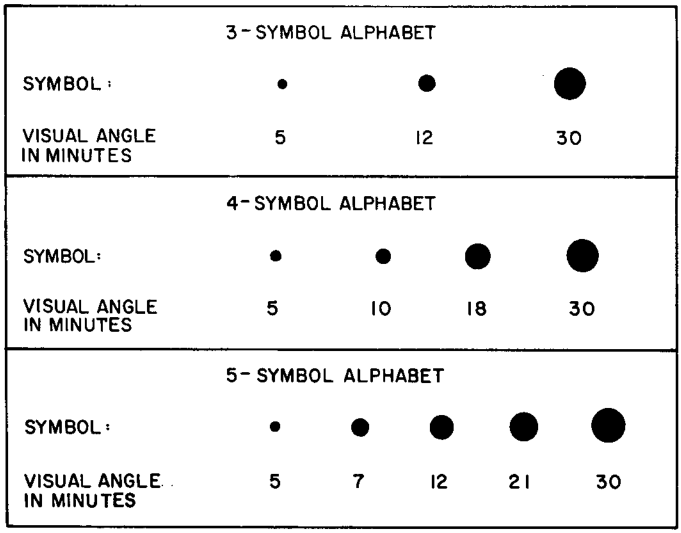


FIGURE 3-37. Maximally discriminable area codes for alphabet sizes of 3, 4, and 5. The 3-, 4-, and 5-symbol alphabets result in less than 1-, 2-, and 5-percent errors, respectively (Muller et al., 1955).

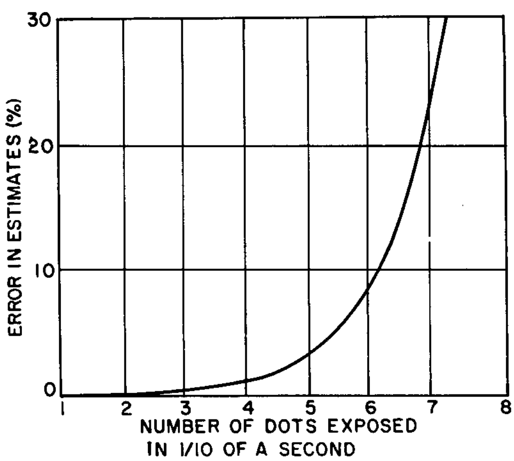


FIGURE 3-38. Accuracy of estimating the number of random dots seen as a function of the number of dots exposed. The viewer has only a fraction of a second to see the dots (Oberly, 1924).

accuracies are for immediate identification of the number of dots. If more time were allowed for observation, the accuracy would be greater, but the time required for identification is usually critical. If immediate identification is required, as many as five or six coding steps could be used to code signals. The data are for dots in random positions. Arrangement in familiar patterns or with uniform spacing improves accuracy even more.

Flash rate coding. Aerial beacons and marine navigational aids often use flashing light frequencies as a method of coding. A particular lighthouse may provide a flashing white light with a frequency of one short flash per second for identification. Flash rate might also be used on a display to represent some characteristic about the target symbol, e.g., target speed. Cohen and Dinnerstein (1958) trained subjects to identify flash rates varying from one flash every four sec. to 12 flashes every sec. Trained subjects were able to use only four rates with reasonably high accuracy. It was found that the maximally efficient frequencies should be equally spaced on a log scale. Because flashing codes are annoying, their application to displays should be limited.

Luminance coding. Under good viewing conditions, no more than four brightness steps can be used for coding. For most practical conditions, only two steps—bright and dim or light and dark—can be used. Coded items often interfere with each other; dimmer targets are often obscured or masked by brighter ones. Also, changes in ambient lighting decrease the accuracy of decoding.

Stereoscopic depth coding. Stereoscopic depth results from fusion of two slightly different pictures. Special equipment is needed to produce and to view two different displays of targets and coded information. A certain proportion of healthy adults (perhaps 10%) do not have stereoscopic depth perception.

Cohen (1955) studied the ability of observers to use binocular disparity (stereo depth) as a coding technique. He found large individual differences in the amount of disparity that the subjects could "fuse," by seeing one object in depth vs. seeing two objects on a flat surface. His data indicate that the number of coded stereo categories is probably limited to four.

Some specific applications of stereo coding may have merit; the use of stereo cues in photo interpretation has been shown to have some value. Signals submerged in noise (such as on radar presentations) could be "raised above" the noise by stereo viewing because noise is randomly located. However, no forceful demonstration supports this hypothesis. For further reading, papers by Vlahos (1965) and Leibowitz and Sulzer (1965) are recommended.

Inclination Coding

The inclination or orientation of a line on a display can be used as a code. An obvious application is to indicate target course or direction. A 24-symbol alphabet of inclination is shown in Figure 3–39 based on the work of Alluisi (1961). These 24 inclinations can be identified with less than 5% error after several hours of training by most adults. The lines can be as short as 0.1 in. but performance is better on longer lines. Using a large inclination alphabet does little to increase the coded information because observers make more errors. However, if two lines of different length are used, such as the hour and minute hands on a standard clock, larger alphabet codes can be read accurately. For additional inclination codes see Alluisi (1961).

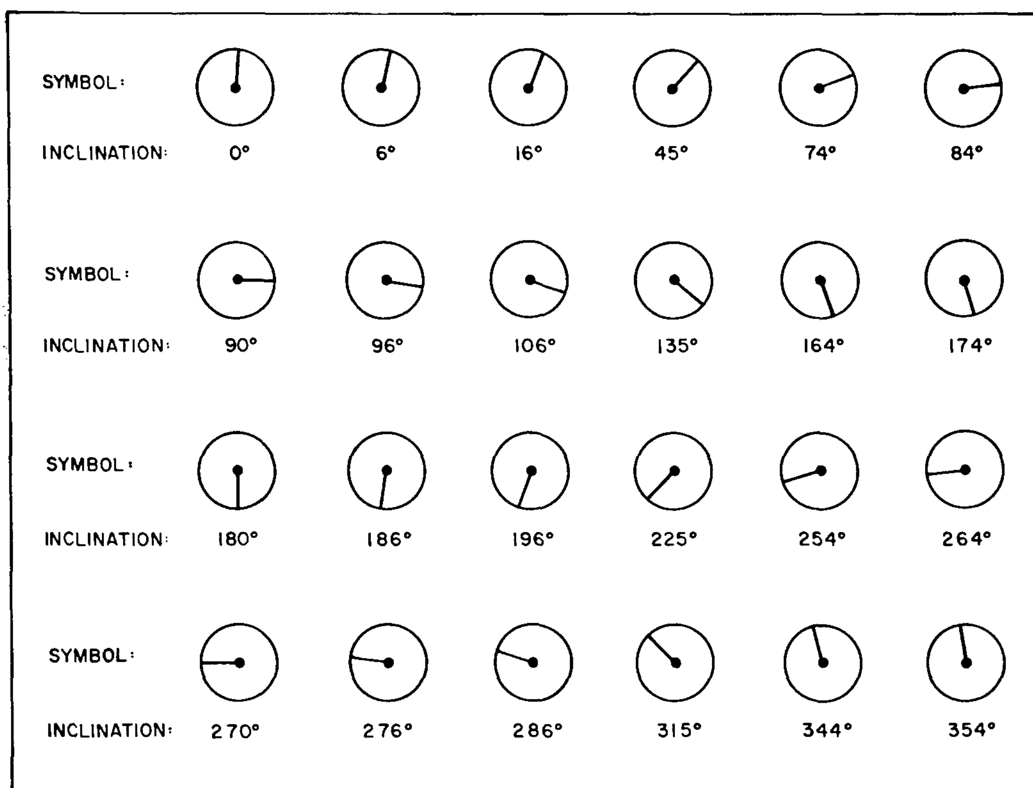


FIGURE 3–39. A 24-symbol alphabet of line inclination. The values used are equally discriminable. (Alluisi, 1961).

Code Compatibility

Coded information can be qualitative or quantitative. Identification and type-of-target information are qualitative. Speed, course, or number of targets are examples of quantitative information. Codes are easier to use and understand if qualitative codes are used for that type of information and quantitative codes are used for quantitative information. Color and geometric and pictorial shapes are usually best for coding qualitative information because the codes use qualitative differences. However, a given geometric shape (e.g., ellipse) can be used to code quantitative information by quantitative changes in ellipse ratio. Coded symbols which vary in magnitude (size, luminance lengths, frequency, etc.) are quantitative because the differences are solely quantitative. Alphanumeric codes can be used either qualitatively or quantitatively.

Compound Coding

A compound code conveys two or more types of information, while a single code, even one having a large alphabet, conveys information on only one parameter of the coded information. For example, the inclination of a line (single code) may convey information on target course. To designate target speed, a compound code could be used by conveying speed by line length. In addition, the color of the line could be used to identify the class of target. The use of compound codes permits more information to be presented on a single display. However, to avoid undue complexity of interpretation, the following recommendations should be noted:

1. Only one kind of information should be coded by one method. Compound coding for only one kind of information usually is less satisfactory than single coding if the single code used is the best available.
2. If two or more kinds of information are to be coded, the same number of coding methods should be used; do not use one coding method to code two or more kinds of information.

3.5 Warning and Signal Devices

Signal or warning displays usually indicate only two information values: safe and unsafe, right and left, on and off. Three-valued information, such as stop, caution, and go is less common. Still other displays, such as lighthouses and aircraft beacons, identify the presence of an obstacle or area of interest.

The three most common devices for presenting such simple two- and three-valued information are signal lights, mechanical "flag" signals, and auditory signals.

3.5.1 Warning Device Standards

Research on vigilance has shown that alertness on a job tends to diminish as time passes during a work period or day. Alertness will become particularly poor for low-probability signals, and most warning signals by their nature do not occur frequently. To get the operator's attention, signals of high attention value are needed. This value increases with size, brightness, loudness, or motion of a signal. But the signal should not be so intense that the operator is blinded or handicapped in any way.

A good warning device meets four requirements:

1. It attracts the attention of a busy or bored operator.
2. It tells him what is wrong or what action to take.
3. It should not prevent his continued attention to other duties.
4. The warning device should not itself be likely to fail or to give false warnings. Failures of the warning device should be easily detected (such as by a press-to-test button).

Grouping of Signals

By grouping signal lights or mechanical "flags" in appropriate patterns the designer can help the operator to learn what is wrong. Such patterns make a different signal easy to detect. A pictorial pattern can be of even more help to the operator. By showing him the positions of switches, valves, etc., as part of a diagram, the effects of their operation are easy to see.

An example of grouping signals is the "Christmas tree" on a submarine that shows the condition of all hatches which must be closed before diving. For each hatch there are two lights: a red one for "open" (i.e., unsafe), and a green one for "closed" (i.e., safe). Before diving, all green lights must be on and all red lights off. The location of a red light in the pattern, plus an identification label, tells the operator which hatch is open.

3.5.2 Selecting Signals

In selecting a warning signal for a particular application, the designer should consider the urgency, the other duties of the operator, and the other warning devices in the station. Unimportant warnings make operators neglect critical ones; too many of the same type are confusing.

Auditory signals should be used only for a few of the most urgent warnings. Such warnings, while attention-getting and independent of visual control, can interfere with speech communication and may be less suitable for indicating what is wrong or what to do.

Signal lights can tell the operator what to do by their location, labeling, color or other coding, but he must be looking in their general direction to notice them. Other visual signals, such as mechanical flags, have low attention value. They are practical for giving "on-off" types of information.

Lights for Warning Signals

A warning light signals a dangerous condition requiring prompt action by the operator. Such lights normally should be red because red means danger to most people. Other signal lights in the operator's vicinity should be of other colors. In addition, the location, luminance, and attention value of lights for warnings should be considered.

Location and identification. Because warning lights become less effective as they are moved out of the center of the field of vision, urgent warnings should always be within 30° of the operator's normal line of sight.

Sometimes many warning or caution signals must be used in a single operator station. This situation not only adds to the operator's identification problem, but it also creates the problem of finding panel space near the operator's normal line of sight. A master warning or caution light is a convenient solution to this problem as illustrated in Figure 3–40. The master light can be located near the operator's line of sight on the

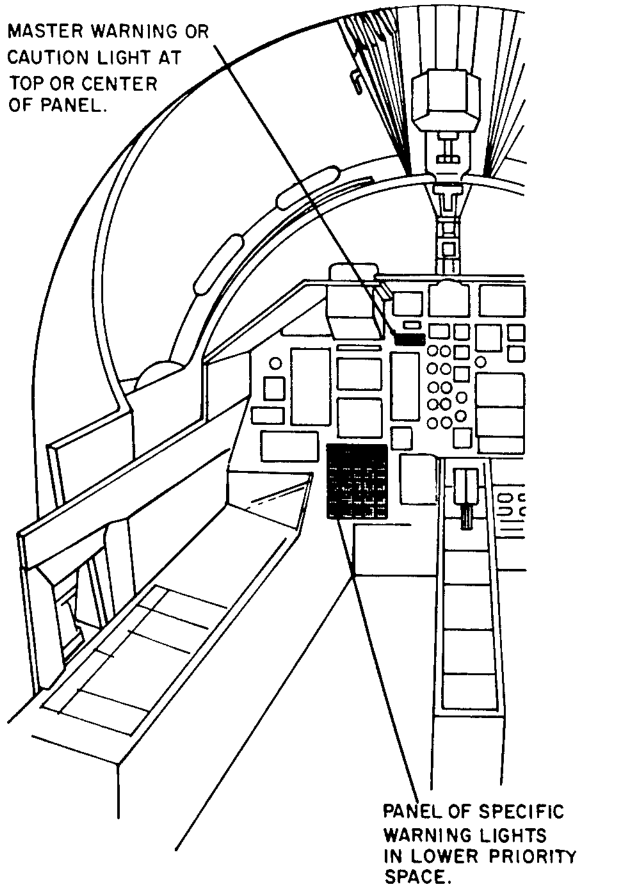


FIGURE 3–40. Application of master warning signal and separate annunciator panel.

instrument panel, and the specific warning panel can be located where space is more readily available. Any time the master light comes on, the operator checks the specific warning panel. For very urgent warnings the master warning light may be supplemented by an auditory warning (Siegel and Crain, 1960). This is particularly advisable if the operator's visual duties could cause him to miss the master light.

Because warning lights call for fast corrective action, their identification must be simple and positive. Ideally, each light should have a unique location and be easily distinguishable from other lights. As a further aid, the warning light may be near or built into the associated corrective control device.

Intensity. Warning lights should be bright enough to stand out clearly against the panel on which they appear under all expected lighting conditions, but they should not be so bright as to blind the operator. In work stations that are darkened at night, provision should be included for dimming the warning lights when other lights are dimmed; this can be accomplished by "tying in" warning lights with the same control used to dim panel or general station lights. In this way, the proper level is provided automatically.

Lighting of words vs. background. Where a warning light includes a word or abbreviation to identify the warning, the lettering should be dark on a lighted background. This will have greater attention value than lighted letters on a dark background. (See Siegel and Crain, 1960.) While this recommendation is most important where verbal information is built into a primary warning light, it is also advisable to light the background rather than the lettering on annunciator panels to help identify the warning message. When a printed word or abbreviation on a signal light merely gives status information (such as "on" and "off"), the lettering should be lighted and the background should be dark.

Size of lettering. Inasmuch as speed of reading is important, any letters or numerals on warning lights should be large. For dark letters on a lighted background viewed at a distance of about 28 in., a letter height of 0.2 to 0.3 in. is recommended.

Attention value. A flashing or interrupted light, while more attention-getting than a steady one, is also more disturbing, so these should be used only for the most urgent warnings. If they are used, however, their rate of flashes should be four per sec., with equal light and dark intervals.

If the operator's head position is relatively fixed, the warning light can be beamed directly at him rather than be diffused, as is usually the case. Although this arrangement is almost certain to get his attention, the possible blinding effect rules it out unless the emergency is great and other methods of presentation are not adequate.

In the same crew station (such as an aircraft cockpit) where lights are used for critical warnings, other lights should not be used for routine signals or status information. (See Crawford, 1962, 1963.) Much of the attention-getting value of warning lights would be lost. Even flashing lights lose their effectiveness if other flashing lights are present.

3.5.3 Caution and Status Signals

Signal lights are used for other purposes than warning; for instance, for indicating minor failures that do not demand immediate action. They can show the operational status of different system components. In these uses, getting attention is less important since the operator usually looks for the lights when he wants information.

The use of lights as caution signals is frequently overdone. A large array can confuse the operator. Some general rules for avoiding the excessive or improper use of such signal lights in a work station are:

1. Use red lights only for warnings.
2. If the station must be operated at times in relative darkness (as in aircraft cockpits, ship radar rooms, etc.), use signal lights only for intermittent and high-priority signals.
3. Avoid using signal lights for information that the operator can get in other ways, such as from control position.

3.5.4 Mechanical Flag Signals

Mechanical "flags," with word or pictorial labels, are often suitable for signalling status or caution, but not for warnings. The "flags" have low attention value, unless they are large and moving, so the operator must look in the direction of the flag to notice it.

As on-off indicators, mechanical flags are often superfluous. In some situations, however, it is desirable to indicate by a positive signal that some component of a system is or is not in operation. For example, inoperative or caged indicators can be clearly identified if mechanical or electrical detection of the component status is practical. In this case, the operator can be alerted by a flag or shutters moving into view such as that shown in Figure 3–41.

3.5.5 Long Distance Signal Colors and Devices

In transportation, visual signals must be effective over long distances where printed messages or other patterned signals are impractical. For such applications major reliance is placed on colored lights, flash rate, and position coding. Light houses, highway signal lights, railway sig-

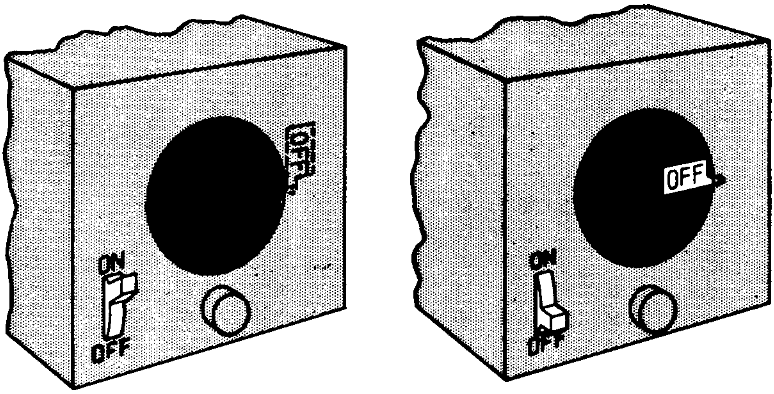


FIGURE 3–41. Use of flags as status indicators.

nals, and exterior lights on aircraft and ships are examples of this use of visual signals.

Color for Signal Lights

Red, green, yellow, and white lights have advantages for long-distance signalling because these colors can be produced with filters that transmit much of the visible light from tungsten lamps. Blue and purple are undesirable because the required filters transmit only a fraction of visible light.

Although yellow and white are visible at maximum range, yellow is easily confused with white. If both yellow and white are needed in the same signalling system, the possibility of confusion can be reduced by using a white of high color temperature, often called "lunar white," or "bluish white." If both yellow and green are used in the same signalling system, the green should be a bluish-green, to minimize the possibility of confusion with yellow.

Recommended choices for signal colors are presented in Table 3–8. For more detailed information and more exact specification of the colors, see U.S. Department of Commerce Handbook 95, 1964, and MIL-C-25050A (1963). See also the section on Visual Coding (3.4) and Figure 3–21.

Recognition Distance for Colored Signal Lights

If it is necessary to design a steady red or green signal light for a required recognition distance, the following formula (I.E.S. Lighting Handbook, 1966) may be used to give an approximation for daylight and clear air conditions:

$$D = 2000\, I, \qquad (3\text{–}7)$$

where D = distance in feet, and
I = intensity in candles, for a similar unit with a clear rather than colored lens.

The above formula applies to red and green signal lights that are steady rather than flashing.

TABLE 3–8. RECOMMENDED COLORS FOR SIGNAL LIGHTS

For maximum distance		For intermediate and near distance
Two colors required	Three colors required	Select number required, beginning at top
Red Green	Red Green White or Yellow	Red Green Yellow Blue White Purple

For yellow and white lights, the distance would be somewhat longer, for blue and purple lights, considerably shorter. For all lights, the recognition distance at night in clear weather will be much greater than that given by the formula. Haze or fog, on the other hand, can greatly reduce the distance.

If signal lights are flashing, the candle power value in the above formula should be reduced from that for a steady light, using the following formula (modified from Wesler, 1960):

$$I_E = \frac{t \times I}{0.09 + t} \qquad (3\text{–}8)$$

where I_E = effective intensity in candles,
I = intensity of steady light in candles, and
t = flash duration, in sec.

This formula assumes insignificant rise and decay times in relation to the period the light is on. It further assumes a flash duration of 0.1 sec. or greater. For other conditions, see Wesler (1960).

3.6 Mechanical Indicators

Mechanical indicators present information symbolically or pictorially by using a moving element, such as a pointer, a pictorial reference marker, or the fluid column in a thermometer. If the indicator has a scale, the scale may be the moving element, with the pointer or reference marker fixed, or, both the scale and pointer may move.

3.6.1 Method of Use

The use to which the operator will put the information presented to him is an important consideration in indicator design. Therefore, an analysis should be made of the type of action the operator will be expected to take while or after receiving information. Generally, indicators are used in one of the following ways:

1. For quantitative reading to an exact numerical value: reading time from a clock, reading heading from a compass, or r.p.m. from a tachometer.
2. For qualitative reading: this means judging the approximate value, trend, rate of change, or direction of deviation from a desired value. It differs from quantitative reading in that the operator does not read an exact numerical value. Examples include noting that a ship has veered to the right of a desired course, or that engine temperature is rising.
3. For check reading: this is verifying that a normal or desired value is or is not being shown. If the reading has deviated from the desired value, the operator might want to look more carefully to decide in which direction the value deviates and whether it is large enough to take corrective action.
4. For setting or adjusting an indicator to a desired value, perhaps adjusting one indicator to match another: setting target bearing and range into a fire-control computer or barometric pressure on an altimeter.
5. For tracking: this is intermittent or continuous adjustment of an indicator to maintain a normal or desired value (compensatory tracking). Holding a constant ship heading in rough sea is an example. To follow a moving target or reference marker with cross hairs is pursuit tracking.
6. For spatial orientation, judging position and movement in one plane or in three dimensions (the operator's own vehicle, a target, the relation between the two, or the location or movement of equipment components are examples). Navigation and fire-control indicators are usually of the spatial-orientation type.

The quantitative, qualitative, and check reading ways refer only to reading the display without considering the response. In the remaining ways, how the operator will respond to the displayed information is considered. Any single indicator usually will be used in more than one way, however.

3.6.2 Selecting Symbolic Indicators

Which one of the many types of mechanical indicators will be best for a particular application depends on the conditions and methods of use.

Table 3–9 lists the relative advantages and disadvantages of the three basic types of symbolic indicators as shown in Figure 3–42. While the direct-reading counter offers reading speed and a minimum of opportunity for error, the

TABLE 3–9. RELATIVE EVALUATION OF BASIC SYMBOLIC INDICATOR TYPES

For—	Counter is—	Moving pointer is—	Moving scale is—
Quantitative reading.	Good (requires minimum reading time with minimum reading error).	Fair	Fair.
Qualitative and check reading.	Poor (position changes not easily detected).	Good (location of pointer and change in position is easily detected).	Poor (difficult to judge direction and magnitude of pointer deviation).
Setting	Good (most accurate method of monitoring numerical settings, but relation between pointer motion and motion of setting knob is less direct).	Good (has simple and direct relation between pointer motion and motion of setting knob, and pointer-position change aids monitoring).	Fair (has somewhat ambiguous relation between pointer motion and motion of setting knob).
Tracking	Poor (not readily monitored, and has ambiguous relationship to manual-control motion).	Good (pointer position is readily monitored and controlled, provides simple relationship to manual-control motion, and provides some information about rate).	Fair (not readily monitored and has somewhat ambiguous relationship to manual-control motion).
Orientation	Poor	Good (generally moving pointer should represent vehicle, or moving component of system).	Good (generally moving scale should represent outside world, or other stable frame of reference).
General	Fair (most economical in use of space and illuminated area, scale length limited only by number of counter drums, but is difficult to illuminate properly).	Good (but requires greatest exposed and illuminated area on panel, and scale length is limited).	Fair (offers saving in panel space because only small section of scale need be exposed and illuminated, and long scale is possible).

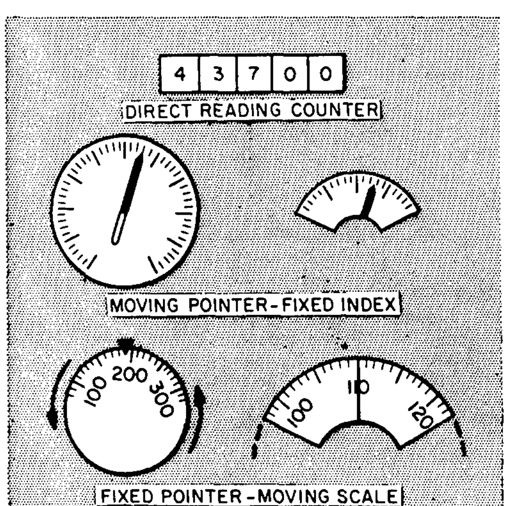


FIGURE 3–42. Basic types of symbolic indicators.

moving pointer type of indicator offers the greatest number of advantages. (For a review of research results see Chambers, 1956.)

Variations of basic types. There are several possible variations of the last two basic types of symbolic indicators:

1. Circular and Curved Scale With Moving Pointer. This design (A in Figure 3–43) is recommended because the circular scale permits a maximum of exposed scale length in a minimum of panel space. A compact indicator case, the long pointer, and rotational movement aid checking and qualitative reading. The circular scale is preferred to the curved scale for most applications.

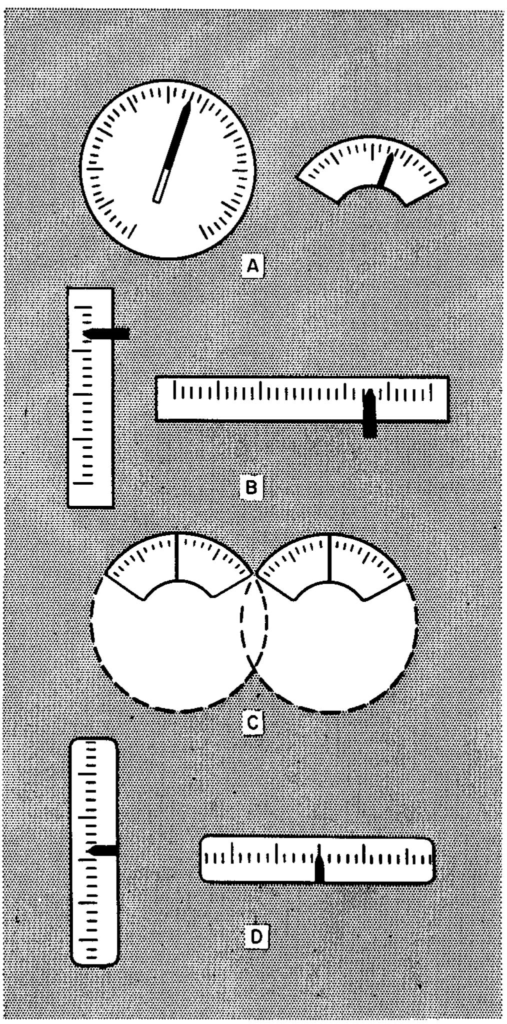


FIGURE 3–43. Variations of moving-pointer and moving-scale indicators.

2. Vertical and Horizontal Straight Scale With Moving Pointer. This design (B) is desirable for short-scale indicators. It permits a saving in front-panel space and provides a good means of pointer alignment and checking. The shorter pointer and lack of rotational movement, however, make it more difficult for the eye to notice change in the position of the pointer.

3. Circular and Curved Scale With Fixed Index. When a fully exposed scale may not be necessary, the partially exposed scale (C) can be generally recommended. This arrangement permits a large range of values in a limited panel space. Overlapping the covered scale portions also saves panel space. However, tracking and orientation indicators, such as magnetic heading indicators, should have the full scale exposed.

4. Vertical and Horizontal Straight Scale With Fixed Pointer. In this design (D), a moving straight scale behind an open window is provided. The moving drum or tape is suitable for presenting a large range of values that are to be read quantitatively.

Long-scale indicators. Conventional methods of presentation are usually inadequate when a large range of values must be displayed on one indicator. An example would be the need for an altimeter with a range of from 0 to 70,000 ft. with reading precision necessitating graduated intervals corresponding to every 20 ft. throughout the entire range. Thus, an indicator with a moving pointer and a fixed scale would require an impractical scale length of about 245 in. Several solutions to the problem have been evaluated and can be listed as follows (see Figure 3–44):

1. Direct-Reading Counter. This design (A) presents a large range of quantitative values, and requires very little panel space. It is not satisfactory, however, for qualitative reading or tracking.

2. Moving Pointer-Counter Combination. This design (B) is a recommended solution for long-scale indicators for check and qualitative reading, but it is slightly inferior to counters alone for quantitative reading (Grether, 1949). For use as an aircraft altimeter, it is advisable to add a single drum, indicating hundreds of feet, to the right of the counter, which indicates thousands (Hill and Chernikoff, 1965).

3. Moving Pointer-Moving Scale Combination. This design (C) is to be avoided because

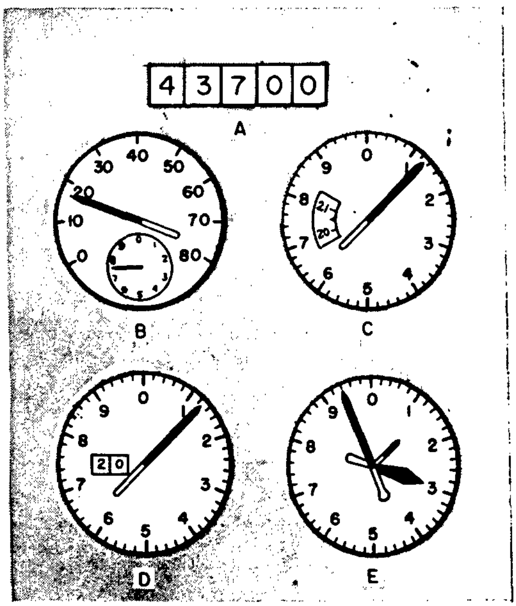


FIGURE 3–44. Indicators for long-scale applications.

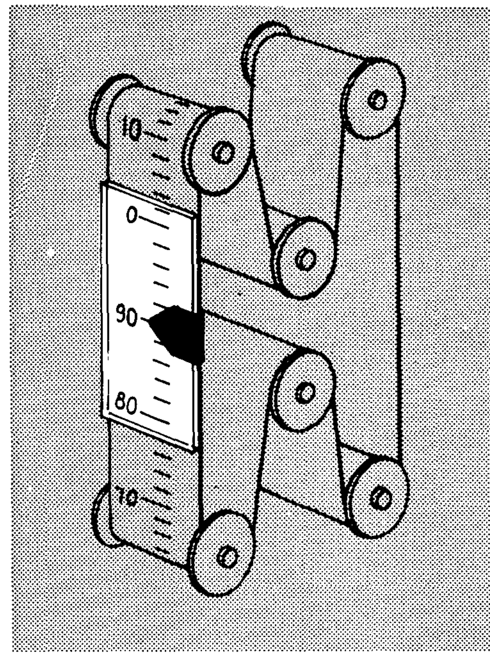


FIGURE 3–45. Use of moving tape as an indicator.

the error probability is high in combining these two types of indications.

4. Subdial. For certain applications it is possible to compress a long scale by means of a vernier subdial. With this design (D), the main dial provides gross values and the subdial permits more precise readings.

5. Multiple Pointers. In this design (E in Figure 44), there is one multirevolution pointer for fine readings and one or more other pointers for gross readings. With this design, the reader must mentally combine several separately indicated values, and the probability of gross reading errors is high (Grether, 1949). Pilot errors in reading altimeters of this type have caused many aircraft accidents.

6. Moving Tape With Fixed Index. A straight scale with a moving tape (see Figure 3–45) can be used for long-scale presentation. Although front panel space is saved with this design, it takes up more area in back of the panel. (For research data on such displays for presenting altitude, see Mengelkoch and Houston, 1957.)

3.6.3 The Design of Symbolic Indicators

A few of the terms to be used in this section and their definitions are as follows (Figure 3–46):

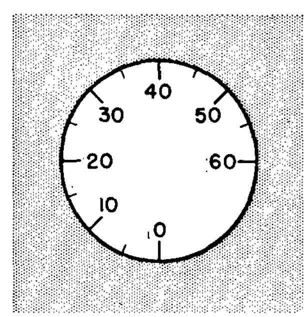


FIGURE 3–46. Explanation of instrument scale terms.

1. Scale Range. This is the numerical difference between the highest and lowest value on a scale.

2. Numbered-Interval Value. This is the numerical difference between adjacent numbers on a scale.

3. Graduation-Interval Value. This is the numerical difference represented by adjacent graduation marks.

Scale selection. Before selecting a scale for a mechanical indicator, a designer should decide

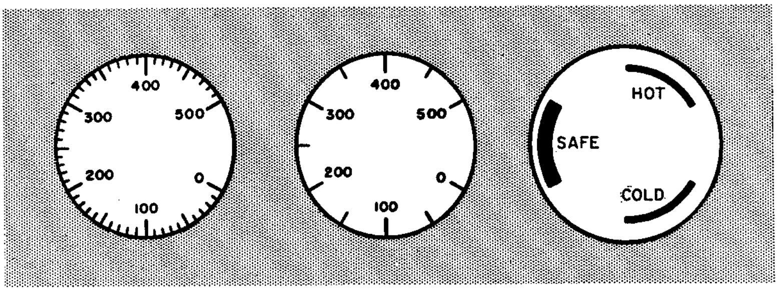


FIGURE 3–47. Illustration of varying degrees of scale complexity.

on the appropriate scale range and should estimate the reading precision required. Figure 3–47 gives examples of different levels of scale preciseness. The designer should select the least precise scale that fulfills the needs of the operator.

If possible, all displays should indicate values in an immediately usable form without mental conversion on the operator's part. An example of transformed scale values can be found in those jet-aircraft-engine tachometers that have been calibrated in percent r.p.m. rather than actual r.p.m. For the pilot, this has several advantages. Maximum r.p.m. differs for different engine models and types. Transforming the scale values into percent r.p.m. relieves the pilot of the necessity of remembering operating r.p.m. values for different engines. In addition, the range from 0 to 100% is more easily interpreted than a range of true values, such as 0 to 8,000 r.p.m., and the smaller numbers on the dial make a more readable scale. In Figure 3–48 the two tachometers illustrate these advantages. The less cluttered dial on the left can be read more precisely than the one on the right.

Interval values. Some combinations of graduation-interval values and scale-numbering systems are more satisfactory than others. The following recommendations will assist the designer in selecting the most readable scale (see Figure 3–49):

1. The graduation-interval values should be one, two, five, or decimal multiples thereof. Graduation-interval values of two are less desirable than values of one or five. Table 3–10 gives examples of good, fair, and poor numerical progressions.

2. There should be no more than nine graduation marks between numbered graduation intervals.

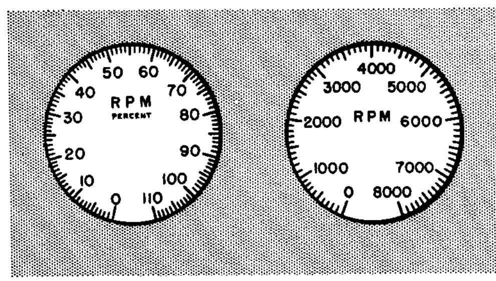


FIGURE 3–48. Comparing percentages and actual values of functions being displayed.

3. Normally, scales numbered by intervals of 1, 10, 100, etc. and subdivided by ten graduation intervals are superior to other acceptable scales.

4. Ordinarily, scales should be designed so that interpolation between graduation marks is not necessary; but when space is limited, it is better to require interpolated readings than to clutter the dial with crowded graduation marks.

With this information in mind, the designer can select the most suitable scale from Figure 3–49. Assuming sufficient space, a scale that is to be read to the nearest 1, 10, or 100, etc. should be selected from those scales with graduation-interval values of 1, 10, or 100. If accuracy to the nearest 0.5, 5, or 50 units is required, scales with 0.5, 5, or 50 graduation interval values should be selected, and so on.

Scale interpolation. Scales that are to be read quantitatively should be designed to be read to the nearest graduation mark. For instance, if we assume a scale range of 50 and a scale that is to be read to the nearest unit, the preferred scale would be numbered by tens with a graduation mark for each unit as shown in Figure 3–50 (A). If the space available for this scale were restricted to 1¾ in., the same scale would appear

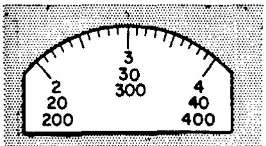
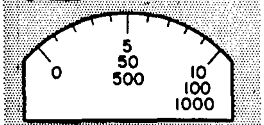
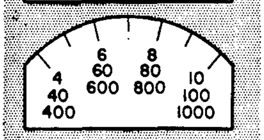
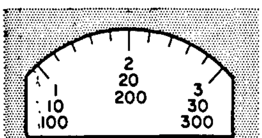
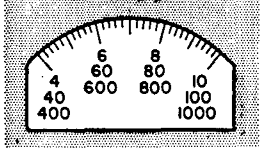
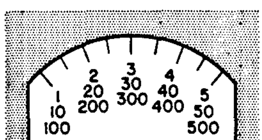
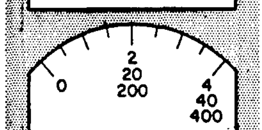
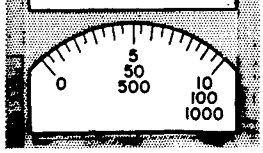
GRADUATION INTERVAL VALUE	RECOMMENDED SCALES	NUMBERED INTERVAL VALUE	GRADUATION MARKS MAJOR	INTERMEDIATE	MINOR
0.1, 1, 10		1, 10, 100	X	X	X
		5, 50, 500	X		X
		2, 20, 200	X	X	
0.2, 2, 20		1, 10, 100	X		X
		2, 20, 200	X	X	X
0.5, 5, 50		1, 10, 100	X	X	
		2, 20, 200	X	X	X
		5, 50, 500	X	X	X

FIGURE 3-49. Recommended scale ranges and interval values.

TABLE 3-10. EXAMPLES OF GOOD, FAIR, AND POOR PROGRESSIONS FOR SCALE NUMBERS

Good					Fair					Poor				
0.1	0.2	0.3	0.4	0.5	0.2	0.4	0.6	0.8	1.0	0.25	0.5	0.75	1.0	
1	2	3	4	5	2	4	6	8	10	2.5	5	7.5	10	
10	20	30	40	50	20	40	60	80	100	25	50	75	100	
100	200	300	400	500	200	400	600	800	1000	250	500	750	1000	
0.5	1.0	1.5	2.0	2.5						0.4	0.8	1.2	1.6	1.8
5	10	15	20	25						4	8	12	16	18
50	100	150	200	250						40	80	120	160	180

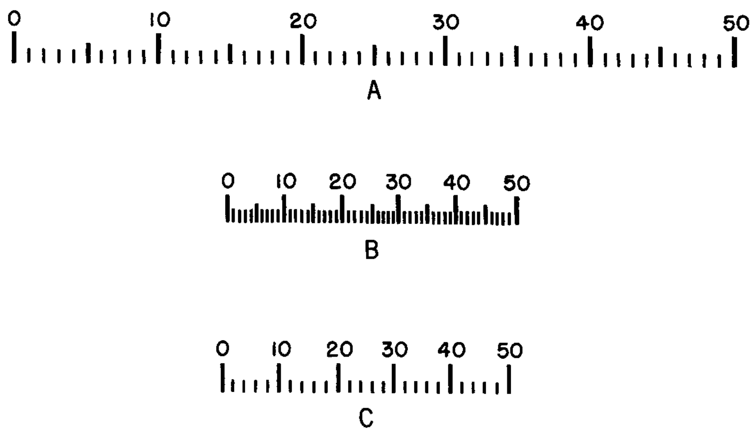


FIGURE 3–50. The effect of scale size and scale range on spacing and interpolation. Scale B is not recommended.

as in Figure 3–50 (B), but the graduation marks on this scale are too crowded to be read accurately and rapidly under low illumination; the midpoints are only 0.04 in. apart, and this is 0.03 in. less than the recommended minimum for a 28-in. viewing distance at low illumination. In such situations, it is better to design a scale in which interpolation is necessary as, for example, in Figure 3–50 (C). This scale has a graduation-mark spacing of 0.08 in., which is acceptable for low illumination. Also, this scale requires only a simple interpolation of one unit between graduation marks. When space is limited, it might be necessary to interpolate in fifths or even tenths, but such interpolation will increase reading errors. (For further information on the size of dials in relation to interpolation, see Murrell et al., 1958; and Churchill, 1956.)

Special scales. Certain applications of scales require unique design features and, in such cases, compromises with the recommendations listed above must be made. Examples of these special cases include multirevolution and nonlinear scales.

1. Multirevolution Scales. The heading indicator in Figure 3–51 (A) contradicts the scale-numbering recommendations in Table 3–10 because it uses the numerical progression of 30, 60, 90, etc. This, however, represents a compromise between good numbering progression and manageable dial size. On the heading indicator, the cardinal points of the compass (north, east, south, and west) serve as anchoring points in interpretation of the indication, and progression by 30's appears to be a good solution when the dial is relatively small. When the dial can be made large enough, however, the numbered graduation intervals should progress by 10's.

2. Nonlinear Scales. Nonlinear scales condense a large range into a relatively small space but in such a way as to permit sensitive readings at certain critical ranges of the scale. In situations where error tolerances are a constant percentage of the indication, logarithmic scales are very suitable if they contain sufficient numbered graduation marks to minimize errors as a result of the linear-scale reading habits of operators. Figure 3–51 (B) illustrates such a scale.

Scale design. For ease of reading, sufficient separation must be maintained between scale indices. In addition, cues should be provided for determining differences between major and minor graduation marks. More specific recommendations for scale dimensions depend on the illumination level at the dial face.

For normal illumination: The following recommendations apply to indicators that are reasonably well illuminated. Assuming high contrast between the graduation marks and dial face, illumination levels on the dial face above 1 ft-L. and reading distances of 13 to 28 in., the following recommendations should be observed (see Figure 3–52):

1. The minimum width of a major graduation mark should be 0.0125 in.

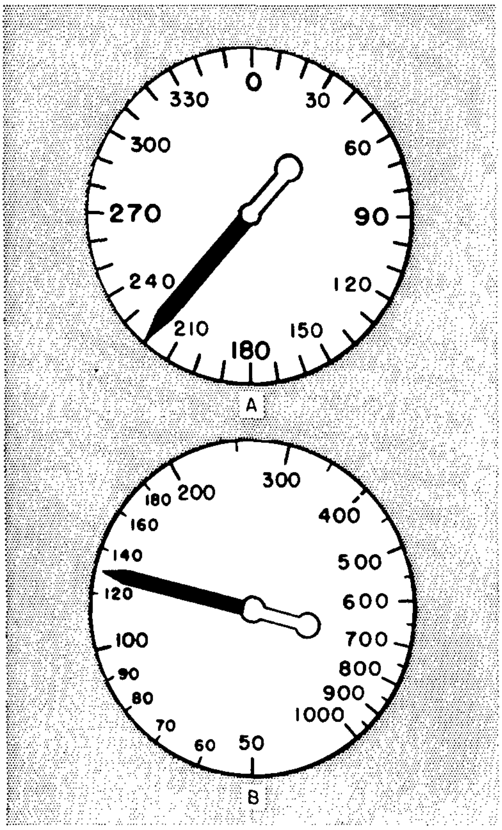


FIGURE 3–51. Illustration of special scales.

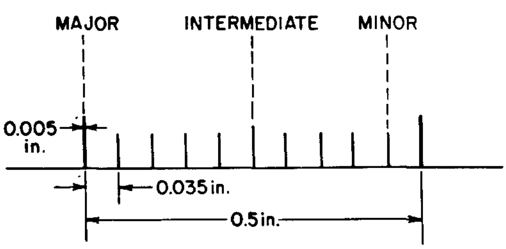


FIGURE 3–52. Recommended minimum scale dimensions for high illumination—above 1.0 ft.-L (28-in. viewing distance).

2. Although graduation marks may be spaced as close as 0.035 in., the distance should never be less than twice the stroke width for white marks on black dial faces or less than one stroke width for black marks on white dial faces.

3. The minimum distance between major graduation marks should be 0.5 in.

4. The height of major, intermediate, and

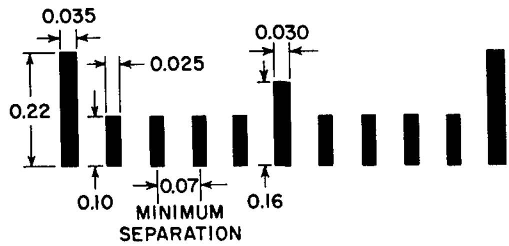


FIGURE 3–53. Recommended minimum scale dimensions for low illumination—0.03 to 1.0 ft.-L (28-in. viewing distance).

minor graduation marks should not be less than 0.22, 0.16, and 0.09 in. respectively.

For low illumination: When indicator scales must be read in other than normal illumination, as for example in a dimly lighted aircraft cockpit or ship bridge at night, special care must be taken to gain maximum readability for the scale design. Under these conditions, the additional aid of varied stroke widths of major and minor graduation marks becomes important.

The recommended minimum dimensions shown in Figure 3–53 are applicable to scale design for low illumination levels and should be followed whenever possible. These dimensions should not be considered as fixed values in the sense that other factors, such as scale size, number of graduation marks, and importance of indication, are not given equal consideration, but as models for relative dimensions. For instance, a reading distance of 28 in. is assumed for the dimensions in Figure 3–53; for other reading distances, a proportional increase or decrease in the recommended scale dimensions is in order.

Numeral and letter size and style. Designers should strive for maximum legibility of numerals and letters on indicator dials, panels, and consoles taking into account space restrictions and range of illumination. Recommendations in the following paragraphs apply to general floodlighted and integrally lighted indicators as well as those viewed under ordinary ambient lighting.

The numeral style defined by Military Standard 33558 (1957) is preferred, although other numerals of the same simple style also are acceptable. (See Atkinson et al., 1952.) The width of all numerals should be 3/5 of the height (see Table 3–11), except for the "4," which should be one stroke width wider than the others, and the "1," which should be one stroke width wide. In

TABLE 3–11. RECOMMENDED NUMERAL AND LETTER HEIGHTS

Nature of markings	Height (in.)*	
	Low luminance†	High luminance‡
Critical markings, position variable (numerals on counters and settable or moving scales)	0.20–0.30	0.12–0.20
Critical markings, position fixed (numerals on fixed scales, control and switch markings, emergency instructions)	0.15–0.30	0.10–0.20
Noncritical markings (identification labels, routine instructions, any markings required only for familiarization)	0.05–0.20	0.05–0.20

* For 28-in. viewing distance. For other viewing distances, increase or decrease values proportionately.
† Between 0.03 and 1.0 ft.-L.
‡ Above 1.0 ft.-L.

addition, the stroke width should be from 1/6 to 1/8 of the numeral height.

Since capital letters are readable at greater distances than lower-case letters, single-word labels or short identification sentences should be in "all caps" instead of "initial caps and lower case." The letter style defined by Military Standard 33558 (1957) is preferred, but commercial types also may be used if they are of the same simple style. (See Brown, 1953.) The width of all letters should be 3/5 of the height (see Table 3–11 for recommended heights) except for the "I," which should be one stroke width, and the "m" and "w," which should be about 1/5 wider than the other letters. Again, the stroke width should be from 1/6 to 1/8 of the letter height.

Scale layout. Numbers should increase in a clockwise direction on circular and curved scales, from bottom to top on vertical straight scales, and from left to right on horizontal straight scales. (See A in Figure 3–54.) Except on multirevolution indicators, such as clocks, there should be a scale break between the two ends of a circular scale (Kappauf, 1951). When the scale has a break in it, the zero or starting value should be located at the bottom of the scale (B), except when pointer alignment is desired for check reading. In this case, the zero or starting value should be positioned so that the desired value is located at the "nine o'clock" position (C). The zero or starting value on multirevolution indicators should be at the top of the scale (D), which is where the operator will expect it to be

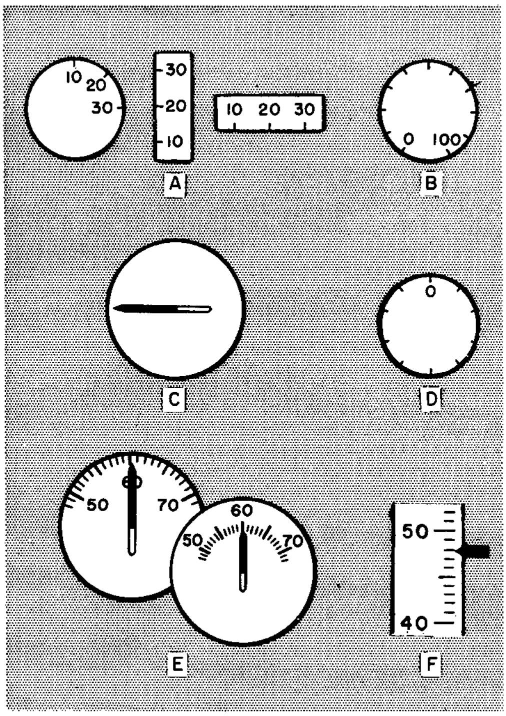


FIGURE 3–54. Different types of scales.

because of his experience in reading clocks, heading indicators, etc.

In general, on circular scales it is better to place numerals inside of the graduation marks to avoid constricting the scale. If ample space exists, however, the numbers should be placed outside of the marks to avoid being covered by the pointer (E). On vertical and horizontal straight scales, the numbers should be located on the side of the graduation marks opposite the pointer, and the graduation marks should be aligned on the side nearest the pointer and "stepped" on the side nearest the numbers (F). It is further recommended that the pointer be to the right of vertical scales and underneath horizontal scales.

Zone marking. Zone markings indicate various operating conditions on many indicators: such as operating range, upper, lower or danger limits, caution, etc. These zone markings can be color or shape coded. Because of frequent need for relocation, they are placed on the indicator's window in preference to the dial face.

Color coding: Figure 3–55 shows recommended colors associated with various operating conditions. If the indicator is illuminated with colored light, particularly red, the designer is cautioned not to use color coding. Colors are not readily distinguishable when illuminated by colored light.

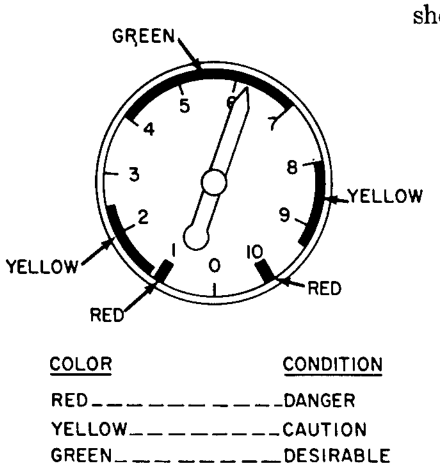


FIGURE 3–55. Color coding of instrument range markers.

Shape coding: Zoning marking shapes can indicate various operating conditions. The shapes in Figure 3–56 (Sabeh et al., 1958) are recommended because they are easy to learn and are distinguishable under low and colored illumination.

Pointer design. While simplicity is the basic principle in pointer design, pictorial shapes can be added for aiding indicator identification or interpretation. (For a review of research data on pointer design, see Spencer, 1963.) Other recommendations for pointer design are listed as follows (see Figure 3–57):

1. Pointers should extend to, but not overlap, the minor scale markings (A).

2. The pointer should be as close to the dial face as possible to minimize parallax (B).

3. For most applications, the section of the pointer from the center of rotation to the tip should be the same color as the dial markings.

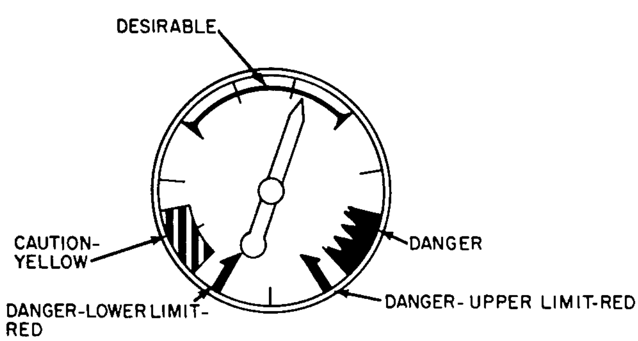


FIGURE 3–56. Shape coding of instrument range markers (Sabeh et al., 1958).

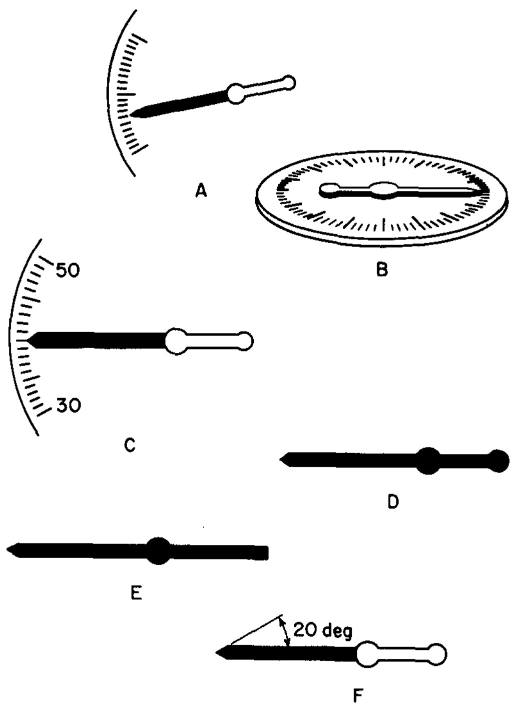


FIGURE 3–57. Pointer design recommendations.

The remaining portion of the pointer should be the same color as the dial face (C).

4. For indicators designed for horizontal pointer alignment, the tail end of the pointer should extend beyond the center of rotation by an amount equal to about one half of the head of the pointer (D).

5. In cases where it is necessary to read the position of the tail of the pointer as well as the head (e.g., heading indicators), the tail should extend beyond the center of rotation by an amount equal to about three-fourths of the head, but the tail should be blunt rather than pointed to avoid confusion with the head (E).

6. Recommended pointer tip angle should be 20° as illustrated in Figure 3–57 (F).

3.6.4 Motion of Moving Element

Other considerations in indicator design concern the direction of motion of the moving element in relation to: (a) the numerical value being indicated, (b) the direction of motion of the associated control, and (c) the direction of motion of the vehicle or component to which the indicator refers. (For a review of research data on direction of motion see Lovelace, 1962.)

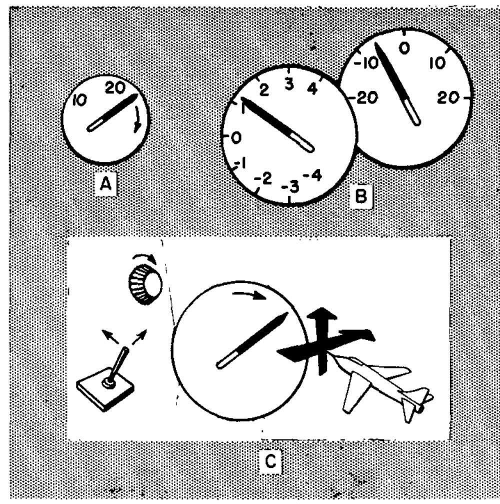


FIGURE 3–58. Recommended direction of motion on circular moving-pointer indicators.

Moving-Pointer, Fixed-Scale Indicators

Circular scales: The magnitude of reading should increase with a clockwise movement of the pointer (see A in Figure 3–58). Where positive and negative values around zero are being displayed, the zero should be located at the "nine" or the "twelve o'clock" position (B). This arrangement provides for pointer increases upward or to the right, which are expected directions. Thus, positive values will increase with clockwise movement of the pointer and the negative values will increase with counterclockwise movement. In addition, clockwise movement of a pointer should result from clockwise movement of the associated knob or crank; movement forward, upward, or to the right of an associated lever; or movement upward or to the right of the associated vehicle or component (C).

Straight scales: The pointer should move up or to the right to indicate an increase in magnitude (see Figure 3–59). This movement should result from: (a) clockwise rotation of the associated knob or crank; (b) movement forward, upward, or to the right of a lever; or (c) movement upward or to the right of the associated vehicle or component.

Moving-Scale, Fixed-Pointer Indicators

Circular scales: Inconsistencies exist between moving circular scales and associated control movements. It can be seen from the following

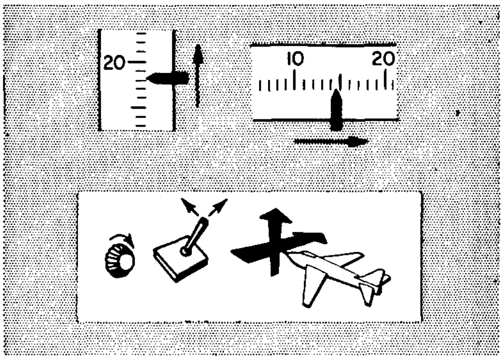


FIGURE 3-59. Recommended direction of motion on straight-scale, moving-pointer indicators.

description of three recommended practices that one of them must be violated when designing circular moving scales:

1. Scale numbers should increase in a clockwise direction around the dial. Thus, values on moving circular scales will increase with counterclockwise movement of the dial face.
2. The direction of movement of the associated control should be compatible with the direction of movement of the dial. Thus, clockwise movement of the control should result in clockwise movement of the dial.
3. Clockwise movement of a control should result in an increase in function.

If the first recommended practice is compromised, operators tend to make final setting errors. If the second practice is compromised, operators are likely to err in the initial direction of turn (Bradley, 1954). If the third practice is compromised, a standardized control-movement–system-movement relationship is violated. (See Chapter 8.)

Because of these incompatibilities, it is recommended that moving-pointer indicators be used in preference to moving-scale indicators wherever possible. If, however, a moving-scale indicator must be used, the following recommended practices in the design and use of circular moving scales will minimize the effects of the ambiguities:

1. The numbers should progress in magnitude in a clockwise direction around the dial face so that counterclockwise movement of the dial face increases the readings. (See A in Figure 3-60.)

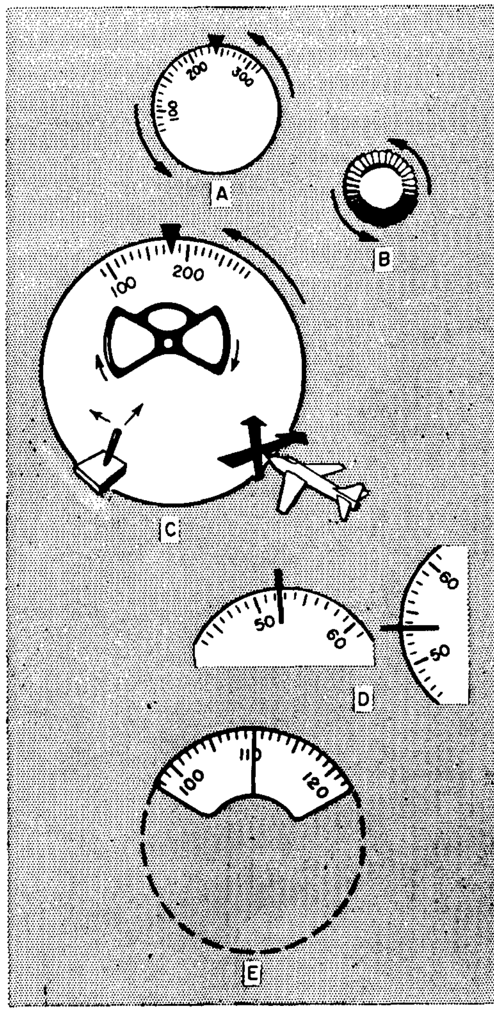


FIGURE 3-60. Recommended direction of motion on circular moving-scale indicators.

2. If the associated control (B) has no direct effect on the behavior of the vehicle (e.g., tuning in radio stations, monitoring electronic equipment, etc.), the scale should rotate counterclockwise (increase) with *counterclockwise* movement of the associated knob or crank.

3. If the associated control (C) does have a direct effect on the behavior of the vehicle (e.g., increases or decreases speed, direction, etc.), the scale should rotate counterclockwise (increase) with *clockwise* movement of the associated knob, wheel, or crank; with movement forward, upward, or to the right of a lever; or with movement forward, upward, or to the right of the associated vehicle or component.

4. The fixed pointer, index, or lubber line should be at the "twelve o'clock" position for right-left directional information, and at the "nine o'clock" position for up-down information (D). For purely quantitative information, either position is acceptable.

5. If an indicator is used for setting, such as for tuning in a desired wavelength, it is usually advisable to cover the unused portion of the dial face (E). In such cases, the open window should be large enough to permit at least one numbered graduation to appear at each side of any setting. If the display is used in tracking, as in the case of heading indicators, the whole dial face should be exposed.

Straight scales: The same direction-of-motion ambiguities exist in these as in circular moving scales, and the same recommendation about using moving-pointer indicators applies. If a moving-scale indicator must be used, the numbers should increase from bottom to top or from left to right (see Figure 3–61), and the scale should move down or to the left (increase) in these situations:

1. When the associated knob or crank is moved clockwise.
2. When the associated lever is moved forward, upward, or to the right.
3. When the associated vehicle or component moves up or to the right.

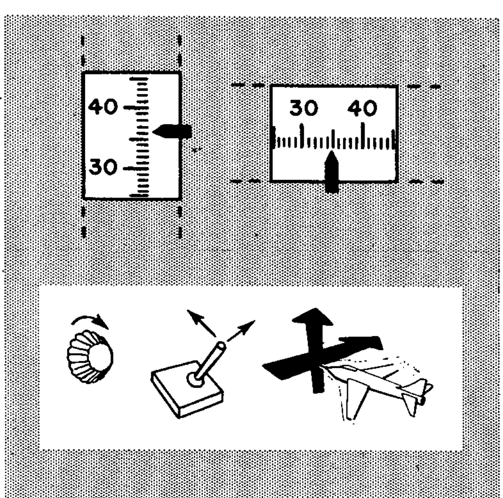


FIGURE 3–61. Recommended direction of motion on straight-scale, moving-scale indicators.

Counter Design

The following design practices are recommended for counter indicators:

1. Numbers change by snap action rather than by continuous movement.
2. Clockwise rotation of a knob or crank should increase the indicated value, and counter numbers should move upward for an increase to be most compatible with the motion of the setting knob or crank. (See A in Figure 3–62.)
3. Do not space numbers further apart than number width or closer together than ¼ number width (B).
4. Avoid windows that limit the viewing angle and increase the space between numbers (C).

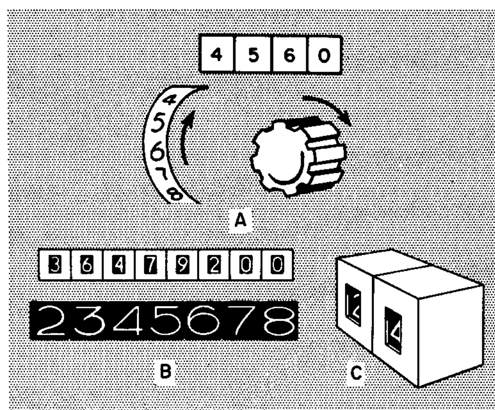


FIGURE 3–62. Recommended direction of motion on counters.

3.6.5 Pictorial Elements of Indicators

Often it is helpful to supplement mechanical indicators with pictorial features. These may consist of a pointer shaped to resemble the object or a component to which the indicator refers, or, a configured background. Such pictorial features can be helpful for position orientation and movements in three-dimensional space. They are also helpful for identifying indicators in terms of the system component to which they refer. Where pictorial elements are used on mechanical indicators, the following precautions should be observed:

1. Generally, pictorial features should be schemes or silhouettes, rather than high-fidelity

portrayals of the things they represent. Efforts to obtain realism can produce false or misleading cues, and may distract attention from the information being displayed.

2. Where a pictorial shape is used as a pointer, it should be clear what part of the figure should be read as the pointer tip. (See Figure 3–63.)

3. Moving elements on the indicator should show position changes of equipment components. In addition, the location and movement of the pictorial part of the indicator should have the same orientation as the equipment component itself. This is illustrated in Figure 3–63 (C) for two aircraft flap-position indicators.

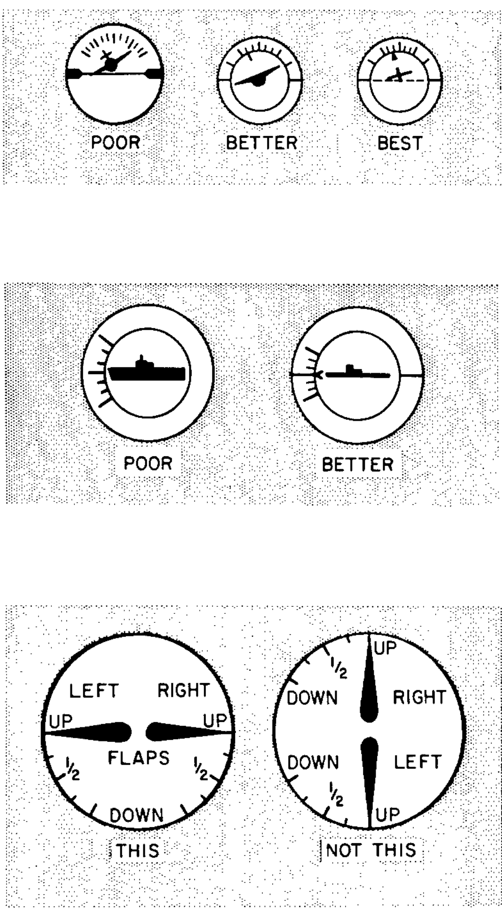


FIGURE 3–63. Use of pictorial elements on indicators for identification and orientation.

Indicator Identification

It is important that the operator have as many cues as possible to identify indicators. Identification can be assisted by labels and by indicator position, shape, and color. Although size and unique configuration assist identification, these cues result from other design considerations.

Labeling. Indicators should be labeled in a simple and direct manner. The following practices are recommended:

1. It is usually best to label indicators in terms of what is being measured and not by the name of the device, as illustrated in the following: use ALTITUDE, not Altimeter, use RPM, not Tachometer, use ACCELERATION, not Accelerometer; use SPEED, not Speedometer.

2. Labels should be as brief as possible, but should not consist of abbreviated words unless the abbreviated form is familiar to all expected operators. Words may be omitted or initials may be used, as in some of the following: use CLIMB, not Rate of Climb, use RPM, not Revolutions per Minute; use MAN. PRESS., not MP; use RANGE, not Rng; use BEARING, not Brng.

3. The label should be placed so numerical designations or graduation marks are not crowded or obstructed. In a circular, moving-pointer, fixed-scale display, the center of the dial face is usually the most appropriate location for the label.

4. Company or trade names should not appear on the visible portion of the display.

Color coding. Color schemes for different indicators can be used to assist in identification, remembering that if indicators have to be read at any time under any illuminant other than white light, color coding should not be used.

Shape coding. Shape coding for indicators has not been thoroughly investigated, but some suggestions can be made from design experience. Differences between rectangular, square, triangular, or round indicators have proven useful for identification. For example, the shapes in Figure 3–64 are suggested by the instrument movement.

Position coding. One of the best means of assisting indicator identification is to maintain a consistent instrument position among various models of the vehicle or system.

3.7 Electronic Displays

Electronic displays provide an excellent technique for presenting information in a variety of

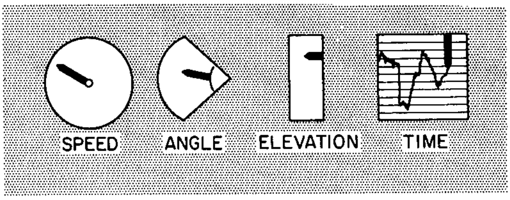


FIGURE 3-64. Examples of shape coding of indicators.

formats so that a single display surface can be time shared by many input devices and sources of information.

3.7.1 Visibility Factors in Electronic Displays

The term *visibility* as used in connection with electronic displays refers to the detection of a target on an electronic scope. Not included are such problems as identification and classification of targets. Visibility problems here are restricted to those situations in which areas are scanned to detect small targets, such as those in air-search and sea-search radar. When radar systems are used for navigation and targeting, large land areas are scanned, but the problem is more one of display resolution than of visibility. Visibility, or probability of target detection, depends on:

1. The size in visual angle of the target.
2. The luminance of the background.
3. The luminance of the target.
4. The length of time the target is present on the scope.
5. The operator's visual-adaptation level.

As applied to electronic displays, background refers to that portion of the scope that includes noise and clutter but no actual target. Noise refers to lighted areas on the tube face not representing reflecting objects. It is usually random and appears as bright spots that change position from sweep to sweep. Clutter refers to signal returns from objects other than targets; i.e., clouds, the sea, etc.

Target size, contrast, and background luminance. The curves presented earlier in Figure 3-13 show the target-to-background contrast required for 99% probability of detection for variously sized circular targets. These data apply to the following situations: (a) the operator is visually adapted to the luminance level of the task: (b) a target that is either brighter or darker than the background: (c) the background luminance is evenly distributed: and (d) the operator is given several seconds for target detection and is alerted to the task.

It is evident from the graph that a reduction in any one factor—background luminance, size or contrast—may be compensated for by an increase in one or more other factors.

Signal duration. The visibility of dim targets is partially a function of the length of exposure time. The relationship between the duration of a flash of light and the relative intensity of the flash required to be seen is shown in Figure 3-65. The curve shows that the eye apparently summates energy up to about 0.1 sec. and that further exposure has little effect on visibility. Of course, if the operator has a search task over a large angular field of view, more search time is required than for visibility summation.

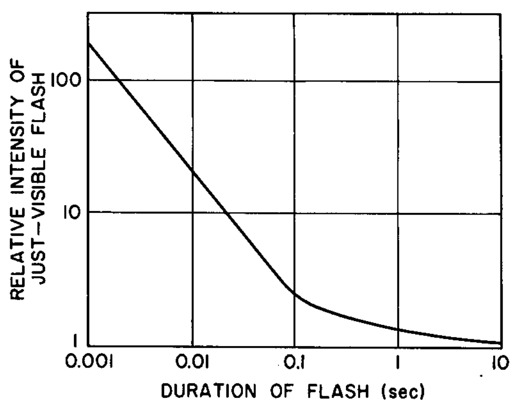


FIGURE 3-65. Data on the temporal summation of the human eye. Intensity x times below 0.1 sec. equals a constant value for threshold visibility. Thus, for extremely short flashes of light high intensities are required for visibility (Blondel & Rey, 1911).

Operator adaptation level. The dark adaptation curves in Figure 3-66 show reduction of the eye's sensitivity to dim visual stimuli immediately after exposure to high luminances. After a period of time in darkness, this sensitivity is regained. The curves show the luminance required for a high-contrast, 10-min. square target to be seen immediately after the eye has been adapted to various levels of higher luminances. After the

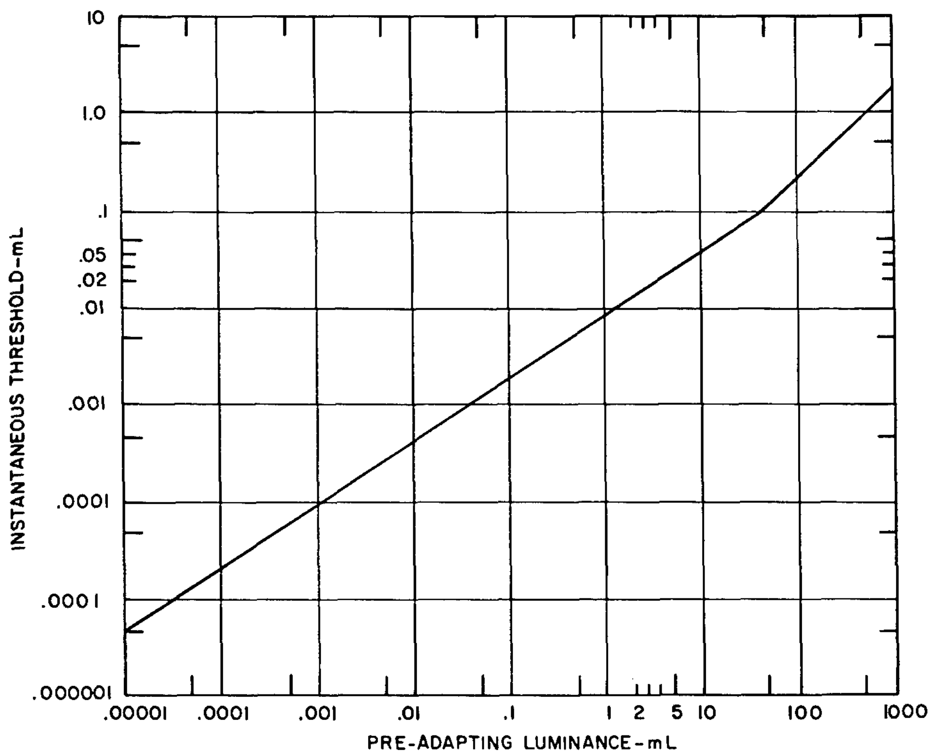


FIGURE 3–66. The relationship of the luminance of a preadapting light upon instantaneous thresholds. The data presented show the luminance required for a 10-min. target to be seen immediately after the eye was adapted to higher luminance levels. (Data from Nutting, 1916.)

eye has been adapted to 1000 mL (typical daytime clear sky luminance), a high-contrast target on a display has to be about 2 mL in luminance to be seen immediately. Based on these and other findings (Hanes and Williams, 1948), the following general recommendations can be made:

1. The visibility of near threshold targets is best when the operator is visually adapted to the luminance of the display.

2. If the operator must perform other visual tasks at higher luminance levels, visibility will not be affected seriously if the higher adapting luminances are not more than 100 times greater than the electronic display luminance.

Duration of pre-exposure luminance. While the eye is less sensitive to dim visual stimuli for some period after having been exposed to relatively high brightness, the data given above show the effect on visual sensitivity of adapting the eye to *high* brightnesses for 5 min. or more. After the eye has been exposed to relatively high brightnesses for about 2.5 min., it reaches a "steady" state of adaptation so that longer periods of pre-exposure have little further effect on its immediate sensitivity, and it is also affected proportionately less by shorter periods. This relationship is shown in Figure 3–67 (Mote and Riopelle, 1953), which shows sensitivity loss vs. pre-exposure time.

In Figure 3–67, for any given exposure duration, the eye's steady-state adaptation level can be obtained by multiplying the corresponding sensitivity loss value by the exposure brightness. The sensitivity loss values are obtained by dividing the exposure time by the steady-state-adaptation time of 150 sec. For example, if the eye is exposed to 2000 mL (daylight sky brightness) for 15 sec., it has a sensitivity loss roughly equivalent to that of being exposed to 200 mL for 150 sec. or more ((15 sec/150 sec) 2000 mL = 200 mL).

Flash blindness. Flash blindness is a transient loss of vision for low-luminance objects following

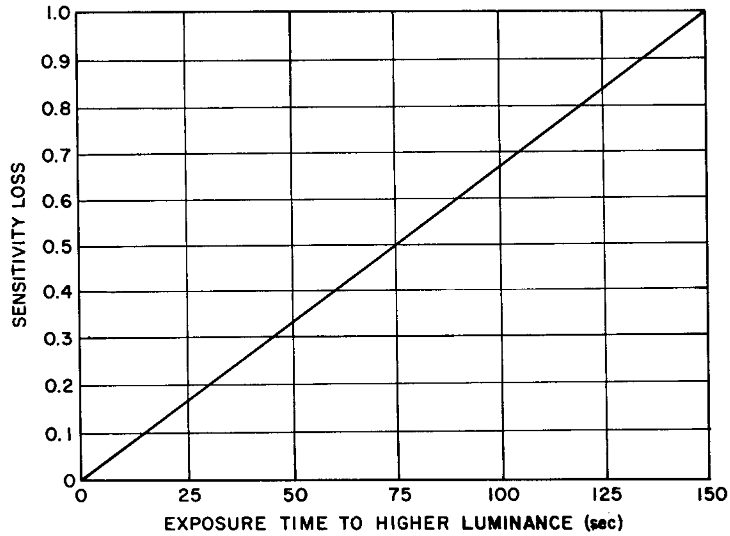


FIGURE 3–67. Relative sensitivity loss to visual stimuli, immediately after exposure of various durations to higher luminances (after Mote and Riopelle, 1953.).

exposure to brief intense light such as lightning, searchlights, and explosions. Data in Figure 3–29, presented earlier, show the time needed to perceive a small target following flash exposure to various intensities lasting 0.9 sec. The visual task was to resolve target detail requiring an acuity of 4 min. visual angle. The luminance needed to resolve this level of acuity is only slightly higher than simple detection visibility thresholds.

3.7.2 Electronic Display Resolution

The quality or resolution of the display system significantly affects information extracted from electronic display images. A low-resolution display is adequate for relatively large target patterns—shore lines, valleys, and industrial complexes—but where the image is small and must be recognized by form contour, high resolution is essential.

Raster scan resolution on video displays. Vertical resolution can be described in terms of the number of raster lines per unit distance. Also, lineal resolution of raster lines is described in terms of the number of resolvable elements per unit length along the lines, established by measuring the minimum separable distance of two distinct points on a video display. The human eye can resolve two contrasting elements if they are separated by as much as 1 min. of arc. At a viewing distance of 20 in. the eye can resolve elements separated by 0.005 in., assuming adequate contrast. A single raster line, 1 in. in length, would require 200 resolvable elements (100 light and 100 dark alternating elements or 100 optical line pairs) to provide a resolution that the eye is capable of perceiving at a viewing distance of 20 in. Also, the number of active raster lines required is at least 200 per in. for the same level of resolution.

Figure 3–68 depicts the considerations in describing raster scan resolution. The simplified drawing assumes electronic parameters that provide an all-or-none response on the display scope. In practice, images are not so sharply separated, adjacent luminous areas being connected by diffuse spatial-luminance transition. Also, signal luminance level is proportional to the extent that the signal image fills the raster scan beam at any point in time. Note that a minimum of three active raster scan lines provides information that two vertically aligned images are present. Because of the Kell effect (the chance that a given detail, smaller in dimension than a scan line, will be scanned by either one or two scan lines) the width of individual scan lines must be 0.5 that of the separation distance

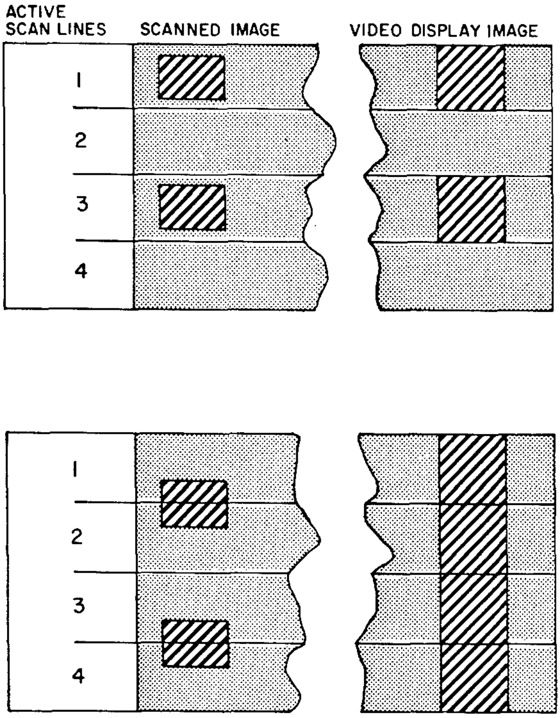


FIGURE 3–68. Simplified illustration of the Kell effect upon distortions to displayed images. The left-hand figures appear on the right as a function of chance raster scan sampling. There is a small inactive scan area, i.e., the active scans almost abut one another.

of two target images to assure display of the two vertically aligned images. Even then, image distortion is evident. Of course, system electronics can sharpen contours by using derivative terms (Brainard and Caum, 1965), but these methods introduce other image distortions.

Figure 3–69 presents a human form at photographic resolution levels and at three levels of video raster scan resolution. The levels of resolution are in terms of the number of active scans composing the human form. The resolution levels are 5.5, 10, and 20 scans.

The original photographic print had a resolution of three optical line pairs per millimeter, about equivalent to 150 video raster scans per inch. The reproduction and printing in this book has reduced the equivalent video resolution to 50 raster lines per in. or 100 lines for the 2-in. human form. It is evident from the figure that 5.5 raster lines provide a poor image, and for this posture, it is an unrecognizable representation of the human form, but 20 lines, although not distinct, reveal a helmeted man holding an object which appears to be a weapon. At 20 lines, each line covers about 1.8 in. of the form.

Resolution on electroluminescent displays. Two-dimensional EL displays are constructed by assembling multiple rows and columns of matrix cells. A thermometer or bar display is the result of generation of either a single row or single column. About 40 cells (20 optical line pairs) to a linear inch, or 1600 cells to one square inch, are presently possible. Thus, EL resolution is about equivalent to commercial television displayed on a 13-in. screen, except that the EL matrix cells are independent, creating possibilities for high contrast and acuity between adjacent cells. At a viewing distance of 20 in. each cell subtends a visual angle of slightly over 4 min. of arc. When a fourfold increase in matrix linear resolution is possible (from 1600 to 25,600 cells to a square in.), EL displays approach resolution capacity of the eye (assuming a viewing distance of 20 in.). The resolution data presented for video raster scan displays are generally appropriate for EL mosaic displays, i.e., the number of raster lines is effectively equivalent to the number of rows (or columns) for mosaic EL elements.

3.7.3 Resolution and Form Recognition

Resolution is a critical variable in form recognition. Assuming an adequate angular subtense of the image, so that the human eye's resolving power is not the limiting factor, the amount of displayed detail will have a significant effect upon the operator's ability to recognize the image. There are several levels of recognition that require mention: (a) *detection* refers only to the observer's ability to perceive the *presence* of an image; (b) *recognition* refers to the observer's ability to perceive sufficient detail to state what the target is, for example, a vehicle, a building, a human, or an airstrip; (c) *identification* refers to perceiving that the image is a Mark IV tank, a Nike missile revetment, an enemy infantryman or a specific airstrip, (this high level of class specification is usually determined by operational requirements and has, otherwise, no rigorous definition); (d) *interpretation* of an image relies on other sources of information or intelligence which assist the observer in identifi-

McDaniel

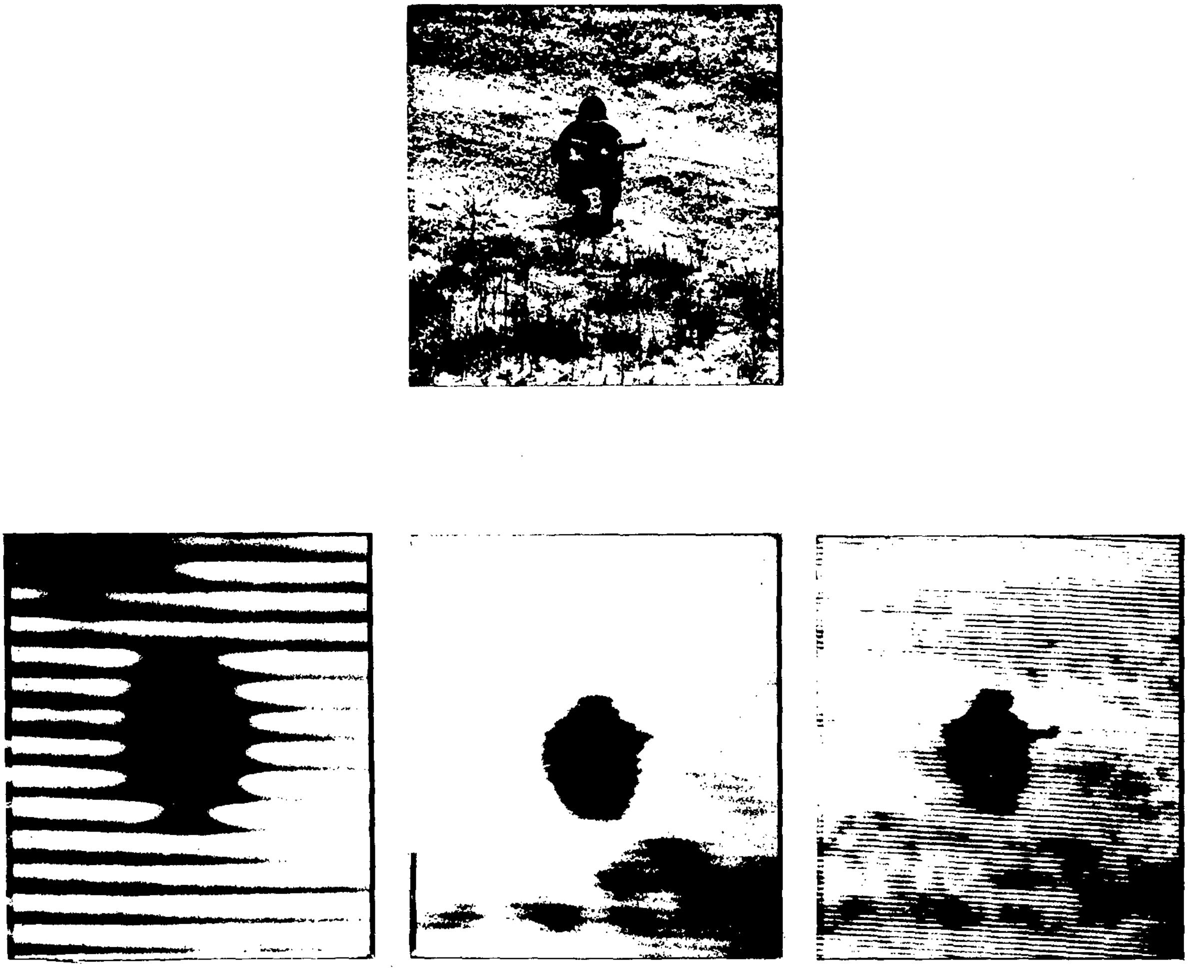


FIGURE 3–69. The appearance of the human form presented at three levels of video raster resolution. The upper figure is the original photograph. From left to right the number of raster lines through the image is 5½, 10, and 20.

cation, (for example, a poorly resolved target, while not permitting the level of visual perception required to identify it on shape characteristics alone may nevertheless be identified accurately by inference).

Video resolution. Figure 3–70 presents data on probability of correct recognition of various types of forms as a function of the number of active raster lines. The numerals and letters used are the Military Standard AN 10400 series. Geometric and pictorial shapes are those used by Sleight (1952). The human silhouettes were of a man with a rifle in various frontal positions such as squatting (as seen in Figure 3–69), aiming, standing, and photographed from above. The forms subtended 30 min. of arc to the viewers at all resolution levels and were randomly located within the scan lines, i.e., there was no compensation for the Kell effect. Elias et al. (1964) performed a similar study using alphanumerics only. They compensated for the Kell effect and found better than 90% recognition at five lines resolution. The appearance of letters at three and five lines of resolution is shown in Figure 3–71.

Brainard et al. (1965) performed a similar study but used targets such as oil storage tanks, bridges, aircraft, and buildings. Their findings were similar. They found low probability of recognition below five raster lines. At nine raster lines, targets were recognized with nearly 90% accuracy. These data must be applied with caution. With figures embedded in cluttered backgrounds and of lower contrast, the probability of correct recognition will be reduced. Also, resolution requirements for less discriminable forms will be greater. (Additional references on resolu-

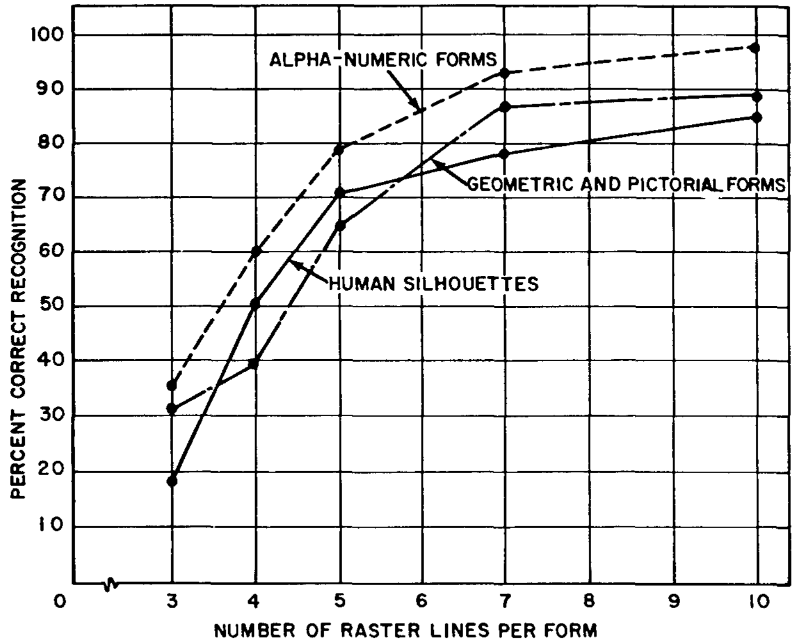


FIGURE 3-70. Recognition of various classes of forms as a function of the number of active raster lines which composed the forms (Baker and Nicholson, 1967).

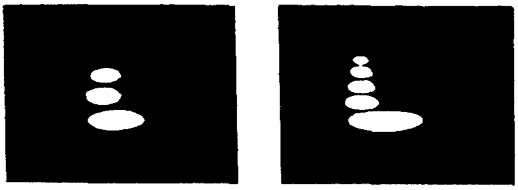


FIGURE 3-71. The appearance of the letter L as it appears at 3 and 5 lines of video resolution (Elias, et al., 1964).

tion and form recognition are Bennet et al., 1963; Jennings et al., 1963; and Williams et al., 1960.)

Electroluminescent and light grid mosaic display resolution. Since EL and other mosaic displays are still in developmental stages, little data on their application to reconnaissance target recognition exist. However, since two-dimensional mosaics are similar to raster scan displays, the previous data on video displays are generally adequate for describing EL resolution in target recognition. A major application for EL displays is the presentation of symbolic information generated from digital processors. In particular, the generation of symbolic forms such as alphanumeric, geometric, and certain pictorial symbols will have wide applicability. The letters of the alphabet and digits 0 through 9 can be reproduced reasonably well using matrix rows which are five cells in height but vary in width. Figure 3-72 illustrates recommended alphanumeric formats. Note that the 5 and S are similar and the Z could be confused with the 2 if the different formats are not learned and remembered. Although the M, W, V, D, and Ø are not good representations, they will not be confused with other symbols. Better alphanumeric symbols can be constructed if seven (or more) cell units in height are possible (Long et al., 1951).

Mosaic displays used to present only numeric information can be designed with seven elements. Highly legible numerics can be presented as shown in Figure 3-73. The numeral 8 uses all seven elements (Fuller, 1965). Fourteen-element displays, as shown in Figure 3-74, can be used to present alphanumeric and special symbols which are easily read. However, the letters O, Q, and D, and number 0 present some problems. These can be resolved by using the following line segments (see Figure 3-75). (See also Bowerman and Wasserman, 1963.)

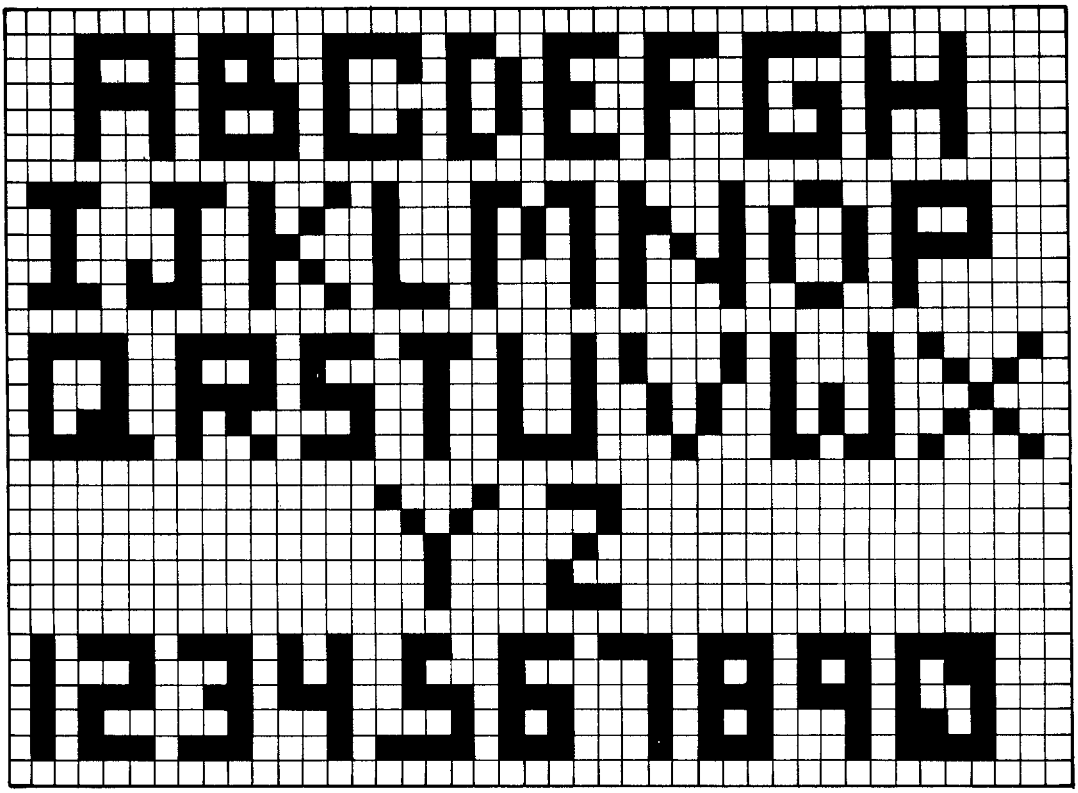


FIGURE 3–72. Recommended letters and numerals composed of cells limited to a resolution of five cells in height.

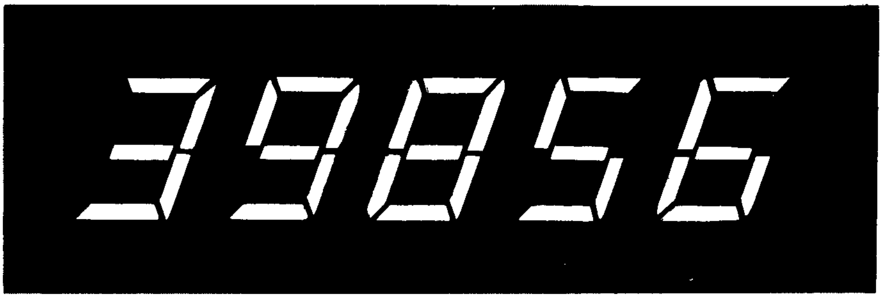


FIGURE 3–73. Seven-stroke segment display format of numeric symbols (Fuller, 1965).

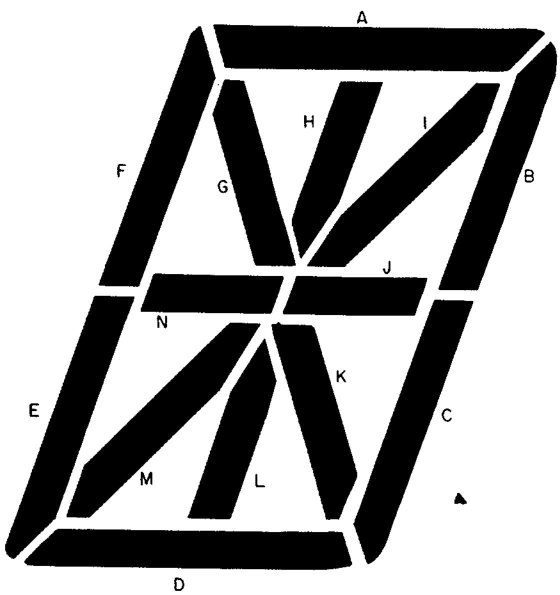


FIGURE 3–74. Fourteen-stroke segment display format for alphanumeric symbols (Bowerman & Wasserman, 1963).

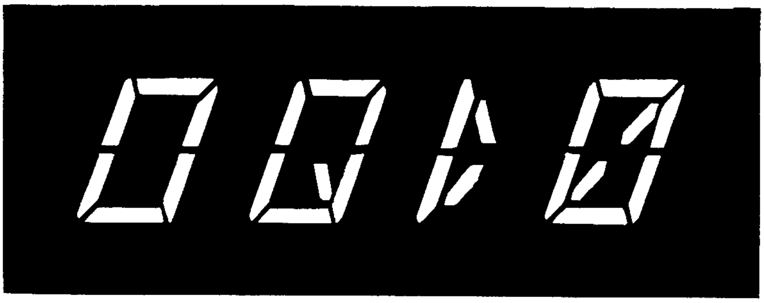


FIGURE 3–75. Use of certain segments for difficult letters.

letter O—Segments ABCDEF
letter Q—Segments ABCDEFK
letter D—Segments EFGM
number 0—Segments ABCDEFIM

3.7.4 Electronic Displays Size and Scale Factors

The determination of optimum display size and scale factor is dependent upon many conditions: (a) object displayed, (b) detail to be seen, (c) viewing distance, and (d) display resolution. Physical constraints on display size, as in aircraft cockpits, will vary with each operational application, but in the following paragraphs only visibility, resolution, viewing distance, and target size are considered.

Visibility. When display size increases, if images do also, signal detectability will be improved even though resolution remains unchanged. The extent of improvement can be determined from curves in Figure 3–13, presented earlier. Since improvement is slight, size is rarely a determining factor. Optical magnification would result in about the same improvement.

Resolution. If the display is capable of more resolution than the eye can perceive, increases in display dimensions or magnification can be advantageous. For example, if a 5-in.-square video display provides 1000 raster lines (500 optical line pairs), the display is capable of resolving detail (such as a line gap) as small as 0.005 in. At a 30-in. viewing distance, the detail would subtend less than 0.6 min. of arc. Based on the acuity data presented earlier in Figures 3–9 and 3–10, it is evident that this detail would be too small to be resolved by an average observer. However, if the 1000-line display were 10 in.

square, the gap would be over 1.1 min. and the display would provide twice the angular subtense of detail. The referenced acuity data indicate that under many conditions this size of detail can be detected. Thus, it is recommended that when *necessary*, the display should be scaled so that detail resolution subtends *at least* 1 min. of arc to the observer's eye. Indeed, 2 min. of arc is a recommended design goal, but there is a disadvantage in exceeding this value. (See Bennett et al., 1967.)

Image size. As indicated earlier (see Figure 3–22), a target image should subtend no less than 12 min. of arc to insure relatively accurate form recognition. Certain highly discriminable symbols, such as letters, numerals, and geometric forms, can be easily recognized under optimal viewing conditions if they subtend as little as 8 min. of arc. However, most target images are not as discriminable and require 12 min. of arc for acceptable form recognition. The value of 12 min. of arc is based on viewing conditions of good contrast, high resolution, and no vibration or buffeting. The scaling function should probably be based on a minimum of 20 min. subtense for the critical targets to meet more operational conditions.

Scale factor is also critically affected by the resolution of the sensor display system. Assume a system capable of providing 1000 active raster scans, or a resolution of 500 optical line pairs, and a surveillance system in which 20-ft. targets must be identifiable on shape characteristics. The display-vs-real world scale factor can be established as follows: The data in Figure 3–70 indicate that target recognition becomes acceptable if about 10 raster lines (5 optical line pairs) are used to display the target image. Thus if a 20-ft. target is scanned by 10 raster lines, ground resolution is 2 ft. per raster line. A 1000-line system would provide a ground range coverage of 2000 ft. For the resolution of the display system to be fully exploited by the observer, displayed detail should subtend at least 1 min. of arc. Assuming a viewing distance of 20 in., raster lines spaced slightly over 0.005 in. will subtend 1 min. of arc per line. However, a ten-line image would subtend only 10 min. of arc, which is below the recommended minimum of 12 min. of arc. Thus, a 1000-line display viewed at 20 in. should have one dimension of at least 6 in.

3.7.5 Indicating Target Position on Electronic Displays

Various methods for indicating position can be evaluated in terms of the accuracy and speed of the estimates. Optimum design within a method can then be found through comparison.

Range markers. Range markers can be placed on polar-coordinate and rectangular-coordinate displays. While the following recommendations apply to range markers on polar-coordinate displays, the more general ones apply to rectangular-coordinate displays as well.

Range-mark-interval values. The range-mark-interval scale should be represented with the numbers progressing by one's, two's or five's and with the appropriate number of zeros following after each digit. In general, progression by one's or ten's (1, 2, 3, or 10, 20, 30, etc.) is superior to progression by two's or five's (Chapanis and Leyzorek, 1950; and Barber and Garner, 1951).

Range-mark separation. Two tasks are involved in the estimation of range: the identification of the nearest inner range-mark value and the interpolation of target position between range marks. For a given scope size, the number of range marks used affects spacing. If the scope size is fixed, a change in the number of range marks affects its numbering system. The effect of range-mark spacing on speed and accuracy is dependent on interpolation accuracy, range-mark identification, and coding.

For interpolation accuracy: An operator is able to interpolate the position of a target between two range marks with the following accuracy:

1. About 50% of the range estimates will be in error by less than 4% of the distance between range marks.
2. About 95% of the range estimates will be in error by less than 14% of the distance between range marks.
3. Operators tend to overestimate the target range by 2% of the interval distance, i.e., there is a positive constant error of 2% of the interval.

The above values apply in the following situations:

1. The distance between range marks subtends a visual angle greater than one degree.

2. The target is a relatively small, well-defined spot.

3. The operator acts as rapidly as possible and attempts to read to the nearest 1% of the interval.

For identification: Gross errors occur in the identification of the range marks adjacent to the target and are multiples of the range-mark-interval value. For example, if the range marks represent intervals of 1000 yd. the gross errors would be of this value. Frequency of occurrence depends on the number of range marks used on a display.

Curves in Figure 3–76 (Baker and Vanderplas, 1956) show the percent error in readings of the total number of range marks used, the data applying to operators reading as rapidly and as accurately as possible. Range marks were scaled so that every fifth ring was discriminably different (thicker) than the others. The graph also shows the time required to determine the range of each signal.

If more than five range marks appear on a display, every fifth mark should be brighter or wider than others. This arrangement assists the operator in identifying any particular mark.

Pointing devices. Accurate and rapid target positioning can be achieved by the use of pointing devices. The pantograph and electronic pointer or stylus are illustrated in Figure 3–77.

The pantograph operator positions a cross hair or pointer, located on the end of an arm, over the target; the position is resolved from the angular position of the arms. The operator positions a hand-held electronic pointer on the target. The X–Y coordinate position can be resolved by two methods. One technique involves a photo-sensitive pointer triggered when the electronic scan line passes through the target. Position can be determined by synchronized timing circuits. The other method uses an electrostatic grid on the display surface. When the pointer contacts the surface of the grid, X–Y position coordinates are resolved from the grid circuits.

Electronic cross hair. An electronic display cursor or marker is positioned over the target, and the cursor X–Y position is controlled by a stick or ball. When the cursor is positioned over the target, X–Y coordinates are resolved from the cursor circuitry.

Positioning accuracy and speed. The accuracy of determining X–Y coordinate position with pointing devices and electronic cross hairs depends on the operator's ability to center a point on the target. For small targets (less than 0.5 in. in diameter), 95% of the positionings will be less than 0.01 in. in error. These accuracies are possible even if only 2 sec. are allowed (see Ford et al., 1949), and if the operator can rest his hand. or arm on the surface when using the pointing or pantograph devices. Also, when using the electronic cursor, high accuracy is possible only if an optimized control is used.

Figure 3–78 shows the amount of time required to determine target position by the various methods. The time required to position elec-

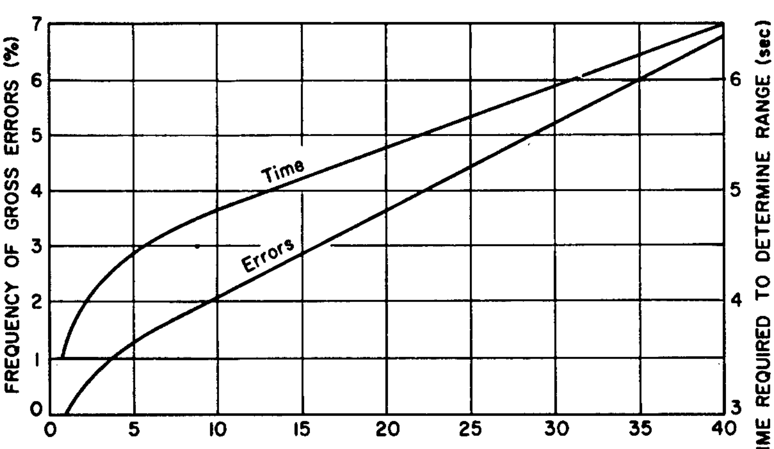


FIGURE 3–76. Frequency of gross errors and speed of readout as a function of the complexity or number of range markers used (Baker & Vanderplas, 1956).

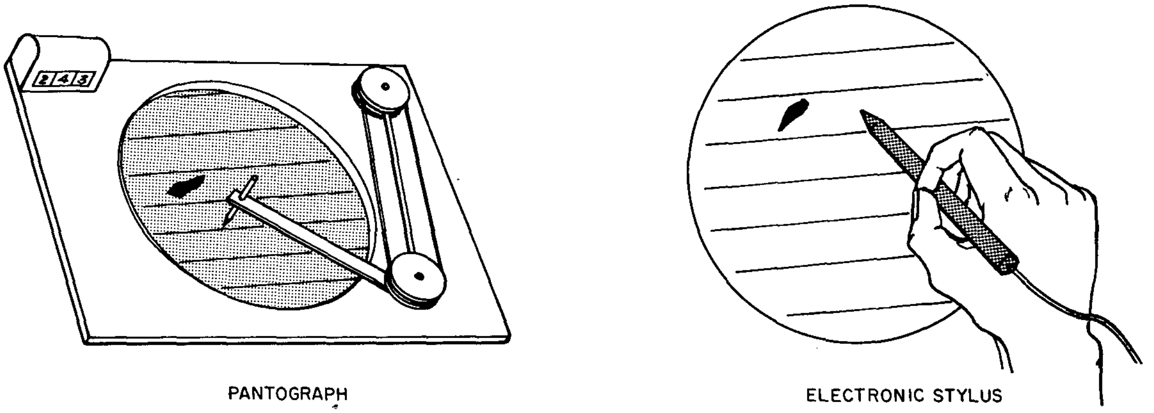


FIGURE 3–77. Two common methods for determining target coordinates.

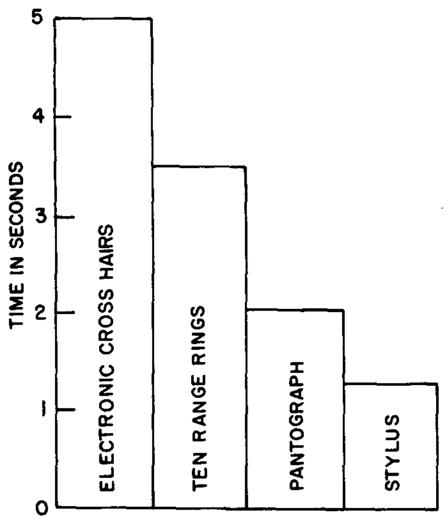


FIGURE 3–78. Average time required to determine target X-Y position as a function of the method employed.

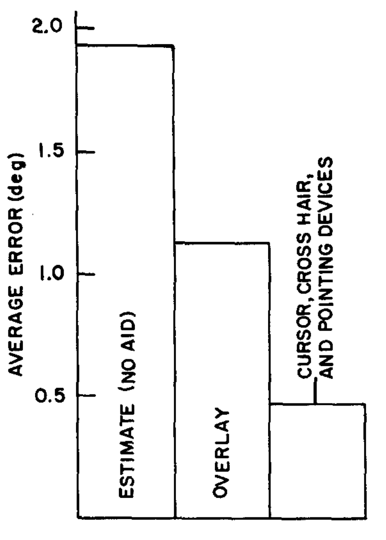


FIGURE 3–79. Accuracy of determining target bearing when using various aiding devices.

tronic cross hairs is highly dependent on the control provided. The data presented are typical of the time required using a proportional rate joy stick control or a ball position control when high accuracy is desired.

Bearing indications. Bearing information can be indicated by:

1. Cursor and bearing dial or counter. A mechanical or electronic cursor is moved by a crank or knob until the signal is bisected. The bearing value is read either from a bearing dial surrounding the scope or from a counter.

2. Bearing dial alone. The operator sights along the target to the bearing dial and estimates the value; no cursor is used.

3. Overlay of radial lines. A transparent overlay is placed over the scope face. Radial lines assist in estimating bearing.

4. Pointing devices. Devices described earlier (pantograph, stylus, and electronic cross hairs) can provide target bearing information by resolving the geometry of the display.

Figure 3–79 shows the approximate accuracy of various methods for indicating bearing. These methods can also be evaluated in terms of the speed and accuracy with which the operator is able to read or relay the information. The position of the target on a plan position display also governs bearing accuracy. When the target is nearer the display's periphery, bearing readout aids are more accurate.

Three-coordinate displays. Three coordinates must be displayed to the pilot in order for him to determine the position of his aircraft: bearing, range, and altitude. A conventional method for such display on two scopes is shown in Figure 3–80 (A). Attempts also have been made to display these three coordinates of information in a single integrated display, but no simple solution exists for three-dimensional displays for accurate read-out of quantitative data.

Another method for displaying three coordinates on a single display is shown in Figure 3–80 (B). Two targets appear for a single object. Range is displayed in the inner circle and altitude is displayed in the outer circle. When bearing is estimated, and range rings are used, bearing, range, and altitude can be orally reported in 9 sec. per target (Gebhard and Bilodeau, 1947). This method requires one-third less time than the two-scope presentation. Although this display results in rapid reading of the space coordinates in distance and angles, it does not result in an accurate perception of the spatial positions.

3.7.6 Electronic Pictorial Flight Displays

During recent years considerable effort has been made to develop flight displays which provide the pilot with information he normally acquires by direct viewing of the outside world. Instrumentation on a plane's attitude, velocity,

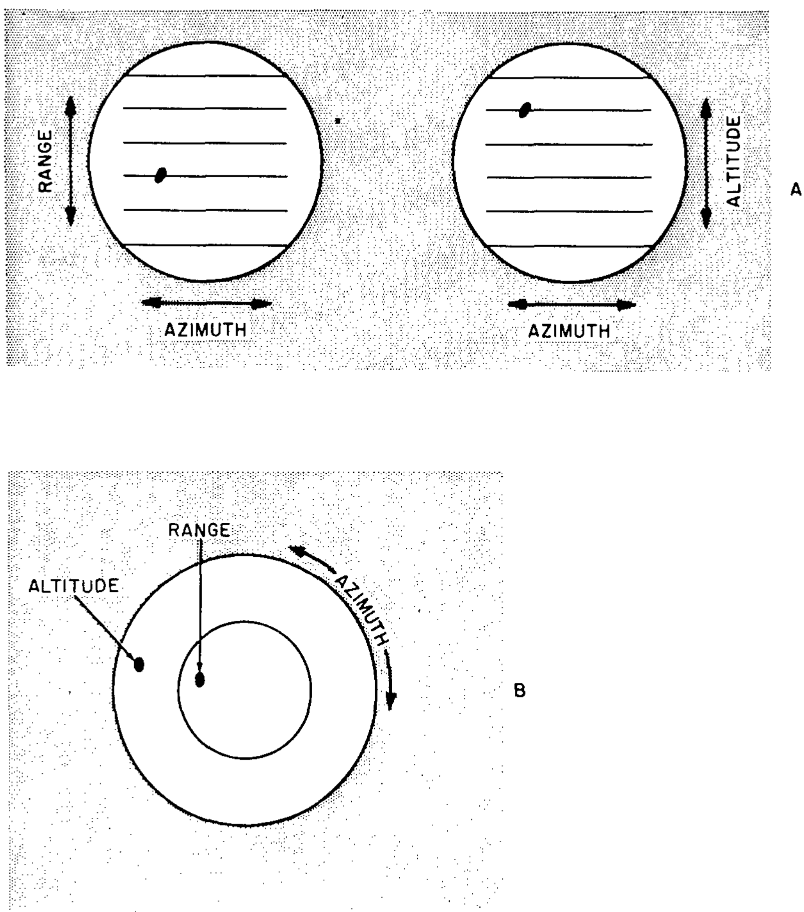


FIGURE 3–80. Two display configurations which present three coordinate target information.

altitude, altitude rates, flight path, relative bearing to reference points, etc., are distributed in various locations on a pilot's display panel. During critical flight maneuvers, the pilot has difficulty in integrating these sources of information.

To make the pilot's assimilation of information on flight parameters easier, a display presenting a pictorial analog of the real world has been developed. Most of this work is now being sponsored by the Joint Army-Navy Aircraft Instrumentation Research (JANAIR) Program.

Contact analog characteristics. The contact analog display presents a point perspective projection on a two-dimensional display surface. The projection is of a three-dimensional model of the real world containing a ground plane representing the local horizontal, and a surface representing the command or desired flight path. Other perspective objects can be presented as required. A unique aspect of the contact analog is that the display of these surfaces follows the laws of geometric perspective and motion perspective. Thus, information is coded in a manner analogous to what the pilot sees in visual flight. (See Figure 3–81.)

As the aircraft continues down the localizer and glide slope path, display slope circles provide perspective and motion cues. Also, the angular subtense and perspective cues of the runway and ground plane are changing consistently with the changing geometry. If the aircraft drifts to the left of the localizer, the displayed localizer path moves to the right and the perspective cues of the runway and ground plane change accordingly. When the aircraft sinks below the command glide slope, the glide slope appears to be passing over the aircraft. In a roll maneuver, the entire contact analog scene rotates as would a real-world scene viewed through the windscreen.

A study comparing a contact analog display in a simulated helicopter approach and hover flight performance with a conventional instrumentation flight performance is documented by Dougherty et al. (1964). In addition to piloting, subjects were required to perform a task of number reading at various frequencies.

The data in Figure 3–82 show superior performance using the contact analog display at all levels of workload on an unrelated task. The authors concluded that pictorial displays are better because: (a) pilots are able to assimilate qualitative information, (b) several parameters can be sampled simultaneously, and (c) peripheral vision is enhanced.

The value and application of contact analog displays will depend on additional data and operational experience. As certain system control functions become more automated, the pictorial displays may aid the operator considerably in overall flight and mission planning. For additional detailed information on contact analog displays, the following references are provided: Carel (1965, 1961, and 1960); Emery and Dougherty (1964).

3.8 Printed Information

Labels, operating instructions, check lists, and display panel operating data give printed information in association with equipment. Applicable design practices are discussed in this section. For recommendations concerning books and similar printed matter, see Tinker (1963). The legibility of instruction decals, check lists, and labels is important, perhaps critical when the operator is pressed for time and when he must operate equipment infrequently.

3.8.1 Design Recommendations for Decals, Check Lists, and Labels

1. Style of Print. Capital letters ("all caps") are recommended for general use, but "initial caps" and lower case letters are permissible. The use of letters of the style found in commercial lettering guides is satisfactory. Stroke-width-to-height ratio should be from ⅙ to ⅛. Letter width-to-height ratio should be 3:5. When space will not permit this optimum, width-to-height ratio may be reduced, but this will require a narrower stroke width to maintain clarity of the letter's identifying features.

2. Size of Print. Recommended letter size depends on the viewing distance and illuminating conditions. Assuming a viewing distance of 28 in. or less and a wide range of illuminating conditions (including illuminances below 1 ft.–c.), letter height should be at least 0.20 in. For less critical functions, or when the illumination is always above 1 ft.–c., letter height may be reduced to 0.10 in.

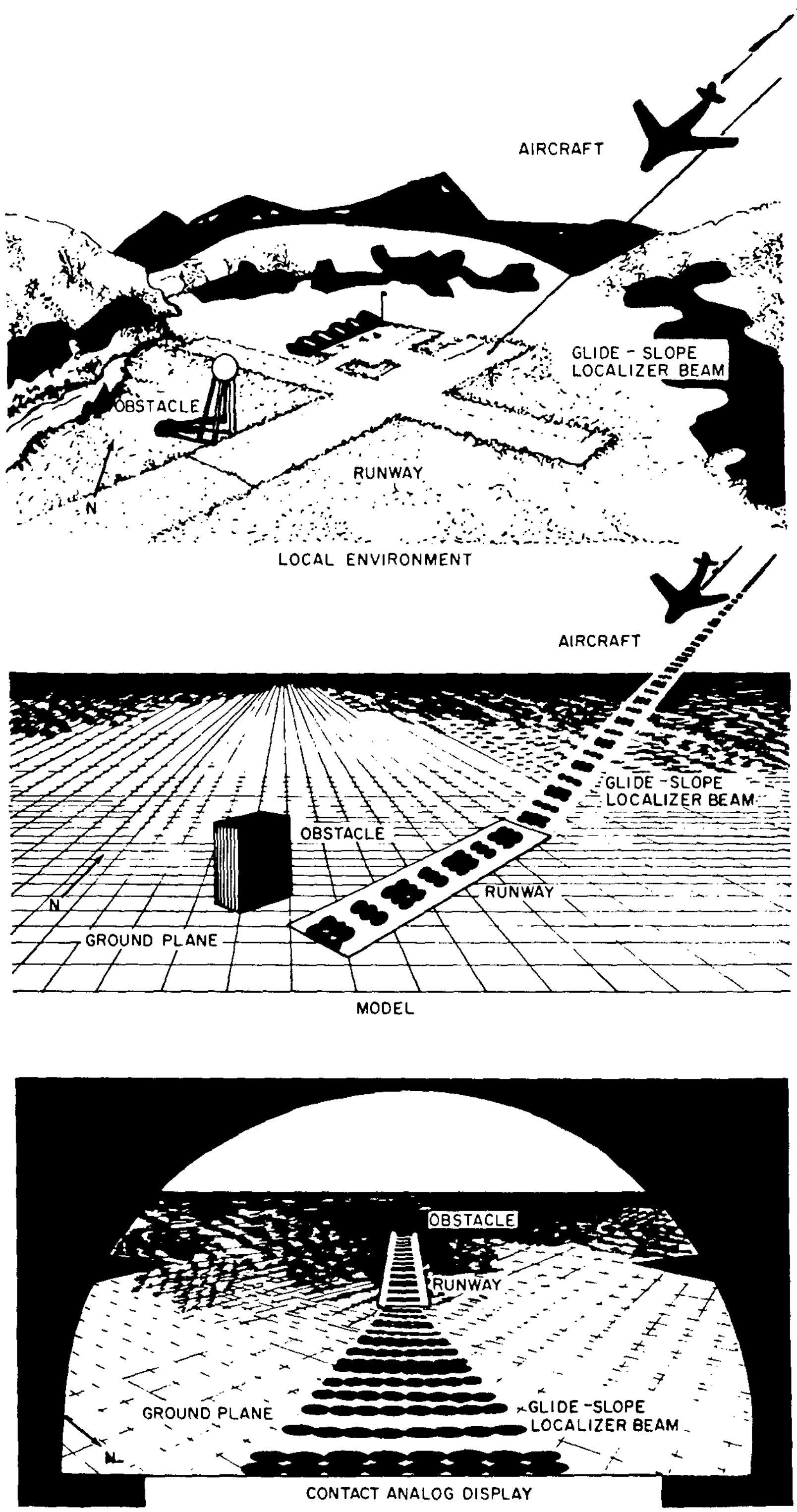


FIGURE 3–81. Relationships between the local environment, the model, and the "contact analog" display (Carel, 1965).

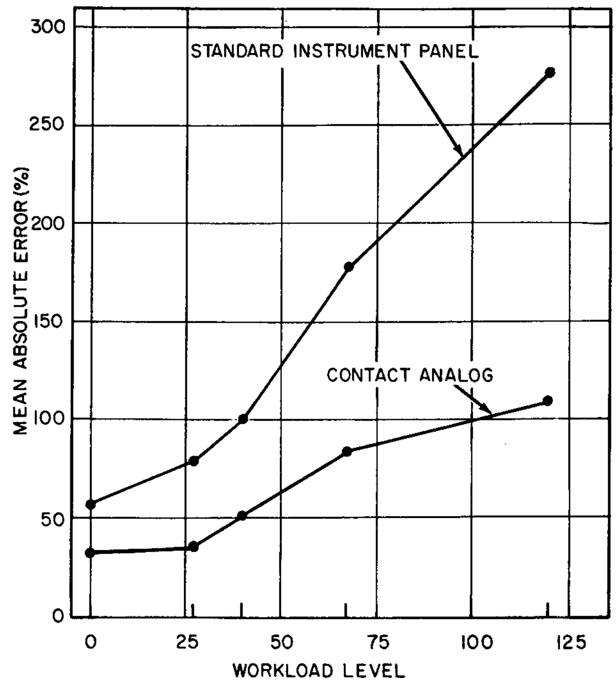


FIGURE 3–82 Simulated flight performance using contact analog and conventional displays for various levels of workload to an unrelated task (Dougherty et al., 1964).

3. Brevity. To save reading time and panel space without distorting intended information, write concise emergency instructions and check lists.

4. Word Selection. Words and sentences are more immediately recognized depending on their degree of familiarity. Therefore, decals, check lists, and labels should be composed of common words, but only if the words will say exactly what is intended. For a list of common words, see Thorndike and Lorge (1944). For particular populations, such as military pilots, common technical terms may be used even though these words occur infrequently in general language.

5. Clarity of Meaning. Instructions should be clear and direct, avoiding complex or lengthy sentence structures (Chapanis, 1965). To be sure a message is not ambiguous, test several different versions on a small sample of people.

6. Contrast. In a task in which dark adaptation is required, letters should be white on a dark background. In tasks where dark adaptation is not essential, the print should be black on a white mat background. Colored print may be used for coding purposes, but colors should be chosen for maximum contrast against the background.

3.8.2 Graphs, Tables, and Scales

Persons who operate equipment frequently use quantitative data obtained from graphs, tables, scales, or simple sliding scale-type computing devices.

Design recommendations. Detailed recommendations for design of scales are presented in the section on Mechanical Indicators. These are applicable also to scales on graphs, nomographs, and slide-rule type computers. The following recommendations apply to graphs and tables (Carter, 1946; Connell, 1947; and Bowen and Gradijan, 1963):

1. A graph is superior to tables or scales if the shape of the function is important in making decisions, or if interpolation is necessary. If interpolation is not required, tables are preferred.

2. Reduce tables to the simplest form con-

sistent with the degree of sensitivity necessary to permit reading without interpolation. For example, if altitude steps of 5000 ft. are not sufficiently sensitive for the accuracy required, more altitude steps should be included.

3. When table columns are long, separate numbers into groups by providing a space between groups of five.

4. Provide at least 0.166 in. between columns that are not separated by vertical rules.

5. Construct graphs so that numbered grids are bolder than unnumbered grids. If ten-grid intervals are used, the fifth intermediate grid should be less bold than numbered grids but bolder than unnumbered grids.

6. Avoid combining many variables into a single multiparametric graph. Generally, the parameters for line graphs should not exceed three. If additional parameters must be shown, use additional graphs.

7. Vertical and horizontal scales on graphs should be adjusted to slightly exceed the maximum ranges of the values to be covered, thereby spreading the plotted data as widely as possible.

3.9 Optical Projection Displays

A common means of presenting information is by projected images. These may be from slides, film strips, vugraphs, motion picture film, opaque pictures, television or other cathode ray tubes, or other image sources. This display method is particularly suited for presenting the same information to a group of observers, those in a command and control center for example. It may also be used for single operators, a common example of which is a gunsight reticle projected onto an aircraft windshield and focused at infinity. Often referred to as a "heads-up" display, this type of image may also be provided by a helmet mounted cathode ray tube and transparent mirror. Still another use of projected displays for single operators is found in flight simulators which include a visual simulation of the outside world.

3.9.1 Advantages and Disadvantages of Projection Displays

Projection displays are particularly suitable for:

1. Group presentation (briefing, training classes, etc.).

2. Pictorial and spatial information (maps, charts, target photos).

3. Presentation of past history as opposed to real time.

4. Presentation of synthetically generated pictures.

5. Simulation of the external visual world (in training simulators).

6. Superposition of data from more than one source.

The major difficulty in the use of projection displays is the need for a darkened room or other control of ambient light. Tolerance limits for ambient illumination, and methods of dealing with the ambient light problem, are considered in this section.

3.9.2 Requirements for Projection Displays

Accepted standards for group viewing of slides and motion pictures are summarized in Table 3–12. More complete information can be found in *IES Lighting Handbook*, 1966, and Eastman Kodak Pamphlet No. S–3.

Seating area and screen size. Seating positions as given in Table 3–12 apply to general audience seating, and are based on considerations of image distortion, light distribution on the screen, and ability of the eye to resolve details. Distances are based on projected image rather than screen width, since the entire screen may not be used. Many military work situations may require deviations from these recommendations because of other equipment, including other displays, among which the operators must divide their attention. When making such deviations, attention must be given to screen reflection and legibility of material to be presented.

Image brightness or luminance. General recommendations for image brightness and light distribution on the screen are given in Table 3–12. These measurements refer to screen brightness with no film in the projector. For motion pictures, the projector should be running while making luminance measurements. Image luminance can be computed from the projector output in lumens, the screen area in sq. ft., and the

TABLE 3–12. GENERAL RECOMMENDATIONS FOR GROUP VIEWING FOR SLIDES AND MOTION PICTURES

Factors	Optimum	Preferred limits	Acceptable limits
Ratio of $\frac{\text{viewing distance}}{\text{image width}}$	4	3–6	2–8
Angle off center line—degrees	0	20	30
* Image luminance—ft.-Lamberts (no film in projector).	10	8–14	5–20
Luminance variation across screen—ratio of maximum to minimum luminance.	1	1.5	3.0
Luminance variation as a function of seat position—ratio of maximum to minimum luminance.	1	2	4.0
Ratio of $\frac{\text{ambient light}}{\text{highest part of image}}$	0	0.002–0.01	0.2†

* For still projections higher values may be used.
† For line drawings, tables, not involving gray scale or color.

gain factor for the screen, by the formula below. For an explanation of gain, see below.

$$\text{Screen luminance (ft.–L.)} = \frac{\text{Projector output (lumens)} \times \text{Screen gain}}{\text{Screen area (sq. ft.)}} \quad (3\text{–}9)$$

In order to see adequate detail and color, lower limits for image brightness must be observed. Upper limits are determined by the level at which some observers will see flicker. For still projections, it may be advisable to exceed these limits to overcome ambient light on the screen.

Ambient illumination. The presence of ambient light, while causing glare and loss of dark adaptation, is normally less serious than the loss of contrast in screen image. In command and control centers, it is not practical to darken the entire room. Tolerable amounts of light on the screen vary with the type of material being projected. Simple black-and-white line drawings, or alphanumeric tables can be projected with screen brightness from ambient light as high as 20% of the highest brightness in the projected image. For photographic materials containing the full range of grays between black and white, and fine shades of color, the proportion should be limited to 1% or less. Where problems are encountered, one or all of the following suggestions can be used to insure adequate contrast.

1. Increase the luminance of the image on the screen (see below).

2. Keep other light in the workspace from reaching the screen, by using light baffles, controlling direction of light, and using black paint to control reflections. Lighting control should be anticipated in the original room layout plan.

3. Use a rear rather than front projection screen.

3.9.3 Projection Screen Types

Front projection screens. The most efficient arrangement for most projection displays is the front projection screen. Ideally it should be made of a material that reflects all light, distributes it uniformly over the angular area occupied by the audience, and wastes no light outside this area. While the ideal material does not exist, a wide variety is available, differing primarily as to selectivity of the light reflection angle. The most critical angular relationships between observers, screen, and projector are shown in Figure 3–83 for both front and rear projection arrangements.

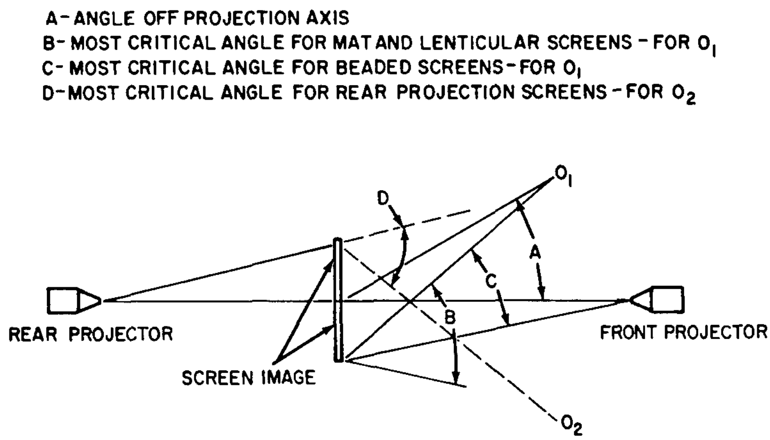


FIGURE 3-83. Critical angles for arrangement of observer (0), screen image, and projector.

The major types of screen surfaces are discussed below. Their reflection and gain characteristics are summarized in Figure 3-84.

Mat screens. A mat screen reflects light evenly in all directions. Therefore, image brightness on the screen will be the same for all observer angles. Most mat screens are about 85 to 90% efficient. For each ft.-c. of projected illumination, screen luminance will be 0.85 to 0.90 ft.-L. This is usually referred to as *gain,* and such a screen is defined as having a gain of 0.85 to 0.90. As this gain factor is increased, the requirements for projector output are reduced.

$$\text{Screen Gain} = \frac{\text{Luminance (ft.-L)}}{\text{Illumination (ft.-c)}} \quad (3\text{-}10)$$

Lenticular and metalized screens. While resembling a mirror, lenticular screens also scatter light. They have a specially prepared pattern of stripes, ribs, or other geometric forms to control the size of a fan-shaped area where most of the light is reflected. They are designed to spread light through a wider angle horizontally than vertically. On these screens the angle of greatest reflection is the reciprocal of the angle of incidence. Compared to a mat screen, lenticular screens give higher brightness or gain. Those designed for group viewing may have a gain of 2 to 4. If designed for a smaller angle, as for individual viewing in a flight simulator, the gain can be much higher.

Beaded screens. Transparent beads return most of the light back toward the source through refraction and internal reflection. The angular spread can be manipulated through the choice of beads, but not through as wide an angle as with lenticular screens. Beaded screens are most suitable for narrow rooms with viewers fairly close to the projection beam. Ambient light's effects will be controlled by its reflection back to the source. Angular spread may be very narrow as shown in Figure 3-84.

Rear projection screens. Rear projections are used for:

1. Work areas requiring high ambient illumination for other activities. Screen materials for rear projection have very low light reflection and appear dark in reflected light, thus ambient light on the viewing side of the screen has relatively little effect on the contrast of the transmitted image.
2. Work situations in which there are physical obstructions to front projection.
3. Portable, self-contained projection units.

Besides minimum reflection or backscattering, rear projection screen materials should achieve a proper balance between light distribution and transmission (gain). Unfortunately, high transmission (gain) can be obtained only at the expense of uniform light distribution over the screen. Curves for typical materials are shown in Figure 3-85. Because of these characteristics, rear projection systems normally give poor light distribution over the screen area. Usually, there is a bright area in that part of the screen on a

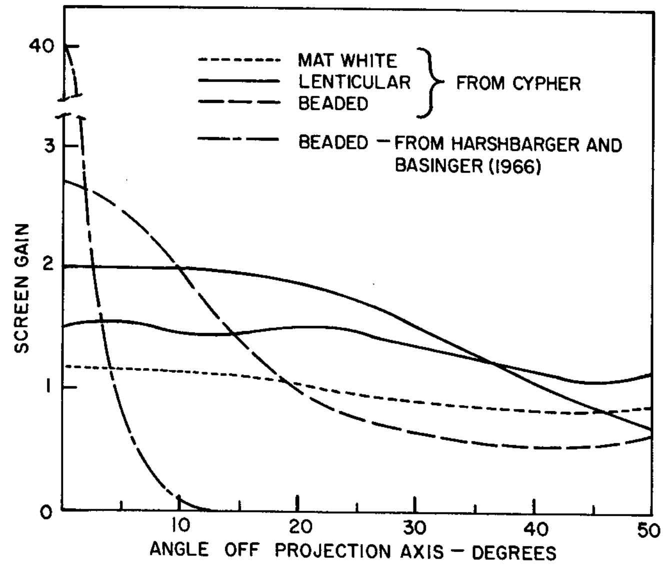


FIGURE 3-84. Gain vs. viewing angle for typical front projection screens.

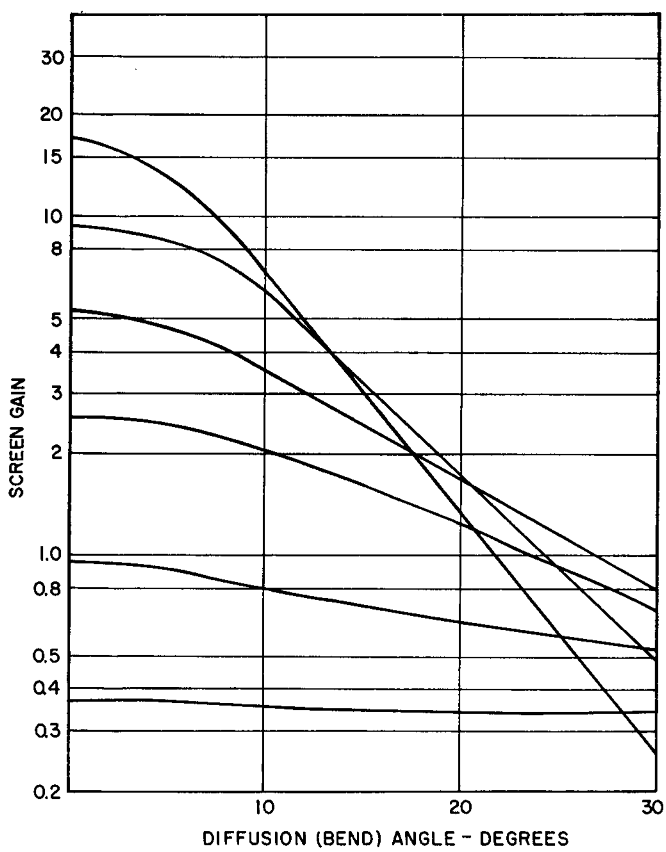


FIGURE 3-85. Gain vs. bend angle for typical rear projection screens (redrawn from Vlahos, 1961).

line between the observer's eye and the projector lens. The light distribution problem can be minimized by using a long focal length projector, and keeping viewers near the projection axis. In a rear projection system, one or more mirrors in the projection path are a means of compressing lens-to-screen distance, and of obtaining right-left image reversal, as is required for handling sound track film.

Although rear projection reduces the problem of ambient light on the viewer's side of the screen, there remains a problem of scattered light from the projector and other lights in the area behind the screen. This can be reduced by mat black paint on walls and other surfaces and by adding light baffles.

3.9.4 Superposition of Projection Displays

Projection displays can be used for the addition of temporary or changing information to a longer term image, e.g., data successively added to a map to show the order of events. Colors may be superimposed with transparent overlays on the source image to obtain intermediate hues for coding purposes. Such superposition at the source gives a subtractive effect. Or, it may be done with two or more projectors aimed at a single screen resulting in an additive mixture. The screen may be either rear or front projection, but both front and rear projection on the same screen is not advisable. Other precautions to be observed in superposition are:

1. Images must be aligned. This is critical, particularly if small details such as alphanumerics must be superimposed. This requires that the projection axes for all image sources be placed as close together as possible.
2. For subtractive superposition (at the source), data must be presented as dark markings on a transparent background.
3. For additive superposition (at the screen) data must be presented as light markings on an opaque background.
4. Light output from different projectors must be balanced.

Multi-color projections. A special case of superposition is the use of three-color additive projection for coding or emphasizing certain data in the display. Optimum colors for coding are red, green and blue obtained either with filters or dichroic mirrors. There is considerable latitude in choice of the particular filters or dichroic mirrors, depending on the source light and other details of the projection system (Riszy, 1965). If the intensity of the three primary colors is balanced, such a system can give seven readily discriminable colors for coding purposes: red, yellow, green, blue-green (cyan), blue-purple (magenta), and white (Riszy, 1965).

3.9.5 Legibility of Projected Data

Since much of the information displayed consists of alphanumeric or similar patterned symbols, some rules for insuring their legibility are the following (for further information, see Ter Louw).

1. Use a Simple Style of Numerals and Letters. Capital letters are preferred over lower case, except for lengthy messages. Use stroke with of 1/6 to 1/8 of numeral or letter height. Width may be still narrower for light markings on a dark background.
2. Size of Numerals and Letters. Use the values in Table 3–13. The three methods of controlling height of numerals and letters give equivalent results.
3. Contrast. May be either dark on light background, or vice versa, except where superposition is used. Avoid colored markings against colored backgrounds of comparable brightness. Maximum contrast aids legibility.
4. Alignment. For superimposed alphanumeric data or other symbols, avoid misregistration exceeding 50% (Snadowsky et al., 1964).

3.9.6 Collimated or Heads-Up Displays

Collimated images, reflected to the observer from a transparent mirror, offer advantages for some special applications. A common and successful application is a sighting device in fighter aircraft. The basic optics for such a sight are illustrated in Figure 3–86. The illuminated aimpoint (cross hair or other symbol) is seen at infinity superimposed over the target area. Motion of either the mirror or the source image can be used to shift the direction at which the aimpoint appears. As can be seen from Figure 3–86, the aimpoint will be visible only as long as the observer's eyes remain within the collimated beam,

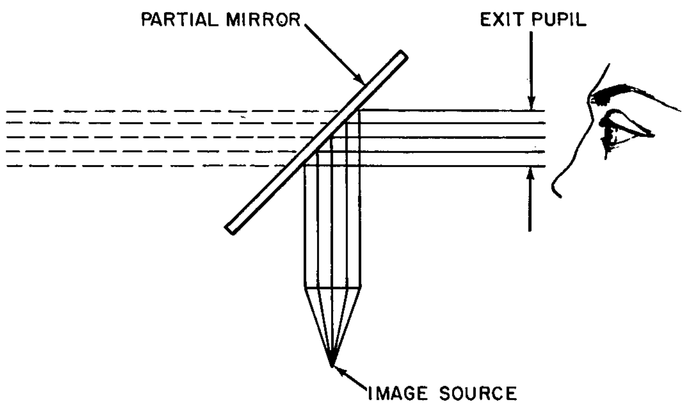


FIGURE 3–86. Basic optics of collimated display.

TABLE 3–13. HEIGHT OF NUMERALS AND LETTERS IN PROJECTED DISPLAYS

	Minimum acceptable	Preferred minimum
Visual angle—minutes	10	15
Letter ht./max. observer distance	1/344	1/230
Letter ht./width of source image*	1/43	1/30

* Assumes maximum observer distance of 8× image width.

or exit pupil. Eye motion within this area does not affect the direction at which the aimpoint is seen.

This type of display also has an application where it is desirable to give the operator visual information as he looks at the external world: during the final phases of an instrument landing, or other flight at low altitude, for example. Here it is helpful to reflect such information as altitude, airspeed, heading, or position from the windshield. In doing so, the temporary losses of external vision while the pilot looks at cockpit instruments are avoided. A related application is to mount a cathode ray tube, collimating lens, and mirror in front of one eye on a helmet. The cathode ray tube display is then visible regardless of head movements.

While these or other collimated displays may be practical and desirable, there are several difficulties:

1. Because of the limited exit pupil area, head motions of the viewer are restricted. It is best if the image is visible to both eyes, requiring an exit pupil in excess of the interpupillary distance (2.84 in. for the 95 percentile). As the eyes move away from the center of the exit pupil area, the edges of the image disappear.

2. Image brightness must be adjustable enough to maintain contrast against anticipated backgrounds. For contrast against sunlit clouds and snow (up to 10,000 ft.–L.), for example, image brightness would need to be 2000 ft.–L. for only 20% contrast. In providing adequate brightness, allowance must be made for only partial reflectance by the transparent mirror. Use of a colored image maximizes contrast against anticipated backgrounds. In addition, excessive image brightness or a large illuminated area interferes with seeing outside.

3. Division of operator attention. Superimposing a collimated image over the operator's field of view does not mean he can give simultaneous attention to both the presented data and other objects in view. Human visual attention is essentially limited to one channel at a time. The collimated image display does, however, avoid the need for the large accommodation and fixation changes otherwise required in shifting from displays in the crew station to viewing of objects outside.

References

Alluisi, E. A. Lineal inclination in encoding information symbolically on cathode ray tubes and similar displays. ASD TR 61–741, 1961, Aeronautical Systems Division, WPAFB, Ohio.

Atkinson, W. H., Crumley, L. M., & Willis, M. P. A study of the requirements of letters, numbers, and markings to be used on trans-illuminated aircraft control panels: V. The comparative legibility of three fonts for numerals. Report No. TED-Namel-609-5, 1952, Naval Air Materiel Center, Philadelphia.

Attneave, F. Some informational aspects of visual perception. *Psychological Review*, 1954, **61**, 183.

Attneave, F. Transfer of experience with a class-schema to identification-learning of patterns and shapes. *Journal of Experimental Psychology*, 1957, **54**, 81.

Baker, C. A., & Grether, W. F. Visual presentation of information. WADC TR 54–160, 1954, Wright Air Development Center WPAFB, Ohio.

Baker, C. A., & Nicholson, R. M. Raster scan parameters and target recognition. Proceedings of the 19th Annual National Aerospace Electronics Conference (NAECON), 1967.

Baker, C. A., & Vanderplas, J. M. Speed and accuracy of scale reading as a function of the number of reference markers. *Journal of Applied Psychology*, 1956, **40**, 307.

Baker, C. A., Morris, D. F., & Steedman, W. C. Target recognition in complex displays. *Human Factors*, 1960, **2**, 51.

Barber, J. L., & Garner, W. R. The effect of scale numbering on scale reading accuracy and speed. *Journal of Experimental Psychology*, 1951, **41**, 298.

Barmack, J. E., & Sinaiko, H. W. Human factors problems in computer generated graphic displays. Report S-234, 1966. Institute for Defense Analyses.

Bennett, C. A., Winterstein, S. H., Taylor, J. D., & Kent, R. E., A study of image qualities and speeded intrinsic target recognition. IBM Report No. 63-535-1, 1963, Oswego, New York.

Bennett, C. A., Winterstein, S. H., & Kent, R. E. Image quality and target recognition. *Human Factors*, 1967, **9**, 5.

Bergum, B. O. Vigilance: A guide to improved performance. Research Bulletin 10, 1963, Human Resources Research Organization.

Berry, R. N., Riggs, L. A., & Duncan, C. P. The relation of vernier and depth discriminations to field brightness. *Journal of Experimental Psychology*, 1950, **40**, 349.

Bishop, A. P., & Crook, M. N. Absolute identification of color for targets presented against white and colored backgrounds. Report No. WADD TR 60–611, 1960, Wright Air Development Division, WPAFB, Ohio.

Blackwell, H. R. Contrast thresholds of the human eye. *Journal of the Optical Society of America*, 1946, **36**, 624.

Blackwell, H. R., & Moldauer, A. B. Detection thresholds for point sources in the near periphery. Project 2455, 1958, University of Michigan Engineering Research Institute.

Blondel, A., & Rey, J. Sur la perception des lumieres breves a la limite de leur portee. *Journal of Physiology* (Paris), 1911, **1**, 530.

Bolin, S. F., Sadacca, R., & Martinek, H. Team procedures in image interpretation. Technical Research Note No. AD–480–533, December 1965, U. S. Army Personnel Research Office (Now, Behavioral Science Research Lab.), Washington, D. C.

Bowen, H. M., & Gradijan. Graphical display of multiparametric information. Part II. Experimental studies of chart design. AMRL TDR–62–115, 1963. Aerospace Medical Research Laboratory, WPAFB, Ohio.

Bowerman, E. R., & Wasserman, M. S. Electroluminescent alphanumeric message display with photoconductor control. Proceedings of the Society for Information Display, 2nd Annual National Symposium, October 1963.

Boynton, R. M., & Bush, W. R. Laboratory studies pertaining to visual air reconnaissance. Report No. WADC–TR–55–304, Part 2, 1957, Wright Air Development Center, WPAFB, Ohio.

Bradley, J. V. Desirable control-display relationships for moving-scale instrument.

WADC–TR–54–423, 1954, Wright Air Development Center, WPAFB, Ohio.

Brainard, R. W., & Caum, K. B. Evaluation of an image quality enhancement technique. AMRL TR 65–143, 1965, Aerospace Medical Research Laboratory, WPAFB, Ohio.

Brainard, R. W., Marshall, R. H., & Hanford, E. C. Resolution requirements for identification of targets in television imagery. Report No. NA63H–794, 1965, North American Aviation, Inc., Columbus, Ohio.

Brown, F. R. A study of the requirements for letters, numbers, and markings to be used on trans-illuminated aircraft control panels: IV. Legibility of uniform stroke capital letters as determined by size and height to width ratio and as compared to Garamond Bold. Report No. TED–NAMEL–609–4, 1953, Naval Air Materiel Center, Philadelphia.

Brown, J. L. The effect of different preadapting luminances on the resolution of visual detail during dark adaptation. Report No. WADC–TR–52–14, 1952, Wright Air Development Center, WPAFB, Ohio.

Brown, J. L. Experimental investigations of flash blindness. *Human Factors*, 1964, **6**, 503–516.

Brown, J. L. Flash blindness: Time required for detection of acuity targets following exposure to short adapting flashes. *Journal of Engineering Psychology*, 1964, **3**, 53.

Burg, A. Visual acuity as measured by static and dynamic tests: A comparative evaluation. *Journal of Applied Psychology*, 1966, **50**(6), 460–466.

Carel, W. L. Human factors on ANIP. NONR 1076(00), 1960, General Electric Co., Ithaca, N.Y.

Carel, W. L. Pictorial displays for flight. JANAIR TR 2732 .01/40, Office of Naval Research, ASTIA Document No. 627669, 1965.

Carel, W. L. Visual factors in the contact analog. NONR 1976 (00) 1961, General Electric Co., Ithaca, New York.

Carter, L. F. A study of the best design of tables and graphs used for presenting numerical data. Report TSEAA–694–1C, 1946, Wright Air Development Center, WPAFB, Ohio.

Chalmers, E. L., Goldstein, M., & Kappauf, W. E. The effect of illumination on dial reading. Report No. AF–TR–6021, 1950, Aerospace Medical Laboratory, WPAFB, Ohio.

Chambers, E. G. Design and use of scales, dials and indicators. APU Report No. 254/56, 1956, Medical Research Council, Cambridge, England.

Chapanis, A. Words, words, words. *Human Factors*, 1965, **7**, 1–17.

Chapanis, A., & Halsey, R. Absolute judgments of spectrum colors. *Journal of Psychology*, 1956, **42**, 99.

Chapanis, A., & Leyzorek, M. Accuracy of visual interpolation between scale markers as a function of the number assigned to the scale interval. *Journal of Experimental Psychology*, 1950, **40**, 655.

Churchill, A. V. The effect of scale interval length and pointer clearance on speed and accuracy of interpolation. *Journal of Applied Psychology*, 1956, **40**, 358–361.

Cobb, P. W., & Moss, F. K. The four variables in the visual threshold. *Franklin Institute Journal*, 1928, **205**, 831.

Cohen, J. Binocular disparity as a coding dimension for pictorial instrument and radar displays. WADC TR 55–393, 1955, Wright Air Development Center, WPAFB, Ohio.

Cohen, J., & Dinnerstein, A. J. Flash rate as a visual coding dimension for information. WADC TR 57–64., 1958, Wright Air Development Center, WPAFB, Ohio.

Connell, S. C. The relative effectiveness of presenting numerical data by the use of scales and graphs. WADC, TSEAA–694–1M, 1947, Wright Air Development Center, WPAFB, Ohio.

Conover, D. W. The amount of information in the absolute judgment of Munsell hues. WADC–TN58–262, 1959, Wright Air Development Center, WPAFB, Ohio.

Crawford, A. The perception of light signals: the effect of the number of irrelevant lights. *Ergonomics,* 1962, **5,** 417–428.

Crawford, A. The perception of light signals: The effect of mixing flashing and steady irrelevant lights. *Ergonomics,* 1963, **6,** 287–294.

Cypher, I. R. *Audio visual guide to better projection.* Morton Grove Ill.: Radiant Mfg. Corp.

Davies, E. B. Contrast thresholds for air-to-ground vision. Technical Report No. 65089, 1965, Royal Aircraft Establishment, Great Britain.

Dougherty, D. J., Emery, J. H., & Curtin, J. G. Comparison of perceptual work load in flying standard instrumentation and the contact analog vertical display. TR D 228–421–019, 1964, Bell Helicopter Co., Fort Worth, Contract NONR 1670 (00).

Duntley, S. Q. The visibility of distant objects. *Journal of the Optical Society of America.* 1948, **38,** 237–249.

Eastman Kodak Co. *Audiovisual projection.* (Pamphlet No. S–3). New York, 1965.

Elias, M. F., Snadowsky, A. M., & Rizsy, E. F. The relation of number of scan lines per symbol height to recognition of televised alphanumerics. Report No. RADC-TDR-64-433, 1964, Rome Air Development Center, New York.

Emery, J. H., & Dougherty, D. J. Contact analog simulator evaluations: Climbout, low cruise and descent maneuvers. TR D 228–421–017, 1964, Bell Helicopter Co., Fort Worth, Contract NONR 1670 (00).

Erickson, R. A. Visual detection of targets: Analysis and review. NAVWEPS Report No. 8617, NOTs TP 3645, 1965.

Foster, H. W. Information displays and information processing tasks. SP–1811, 1964, System Development Corp.

Ford, A., Rigler, D., & Dugan, G. E., Pantograph radar tracking: Point centering experiments. Report No. TR 5969, 1949, Aerospace Medical Laboratory WPAFB, Ohio.

Fuller, P. R. Human engineering study of electroluminescent displays. I. Time and accuracy of reading seven-stroke EL numeric compared with back projected regular shaped arabic numerals. Engineering Memorandum Report 182, 1965, Lear Seigler Inc., Grand Rapids, Mich.

Gebhard, J. W., & Bilodeau, E. A. Appraisal of an experimental PPL presenting simulated bearing, range and height information. Report No. 166–1–37, 1947, System Research Field Laboratory, Johns Hopkins University, Jamestown, R. I.

Green, B. F., & Anderson, L. K. Color coding in a visual search task. *Journal of Experimental Psychology,* 1956, **51,** 19.

Green, B. F., McGill, W. J., & Jenkins, H. M. The time required to search for numbers on large visual displays. TR No. 36, 1953, MIT Lincoln Lab.

Grether, W. F. Psychological factors in instrument reading: I. The design of long-scale indicators for speed and accuracy of quantitative readings. *Journal of Applied Psychology,* 1949, **33,** 363–372.

Hanes, R. M., & Rhoades, M. V. Color identification as a function of extended practice. *Journal of the Optical Society of America,* 1959, **49,** 1960.

Hanes, R. M., & Williams, S. B. Visibility on cathode ray tube screens. *Journal of the Optical Society of America,* 1948, **38,** 36B.

Harshbarger, J. H., & Basinger, J. D. Test and evaluation of electronic image generation and projection devices. Vol. III. Evaluation of projection screens. Technical Report TR–65–116, 1965, Aerospace Medical Research Laboratories, USAF, WPAFB, Ohio.

Hecht, S., Ross, S., & Mueller, C. G. The visibility of lines and squares at high brightness. *Journal of the Optical Society of America* 1947, **37,** 500.

Hill, J. H., & Chernikoff, R. Altimeter display evaluation: Final Report No. 6242, 1965, U. S. Naval Res. Labs.

Hill, N. E. G. The recognition of colored light signals which are near the limit of visibility.

Proceedings of Physiological Society (London), 1947, **59**, 560.

Howell, W. C., & Fuchs, A. H. The study of graphic languages. RADC–TR–61–76., 1961, Rome Air Development Center, Rome New York.

Illuminating Engineering Society. *Lighting handbook: The standard lighting guide* (4th ed.), IES, New York, 1966.

Ireland, F. H., Kinslow, W., Levin, E., & Page, D. Experimental study of the effects of surround brightness and size on visual performance. Report No. AMRL–TR–67–102, 1967. Aerospace Medical Research Lab., WPAFB, Ohio.

Jennings, L. B., Meeker, F. B., Pravor, G. A., & Cook, R. N. Ground resolution study final report. RADC–TDR 63–224, 1963, Rome Air Development Center, Rome, New York.

Kappauf, W. E. Design of instrument scales for maximum legibility: V. Origin location, scale break, number location, and contrast direction. USAF Aero Med. Lab. AF–TR–6366, 1951, WPAFB, Ohio.

Kelley, C. R. Further research on the predictor instrument. TR 252–60–2., 1960, Office of Naval Research.

Kelley, C. R. *Manual and automatic control.* New York: John Wiley & Sons, 1968.

Kristofferson, A. B. Visual detection as influenced by target form. In J. W. Wulfeck & J. H. Taylor (Eds.), *Form discrimination as related to military problems.* NRC Publication No. 561, 1957.

Lamar, E. S., Hecht, S., Schlaer, S., & Handley, C. D. Size, shape and contrast in detection of targets by daytime vision. I. Data and analytical discussion. *Journal of the Optical Society of America,* 1947, **47**, 531–545.

Leibowitz, H. W., & Sulzer, R. L. An evaluation of three-dimensional displays. Report of Working Group 6, Armed Forces National Research Council Committee on Vision, 1965.

Long, E. R., Reid, L. S., & Queal, R. W. Factors determining the legibility of letters and words derived from elemental printers. Wade Tech Reports Nos. 5922–5924, 1951.

Lovelace, N. E. Direction-of-motion stereotypes: A review. *Ergonomics,* 1962, **5**, 357–383.

Lythgoe, R. J. The measurement of visual acuity. Rep. No. 173, 1932, Medical Res. Council, London, England.

Mackworth, J. F. Perceptual coding as a factor in short term memory. *Canadian Journal of Psychology,* 1966.

Mandelbaum, J., & Sloan, L. L. Peripheral acuity with special reference to scotopic illumination. *American Journal of Opthalmology,* 1947, **30**, 581.

McKechnie, D. Effects of target specific briefing information on location and identification of targets from SLR imagery. Technical Report, 1967, Aerospace Medical Res. Labs., WPAFB, Ohio.

Mengelkoch, R. F., & Houston, R. C. Investigation of vertical displays of altitude information: AF WADC–TR–57–384; II. 57–385; III 57–549., 1957. Wright Air Development Center, WPAFB, Ohio.

Middleton, W. E. K. *Vision through the atmosphere.* Toronto: Univ. of Toronto Press, 1958.

MIL–C–25050A. Colors, aeronautical lights and lighting equipment, general requirements for. 1963.

MIL–L–650 3 E (USAF). Lighting equipment, aircraft, general specification for installation of. 1966.

MIL–P–7788D. Panels, information, integrally illuminated. 1967.

MIL STD 33558. Numerals and letters, aircraft instrument dial, standard. 1957.

Miller, J. W., & Ludvigh, E. The effect of relative motion on visual acuity. *Survey Opthamology,* 1962, **7**(2). 83–116.

Moler, C. G. Helicopter armament program air-to-ground target detection and identification. TM 1–62, 1962, Human Engin. Lab. Aberdeen, Md.

Moon, P., & Spencer, D. D. Visual data applied to lighting design. *Journal of the Optical Society of America*, 1944, **34,** 605.

Mote, F. A., & Riopelle, A. J. The effect of varying the intensity and duration of pre-exposure upon subsequent dark adaption in the human eye. *Journal of Comparative and Physiological Psychology*, 1953, **46,** 49.

Muller, P. F., Sidorsky, R. C., Slivinske, A. J., Alluisi, E. A., & Fitts, P. M. The symbolic coding of information of cathode ray tubes and similar displays. Report No.WADC–TR–55–375, 1955, Wright Air Development Center, WPAFB, Ohio.

Munsell, A. H. *Book of color.* Baltimore: Munsell Color Book Co., 1942.

Murrell, K. F. H., Lurie, W. D., & McCarthy, C. The relationship between dial size, reading distance and reading accuracy. *Ergonomics*, 1958, **1**(2), 182–190.

Nutting, P. G. Effects of brightness and contrast in vision. *Transcriptions of the Illuminating Engineering Society*, 1916, **11,** 939.

Oberly, H. S. The range of visual attention, cognition and apprehension. *American Journal of Psychology*, 1924, **35,** 332.

Peskin, J. C., & Bjornstad, J. The effects of different wavelengths of light on visual sensitivity. Memo Report MLREXD–694–93A. 1948, Aerospace Medical Research Lab., WPAFB, Ohio.

Peterson, D. J., & Smith, H. A. Development of high contrast electroluminescent techniques for aircraft displays. Project No. 6190, Technical Paper, 1965, Air Force Flight Dynamics Lab.

Rizsy, E. F. Color specifications for additive color group displays. Technical Report No. RADC–TR–65–278, 1965, Rome Air Development Center, Rome, New York.

Sabeh, R., Jorve, W. R., & Vanderplas, J. M. Shape coding of aircraft instrument zone markings. Technical Note No. 57–260, AD118343, 1958. Aerospace Medical Research Lab., WPAFB, Ohio.

Siegel, A. I., & Crain, K. Experimental investigations of cautionary signal presentations. *Ergonomics*, 1960, **3**(4), 339–356.

Simon, C. W., & Craig, D. W. Effects of magnification and observation time on target identification in simulated orbital reconnaissance. *Human Factors*, 1965, **7**(6) 569–583.

Sleight, R. S. The relative discriminability of several geometric forms. *Journal of Experimental Psychology*, 1952, **43,** 424.

Smith, S. L., & Thomas, D. W. Color versus shape coding in information displays. *Journal of Applied Psychology*, 1964, **48,** 137.

Snadowsky, A. M., Rizsy, E. F., & Elias, M. F. Misregistration in color additive displays. Report No. RADC–TDR–64–488, 1964, Rome Air Development Center, Rome, New York.

Snyder, H. L. Visual aspects of low level flight. ONR symposium on visual and display problems related to flight at low altitude, Autonetics, X4–610/3111. 1964.

Spencer, J. Pointers for general purpose indicators. *Ergonomics*, 1963, **6**(1), 35–49.

Steedman, W. C., & Baker, C. A. Target size and visual recognition. *Human Factors*, 1960, **2**(3), 120–127.

Taylor, J. H. Contrast thresholds as a function of retinal position and target size for the light-adapted eye. REF. 61–10, 1961, Scripps Institution of Oceanography.

Ter Louw, A. L. Legibility standards for projected material. Eastman Kodak Pamphlet No. S–4, Eastman Kodak Co.

Thorndike, E. L., & Lorge, I. *Teachers word book of 30,000 words.* New York: Teacher's College, Columbia University, 1944.

Tinker, M. A. *Legibility of Print.* Ames, Iowa: Iowa State University Press, 1963.

United States Department of Commerce, National Bureau of Standards. United States standard for the colors of signal lights. (Handbook 95) Washington, D.C.: NBS, 1964.

Van Ausdall, B. A. & Self, H. C. Effects of display polarity on target detection with side-looking radar. AMRL Tech. Rep. 64–82, 1964. AD 609–246, Aerospace Medical Research Laboratory, WPAFB, Ohio.

Vlahos, P. Selection and specification of rear projection screens. *Journal of the Society of Motion Picture and Television Engineers,* 1961, **70,** 87–95.

Vlahos, P. The three-dimensional displays: Its cues and techniques. *Information Displays,* 1965, **6,** 10.

Wesler, J. E. The effective intensity of flashing lighted aids to navigation. Sixth International Technical Conference on Lighthouses and Other Aids to Navigation. Report No. AD 242046, 1960, U. S. Coast Guard, Washington, D. C.

White, C. T. & Ford, A. Eye movements during simulated radar search. *Journal of the Optical Society of America,* 1960, **50**(9), 909–912.

Whiteside, T. C. D. The problems of vision in flight at high altitude. Agardograph. London: Butterworth's Scientific Publication, 1957.

Williams, A. C., Jr., Simon, C. W., Hangen, R. & Roscoe, S. N. Operator performance in strike reconnaissance. WADC TR 60–521, 1960, Wright Air Development Center, WPAFB, Ohio.

Woodson, W. E. & Conover, D. W. *Human engineering guide for equipment designers.* Berkeley: University of California Press, 1964.

Selected Readings

Handbook of the Illuminating Engineering Society. New York, 1966.

National Research Council. A survey report on human factors in undersea warfare. Prepared by the Panel on Psychology and Physiology, Committee on Undersea Warfare, 1949, Supplement 1954. National Research Council, Washington, D. C.

Stevens, S. S. (Ed.) *Handbook of experimental psychology.* New York: John Wiley and Sons, Inc., 1951.

Chapter 4

Auditory and Other Sensory Forms of Information Presentation

Bruce H. Deatherage

*Applied Research Laboratories, The University of Texas at Austin
Austin, Tex.*

In addition to visual presentation of information, the human operator may receive information through other sensory channels: hearing, smell, somesthesis, kinesthesis, and balance. Each is a potentially useful channel having advantages and limitations. This chapter emphasizes hearing and the presentation of auditory signals. Guidelines are provided for when and how to present each type of auditory or other sensory information. The perception of sound, the problems of masking and its reduction, and the choice of single and multiple signals are covered. Some information is given on other senses as media for information presentation. Speech as a mode of information presentation is discussed in detail in the following chapter.

CONTENTS

Armand N. Chambers contributed material to Section 4.8 of this chapter.

This chapter is based in part on material presented in Chapter 3 of the previous *Guide*. It was reviewed by Donald E. Broadbent and Gilbert C. Tolhurst.

4. Auditory and Other Sensory Forms of Information Presentation

4.1 The Use of Auditory Presentation

Some signals are better suited for auditory than for visual presentation and vice versa. The choice of which sense to use depends on: (a) the nature of the signal, (b) the conditions under which it must be received, and (c) the characteristics of the person involved. Table 4–1 summarizes situations in which one form is preferred over another.

TABLE 4–1. WHEN TO USE THE AUDITORY OR VISUAL FORM OF PRESENTATION

Use auditory presentation if:	Use visual presentation if:
1. The message is simple.	1. The message is complex.
2. The message is short.	2. The message is long.
3. The message will not be referred to later.	3. The message will be referred to later.
4. The message deals with events in time.	4. The message deals with location in space.
5. The message calls for immediate action.	5. The message does not call for immediate action.
6. The visual system of the person is overburdened.	6. The auditory system of the person is overburdened.
7. The receiving location is too bright or dark-adaptation integrity is necessary.	7. The receiving location is too noisy.
8. The person's job requires him to move about continually.	8. The person's job allows him to remain in one position.

4.1.1 Signals Suited to Auditory Presentation

Auditory presentation is preferred over visual presentation:

1. For signals of acoustic origin. Ingenious visual displays for speech have been devised, but none is likely to supplant hearing except when deafness or intense noise conditions render hearing useless.

2. For warning signals. A visual warning signal must be seen in order to warn. On the other hand, hearing is omnidirectional and cannot be involuntarily shut off. It is therefore best for calling attention to imminent or potential danger.

3. For situations when too many displays are visually presented—in piloting an airplane, for example.

4. For presenting information independently of head orientation—as when duties require body movement or head turning.

5. For situations when darkness limits vision or makes seeing impossible.

6. For conditions of anoxia in high altitudes or high positive g forces; auditory sensitivity is more resistant to anoxia than is visual sensitivity. A man suffering from oxygen deficiency may have his vision seriously impaired, but he can still hear signals.

7. When signals must be distinguished from noise. The ear acts as a frequency analyzer, making it an effective detector of periodic signals in noise. Even when it is considerably weaker than the background noise, if the signal is a sinusoid (pure tone) or a combination of sinusoids (complex tones), the ear can detect it. The ear also efficiently detects periodic modulation in the very-low-frequency range and responds to variations in intensity or frequency.

4.1.2 Choosing the Form of Auditory Presentation

Use tonal or noise signals, rather than speech:

1. For simplicity.

2. When listeners are trained to understand coded signals.

3. For designating a point in time that has no absolute value.

4. When immediate action is desired.
5. In conditions unfavorable for receiving speech messages. (Tonal signals can be heard at noise levels where speech is unintelligible.)
6. When security of the message is desired; coded tonal or noise signals may be used.
7. If speech communication channels are overloaded.
8. If speech will mask other speech signals or annoy listeners for whom the message is not intended.

Use speech rather than tonal or noise signals under these conditions:

1. For flexibility.
2. To identify a message source.
3. When listeners are without special training in coded signals.
4. There is a necessity for rapid two-way exchanges of information.
5 The message deals with a future time requiring some preparation. (Example: The countdown preparatory to firing a missile—tonal signals could be miscounted.)
6. Situations of stress might cause the listener to "forget" the meaning of a code.

4.1.3 Some Common Uses for Auditory Presentation

Speech is best for transmitting urgent messages, since the maximum transmission rate of speech is 250 words per minute (wpm). Morse code, on the other hand, is intelligible under low signal-to-noise ratios, but the maximum rate of transmission is about 30 wpm.

Noise signals are best for sea and air navigation. Lighthouse diaphones that pulse messages in Morse code to indicate position are audible for long distances; whistling and bell buoys can locate channels and shoals; radio-range signals and fan-marker radio beacons can mark airways. Other types of signals, detecting, echo ranging, warning, and alarm signals, are commonly auditory.

Table 4–2 summarizes the principal characteristics and special features of different types of auditory alarm and warning signals. Examples given are for horns, whistles, sirens, bells, buzzers, chimes, gongs, and oscillators.

Design recommendations. The following principles for selection and design of alarm and warning signals should be observed (see also Table 4–3):

1. At a minimum, use sounds having frequencies between 200 and 5000 Hz, and if possible, between 500 and 3000 Hz, because the human ear is most sensitive to this middle range.
2. Use sounds having frequencies below 1000 Hz when signals must travel long distances (over 1000 ft.) because high frequencies are absorbed in passage and hence cannot travel as far. Figure 4–1 shows attenuation of sounds of various frequencies in calm air for distances from 10 to 10,000 ft. under conditions free from the effects of reflecting surfaces and obstacles.
3. Use frequencies below 500 Hz when signals must bend around obstacles or pass through partitions. (See Figure 4–2.)
4. In noise, signal frequencies different from those most intense frequencies of the noise are best in order to reduce masking of the signal.
5. Use a modulated signal to demand attention. Intermittent beeps repeated at rates of one to eight beeps per second, or warbling sounds that rise and fall in pitch are seldom encountered, and are therefore different enough to get immediate attention. If speech is necessary during an alarm, use an intermittent, pure-tone signal of relatively high frequency.
6. Use complex tones rather than pure sinusoidal waves, because few pure tones can be positively identified but each complex sound is noticeably different from other sounds.

4.1.4 When to Use Auditory Displays

Certain considerations are helpful in deciding when to use an auditory display for spatial information:

1. Use auditory displays to relieve the eyes. Although the eye is better for spatial discrimination, it can look in only one direction at a time. In general, auditory spatial displays are recommended only when the eyes are fully engaged and additional spatial information is needed.
2. Use auditory displays (other than speech) to present restricted information, such as the following:

TABLE 4–2. TYPES OF ALARMS, THEIR CHARACTERISTICS AND SPECIAL FEATURES

Alarm	Intensity	Frequency	Attention-getting ability	Noise-penetration ability	Special features
Diaphone (foghorn).	Very high	Very low	Good	Poor in low-frequency noise. Good in high-frequency noise.	
Horn	High	Low to high.	Good	Good	Can be designed to beam sound directionally. Can be rotated to get wide coverage.
Whistle	High	Low to high.	Good if intermittent.	Good if frequency is properly chosen.	Can be made directional by reflectors.
Siren	High	Low to high.	Very good if pitch rises and falls.	Very good with rising and falling frequency.	Can be coupled to horn for directional transmission.
Bell	Medium	Medium to high.	Good	Good in low-frequency noise.	Can be provided with manual shutoff to insure alarm until action is taken.
Buzzer	Low to medium.	Low to medium.	Good	Fair if spectrum is suited to background noise.	Can be provided with manual stutoff to insure alarm until action is taken.
Chimes and gong.	Low to medium.	Low to medium.	Fair	Fair if spectrum is suited to background noise.	
Oscillator	Low to high.	Medium to high.	Good if intermittent.	Good if frequency is properly chosen.	Can be presented over intercom system.

TABLE 4–3. SUMMARY OF DESIGN RECOMMENDATIONS FOR AUDITORY ALARM AND WARNING DEVICES

Conditions	Design recommendations
1. If distance to listener is great—	1. Use high intensities and avoid high frequencies.
2. If sound must bend around obstacles and pass through partitions—	2. Use low frequencies.
3. If background noise is present—	3. Select alarm frequency in region where noise masking is minimal.
4. To demand attention—	4. Modulate signal to give intermittent "beeps" or modulate frequency to make pitch rise and fall at rate of about 1–3 cps.
5. To acknowledge warning—	5. Provide signal with manual shutoff so that it sounds continuously until action is taken.

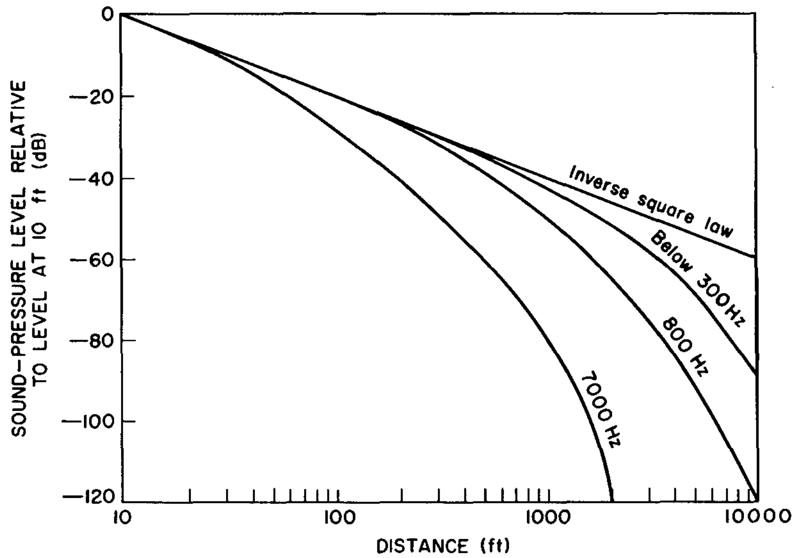


FIGURE 4–1. Sound pressure level attenuation as a function of the distance of transmission in calm air.

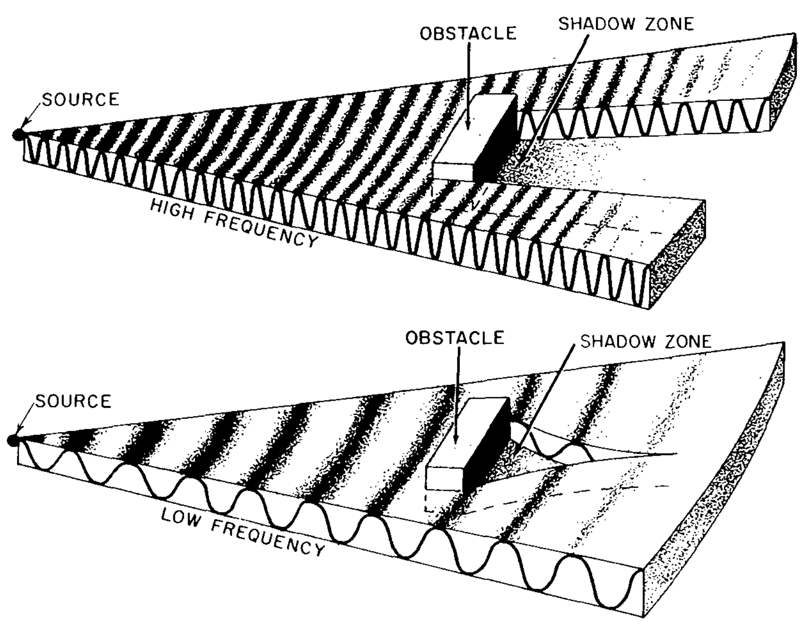


FIGURE 4–2. High-frequency sounds (wavelength shorter than width of obstacle) are greatly attenuated in shadow zone. Barriers based on this principle are sometimes used for control of high-frequency noise.

(*a*) "Yes-no" information and indications of amount or degree. Auditory displays can represent error or deviation from a course, speed, attitude, or other "normal" condition.

(*b*) Continuous information. For example, radio-range signals present practically continuous information about one kind of event—the course the aircraft is flying.

(*c*) Automatic information—recorded word signals as from an automatic annunciator.

3. Use auditory displays of tonal or noise signals when speech channels are already fully employed. Most of the auditory displays that utilize tonal signals can be heard through speech, and, conversely, speech can be understood while hearing the tonal signals over the same receiving system.

Radio-range signals are a good example of all of the above points. They relieve the pilot's eyes, present one restricted kind of information (course), and do this continuously while giving both "yes-no" and amount of error in course. They present information automatically, and avoid overburdening the speech channel while making it possible for the pilot to hear both speech communication and range signals.

Design recommendations. When designing auditory displays for spatial orientation:

1. Confine representation to a single dimension; multidimensional displays are less effective than their visual counterparts.

2. Provide a standard stimulus to represent the "normal," then make abrupt changes to indicate departures from the normal. Human listeners are sensitive to frequency or intensity changes but poor at identifying a unique signal.

3. Provide changes in intensity rather than frequency as a spatial cue. Because everyone with normal hearing can detect changes in intensity, it is easier to control these changes.

4. Use intermittent to repeated changes in a signal rather than a single change followed by a continuous signal. The ear is much more likely to detect changes in a signal occurring every second or two than at longer intervals.

5. If absolute identification is required, limit the number of signal categories to four because listeners cannot identify correctly more than a few different intensities, pitches, or interruption rates.

6. The following "natural" relationships between auditory signals and the dimensions they represent are quickly learned or are perceived with little training:

(*a*) Binaural intensity differences serve to localize (in bearing) the direction of a sound.

(*b*) Pitch differences naturally represent up and down (high and low pitch). To indicate climb or upward pointing, raise the pitch. Combined with binaural changes in pitch from one ear to the other, "left wing high" for instance, can be represented.

(*c*) A slow interruption rate is a natural indication of speed—an increase or decrease in interruption rate is immediately perceived as a change in speed (or rate) of interruption.

4.1.5 Target Tracking

Auditory signals have proved satisfactory for compensatory tracking. Tracking is controlling an instrument to maintain a desired value (compensatory tracking) or following a moving reference marker (pursuit tracking). Compensatory tracking requires a display in direction and magnitude of the error.

In one system displaying error direction only, a 400-Hz tone was switched from one ear to the other when the operator veered off target. When he centered his joy-stick control "on target," he heard nothing in either ear. This method of auditory presentation was compared with visual tracking in which one of two lights signaled "off target" to the left or right. The two kinds of presentation were equally good for both simple courses (two-cycle-per minute sinusoid) and complex courses (combinations of 2-, 6-, and 15-c.p.m. sinusoids), (Humphrey and Thompson, 1952).

In a more complex system displaying both direction and magnitude of error, auditory tracking performance was not as good as with a visual presentation (Humphrey and Thompson, 1953). A continuous tone was presented in both ears for "on target," shifted to the left or right when "off target," while simultaneously interrupted at a rate proportional to the error in displacement. When compared with a corresponding visual presentation in the form of a spot on an oscilloscope, the auditory presentation of this system was inferior. But the operator was permitted to expand the visual presentation by adjusting a gain control. Had a similar adjustment of the auditory presentation been provided, the results probably would have been more favorable to the auditory method.

Although visual displays are preferred for presenting most radar information, auditory presentation has advantages for rather special purposes:

1. To detect the presence of radar signals. The pulse-repetition frequency often is in the audible frequency range. When it is, the train of radar impulses can be used to detect the presence of radar signals.
2. To identify particular radars. With training, a listener can identify such characteristics of a radar signal as repetition rate.
3. To detect a target through nonelectronic jamming.

4.2 Perception of Sound Signals

4.2.1 Characteristics of Sound Signals

Sound waves are a particular kind of wave, called elastic, which are propagated in media, such as air and water, characterized by mass and elasticity. Momentum is transferred from particle to particle in the medium, and particles tend to "spring" back to their original positions. The physical parameters of sound are *frequency*, *amplitude*, *duration*, and *waveform*. The *phase* relations among components of a complex sound have some importance, too. Each of these physical dimensions has some subjective aspect.

Frequency. Frequency is expressed in hertzes (Hz, or cycles per second). A pure tone will have a simple sinusoidal waveform, and complex sounds can be analyzed into the set of sinusoids that make them up. A convenient form for this is $a_1 \sin 2\pi f_1 t + a_2 \sin 2\pi f_2 t + \ldots$. The constants a_i represent maximum amplitude of particle displacement, and f_i is frequency. There is nothing in this expression denoting duration, but t is necessary to calculate instantaneous amplitude. Neither is there any factor for phase shown. Phase will be discussed a little later, but duration will not be discussed since its meaning is nearly self evident.

Amplitude. Since sounds are pressure waves varying around the normal air pressure, the amplitude measured as *average* pressure does not convey any meaning. The root-mean-square (r.m.s.) pressure is therefore usually used to describe the magnitude of a sound.

Because man can hear sounds over an extremely wide range of sound pressures (from about 0.0001 to 1,000 μbar), instead of using the value of sound pressure directly, it is customary to deal with sound-pressure level (SPL) defined by

$$L = 20 \log(p_1/p_0) = 10 \log(p_1/p_0)^2 \quad (4\text{–}1)$$

where L is the sound-pressure level of sound pressure p_1, and p_0 is a reference pressure.

The unit of sound-pressure level is the decibel (dB). The dB is a relative unit; it refers to a ratio of sound pressures (e.g., p_1 and p_2). Thus, the difference between two sound-pressure levels L_1 and L_2 is

$$\begin{aligned} L_2 - L_1 &= 20 \log(p_2/p_0) - 20 \log(p_1/p_0) \quad (4\text{–}2) \\ &= 20 \log(p_2/p_1). \end{aligned}$$

Note that the reference pressure (p_0) does not affect the difference, provided that the same reference pressure is used in both expressions for the two levels.

The pressure $p_0 = 0.0002$ μbar has been adopted as a standard reference pressure, and it is understood, ordinarily, when a sound-pressure level is stated. Practically all commercially available sound-measuring equipments are calibrated to read sound-pressure level based on this reference pressure.

The standard reference pressure of 0.0002 μbar was adopted because it is very close to the lowest level (absolute threshold) the human observer can hear when the signal is a 1000-Hz pure tone. Figure 4–3 shows typical SPL's for some common sounds. Also shown are the corresponding r.m.s. sound pressures and intensities for sound waves in air.

Waveform. Complex sounds can be either periodic or aperiodic. Periodic sounds tend toward tonality or musicality, while aperiodic sounds are impulsive, like clicks or booms, or "noisy" like the hiss of an air jet or the roar of a rocket engine. A convenient way for describing or characterizing a complex sound is by its waveform.

Waveform is most conveniently specified in terms of the various frequency components that

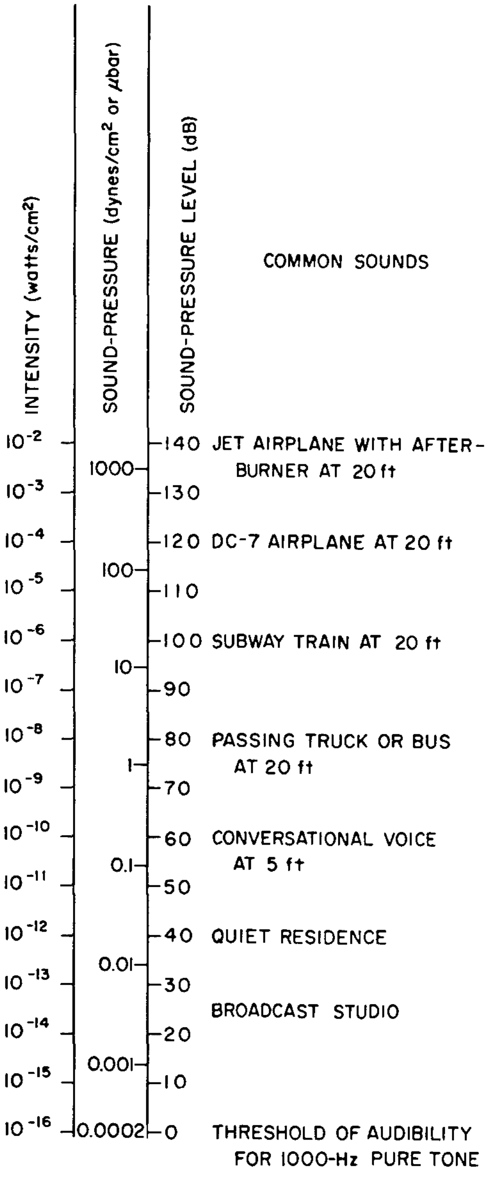


FIGURE 4–3. Typical SPL's for some common sounds.

comprise the sound. Thus, sounds having a precisely periodic waveform are composed of a series of single-frequency components located at discrete frequencies and integrally related. Aperiodic sounds, similarly, might consist of discrete components, or they might be composed of a continuum of frequency components distributed over a broad band; the former are said to have a line spectrum, the latter a continuous spectrum.

Phase. Phase relations among components of periodic complex sounds are useful for identifying sound sources. The utility of phase differences seems to lie in the listener's ability to use them in identifying a source of sound. Skilled blind travelers, for example, make good use of natural sounds and echoes for getting about in their daily lives. This is partially due to the fact that the two ears are separated. If the head is turned from the sound source, the sound must travel further to reach one ear than it does for the other, resulting in a phase difference. However, man does have some sound localization ability under monaural conditions.

4.2.2 Methods of Presenting Sound Signals to the Ears

The fact that man has two ears can be put to good use in designing auditory-presentation systems. Sometimes, in monitoring situations, the two ears can be used relatively independently of each other. In other cases, by adjusting the interaural phase and amplitude of signals, the two ears can use directional information effectively (Licklider, 1951). These are examples of the use of binaural presentation. Monaural presentation is used where it is desirable to leave one ear uncovered for hearing ambient sounds.

Binaural Presentation

The fact of man's having *two* ears provides opportunities for improving the detectability of signals as compared with monaural cases. Stimulating both ears identically is called diotic stimulation, whereas stimulating with an interaural phase or amplitude difference, or both, is called dichotic stimulation. Diotic presentation lowers the threshold of hearing over that for strictly monaural presentation, but still more improvement can be had by taking proper advantage of dichotic presentation where differences in amplitude and phase are used. The basic fact is that a signal presented together with an unwanted sound, or masking noise, is easier to detect if the signal appears in one apparent azimuthal location and the noise in another. This difference in location can be achieved by manipulating the interaural amplitude or phase, or both, of the signal and noise independently. Some gain is achieved also if the two ears receive from independent noise sources and the same signal source.

Increasing detectability. Although two ears can permit a "release from masking" in the presence of radio static or sonar water noise, two ears have little advantage because of the impossibility of independent control over signal and noise phases. The aim should be to obtain control by use of widely separated directional radio antennas, or by sonar transducer arrays, or by multiplexing. If inversion of the interaural phase of a low-frequency tonal signal (or aperiodic signals like speech) is possible, leaving the noise unchanged, a gain of as much as 14 to 16 dB can be achieved (Hirsh, 1948; Jeffress et al., 1952, 1962; Deatherage, 1966).

Binaural listening to a noise-free electrical signal that is presented in a strong acoustic noise is possible, i.e., listening with earphones in an aircraft. Here, inverting the connections to one earphone to put the signal 180° out of phase between the ears is recommended. If the ambient noise is of low frequency, irrespective of location, it tends to have the same phase at both ears. Thus, reversing the phase of the signal at one ear produces some release from masking. On the other hand, if the noise is made up largely of high frequencies, there *can* be substantial phase differences across the head with little gained by manipulating the signal's phase.

Various unsuccessful attempts have been made to increase signal identifiability by presenting a signal of few frequencies to one ear, and another one of a few different frequencies to the other ear. It appears that part of what one gains through separation may be lost through uncertainty about "where to listen," or through divided attention between competing messages. Hence, this method is to be used only where signals are infrequent and two signals are unlikely to occur simultaneously.

Providing auditory perspective. Two methods can be used to preserve the azimuth directional property of a sound picked up by microphones. One is by mounting a pair of microphones on a dummy head and feeding their output through separate channels to respective earphones. This method provides the aural illusion that the listener is located at the position of the microphones. A second method, known as stereophonic presentation, accomplishes an analogous result using multiple microphones and speakers to replicate the original sound field (Snow, 1953).

Artificial directional effects can be obtained by controlling intensity and time delay at earphones or loudspeakers. An interaural time delay between clicks of only 10 μsec. will produce barely perceptible shifts in localization (Klumpp and Eady, 1956), and a time delay of 0.6 msec. will completely lateralize a click, that is, the click will be perceived as being perpendicular to an ear. The sound will continue to remain lateralized for time differences up to about 2 msec., beyond which it will tend to be heard separately in the two ears as two clicks. In general, complex sounds, such as speech, clicks, or broadband noise, are localized much better than are pure tones.

4.2.3 Bone Conduction

In some situations, it is convenient to bypass the normal pathway of the external ear canal and introduce sound into the inner ear through the bones of the skull by use of bone-conduction vibrators. The vibrator is held in contact with the mastoid or forehead or any other position on the head with but a slight reduction in sensitivity. In fact, vibratory energy can be conducted to the inner ear from more remote locations on the body, i.e., elbow, sternum, knee, etc., but with greater transmission loss.

Bone conduction can be, and usually is, used in underwater hearing where use of underwater earphones or loudspeakers makes a poor match between the low acoustic impedance of the eardrum and the high impedance of the water. A lowered sensitivity (about 50 dB) of the ear to sound (Hamilton, 1957) is the result. When the head is immersed in water, sound signals can be presented through bone conduction vibrators because it appears that underwater sound is conducted to the inner ear about as readily by bone conduction as by the normal eardrum-middle-ear path.

By bone conduction, persons hear frequencies not usually audible by air conduction. If sufficiently powerful, frequencies even higher than 100 kHz can be perceived, but are heard as the highest pitch audible by air conduction. This power can cause permanent damage to hearing, however, and the damage can be induced without warning pain (Deatherage et al., 1954).

Frequent reports, especially in the popular press, of the "direct hearing by the brain" of RF electromagnetic energy require some explanation. If the RF energy is amplitude modulated at sonic frequencies, the tissues of the body are driven to propagate *acoustic* energy, and thus hearing may result (von Gierke and Sommer, 1964). Considerations of power requirements may often dictate that other means of communicating the signal be chosen. It must clearly be recognized that at least some residual hearing be present for hearing sensations to result from this method of stimulation. Profoundly deaf persons cannot be made to hear through this method, since its mechanism is *acoustic* and not direct stimulation of the afferent nerves or brain.

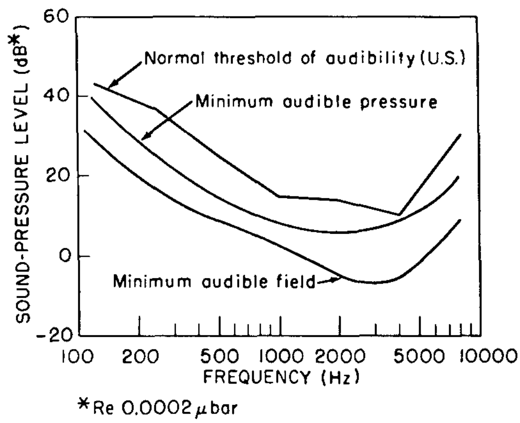


FIGURE 4-4. Three absolute threshold curves for pure tones.

4.2.4 Absolute Threshold of Hearing

The absolute threshold of hearing is the minimum sound-pressure level of a specified sound that is required to elicit the sensation of hearing in a specified fraction of trials (ordinarily 50%) with no detectable airborne masking noise present. The value of the absolute threshold depends on the type of sound; its frequency; duration; repetition rate, etc.; method of presentation, whether by loudspeaker or headset, binaural or monaural, etc.; and who the listener is, with respect to age, past exposure to noise, listening experience, history of ear trouble, etc.

Pure-Tone Thresholds

There are three generally accepted absolute thresholds for pure tones: the Minimum Audible Field (MAF), the Minimum Audible Pressure (MAP), and the Normal Threshold of Audibility (NTA). These are all shown as a function of frequency in Figure 4–4.

1. Minimum Audible Field is defined as the sound-pressure level (SPL) at the absolute threshold of a young but trained listener with no history of ear trouble. The listener is facing the source in a free-sound field (open air or anechoic room) and SPL is measured at the point where the center of the head would be.

2. Minimum Audible Pressure is the sound-pressure level at the absolute threshold of a young, trained listener with no history of ear trouble. Sound is presented by earphone to a listener in a quiet room. SPL is measured in the ear canal close to the eardrum, under the earphone. The slight elevation (greater SPL) of this threshold over the MAF is caused by the elimination of the effects of acoustic diffraction around the head and by the physiological noise generated in the ear cavity when it is enclosed by an earphone.

3. Normal Threshold of Audibility is the modal value of the minimum sound-pressure level at the entrance to the ear canal that can be heard by a large sample of young (18 to 30 yr.) and unpracticed listeners wearing earphones in a quiet room. This was the accepted American standard for the calibration of audiometers. Because young, *practiced* listeners can hear much better than at this threshold, there was a strong movement to accept the British Normal Threshold, which is much closer to the MAP. Most persons in the United States have adopted the International Standards Organization's standard of 1964 (ISO–1964), one very close to the British Normal Threshold, although a few still use the American Standards Association standard of 1951 (ASA–1951); however, they must identify the standard used.

Effect of Age on Hearing

Hearing sensitivity at high frequencies tends to decrease with age, particularly for men. Figure 4–5 (American Standards Association, 1954) shows the average hearing loss for pure tones at different frequencies for men and women of different ages. Individual people might vary markedly from these curves.

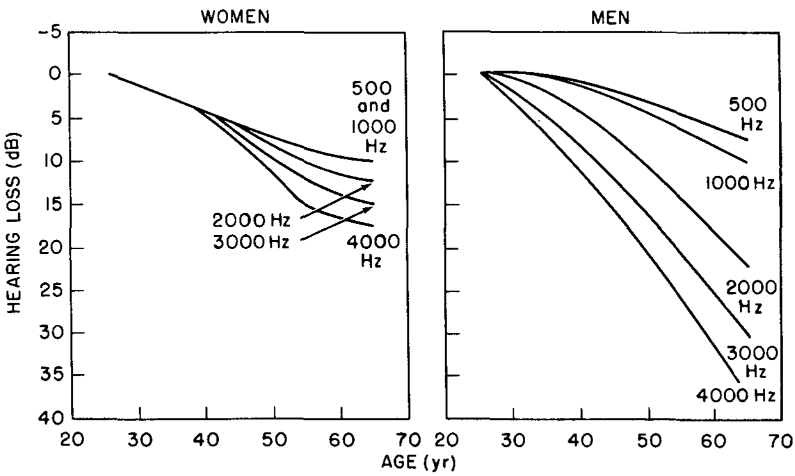


FIGURE 4–5. Hearing loss for females and males as a function of aging. (ASA, 1954)

Binaural vs. Monaural Thresholds

The binaural threshold for ears of equal sensitivity is about 2 to 3 dB below the monaural threshold. If ears of unequal sensitivity differ by more than about 6 dB in their monaural thresholds, the binaural threshold is essentially that of the better ear. On the average, the better ear differs from the mean of the two ears by about 2 dB at frequencies below 1000 Hz and up to 6 dB at 10,000 Hz (Fletcher, 1953).

4.2.5 Loudness and Pitch

Loudness is the magnitude of the auditory sensation elicited by a sound stimulus. It is a subjective response to a sound and depends primarily on the changes in, or the relative magnitudes of, sound pressure and, secondarily, on the frequency, duration, and spectrum of the sound. Although a subjective response, loudness can be quantitatively evaluated.

The loudness-level scale. Two sounds can be equated by alternately listening to one sound and then another and adjusting levels until the two seem equally loud. This technique of loudness balancing has been used to construct a loudness-level scale for quantitative evaluation. The unit of this scale is the phon. The loudness level of a sound in phons is defined as the sound-pressure level of a 1000-Hz pure tone that sounds equally loud. Equating the loudness of pure tones of various frequencies and levels yields the equal-loudness contours shown in Figure 4–6.

The Loudness Scale

These contours provide data for equally loud tones, but not for different magnitudes when loudnesses are unequal. For this, a loudness scale is used, the unit of which is the sone. Unit loudness on this scale is the loudness of a 1000-Hz tone 40 dB above absolute threshold. Sounds are then scaled in terms of their loudnesses judged as multiples or fractions of this reference sound. Thus, a sound ten times as loud as the reference sound has a loudness of 10 sones. Although loudness judgments are difficult to make, the relation between loudness (in sones) and loudness level (in phons) can be shown by the line in Figure 4–7. This shows that loudness doubles for each 10-phon increase in loudness level (Stevens, 1955).

Calculation of Loudness

The loudness of sounds can be calculated from measures of sound pressure level. Three measures of loudness deserve consideration: Speech Interference Level (SIL), Loudness Level (LL), and Perceived Noise Level (PNdB).

1. SIL is measured in decibels, dB, and is the average of the sound pressure levels in each of the octave frequency bands, 600 to 1200 Hz 1200 to 2400 Hz, and 2400 to 4800 Hz. The result yields a prediction of the interfering effect of the noise on speech.

2. LL is defined as the SPL of a 1000-Hz tone that sounds equal in loudness to the sound

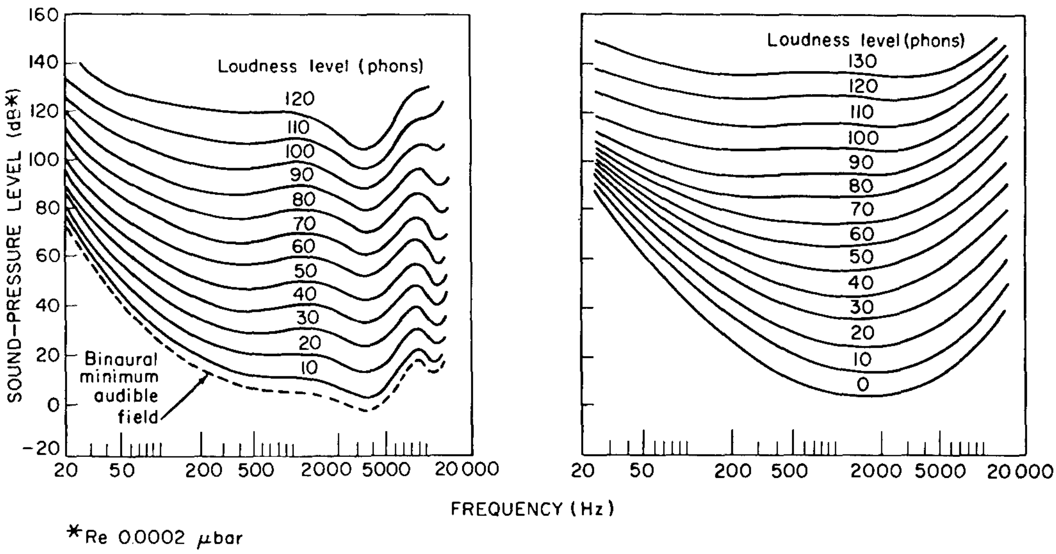


FIGURE 4–6. Equal-loudness contours for pure tones. For left position data, listener is in free field facing source; sound pressure is measured at position of listener's head center (Robinson and Dadson, 1957). For right position data, listener is wearing earphones; sound pressure is measured under earphones. (Fletcher and Munson, 1933.)

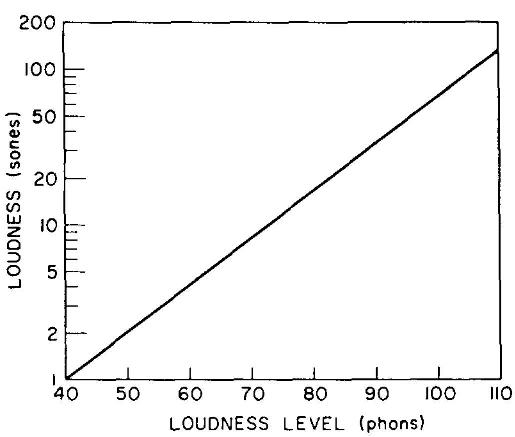


FIGURE 4-7. The relation between loudness (sones) and loudness level (phons) (Stevens 1955).

or noise being rated. LL may be calculated. (See Beranek, 1960.)

3. Perceived-noise level is measured in units symbolized as PNdB, and is defined as the SPL of a band of noise from 910 to 1,090 Hz that sounds as noisy as the sound or noise of comparison. In a manner similar to the plotting of equal-loudness contours in Figure 4–6, equal-annoyance contours may be plotted. (See Beranek, 1960, for the method for computing PNdB.)

Calculating loudness of complex sounds is difficult because one component may mask another. If we add two sounds whose loudness in sones is known, their loudness is additive when they are far apart in frequency, but when they are close, one sound reduces the loudness of the other. The frequency analysis performed by the ear is like that of a band-pass filter. Such a filter, with its limits set to include the frequency of the signal, rejects noise outside of these limits, and increases the signal-to-noise (S/N) ratio, thereby making the signal more audible. The ear is capable of doing a similar thing. The width of its "filter," called the critical band, varies from about 50 to over 200 Hz (Fletcher, 1953). At a frequency of 800 Hz, the critical band is approximately 50 Hz wide. It should be noted, however, that the ear can distinguish the difference in pitch of two different frequencies which are both within a critical band. (See Scharf, 1966.)

Calculating a loudness index by using critical bands is complex; however, a simpler method yields a close approximation (Beranek, 1960). Still another roughly approximate method is simply to read the level on the A scale of a sound level meter in place of calculating PNdB. The latter method is obviously recommended only where rank ordering or similar approximations are adequate but is becoming more popular as appropriate correction factors are determined.

Pitch

Pitch, like loudness, is a subjective attribute of sound, determined primarily by frequency, but also affected by loudness, spectrum, etc. A scale for the quantitative rating of the magnitude of pitch in a manner similar to that described above for loudness has been established. The unit of this scale is the mel, which is defined as the pitch of a 1,000 Hz pure tone at a level 40 dB above absolute threshold.

Figure 4–8 (Stevens and Volkmann, 1940) shows the dependence of pitch on frequency. The data in this figure can be used, for example, to select the frequencies of a set of pure tones so that the intervals between them are perceived as equal in pitch.

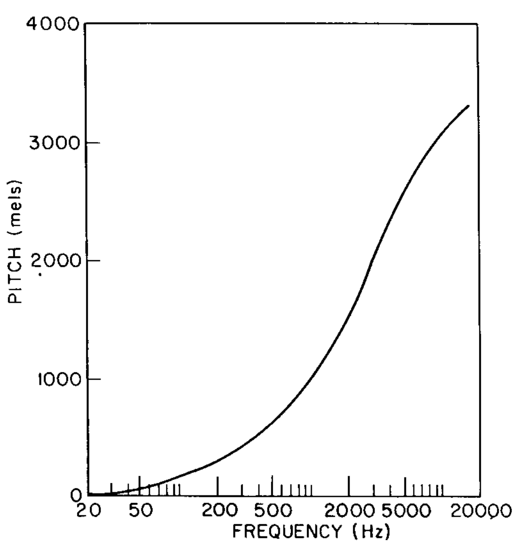


FIGURE 4–8. Dependence of pitch on frequency (Stevens and Volkmann, 1940).

4.3 Signal Processing and Control

In designing an auditory presentation, consider the following:

1. Signals and unwanted noises that might be picked up by the system.
2. Choice of ways in which the system can process signals and noise.
3. Method by which the processed information can be presented to the listener.

4.3.1 Signal and Noise Relationships

Noise is usually the major limiting factor in signal processing because few environments are free of it. The designer should look for ways to separate signal and noise. In attempting to do this, he will find the two are related as follows:

1. Sometimes noise is fixed by the environment, but the designer can select the signal to be used. For example, in sending radio-range signals to a pilot, the noise of the aircraft can be considered to be relatively fixed; hence the designer can choose the frequency which will be minimally masked by the noise.
2. In other instances, both signal and background noise are determined by the listening environment and not by the designer. For example, in passive sonar, the exact nature of sounds emitted by ships of unknown identity is indeterminate. Here, data on the background noise and the probable nature of the signals must be used to determine how the signal should be processed. Often, adjustable controls should be provided so the operator can select a spectrum combination different from both signal and noise most favorable for detection.
3. In still other cases, the signal is fixed but the noise can be controlled either by reducing its overall level or by appropriate networks that reduce the noise level in those regions of the spectrum containing the signal. For example, muff-type earphones can be used to prevent environmental noise from masking signals heard over a headset. Electrical filters can be employed to reduce the interference of static and radio noise with Morse code signals. Peak clippers can be used to reduce the high peak amplitudes of impulse noise such as that from automotive ignition systems.

4.3.2 The Masking of Sound by Noise

When noise is mixed with a signal, the threshold for hearing that signal is raised. This phenomenon is called masking, and the elevated threshold is known as the masked threshold.

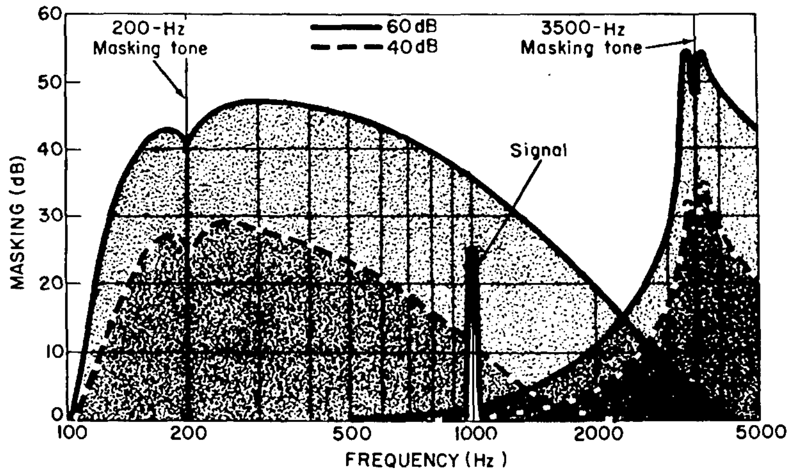


FIGURE 4–9. Masking produced by pure tones.

The term, *threshold* is used here in a general sense.

(There is good reason to question the meaning of the term *threshold* as a useful way of describing responses to masked signals. Indeed, even the absolute threshold probably should be thought of as a *masked* threshold in which the masking noise is noise internal to the receiver. A full discussion of this point is in Section 4.7., The Theory of Signal Detectability.)

Monaural Masking

The masking effects of a pure tone, or of a noise with a strong component in one part of the spectrum, are different from those of a narrow- or wideband noise. Thus, masking by each of these will be discussed separately in the paragraphs that follow.

Masking by pure tones. The masking effect of a pure tone, or of a noise with a strong tonal component, is greatest near the particular tonal frequency but also extends to signals on both sides of the masking tone. (See Figure 4–9.) Masking produced by a 200-Hz pure tone at 60 dB and 40 dB is shown by solid and dashed curves at left. Signal, shown in middle, is masked by a 60-dB tone but is audible above a 40-dB tone. Solid and dashed curves at right are for a 3500-Hz pure tone at 60 and 40 dB. Because masking does not tend to spread downward in frequency, a 1000-Hz signal remains unmasked by a 3500-Hz tone. Likewise, the masked threshold of a signal is raised more by frequencies in its vicinity and is raised more by frequencies below than by those above that of the signal. At relatively high intensities, however, the masked threshold of signals with some integral multiple of the masking tone is raised more than those having no harmonic relationships to the masking tone.

Curves of masking vs. frequency for masking by pure tones of various frequencies and levels are shown in Figure 4–10. The presence of audible beats at frequencies very near the masking tone and its harmonics can increase the audibility of a tonal signal and, therefore, sharply reduce effective masking at these frequencies. The "dip" in the curve disappears for signals too brief to set up beats. Note that masking tends to be confined to the vicinity of the masking tone for tones of low level (20 to 40 dB), whereas at high levels considerable masking occurs, extending particularly to frequencies above the masking tone.

Masking by narrowband noise. Effective masking vs. frequency range curves, for masking by a narrow band of noise, are shown in Figure 4–11 (Egan and Hake, 1950). These are somewhat similar to the curves for masking by pure tones except that the sharp dips caused by beats are absent.

Masking by wideband noise. The masking effects of a wideband noise extend over and somewhat beyond the entire spectrum of the masking noise.

Figure 4–12 (Hawkins and Stevens, 1950) shows masked thresholds of pure tones when

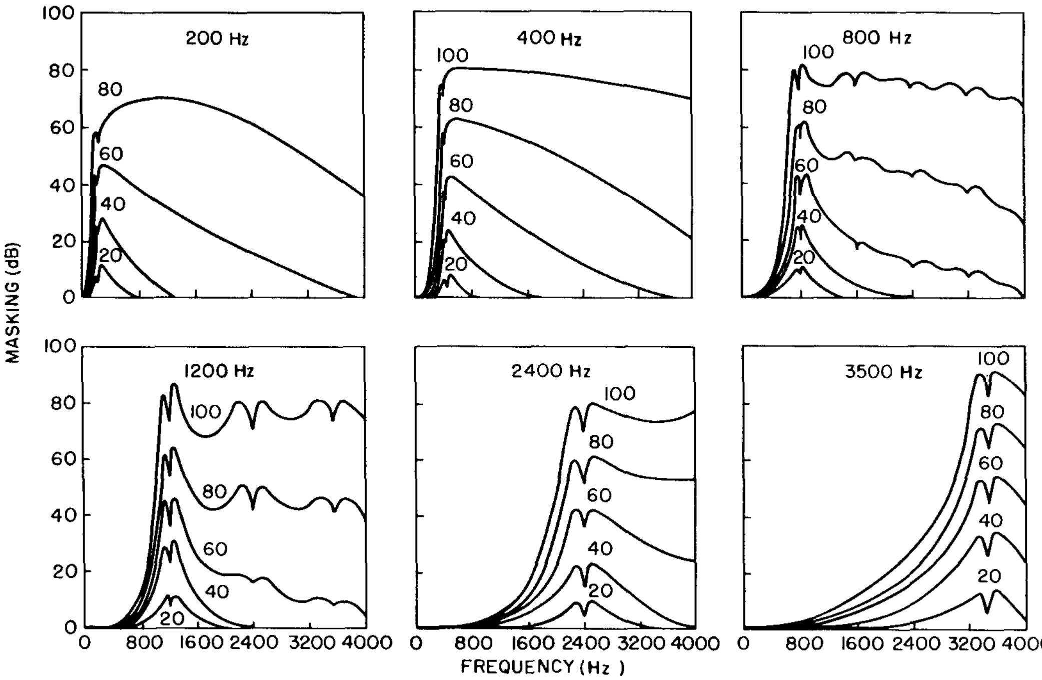


FIGURE 4–10. Masking as a function of frequency for masking by pure tones of various frequencies and levels. Number at top of each graph is frequency of masking tone. Number on each curve is level above threshold of masking tone. (Wegel and Lane, 1924.)

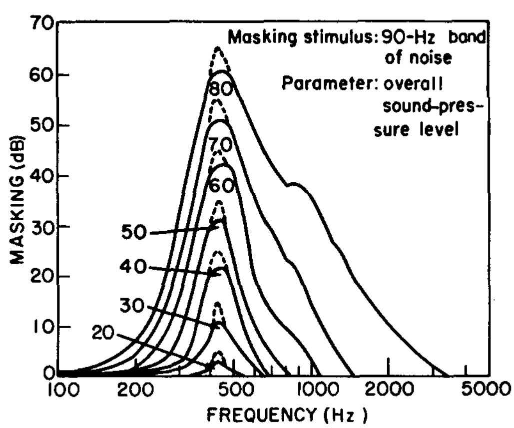


FIGURE 4–11. Curves of masking vs. frequency for masking by a narrow band of noise (Egan and Hake, 1950).

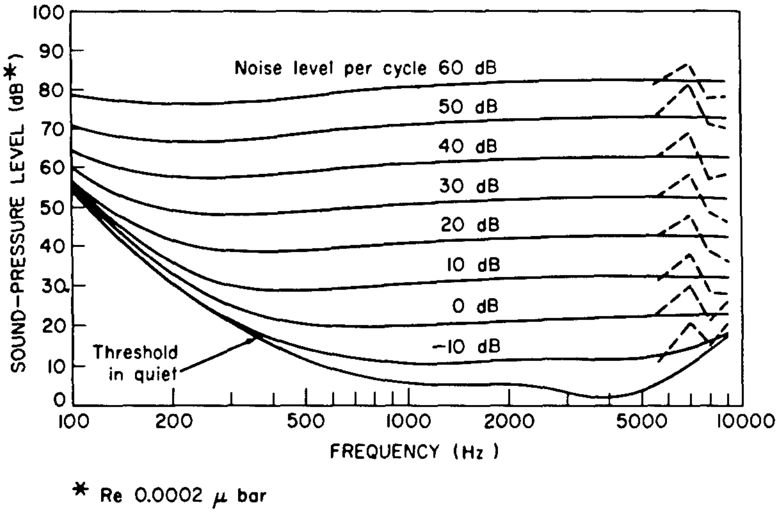


FIGURE 4–12. Masked thresholds of pure tones when masked by a wide-band noise of uniform spectrum (Hawkins and Stevens, 1950).

masked by a wideband noise of uniform spectrum (white noise). The linear increase of masking with the sound pressure level of the masking noise is apparent from the regular spacing of the contours.

Predicting the masked threshold. The masked threshold of a pure tone masked by wideband noise can be predicted if the spectrum level of the noise is known at the frequency of the tone. In such a prediction, it may be assumed that the masking is caused by noise components, the frequencies of which lie in a band near that of the tone and, in fact, lie within the same critical band as the tone. When used to predict masking, the critical bandwidth is so defined that the SPL of the noise in the critical band is just equal to the SPL of the tone at its masked threshold. Figure 4–13 shows the generally accepted values of critical bandwidth as a function of frequency.

To predict a masked threshold at a frequency f by this method, proceed as follows:

1. Measure the spectrum level of the wideband masking noise at f.
2. Correct this measured level to the level in the critical band at f by adding the $\log_{10}$ of the critical bandwidth. This correction can be read directly from the left-hand ordinate of Figure 4–13.
3. This corrected value is the masked threshold at f if the value is more than 20 dB above the absolute threshold at f. If it is less than 20 dB, a correction must be made for non-linearity in the masking vs. noise-level function near the absolute threshold.

Correcting for nonlinearity. Masking at any particular frequency is a linear function of the level of noise in the critical band at the frequency, except for the "toe" of the function which, at very low noise levels, flattens out and becomes asymptotic to zero. The level of noise in a critical band above the absolute threshold of a pure tone at the center frequency of this band is known as the effective level (Z) of the noise at the center frequency.

If masking (M) is plotted as a function of Z for a number of different frequencies, the plotted points form the curve shown in Figure 4–14. Note that, when M is greater than about 20 dB, $M = Z$, and that, when M is less than 20 dB, $M > Z$. The magnitude of the correction that must be added to the level of the noise in a critical band to predict the masked threshold is the value of the difference $M - Z$, which can be obtained from Figure 4–14.

Interaural Masking

The statements made above apply to monaural masking—when signal and noise together reach the ear or ears from the same source. In interaural masking, when the signal is fed into one ear and the noise into the other, different rules apply.

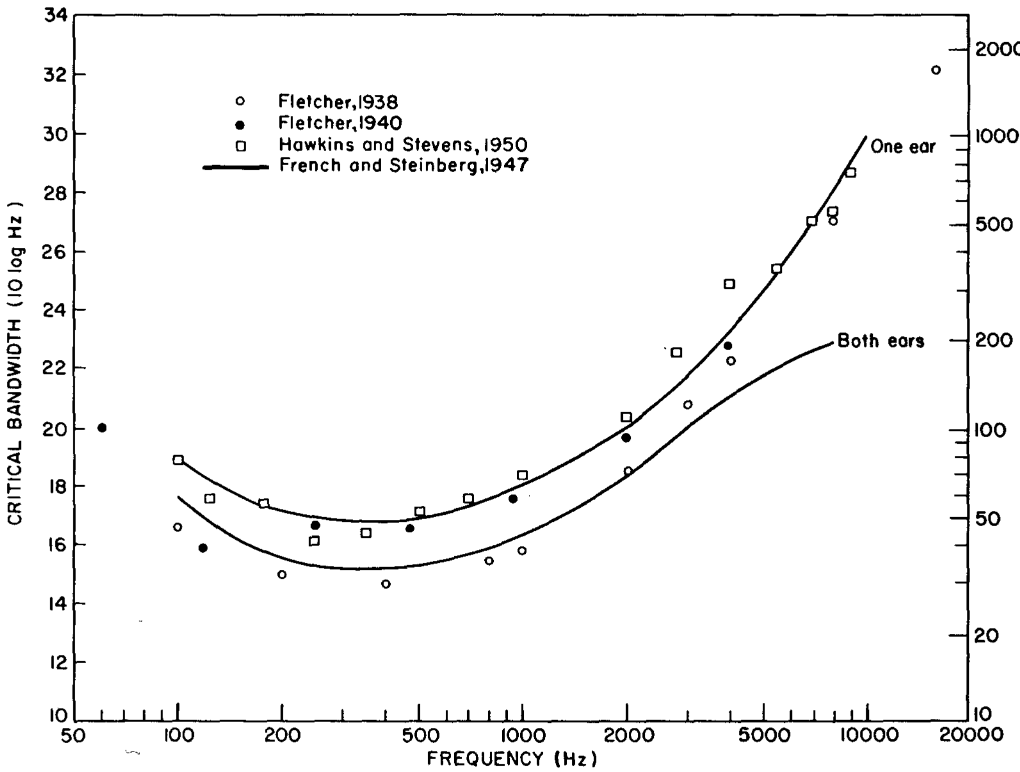


FIGURE 4–13. Generally accepted values of critical bandwidth as a function of frequency.

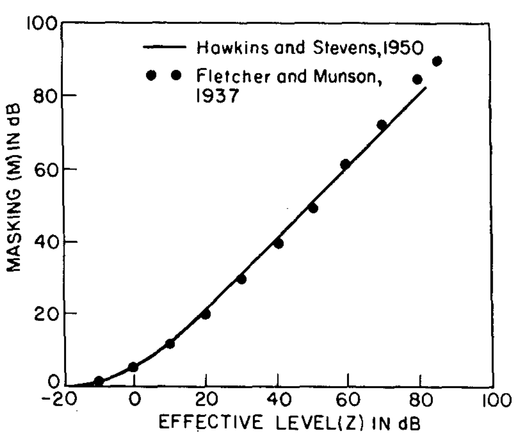


FIGURE 4–14. Masking (M) plotted as a function of the effective level (Z) for a number of different frequencies.

Little masking of the signal occurs when the SPL of the noise is relatively low, below 40 or 50 dB, because the listener is able to distinguish sounds heard separately by the two ears. Masking occurs only when the SPL of the noise is sufficiently high for the sound to be conducted through the bone of the skull to the opposite ear (Fletcher, 1953), which might be at SPL differences of 50 dB or higher. Such masking, however, becomes a case of monaural masking with the head serving as an attenuator. Thus, interaural masking becomes a problem only when earphones are used, and then only when the sound in one greatly exceeds that in the other. The effective level of the masking noise in one ear must be about 50 dB or more above the signal in the other ear to produce interaural masking.

Binaural Masking

Man also uses his two ears to obtain a release from masking. Certainly the techniques of stereophonic listening, increasingly familiar to the general public, are of value in obtaining release from masking, in permitting increases in the number of voice channels which may be operative simultaneously, and in compensating for degradation of detectability in those cases where the signal is of uncertain frequency.

As Jeffress et al. (1956) have pointed out, all binaural stimulus conditions produce "summation," and present the possibility, therefore, for improving the detectability of signals.

Because a change in signal level of as little as 3 dB can change the percentage of correct judgments from just above chance to very near 100%, an effective change of as little as 1 dB may be valuable. Signals in a noisy channel fed to one ear, with the other ear receiving a band of noise 100 cycles wide which has no relation to the noise or signal in the first ear, will yield about 1 dB of release from masking. (See Dolan, 1965.)

The Masking Properties of Common Noises

Earlier in this section, in Predicting Masked Threshold, a method was presented for calculating the masking properties of noises whose spectrum level was known. Chapter 5 presents means for computing or estimating the amount of interference to speech for various noises, or the interfering qualities of speech on speech. (See Section 5.4.)

If a noise has a spectrum level below 80-dB SPL at the frequency of a certain signal, the calculation is uncomplicated. High-frequency bands of noise with spectrum levels above 80-dB SPL mask lower frequencies, and may affect speech perception or the detection of tonal signals. The phenomena of *remote masking* and *temporary threshold shifts* are more complex. Bilger (1959) has shown how the different types of masking add together. Above 80-dB SPL, each 1 dB of added noise brings about 2 dB of masking for tones lying below the noise band.

There are differences in the masking properties and aftereffects between steady noises and impulse noises. Machines like typewriters and computer printers emit both kinds of noise, and the printers, at least, are noisy enough to interfere with the perception of speech. Ear defenders which usually filter out the high-frequency components of noise improve speech perception markedly under such conditions.

Impulse noises like rifle or mortar fire result in additional effects. When these noises are near, they are intense enough to induce some temporary threshold shift (Ward et al., 1958; Ward et al., 1961; and Ward, 1962). It is commonly assumed that temporary shifts of threshold caused by exposure to noise will ultimately produce permanent shifts. Temporary threshold shifts are not uniform across the audible range, with high frequencies being more affected than low, and the vicinity of 4 kHz being affected more readily.

4.4 The Reduction of Masking

Two types of masking can be reduced by filtering: (a) the masking produced by components within a critical bandwidth centered at the signal frequency (called direct or centered masking) and (b) the masking effect of pure tones or tonal noises on signals lying outside the critical band (called remote or displaced masking). These are discussed in the following paragraphs.

4.4.1 The Reduction of Direct Masking

Very narrow band-pass filters that reject noise within the critical band can be used to reduce the masking of a wideband noise on a tone that lies at a frequency within the noise spectrum. For such a band-pass filter to be of value, however, it must be narrower than the critical band. Otherwise it only rejects noise that has no masking effect. Up to a certain point, the narrower the filter, the greater the reduction of masking noise for the particular tonal signal (Schafer et al., 1950).

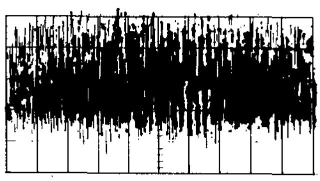
On the other hand, the band-pass filter should not be made too narrow because the noise passing through a very narrow filter takes on a tonal quality that is difficult to distinguish from the signal itself. Furthermore, if the pass band is narrower than the reciprocal of the signal duration, it begins to reduce signal strength. But, because the band-pass filter must be quite narrow to effect any improvement in audibility, it can be used only with signals that are very stable in frequency.

4.4.2 The Reduction of Remote Masking

Filters that reject noise in the spectrum outside the signal and its critical band can be used to advantage whenever the masking noise is of such high amplitude that it causes remote masking at the signal frequency. The removal of such noise by a filter reduces the remote masking and increases signal detectability. Because remote masking is more effective when the masking noise has a lower frequency than the signal, the greatest benefit of a filter is realized when it is used to reject noises at frequencies below that of the signal.

To reduce remote masking, a filter should be designed to reject the components of noise causing the remote masking without appreciably reducing the amplitude of the signal. Thus, in Figure 4–15, a 500-Hz high-pass filter could remove the 60 dB, 200-Hz noise component, with its remote masking, so that the 1000-Hz signal would become audible. Figure 4–15 illustrates a similar effect which would be achieved by a band-pass filter centered on the frequency of the signal.

UNFILTERED



FILTERED

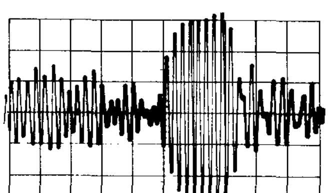


FIGURE 4–15. Oscillograms of signal consisting of 10-msec. pulse of 800-Hz tone buried in a wideband white noise. Effective filter action makes signal clearly visible on oscilloscope after filtering signal and noise by 200-Hz band-pass filter.

4.5 The Reduction of Loudness

Under certain circumstances, filters can be used to improve signal detection by lowering the overall noise level within a system, permitting the operator to increase the gain so that the signal is at the optimum level for detection. The circumstances are as follows:

1. When the noise is greatest in bands not containing the signal, filters can be used to reduce it separately from the signal.
2. When the noise is distributed over a wide band of frequencies, and the signal can occur at any frequency in the band, an equalizer can be used to shape the noise spectrum so that its masking is about the same at all frequencies.

Wideband noises that are loud enough to raise the masked threshold more than 20 dB at all frequencies are likely to be uncomfortably loud. The resulting discomfort causes the operator to turn the gain down, thereby reducing the signal level to an intensity less than 20 dB above the absolute threshold. This is particularly probable when the signal is relatively low in frequency, and the noise contains much high-frequency energy.

Thus, the overall noise level can be reduced by inserting filters or equalizers in the signal processing system. By reducing the noise level, such filtering or equalizing networks permit the operator to increase his gain without making the noise uncomfortably loud. The increased gain brings the signal level into the optimum-detection zone of 20 to 80 dB above absolute threshold. As a result, his ability to detect the signal is improved, as illustrated in Figure 4–16. The spectrum at the top shows signal in the presence of wideband, high-frequency noise, which, though not masking the signal, is so loud (unfiltered) that the operator has reduced his gain control to the point where the signal is below absolute threshold and, hence, inaudible. In the center case, a 2000-Hz low-pass filter has allowed the operator to increase his gain control to a point where noise is as loud as previously (equal gray area), but signal is now raised above absolute threshold and becomes audible. The lower spectrum has a 500-Hz low-pass filter to further increase gain and signal audibility.

4.5.1 Signal Level Control

Any signal used in an auditory display must exceed the absolute auditory threshold to be heard, even under the ideal circumstance of an attentive listener and quiet surroundings. Requirements are actually more stringent because the listener must make critical discriminations

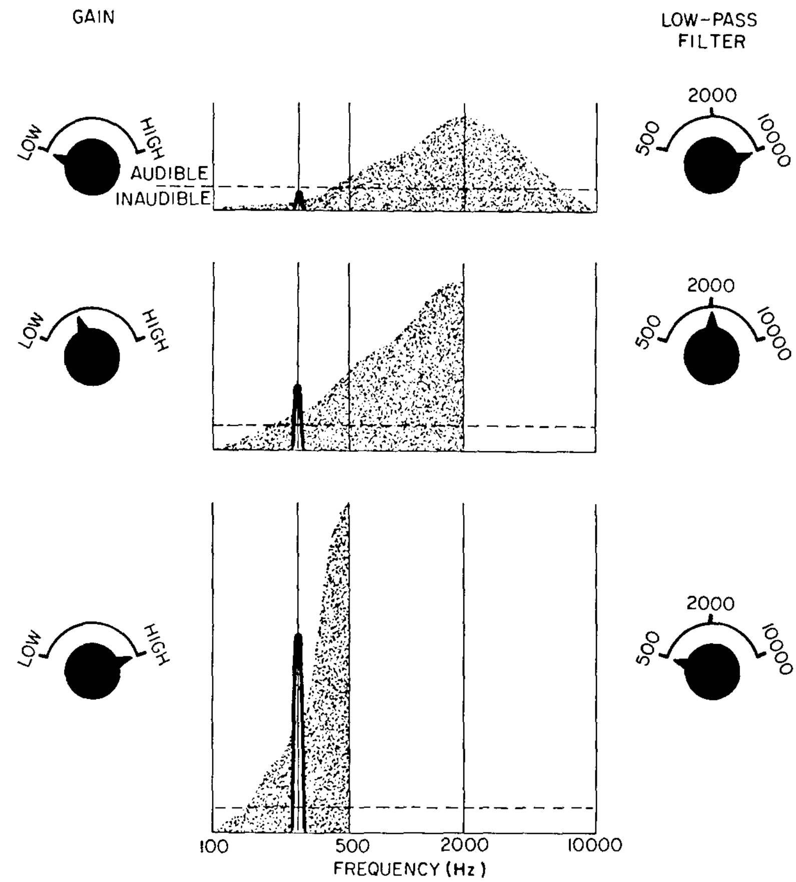


FIGURE 4–16. Use of filters to reduce loudness of background noise.

between characteristics of sounds above the masking level of any noise environment.

4.5.2 Minimum Signal Level

Ordinarily it is necessary to present sounds at levels well above the absolute threshold. Discrimination of small changes in signal intensity and pitch is performed best at levels more than 60 dB above the absolute threshold. In the presence of noise, the signal must exceed its masked threshold, preferably by at least 15 dB for good discrimination.

4.5.3 Maximum Signal Level

High SPL's impair hearing or create discomfort or pain. Individuals differ in levels they will tolerate, but enough data are available to establish the following working rules:

1. To avoid feelings of discomfort, do not expose individuals to SPL's above 120 dB. These cannot be tolerated for long periods of time.
2. For most people, levels above 135 dB cannot be tolerated for even brief periods.
3. Exposures to intense sounds should be as brief as possible. Damage-risk criteria for different exposure times are shown in Figure 4–17. The human ear can stand extremely intense sounds for a few seconds (e.g., 130 dB for 10 sec.) without lasting effects; prolonged exposure to intensities of approximately 90 dB can cause hearing loss. Figure 4–18 shows damage-risk criteria for eight hours of continuous exposure for different age groups and different frequencies. Figure 4–19 shows differences between damage-risk criteria and thresholds for audibility levels.
4. Use of high-frequency sounds should be avoided. With low-frequency sounds (500 Hz and below), there is less risk of hearing damage.

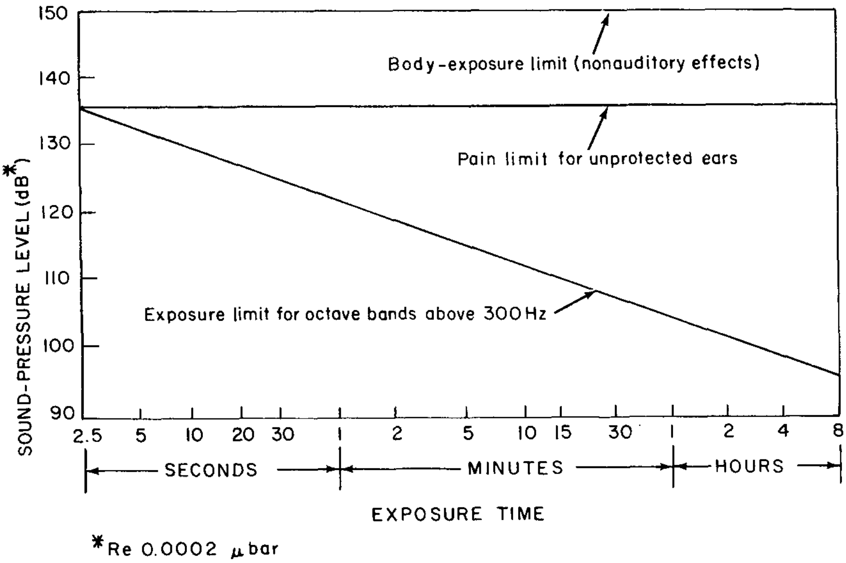


FIGURE 4–17. Damage risk criteria for various exposure times up to 8 hr. Pain limit for unprotected ears is shown at 135 dB. When ear protectors are used, sound pressure level in sound field can exceed these criteria by amount of attenuation provided by protectors. Body-exposure limit at 150 dB is point at which potentially dangerous non-auditory effects occur. This level should not be exceeded in any case (Eldred et al., 1955).

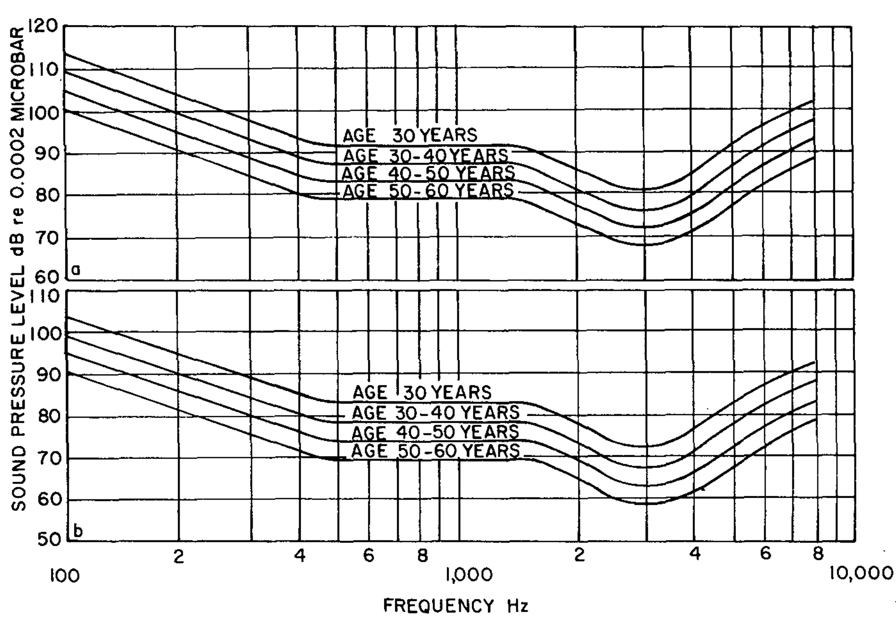


FIGURE 4–18. Damage-risk criteria for (a) wide band noise measured by octave, 8 hr. continuous exposure; and (b) pure tones or critical bands of noise. The parameter is age From *Noise Reduction* by L. L. Beranek. Copyright 1960 by McGraw-Hill Book Company. Used with permission of McGraw-Hill Book Company.

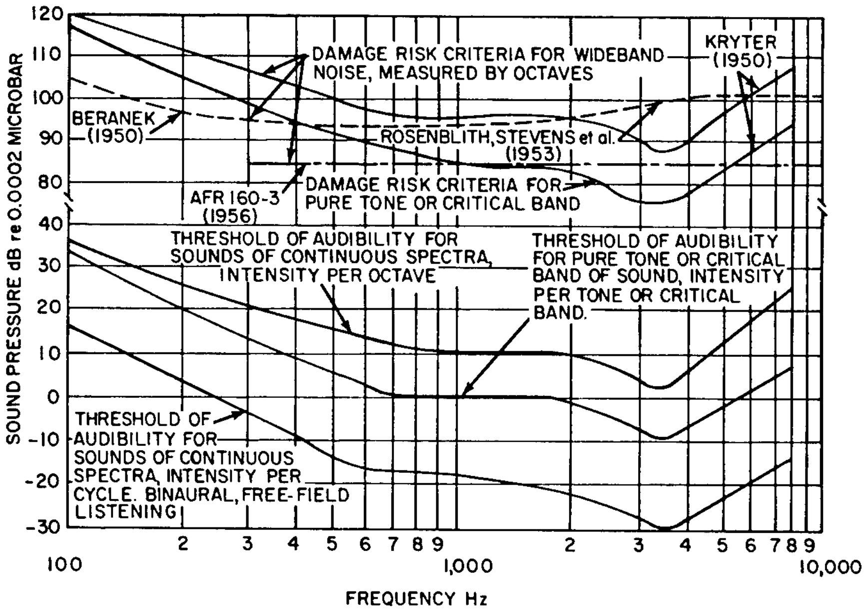


FIGURE 4–19. Showing various thresholds of audibility curves and damage-risk criteria. The damage-risk criteria of Beranek, Rosenblith and Stevens are the same except at the higher frequencies. From *Noise Reduction* by L. L. Beranek. Copyright 1960 by McGraw-Hill Book Company. Used with permission of McGraw-Hill Book Company.

4.5.4 Optimum Signal Level

A typical auditory task is to detect small changes in intensity or frequency. This can be done best at comfortable, but fairly loud, listening levels. The optimum level, however, depends on whether the individual is listening in quiet or in noise. For listening in quiet:

1. If the task is to detect changes in intensity, set the signal level 60 dB or more above the absolute threshold (Riesz, 1928). Because threshold varies with frequency, this will result in a higher SPL at low and high frequencies than at medium frequencies between 1000 and 4000 Hz.

2. If the task is to detect changes in frequency, set the signal at least 30 dB above absolute threshold (Shower and Biddulph, 1931).

3. For the most comfortable listening level, set the signal about 40 to 50 dB above the absolute threshold (Pollack, 1952). Comfortable listening, as with ability to detect intensity or frequency changes, depends on the region excited within the frequency spectrum.

4. For listening in noise, a convenient rule of thumb for specifying the optimum signal level at the entrance to the ear canal is to select one midway between the masked threshold and 110 dB.

4.5.5 Dynamic Range and Volume Control

Signal level at the ear can vary between the absolute or masked threshold as a lower limit and the threshold of discomfort or pain as the upper limit. The "useful dynamic range" of acceptable signal levels depends somewhat on the frequency of the signal and on the background noise spectrum. (See Figures 4–20 and 4–21.) For detection and effective identification, auditory signals must be kept within it. This can be accomplished as follows:

1. A signal can be controlled at its source by using suitable metering equipment. In radio broadcasting, an announcer regulates his voice signal by watching a VU meter, which was designed to have temporal response characteristics assumed for the human ear, and indicates the signal level in decibels. By controlling his voice as he talks, so the needle regularly hits a constant value, the announcer can keep his voice

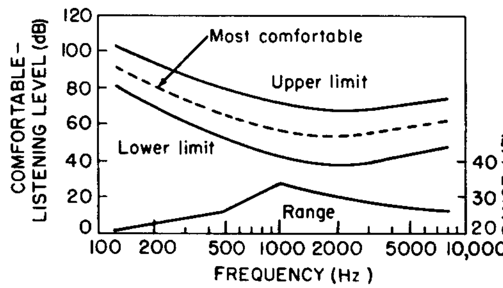


FIGURE 4-20. Comfortable listening levels for pure tones. Dashed curve shows mean level judged most comfortable by 33 observers while listening binaurally to pure tone. Upper solid curves show maximum and minimum levels that were still considered comfortable by these listeners. Lower solid curve, reading right-hand ordinate, shows range between upper and lower limits for comfort (Pollack, 1952).

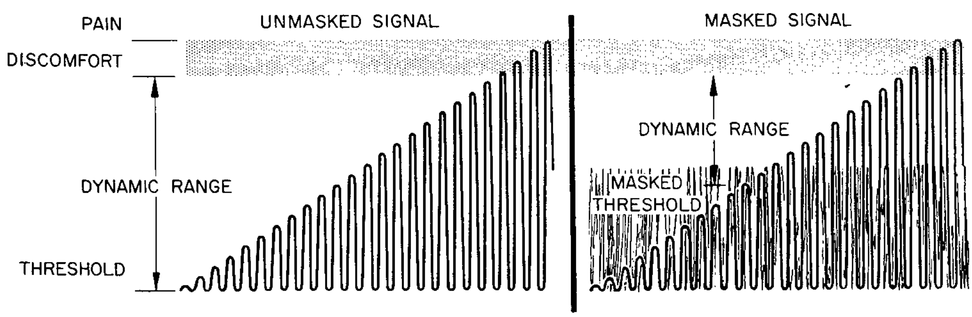


FIGURE 4-21. Effect of masking noise in reducing dynamic range.

signal within the dynamic range of his listeners (and also that of the radio carrier wave).

2. The signal level can be set by the listener. If the signal is reasonably constant in level, the listener can set a volume control to a desired listening level. Such controls can be continuous, as they are in phonograph, radio, and television sets, or they can be stepped, as they frequently are in electrical attenuators. Incremental changes in level sounds are virtually continuous when successive steps are 1 dB or less; 2 dB steps are slightly noticeable, but steps greater than 5 dB are used only in special circumstances. This manual method of setting the volume control is satisfactory when the available dynamic range is fairly large (40 dB or more), in which case deviations of 10 dB or more from the optimum listening level can be tolerated.

3. Signal level can be brought within the dynamic range by a circuit known as a compressor (see Figure 4–22), a device particularly suitable for cases in which the signal unavoidably varies in strength more than the acceptable dynamic range. The compressor reduces the amplitude of strong signals while leaving weaker signals nearly unaltered, and thus reduces the range of signal variation. When the average level of the signal varies widely, it might be necessary to supplement a compressor with either an AVC (see below) or a manual volume control. Because a well-designed compressor does not distort the signal's waveform, it is used extensively in recording and broadcasting. To avoid distortion, the attack and release times of the compressor action must be compatible with the dynamic properties of the source material. For speech and music, a fast attack time of about 10 msec. combined with a slow release time of the order of 1 sec. has been found to be satisfactory (Hathaway, 1950). The rapid attack time enables the circuit adjust immediately to the sudden onset of a new signal, and the slow release time tends to stablize the gain during reception of natural sounds that are inherently intermittent, such as speech and music.

4. The signal level can be set by an automatic volume control (AVC) circuit, also called automatic gain control (AGC), in the signal processing system. (See Figure 4–22.) Some sort of automatic control becomes critical when high ambient noise levels reduce the usable dynamic range to 20 or 30 dB or less. Most radio receivers have AVC circuits that maintain a relatively constant volume by controlling the amplifica-

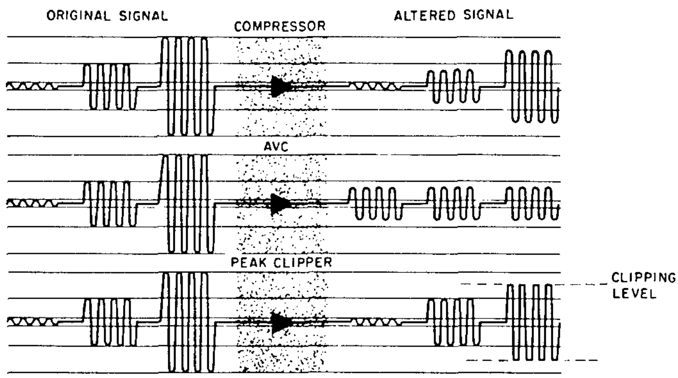


FIGURE 4–22. Signal level brought within the dynamic range by means of a compressor.

tion of the carrier wave. (See Chapter 5 for a more detailed discussion of AGC.)

5. Sometimes, a peak clipper can replace a compressor to bring signals within the usable dynamic range. Peak clippers can be used with speech to keep instantaneous speech wave peaks from exceeding the threshold of discomfort. (See Chapter 5.) Provided the signals are clipped no more than 20 dB, word by word assessment of speech remains about as effective as unclipped speech because little intelligibility is contained in the clipped-off peaks. Peak clipping markedly distorts a signal, however. Its use should be limited to cases where distortion can be tolerated; for example, vocoder speech communication systems but not entertainment systems.

4.6 The Selection of Signals

4.6.1 Single Signals

In many auditory presentation systems the operator can select the kind of signal he uses. For example, a sonar operator using an active sonar mode can adjust the frequency or duration of his signal by adjusting controls at his disposal, or a Morse-code operator can adjust the beat-frequency oscillator on his receiver to select his listening frequency. The selection of signal parameters depends on the nature of the change the operator must detect. Changes either in frequency or in intensity (usually stated as sound pressure level) that are just perceptible (or just noticeably different) depend on the sound pressure level, frequency, duration, and method of presenting the differences.

Detecting frequency changes. Figure 4–23 (Shower and Biddulph, 1931) shows just-noticeable-differences (JND) in frequency for pure tones at various levels above threshold. Note that the smallest JND is about 2 or 3 Hz and is found at frequencies below 1000 Hz, whereas above 1000 Hz, the JND is about 0.3% of the frequency.

Small changes in pitch are heard best at comfortably loud levels (30 dB or more above threshold) and at durations in excess of 0.1 sec. The dependence on duration of the JND in frequency is shown in Figure 4–24; the JND is smaller for a change repeated cyclically than for a single occurrence. A rate of 2–3 repetitions per second is optimum.

If the operator is to detect a change in frequency, a compromise often must be made between the frequency at which changes are most easily detected and masking effects of ambient noise. The smallest changes in frequency can be detected at low frequencies, but ambient noise is usually greatest at low frequencies (Beranek, 1957). Thus, the best compromise is between 500 and 1000 Hz.

Detecting intensity changes. Figure 4–25 (Riesz, 1928; Miller, 1947a), shows JND in sound pressure level for pure tones of various frequencies and for wideband noise. Note that the smallest changes are detectable at the higher intensity levels (about 60 dB or more above threshold). If the changes in sound pressure level occur cyclically, the JND is smallest at a rate of 2 or

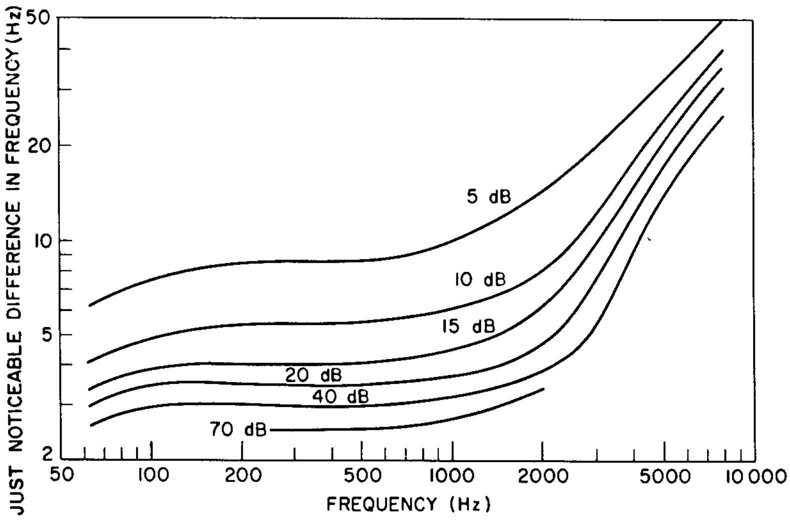


FIGURE 4–23. JND in frequency for pure tones at various levels above threshold (Shower and Biddulph, 1931).

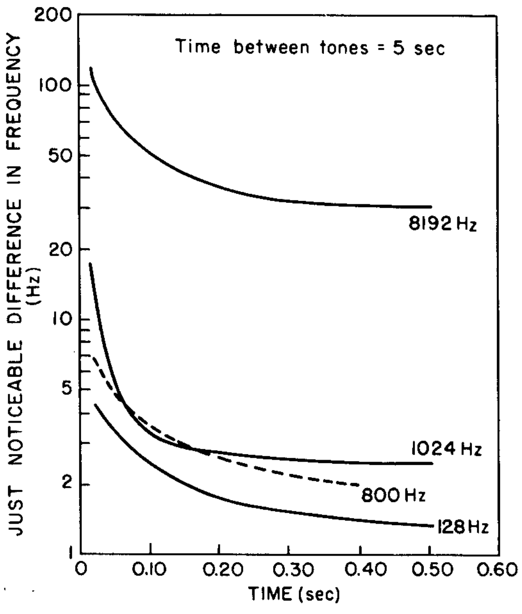


FIGURE 4–24. Dependence on duration of the JND in frequency.

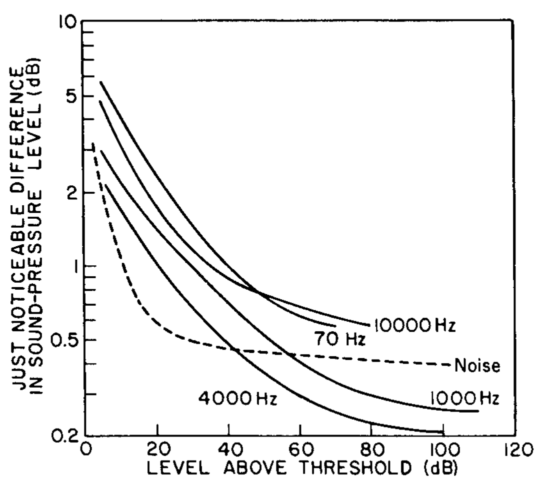


FIGURE 4–25. JND in sound-pressure level (SPL) for pure tones of various frequencies and for wide-band noise (Riesz, 1928, and Miller, 1947a).

3 repetitions per second. Also, Figure 4–26 shows that the JND in sound pressure levels for bands of noise is dependent on bandwidth as well as on frequency. Note that smaller differences can be perceived for wider bands.

Changes in the intensity of a pure tone are most easily detected at frequencies between 1000 and 4000 Hz. If the signal has a wide continuous spectrum, such as random noise, and the operator must detect changes in intensity, a listener can more easily detect changes in a wide rather than a narrow band of the noise. (See Figure 4–26.) Probably the listener hears more random fluctuation in the narrow band intensity than he does in that of the wide band; hence, he is less able to detect a small shift in the intensity of the narrow band. (See Figure 4–27.)

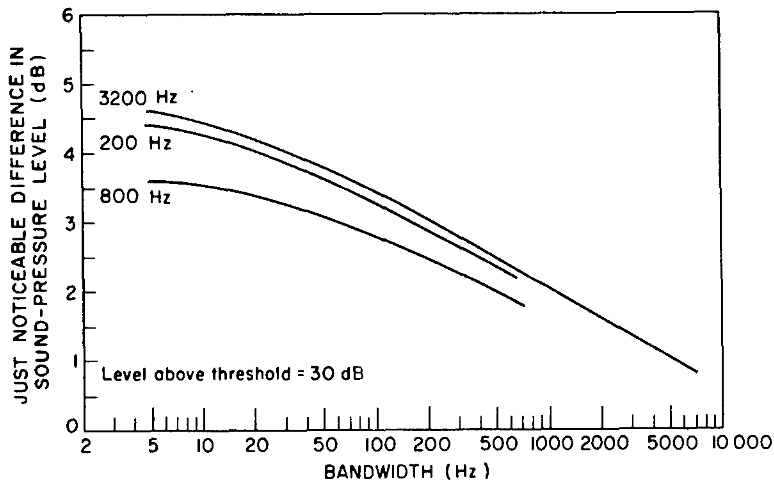


FIGURE 4–26. JND in SPL for bands of noise is dependent on bandwidth as well as frequency.

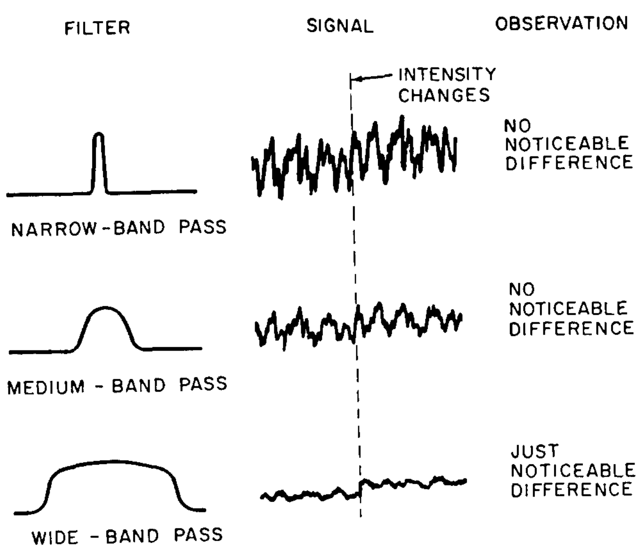


FIGURE 4–27. Selection of signals.

Selecting Tonal Duration

A sound's audibility depends on its duration because the ear's response is not instantaneous. Rather, in the case of pure tones, it takes 200 to 300 msec. to build up (Munson, 1947) and approximately 140 msec. to decay (Stevens and Davis, 1938), although wideband noises build up and decay somewhat faster. Consequently, sounds of less than 200 to 500 msec. in duration do not sound so loud and are not so audible in noisy backgrounds as are sounds of longer duration.

A slight gain in detectability accrues as the signal's duration exceeds 500 msec. The detection process seems to fluctuate because of variable properties of the background noise, or in the signal, or because of the listener's acuity. Detectability increases with signal duration, but any gain beyond a few seconds is generally quite small.

In the presence of masking noise, a pure tone's duration should be no shorter than 300 msec. For tonal signals briefer than 300 msec., the product of time and intensity required for threshold is constant, so signal intensity must be proportionately increased.

For short-duration sounds, the above rule can be generalized further, i.e., the time integral of the intensity is a constant, which provides a rule for determining an equivalent intensity and duration for sounds having nonrectangular envelope shapes. (See Figure 4–28.) The effect of tonal duration on the masked threshold for bursts of tones of various frequencies having rectangular envelopes is shown in Figure 4–29 (Garner and Miller, 1947).

Tones should last from 10 to 100 msec. if an operator must detect changes in pitch (Turnbull, 1944). Pitch sensitivity is slightly impaired for tones between 50- and 100-msec. and deteriorates rapidly as tonal duration is shortened below 50 msec. Tones shorter than 10 msec. have very little pitch (Doughty and Garner, 1947), because, as a tone pulse is shortened, its spectrum spreads out and becomes more and more like a noise. The rule is that the width of the resulting spectrum, at its half-power points, is equal to the reciprocal of the duration ($\Delta f = 1/T$), where Δf is the effective bandwidth and T is the duration of the tone pulse. (See Figure 4–30.)

4.6.2 Multiple Signals

In many auditory signaling systems, it is necessary for a listener to monitor several different channels, either simultaneously or at different times. In selecting signals for such systems, certain factors should be considered. These are discussed in the following paragraphs.

Using Multiple Frequencies

When a listener must monitor several channels simultaneously but attend to only one channel at a time, channels should be spaced at different frequencies. (See Figure 4–31.) In addition, channel frequencies should be spaced as widely as possible throughout the available spectrum to minimize signals being masked from one channel by signals from another and to make channel identification easier. The designer should avoid simple multiples or submultiples

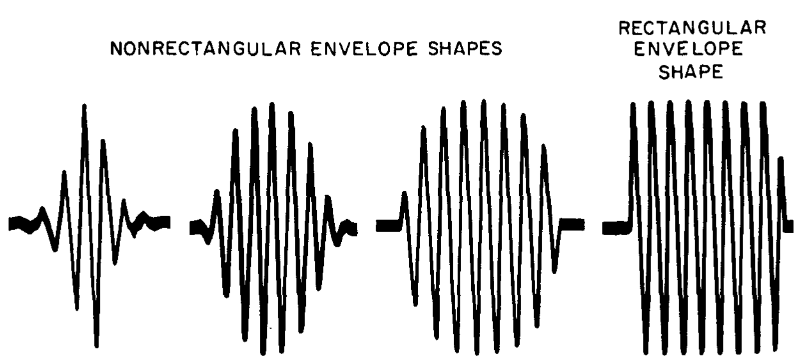


FIGURE 4–28. Determining an equivalent intensity and duration for sound having nonrectangualr envelope shapes.

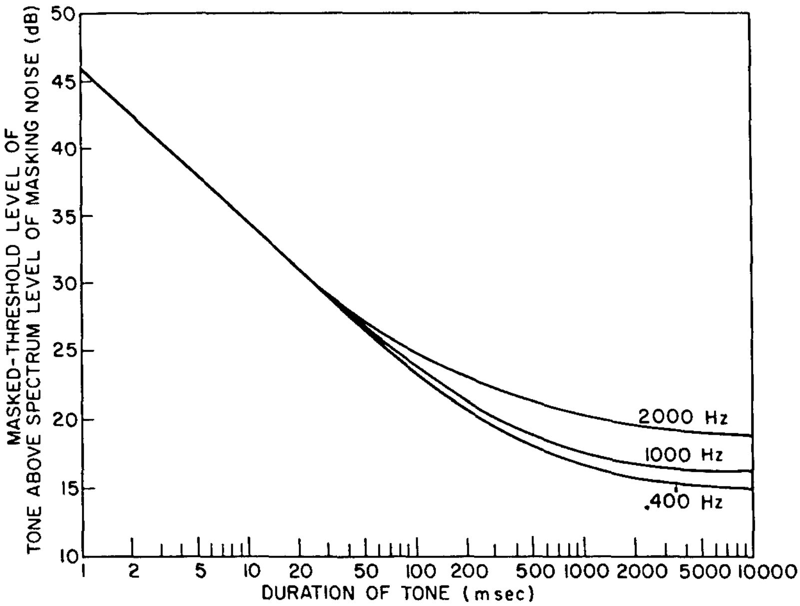


FIGURE 4–29. Effect of tonal duration on the masking threshold for bursts of tones of various frequencies having rectangualr envelopes (Garner and Miller, 1947).

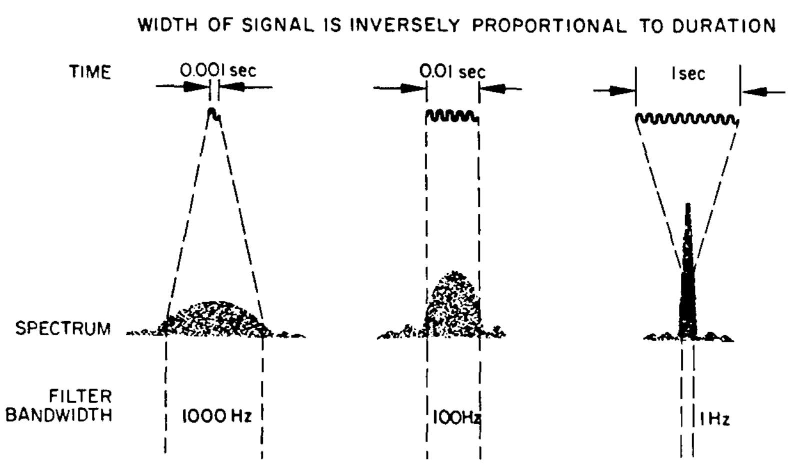


FIGURE 4–30. Selection of signals.

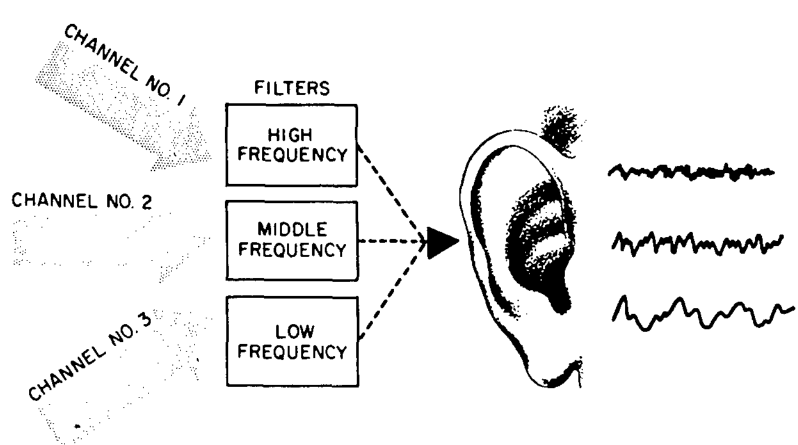


FIGURE 4–31. When a listener monitors several channels at one time, channels should be spaced at different frequencies.

of other frequencies, e.g., 2:1, 3:1, etc., because they are more easily confused.

Sequencing the Signals

When several active channels must be monitored in which urgent messages come in on two or more channels at the same time, it is best to provide for separate operators for each channel. Messages which are not urgent can be stored and sequenced. A system known as RODMEX (Readout-on-Demand Message Expeditor) records incoming messages on magnetic tape and actuates indicator lights to show that a message is stored. The operator can listen to these messages in any sequence he desires simply by pushing appropriate control buttons. Laboratory tests simulating a typical competing message situation showed that elimination of repeats by RODMEX resulted in a net gain in time and accuracy when compared with direct listening (Bertsch, 1956).

Identifying the Signals

Often a listener must rapidly and correctly identify which channel is presenting a message to him. The ability to make such an identification depends on parameters available for channel coding.

One- and Two-Dimensional Coding

If only one parameter can be used for coding the channels, no more than four or five channels can be correctly identified all the time. If intensity is the parameter, the limit is four; if frequency is the parameter, the limit is five.

In information-theory terms, coding by intensity alone yields 1.7 bits per stimulus, corresponding to nearly perfect identification of four levels. Coding by frequency alone yields about 2.3 bits, corresponding to perfect identification of five frequencies. To attain this performance, however, the stimulus range and intervals must be chosen carefully and the listeners must be well trained (Pollack, 1953).

Two-dimensional coding further increases the number of absolute identifications possible. Combinations of intensity and frequency, each divided on a five-step scale, have been shown to yield 3.1 bits per stimulus, corresponding to perfect identification of slightly over eight channels.

Multidimensional Coding

With additional parameters combined to code channels, a limitless number of signals can be identified. For example, a listener can recognize hundreds of different voices and thousands of different sounds because of the many complex characteristics serving as identification cues.

Experiments indicate that it is better to use more parameters with fewer steps, e.g., a two-step, eight-parameter signal yielded about the same performance (6.9 bits per stimulus) as a five-step six-parameter signal (7.1 bits per stimulus). The parameters used in these tests were direction, tonal frequency and level, noise frequency and level, repetition rate, on-off time fraction, and duration. (See Pollack and Ficks, 1954.) When as many as six or eight parameters are used, about three steps per parameter are optimal.

4.7 The Theory of Signal Detectability

In psychophysics prior to the 1950's, "thresholds" of detectability for the various sensory modalities were defined as the signal intensities at which a subject "just detected" or "just failed to detect" 50% of the signals presented. By varying the level of the signal, one could generate a psychometric function, such as those shown in Figure 4–32, which presents "percent correct responses" as a function of signal level. The three functions shown are similar to those which can be obtained from a single subject who is instructed to (a) "be sure not to miss any signals," (b) "just get as many as you can without worrying about it," and (c) "be very sure a signal is present before you say so." As may be seen, the more "strict" his instructions, the fewer correct responses he makes at a given signal level. The attitude of the observer or the instructions he receives are thus shown to greatly influence his "threshold," as classically defined. Any comparison of two or more types of signals or displays is effectively ruled out by such phenomena, because it is impossible to determine whether differences in performance

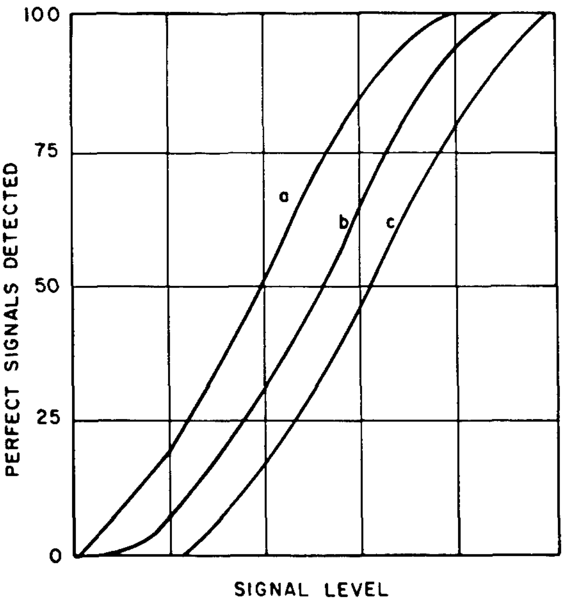


FIGURE 4-32. Psychometric functions for three signal levels.

are caused by a change in ability to detect signals or by a change in "criterion" or attitude on the part of the observer.

The Theory of Signal Detectability (TSD) offers a technique to separate the effects of the observers' criteria and of the detectability of signals, and to determine the magnitude or relative value of each. As set forth by Swets, Tanner, and Birdsall (1961) and by Green (1960) among others, TSD assumes that the observer who attempts to make a binary decision concerning the presence or absence of a signal on a a given trial obtains his information in the presence of noise. The noise may be part of the natural environment, it may arise within the nervous system of the observer, or it may be produced by the experimental apparatus, intentionally or unintentionally, or both.

The simplifying assumption most often made is that the noise is random; an assumption which adequately represents the data, even though the noise may depart from the normal in some instances. Jeffress (1964) discusses one such case. At any rate, the noise is continuously varying from trial to trial whether or not a signal occurs.

Another assumption is that, in its effects on the observer, sensory excitation may be regarded as a unidimensional variable. The observer is, theoretically, aware of the probabilities of each excitatory state possible during observation intervals containing signals and during those containing no signals. His decision as to whether or not a signal occurred during a given interval is based on a ratio of these two probabilities, a likelihood ratio. Because of the continuous variation in the noise, either with or without an added signal, some noise will appear "signal-like" to the observer and some signals will appear to be "noise." The criterion adopted by an observer serves as a decision rule to determine what ratio of signal-like and noise-like qualities of a sound he requires to judge that a signal occurred.

One consequence of the adoption of a criterion is that some signals will be judged, erroneously, to be "non-signals" and some noise will be called "signal." The manner in which incorrect responses or "false alarms" is handled constitutes one of the basic distinctions between TSD and most older psychophysical techniques. With classical methods, a response of "signal" when no experimenter-induced signal occurred was counted as "guessing" and prompted either a stern lecture from the experimenter to stop making "mistakes," or a mathematical "correction for guessing," or both of these. TSD, on the other hand, makes use of such false alarms to infer something about the nature of the stimulus conditions which obtain for the observer.

Given the assumptions that the noise under consideration is randomly distributed and that the addition of a signal results in a similar distribution with a higher mean value, Figure 4-33 serves to illustrate the populations from which a subject must draw when he makes his decision to say "yes" or "no" in a detection task.

The two probability density functions represent the possible values of noise and noise-plus-signal which the subject might receive when, for example, the noise is set at some r.m.s. level and the signal is, in the absence of noise, set at a desired voltage. The two calibrated values represent the means of the two distributions. The instantaneous value of the noise is certainly expected to fluctuate about its mean. Addition of the signal results in a similarly variable distribution. The distance between the

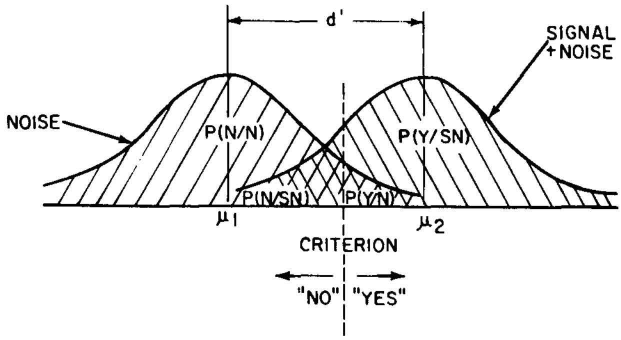


FIGURE 4–33. Distribution of "yes-no" responses of observers in a detection task.

means of the two distributions, which is given the notation d′, is a measure of how detectable the signal would be to the "ideal observer" postulated by TSD. (The ideal observer is assumed to have all pertinent information concerning the stimulus configuration and a perfect memory for waveforms from one interval to the next.)

In a binary decision task involving the illustrated distributions, the possible response alternatives can be presented in the following 2×2 matrix. (See Figure 4–34.)

		STIMULUS	
		NOISE	SIGNAL +NOISE
RESPONSE	"YES"	P(Y/N)	P(Y/SN)
	"NO"	P(N/N)	P(N/SN)

FIGURE 4–34. 2×2 matrix of response alternatives.

P(Y|N)+P(N|N) = 1.0 and P(Y|SN)+ P(N|SN) = 1.0, independently of the criterion adopted by the observer. His criterion *does* determine the relative values of the cells P(Y|N) and P(Y|SN), which, by convention, are the two which are compared to calculate d′. (Notice that one could look at the distribution of "no" responses and obtain the same results, quite unlike the information which traditional psychophysics extracted from negative responses.) Only when the entries in these two data cells are equal does d′ = zero. As the observer adopts a more lax criterion, the values of both of the cells increase; as he becomes more strict, they decrease. It is important, too, that the increase and decrease are not simply linear, as would be predicted by a "threshold" theory. Such a theory would state that, below threshold, the subject who was told to get more signals would simply guess and add proportionally the same amount to both cells. To become more strict, he would subtract from both cells in the same manner. Swets (1961) presents a discussion of data which show that such is not the case, that the changes in the entries in the matrix are more in line with what one would expect if the criterion line in the previous figure were moved to the left or right.

It may be seen that even an ideal observer could not, in the situation illustrated, get 100% hits and no false alarms. If he were instructed to "Get as many signals as you can," he would set his criterion to the left (deeper into the noise distribution) and he would respond "yes" to all values to the right of that. He would, in other words, correctly detect many of the signals; but would also make a high number of false alarms. If he were told, "Be certain a signal is present before you say yes," he would move his criterion to the right. As a consequence, he would make fewer false alarms; but his hit rate would also decrease. Only if the two distributions were completely separated, as would occur with an extremely high signal level, would an ideal observer be able to get all of the signals and make no false alarms. (See Figure 4–35.)

The values and costs of making correct and incorrect responses also determine the criterion

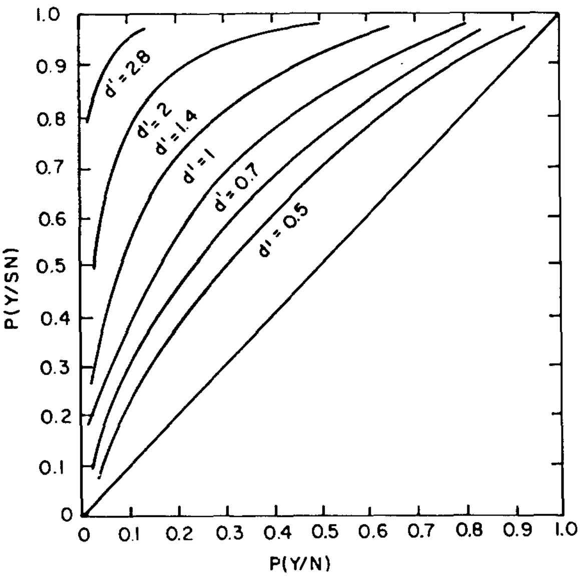


FIGURE 4–35. Receiver operating characteristic curves From *Signal Detection and Recognition by Human Observers* by John A. Swets. Copyright 1964 John Wiley & Sons, Inc. Used with permission of John Wiley & Sons Inc.

the ideal observer chooses. If, for example, saying "yes" means a complex weapons system with political overtones will be set into operation, the pressure will be in the direction of a more strict criterion; not many hits, but not many false alarms. If failure to make a detection will result in the destruction of the operator and his crewmates, while a false alarm will result in expenditure of a fairly "cheap" armament, the observer will tend to adopt a more lax criterion: more hits, but more false alarms.

Beneath all of the criterion shifts discussed, d′, the detectability of the signal, has not changed. It is this ability to differentiate between the criterion selected by an observer and the detectability of a signal that makes TSD a powerful base from which to examine psychophysical phenomena.

In application, human observers are presented with stimuli and their responses recorded and tabulated. From tables of d′ which are available, such as that prepared by Elliott (1959), the value associated with the hit and false alarm probabilities of the data is obtained. This is equivalent to stating that, "if an ideal observer had made such scores, the separation between the means of the two distributions would have been . . . ." The performance of the human observer is thus compared to that of a hypothesized ideal: another aspect of TSD which makes it a useful technique.

A plot of several receiver operating characteristic, ROC, curves is shown in Figure 4–35 and d′ is the parameter. For each d′ a signal-to-noise ratio may be determined, and an examination of the figure makes clear the relation between increasing d′, or signal-to-noise ratio, and the increasing probability of a "hit," P(Y | SN), for a fixed false alarm rate, P(Y | N).

Several points in the previous discussion are of particular importance in the area of human performance relative to various display systems and formats which might occur in sonar systems. First, the most common descriptor of a system by which it is compared to another is often the "recognition differential." This is nothing more than a technical name for the traditional "threshold," the signal level required to produce 50% detections. The previous discussion of TSD should make clear that such a concept is not justified on the basis of human performance. The differences in "recognition differential" between two systems might be a consequence of a change in criterion, rather than of a change in detectability produced by one of the sets.

Another point is that the traditional threshold is not an adequate measure of the capabilities of operators relative to even *one* single display or format. In the literature one may still read reports concerning the discovery of "two groups of observers": one, a "good" group which detects many of the signals (measured as percent of signals detected) and another "poor" group which detects few of the signals. The two groups may change percent correct in an orderly fashion when signal levels are altered, one remaining "superior." Such results clearly call for examination in terms of observers' criteria.

Still another result which should call for a consideration of criterion effects is that of *no* difference between displays, as shown by percent detections. Percent correct responses can remain unchanged in two conditions, while false alarm rates change markedly.

From a consideration of the relative capabilities of the two techniques, threshold determination and TSD, it should be clear that the former is not merely a weaker or a less desirable one for evaluating displays or human performance, but that it is altogether undesirable. The information ignored (i.e., false alarms) is seen to be extremely important in determining the detectability of signals. In nearly every case but those in which only crude measures of level or ability are desired, TSD would be the preferred technique.

4.8 Other Sensory Systems

Audition and vision are the two most frequently used sensory channels for transmitting information. This is indicated by the rich vocabulary for describing their phenomena. Because of man's exposure to exotic environments and to increased information processing, other sensory input channels are now receiving much needed attention. These channels are: (a) somesthetic sensing, i.e., touch, temperature, and pain; (b) the sense of smell, (c) the sence of balance or vestibular sense; plus (d) sensations of position and movement or kinesthesis, all of which contribute to a knowledge of the immediate environment.

4.8.1 Somesthetic Senses

The somesthetic or skin senses include: (a) tactual or touch sense, (b) temperature, and (c) pain. Of these, the tactual or touch sense provides the most information for task performance; the others are primarily sources of information used for body protection.

Touch sensitivity is dependent on deformation of the skin. The rate at which the skin is deformed is important in determining thresholds. For example, the absolute threshold for touch is lower as a stimulator is pressed against the skin more rapidly than if the pressure is applied slowly. In fact, if the stimulator is applied slowly enough, the person will be unaware of the pressure. Also, once a constant pressure has been reached, the sense will adapt and the person will cease to be aware of the contact; for example, the person may be unaware of his clothing until he moves and creates new pressure gradients. The absolute threshold can vary with the part of the body stimulated, e.g., the palm of the hand is more sensitive than the back of the hand. The perception of two or more pressure points as separate, or the *difference thresholds* for touch, are more difficult to establish but, in general, the separation increment is smaller for that body surface which displays the higher absolute sensitivity. Sensitivity to vibratory stimuli represents another dimension of touch sensitivity. The maximum sensitivity occurs at about 250 Hz. The upper limit for the perception of vibration of the skin has not been clearly established because of the difficulty of moving the skin at higher frequencies. In addition, the skin can respond to complex touch stimuli, for example, texture (roughness) and shape.

Sensing of temperature or of pain have little significance as sources of information for the performance of work, except the latter which may degrade it; with excessive temperatures, discomfort outweighs motivation.

Factors Which Influence Somesthetic Sensing

As with man's other senses, the somesthetic senses are influenced by a great many factors. Sensitivity to touch varies considerably from individual to individual and, as already noted, there is considerable variation in sensitivity with the part of the body stimulated. Also, extremes of temperature affect sensitivity. One only need recall a time when his hands were numb from cold to know that touch sensitivity is degraded by low temperatures.

The most significant consideration about man's tactual sensing capabilities for design purposes has to do with the design of controls that are easily identified by touch. The design of gloves which does not excessively reduce tactual sensitivity is also of interest.

Possibilities exist for training a person to interpret tactual stimulation (in much the same manner as Morse code), thereby adding another channel by which communication could be provided (Hennessy, 1966).

4.8.2 Olfaction and Gustation

The *absolute threshold* for the olfactory or smell sense has been estimated to be 10,000 times as sensitive as taste, but difference thresholds are not as sensitive. The fundamental odors frequently are considered to be "fragrant," "spicy," "resinous," "burnt," and "putrid," although other classification schemes are also used. However, smell stimuli are always mixtures of a few basic odors. They may also involve mechanical and chemical stimuli as well as those carried by the air. For example, the odor of ammonia is due to excitation of pain receptors as well as those of smell. Stimulus thresholds vary with the substance; for example, only 0.00004 mg/liter of air for artificial musk can be detected, whereas 5.83 mg/liter of air is required for the detection of ethyl ether. The *difference thresholds* are ordinarily quite large for the olfactory sense.

The gustatory or taste sense is mediated by taste buds in the tongue. The fundamental taste qualities are usually considered to be "salt," "sour," "sweet," and "bitter." All other tastes are combinations of these.

Factors Which Influence Sensing

Olfactory and gustatory sensitivity varies considerably from individual to individual. Both show nearly complete *adaptation* with continued exposure to the same substance. Respiratory infections which block the nasal passageways can effectively eliminate olfaction. Taste sensitivity can be affected by the portion of the tongue stimulated, the temperature of the substance, and internal factors such as salt deficiencies.

The use of the olfactory sensitivity has some application in the detection of hazardous conditions, such as fumes of toxic gases. However, the sense of smell should not be depended upon as a reliable source of information. The use of appropriate sensing instruments is a much safer approach. In other situations, the first indication of an equipment malfunction may be through the smelling of burning insulation. However, this sensory capability is hardly reliable enough to use as a basis for design. To date, no use has been found for the sense of taste in the design and operation of systems, although this hardly denies its importance in everyday life.

4.8.3 Kinesthetic and Vestibular Senses

Types of Information

The capabilities of the kinesthetic and vestibular senses will be considered together since they jointly account for man's ability to perceive: (a) the position and orientation of his body and limbs; (b) the movement of his body and limbs; (c) the position or attitude of vehicles with man in the vehicle; and (d) the movement of vehicles, again with man in the vehicle. Both of these senses take on added importance in the absence of, or with reduced, visual information, although as will be noted later, these sometimes can provide erroneous information and may conflict with visual information.

Vestibular Sense

Having evolved over millenia in a gravitational environment, the vestibular organs and semicircular canals are quite specialized. In the organs of the inner ear are little "weights" called *otoliths*. Because of inertia, they tend to remain stationary when the fluid in the semicircular canal is displaced in response to changes in linear and angular velocity. This in turn triggers neural impulses. These impulses are "processed and interpreted" to be used for such things as postural control and so on. As in the case of temperature, the vestibular sense is a poor candidate through which to transmit coded information.

The kinesthetic sense, or muscle sense, provides important knowledge about the position and movement of the body. For example, even with his eyes closed, a person can walk, sit down, get up, and so forth, although not as well as when he also has visual cues. The kinesthetic sense provides information on the position of the limbs, how far they have been moved, and the posture of the body as a whole. For example, this sense provides important information for the "blind" positioning of controls (i.e. "positioning without visual cues") and performance of repetitive, continuous, and serial movements. Kinesthesis contributes to judging accurately the weight of different objects when they are lifted. In addition, the sense provides information about changes in orientation and equilibrium, either by detecting changes in the position of body members as caused by external forces, or by detecting reflex changes in the muscle system which maintain posture. For the most part, these stimuli provide information only that some change has occurred and visual cues then must be relied upon to determine the exact nature of the change.

The unique characteristic of kinesthesis is that its stimulation originates within the body itself, and not from external stimulation as do most of the other sensory channels. Kinesthetic stimuli are always present, but usually without full awareness of them by the person. Their sensation qualities are limited and relatively poorly localized. As such they cannot be quantified to the same extent as visual and auditory stimuli. Attempts at quantification are usually based on measurement of a movement threshold at which a displacement or velocity can be detected. The *absolute threshold,* whether measured in terms of minimum angular displacement that can be detected (rate of movement being held constant) or in terms of the minimum velocity of motion that is discriminable, is most sensitive for the shoulder joint and is least sensitive for the ankle.

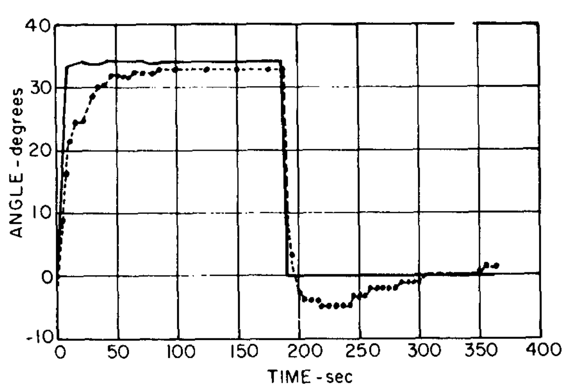


FIGURE 4–36. Perception of the vertical by the vestibular sense (Graybiel, 1952).

The vestibular sense also provides information about position and movement of the body as a whole. Along with the kinestietic sense, it provides information on the static and dynamic equilibrium of the body in standing and walking. Displacement of the body from the vertical also is perceived by this sense (see Figure 4–36); but more directly, it provides information to the person about changes in velocity or acceleration, although the delay in perception is greater than that for kinesthetic sense. For example, with an angular acceleration of 10°/sec², motion per-

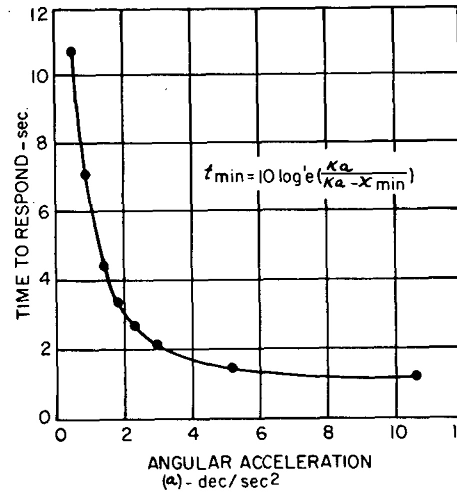


FIGURE 4–37. Thresholds for sensing rotation (Guedry and Richmond, 1957).

ception occurs in about 1 sec., but, if the angular acceleration is only about 0.5°/sec², it may take the individual as long as 10 to 12 sec. to perceive the motion (see Figure 4–37). The absolute threshold for perception of motion by the vestibular sense is usually considered to be between 0.1° and 0.5°/sec. It is important that the sensations provided by the vestibular sense not be in conflict with visual or kinesthetic sensations, since such conflict can lead to disorientation. Rotation of the body, tilting the head when the body is rotating, rotating of the body opposite from that of a vehicle on which the person is riding, or vertical oscillation (such as that encountered in softly sprung cars or on ships) can produce profound disorientation and often motion sickness.

Factors which Influence Kinesthetic and Vestibular Sensing

Although there are individual differences in sensitivity to kinesthetic stimuli, the most important source of variation is the result of man's ability to learn to interpret these cues accurately. For example, with practice an individual can learn to position a control quite accurately without visual cues. The absence of the earth's normal gravitational field, such as the weightlessness of space flight or a reduced gravitation (such as the lunar surface) results in the reduction or loss of many kinesthetic cues.

For the most part, the capabilities of the kinesthetic and vestibular senses are relatively insignificant for design purposes. Probably the most noticeable exception for the kinesthetic sense is in the design of controls where this sense aids man in positioning controls in the absence of visual cues. Also, both senses provide some information for the attitude and change of motion of vehicles, although it is doubtful if this has any direct implications for the design. The important consideration in designing for vehicle operation is to avoid rotations or oscillations which are conflicting or disturbing to the person or may cause motion sickness. If this cannot be done completely, suitable training may be necessary to allow the person to become as habituated as possible to these motions. The latter may have significant implications for the design and operation of rotating space stations, where rotation is employed as a means of producing artificial gravity.

References

American Standards Association. *The relations of hearing loss to noise exposure.* New York: ASA, 1954.

Beranek, L. L. Revised criteria for noise in buildings. *Noise Control,* 1957, **3**(1), 19.

Beranek, L. L. *Noise reduction.* New York: McGraw-Hill, 1960.

Bertsch, W. F. Effects of two message-storage schemes upon communications within a small problem-solving group. *Journal of the Acoustical Society of America,* 1956, **28,** 550.

Bilger, R. C. The additivity of different types of masking. *Journal of the Acoustical Society of America,* 1959, **31,** 1107–1109.

Deatherage, B. H. Examination of binaural interaction. *Journal of the Acoustical Society of America,* 1966, **39,** 232–249.

Deatherage, B. H., Jeffress, L. A., & Blodgett, H. C. A note on the audibility of intense ultrasonic sound. *Journal of the Acoustical Society of America,* 1954, **26,** 582.

Dolan, T. R., Deatherage, B. H., & Hafter, E. R. Masking-level differences for monaural signals binaurally masked. *Journal of the Acoustical Society of America,* 1965, **38,** 928(A).

Doughty, J. M., & Garner, W. R. Pitch characteristics of short tones: I. Two kinds of pitch threshold. *Journal of Experimental Psychology,* 1947, **37,** 351.

Egan, J. P., & Hake, H. W. On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America,* 1950, **22,** 622.

Eldred, K. M., Gannon, W. J., & von Gierke, H. Criteria for short time exposure of personnel to high intensity jet aircraft noise. Report No. WADC–TN–55–355, 1955, Wright Air Development Center, WPAFB, Ohio.

Elliot, P. B. Tables of d′. Technical Report No. 97, 1959, University of Michigan Electronic Defense Group.

Fletcher, H. Loudness, masking, and their relation to the hearing process and the problem of noise measurement. *Journal of the Acoustical Society of America,* 1938, **9,** 275.

Fletcher, H. Auditory patterns. *Reviews of Modern Physiology,* 1940, **12,** 47.

Fletcher, H. *Speech and hearing in communication.* New York: D. Van Nostrand Co., Inc., 1953.

Fletcher, H., & Munson, W. A. Loudness, its definition, measurement, and calculation. *Journal of the Acoustical Society of America,* 1933, **5,** 82.

Fletcher, H., & Munson, W. A. Relation between loudness and masking. *Journal of the Acoustical Society of America,* 1937, **9,** 1.

French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America,* 1947, **19,** 90.

Garner, W. R., & Miller, G. A. The masked threshold of pure tones as a function of duration. *Journal of Experimental Psychology,* 1947, **37,** 293.

Graybiel, A. Oculogravic illusion. *AMA Archives of Opthalmology,* 1952, **48,** 605–615.

Green, D. M. Psychoacoustics and detection theory. *Journal of the Acoustical Society of America,* 1960, **32,** 1189–1203.

Guedry, F. E., Jr., & Richmond, G. Differences in response latency with different magnitude angular accelerations. USA–MRL Report No. 301, 1957, Army Medical Research Labs., Fort Knox, Kentucky.

Hamilton, P. M. Underwater hearing thresholds. *Journal of the Acoustical Society of America,* 1957, **29,** 792.

Hathaway, J. L. Automatic audio gain controls. *Audio Engineering,* 1950, **34**(9), 16.

Hawkins, J. E., & Stevens, S. S. The masking of pure tones and of speech by white noise. *Journal of the Acoustical Society of America,* 1950, **22,** 6.

Hennessy, J. R. Cutaneous sensitivity communications. *Human Factors,* 1966, **8**(5), 463–469.

Hirsh, I. J. The influence of interaural phase on interaural summation and inhibition. *Journal of the Acoustical Society of America,* 1948, **20,** 536.

Humphrey, C. E. Auditory displays: 1. Spatial orientation by means of auditory signals—an exploratory study. Report No. APL/JHU–TG–122, 1952, Applied Physics Lab., The Johns Hopkins University, Silver Spring, Maryland.

Humphrey, C. E., & Thompson, J. E. Auditory displays: I. Comparisons of auditory and visual tracking in one dimension: B. Discontinuous signals, complex course. Report No. APL/JHU–TG–147, 1952, Applied Physics Lab., The Johns Hopkins University, Silver Spring, Maryland.

Humphrey, C. E., & Thompson, J. E. Auditory displays: II. Comparison of auditory and visual tracking in one dimension: A. Discontinuous signals, simple course. Report No. APL/JHU–TG–146, 1952. Applied Physics Lab., The Johns Hopkins University, Silver Spring, Maryland.

Humphrey, C. E., & Thompson, J. E. Auditory displays: II. Comparison of auditory tracking with visual tracking in one dimension: C. Continuous signals, simple, intermediate, and complex courses. Report No. APL/JHU–TG–194, 1953, Applied Physics Lab., The Johns Hopkins University, Silver Spring, Maryland.

Jeffress, L. A. Stimulus-oriented approach to detection. *Journal of the Acoustical Society of America,* 1964, **36,** 766–774.

Jeffress, L. A., Blodgett, H. C., & Deatherage, B. H. The masking of tones by white noise as a function of the interaural phases of both components. I. 500 Cycles. *Journal of the Acoustical Society of America,* 1952, **24,** 523–527.

Jeffress, L. A., Blodgett, H. C., & Deatherage, B. H. Masking and interaural phase II. 167 Cycles. *Journal of the Acoustical Society of America,* 1962, **34,** 1124–1126.

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., & Wood, C. L. Masking of tonal signals. *Journal of the Acoustical Society of America,* 1956, **28,** 416–426.

Klumpp, R. G., & Eady, H. R. Some measurements of interaural time difference thresholds. *Journal of the Acoustical Society of America,* 1956, **28,** 859.

Licklider, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley and Sons, 1951.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. *Journal of the Acoustical Society of America,* 1947a, **19,** 609.

Munson, W. A. The growth of auditory sensation. *Journal of the Acoustical Society of America,* 1947, **19,** 584.

Pollack, I. Comfortable listening levels for pure tones in quiet and noise. *Journal of the Acoustical Society of America,* 1952, **24,** 158.

Pollack, I. The information of elementary auditory displays. *Journal of the Acoustical Society of America,* 1953, **25,** 745.

Pollack, I., & Ficks, L. Information of elementary multidimensional auditory displays. *Journal of the Acoustical Society of America,* 1954, **26,** 155.

Riesz, R. R. Differential intensity sensitivity of the ear for pure tones. *Physiological Review,* 1928, **31,** 867.

Robinson, D. W., & Dadson, R. S. Threshold of hearing and equal-loudness relations for pure tones and the loudness function. *Journal of the Acoustical Society of America,* 1957, **29,** 1284.

Rosenblith, W. A., & Stevens, K. N. Noise and man: Vol. II of Handbook of acoustic noise control. Report No. WADC–TR–52–204, 1953, Wright Air Development Center, WPAFB, Ohio.

Schafer, T. H., Gales, R. J., Shewmaker, C. A., & Thompson, F. O. The frequency selectivity of the ear as determined by masking experiments. *Journal of the Acoustical Society of America,* 1950, **22,** 490.

Sharf, B. Critical bands. In *Foundations of modern auditory theory.* Syracuse University, Laboratory of Sensory Communication, Syracuse, New York, 1966. (pre-publication)

Shower, E. G., & Biddulph, R. Differential pitch sensitivity of the ear. *Journal of the Acoustical Society of America,* 1931, **3,** 275.

Snow, W. B. Basic principles of stereophonic sound. *Journal of the Society of Motion Picture and Television Engineers,* 1953, **61,** 567.

Stevens, S. S. The measurement of loudness. *Journal of the Acoustical Society of America,* 1955, **27,** 815.

Stevens, S. S., & Davis, H. *Hearing, its psychology and physiology.* New York: John Wiley and Sons, 1938.

Stevens, S. S., & Volkmann, J. The relation of pitch to frequency: A revised scale. *American Journal of Psychology,* 1940, **53,** 329.

Swets, J. A. Is there a sensory threshold? *Science,* 1961, **134,** 168–177.

Swets, J. A. *Signal detection and recognition by human observers.* New York: John Wiley and Sons, 1964.

Swets, J. A., Tanner, W. P., & Birdsall, T. G. Decision processes in perception. *Psychological Review,* 1961, **68,** 301–340.

Turnbull, W. W. Pitch discrimination as a function of tonal duration. *Journal of Experimental Psychology,* 1944, **34,** 302.

von Gierke, H. E., & Sommer, H. C. Hearing and feeling sensations in electric fields. *Journal of the Acoustical Society of America,* 1964, **36,** 1997(A).

Wallach, H., Newman, E. B., & Rosenzweig, M. R. The precedence effect in sound localization. *American Journal of Psychology,* 1949, **62,** 315–336.

Ward, W. D. Effect of temporal spacing on temporary threshold shift from impulses. *Journal of the Acoustical Society of America,* 1962, **34,** 1230.

Ward, W. D., Glorig, A., & Sklar, D. C. Dependence of temporary threshold shift at 4kc on intensity and time. *Journal of the Acoustical Society of America,* 1958, **30,** 944.

Ward, W. D., Selters, W., & Glorig, A. Exploratory studies on temporary threshold shift from impulses. *Journal of the Acoustical Society of America,* 1961, **33,** 781.

Wegel, R. L., & Lane, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. *Physiological Review,* 1924, **23,** 266.

Selected Readings

Beranek, L. L. *Noise reduction.* New York: McGraw-Hill, 1960.

Licklider, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley and Sons, 1951.

Stevens, S. S. & Davis, H. *Hearing, its psychology and physiology.* New York: John Wiley and Sons, 1938.

Chapter 5

Speech Communication

Karl D. Kryter

Stanford Research Institute
Palo Alto, Calif.

Criteria for equipment design to enable improved human speech communication are the major concern of this chapter. Measurements of fundamental speech sounds, their pressure, range, and level, as well as ways of measuring intelligibility through actual testing or through calculation are presented. Problems of noise and reverberation, and protection against them are also discussed as well as how components can aid intelligibility through gain control and peak clipping. Special design problems covering requirements of multichannel listening, communication in unusual environments, through masks, and using bandwidth compression are all considered, concluding with the human criteria of communication performance, trainability of persons, language factors influencing intelligibility, and the necessity for feedback.

CONTENTS

This chapter was reviewed by John W. Black.
The material in this chapter is a revision of Chapter 4 of the previous *Guide*.

5. Speech Communication

5.1 Speech Communication

As in other human engineering design procedures, those for speech communication follow a basic pattern. One of the first steps is to establish what the system will require. Occasionally some special speech situation must be included in preliminary considerations, such as requiring communication through a facial mask, or at high altitudes, or under water, or one requiring simultaneous transmission from several sources to a single listener.

Next, performance criteria are established. Will naturalness or quality of received speech be the desirable factor? Or will identifiability of the talker be most important? The main criteria usually involve intelligibility of the message and much attention is paid in this chapter to the question: How can the designer maximize the intelligibility of speech in a communication system?

The third basic step in establishing desirable characteristics of communication equipment is the environment in which the communication system will operate. It may be one where noise or reverberation are the problems. Noise is a major consideration throughout the chapter.

Next the designer must specify characteristics of the components of a communication system, i.e., criteria for the microphones, transmitters, receivers, and headsets. He must also determine the correct bandwidth. Because many individual voice channels may be used in a network, special speech processing techniques which reduce the bandwidth normally required for speech should be considered. Techniques for bandwidth compression are discussed.

Finally, the design engineer prepares at least one workable configuration, evaluates its effectiveness, and estimates its cost. By manipulating variables to maximize the difference between performance and cost, he can find the most effective design.

5.2 Characteristics of Speech

5.2.1 Representations of Speech Waves

The speech wave can be regarded as a function of time only (waveform), as a function of frequency only (spectrum), or as a function of both time and frequency (intensity-frequency-time pattern).

Time representation. The waveform function relates instantaneous speech pressure to time. Speech pressure is the measured force exerted by the speech wave on a unit area at an arbitrary, but specified, location with respect to the talker—usually normal to the line from the speaker's lips and 1 m. away.

Frequency representation. The speech spectrum corresponds to and is a transformation of the pressure waveform. It is a complex function of frequency only and can be resolved into two parts. One part gives the amplitudes of the various frequency components of the speech signal; the other gives the phase angles of those components at some arbitrary instant in time.

The distribution of pressure amplitudes as a function of frequency can be measured and depicted in several ways. The most widely used method is to transduce the acoustic speech signal into an electrical signal and pass it through a band-pass filter. To obtain the root-mean-square (r.m.s.) pressure:

1. The voltage output of the filter is squared and integrated over time.

2. The integral is divided by the duration of the period of integration, and the square root of the result is taken.

This computation yields the r.m.s. voltage in the filter pass band, which translates into an r.m.s. pressure using a calibration curve for the transducer in the analyzing system.

The numerical value of the r.m.s. pressure is the spectral coefficient corresponding to the center frequency of the analyzing filter. When octave-band or half-octave-band filters are used, stable spectra are obtained with integration periods as short as 1 to 2 min.

Octave-band-spectrum. When filters one-octave wide are used, with spectral coefficients related to the center or boundary frequencies of the octave, the spectrum is called the octave-band. The spectral coefficients are transformed to a logarithmic scale and are expressed in decibels (dB) relative to some convenient reference value such as microbars (μbar) or microbars per volt (μbar/v).

Spectrum level. To obtain the pressure level per cycle, divide the r.m.s. pressure in each band by the width of the band in cycles per second, often designated as Hertz (Hz). When the quotient is translated into decibels, the result is called the spectrum level. The relationship among octave, one-half-octave, one-third octave, and spectrum level is shown in Figure 5–1. The octave-band level is arbitrarily taken to be 0 dB.

Overall level. It is usually helpful to give an indication of the overall level of speech. "Overall" indicates that the measurement of an unfiltered speech wave is made over the entire audio-frequency range.

Results of measurements. The octave-band spectrum and curves relating spectrum level to frequency are shown in Figure 5–2. The right-hand ordinate is appropriate for the octave-band spectrum and the left-hand ordinate is appropriate for the spectrum-level curves.

The solid curve in Figure 5–2 is an idealized long-term average based on several experimental determinations in which octave-band and half-octave-band filters were used (French and Steinberg, 1947). The irregular dashed line represents the spectrum level determined in a typical set

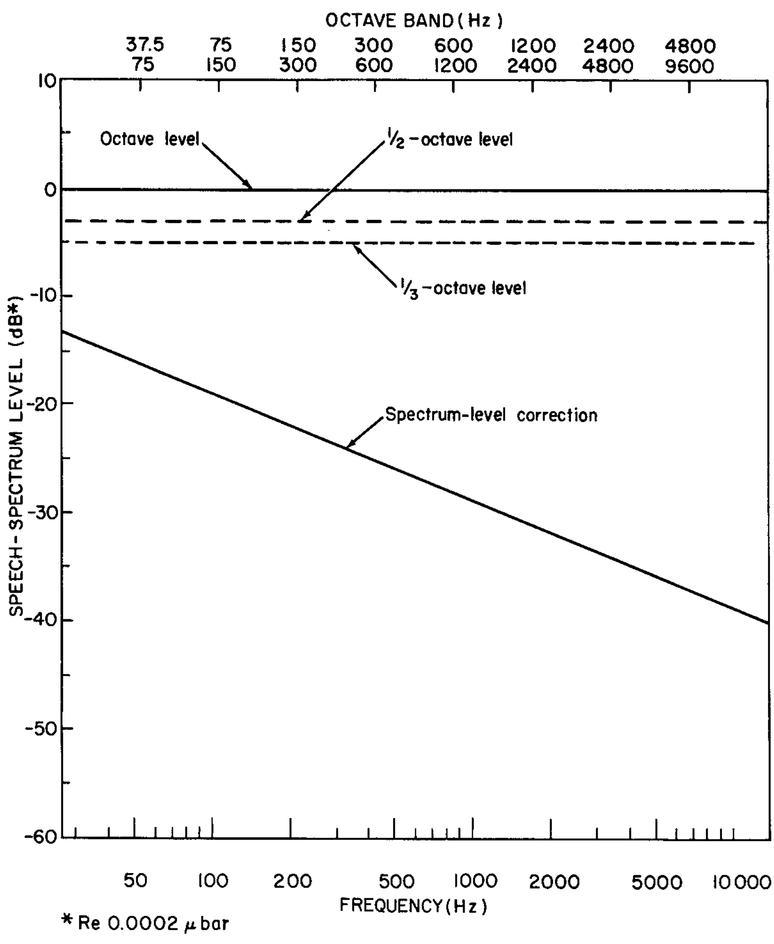


FIGURE 5–1. The relationship among octave, one-half octave, one-third octave, and spectrum level.

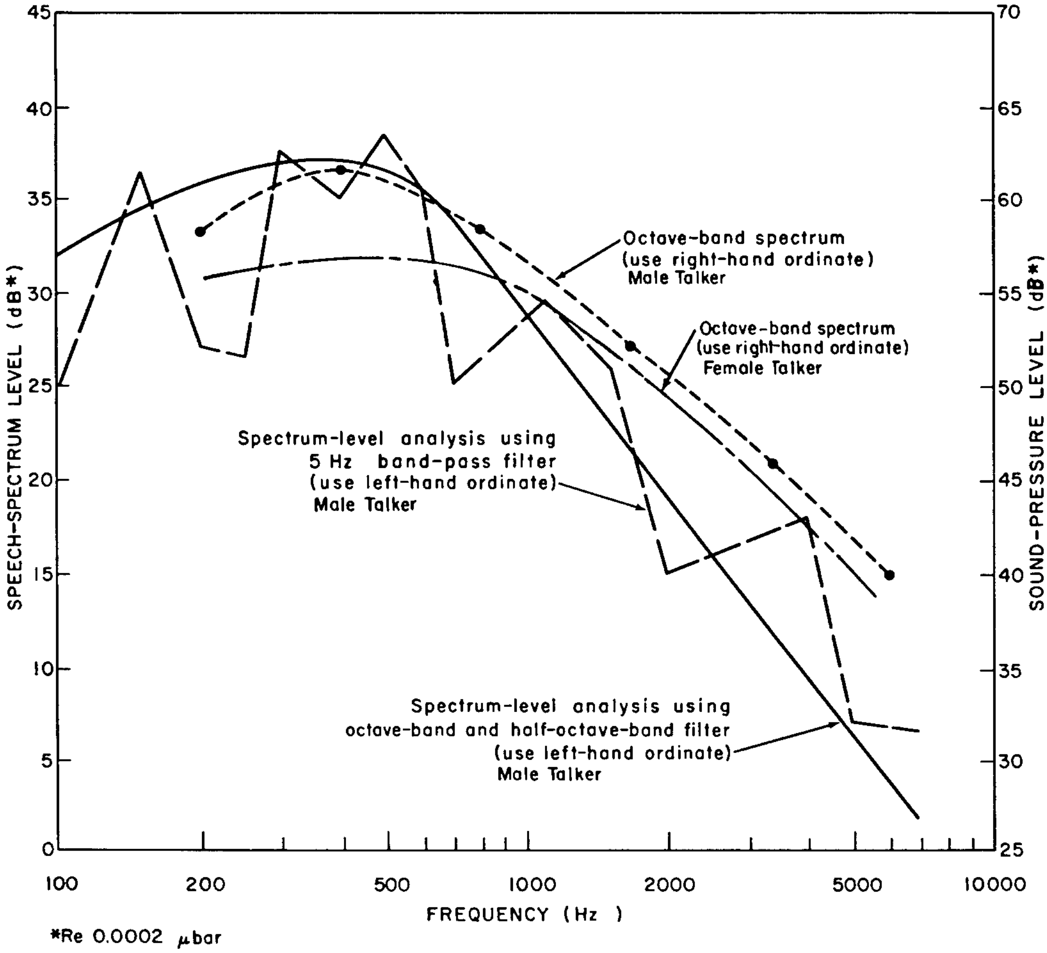


FIGURE 5–2. The octave band spectrum and two curves relating to spectrum level to frequency.

of measurements made on a 3-min. sample of speech by a single talker with a 5-Hz filter set at various places along the frequency scale (Stevens et al., 1947). It illustrates that any particular sample of speech is likely to depart considerably from the long-term average.

The octave-band spectrum shown in Figure 5–2 is an average based on three independent studies of 17 male talkers who used a moderate level of voice effort. The overall speech level is 65 dB relative to 0.0002 μbar. The spectrum produced by female speakers is roughly 2 dB lower in level in the frequency region above about 800 Hz, but averages about 5 dB less in the region from about 200 to 800 Hz. As a result, for a comparable level of effort, the female speech signal measures overall about 3dB less than that for male talkers.

The spectra of speech shown in Figure 5–2 are "long-time spectra." Systematic changes in the function relating spectrum level to frequency would not have been observed had the duration of the measurement been decreased, but time does not appear as an independent variable.

It may be important to examine changes of speech pressure with time while simultaneously retaining the analysis of the speech wave into several or many bands of frequency. One way of accomplishing this is to divide the speech signal into a number of frequency bands (by means of band-pass filters) and then to divide the component signals—the individual signals in the several bands—into fractions of a second.

Measurements have been made in this manner with octave-band and half-octave-band filters. The maximum instantaneous pressure in each

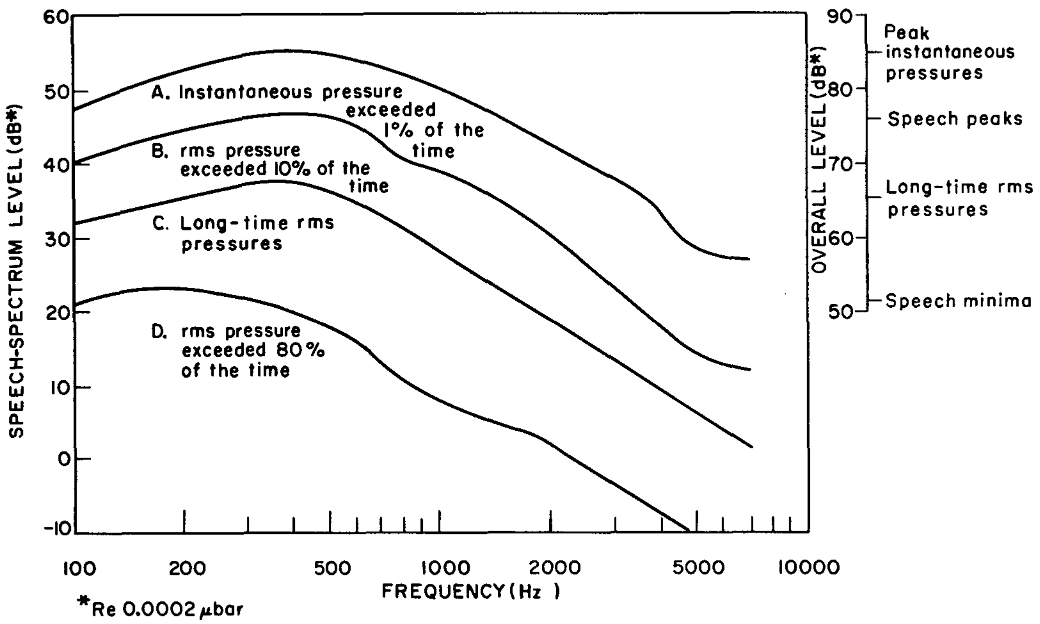


FIGURE 5–3. Relation of sound-pressure level to frequency (Dunn and White, 1940).

⅛-sec segment, and, also the r.m.s. pressure in each ⅛-sec segment, are determined. Spectrum levels are derived by dividing the squares of the instantaneous pressure and the r.m.s. pressure by the filter bandwidth and then converting the quotients into decibels.

Four curves relating sound-pressure level to frequency are shown in Figure 5–3 (Dunn and White, 1940). Curve A shows for each frequency band of 1–Hz width the instantaneous pressure that was exceeded in only 1% of the ⅛-sec. intervals. In a sense, this curve is a "peak-instantaneous-pressure" curve. Curve B shows for each frequency band the r.m.s. pressure that was exceeded in only 10% of the ⅛-sec. intervals. We will call this one the curve of "speech peaks." Curve C is the long-time r.m.s. pressure corresponding to the solid curve shown in Figure 5–2.

Curve D in Figure 5–3 shows the r.m.s. pressure for each frequency band that was exceeded in 80% of the ⅛-sec. intervals. Inasmuch as about one-fifth of ordinary conversational speech is dead time, this lowermost curve represents, in a sense, the r.m.s. pressure of the weakest sounds. This curve is the "speech minima" curve. At the right-hand side of Figure 5–3 are represented the corresponding overall levels—the values for unanalyzed, unfiltered speech. Although the instantaneous speech pressure is usually below the r.m.s. pressure, the speech wave occasionally makes rather extreme excursions. One problem in designing a speech-communication system is whether to provide for faithful reproduction of these extremes of speech waves. (See discussion on peak clippings, Section 5.5.2.)

5.2.2 R.M.S. Pressure of Fundamental Speech Sounds

By making the intervals of time over which the squared pressures are averaged correspond directly to the intervals during which the specific vowels and consonants are spoken, one can relate the measurements to the individual sounds as uttered. The r.m.s. pressures as measured by telephone engineers are shown in Table 5–1. Values are expressed in decibels relative to the r.m.s. pressure of the weakest speech sound (the initial consonant, "th," of "thin").

5.2.3 Dynamic Range

Speech that is too soft will be masked by noise. Speech that is too loud will overload the communication system. Dynamic range is the difference, in decibels, between the pressure level

TABLE 5–1. TYPICAL R.M.S. PRESSURE LEVELS OF THE FUNDAMENTAL SPEECH SOUNDS

Key word	Sound*	Pressure level (dB)	Key word	Sound*	Pressure level (dB)
talk	o′	28.2	chat	ch	16.2
top	a	27.8	me	m	15.5
ton	o	27.0	jot	i	13.6
tap	a′	26.9	azure	zh	13.0
tone	o	26.7	zip	z	12.0
took	u	26.6	sit	s	12.0
tape	ˈa	25.7	tap	t	11.7
ten	e	25.4	get	g	11.7
tool	u	25.0	kit	k	11.1
tip	i	24.1	vat	v	10.8
team	e	23.4	that	th	10.4
err	r	23.2	bat	b	8.0
let	l	20.0	dot	d	8.0
shot	sh	19.0	pat	p	7.7
ring	ng	18.6	for	f	7.0
me	m	17.2	thin	th	0

* Spoken by an average talker at a normal level of effort (Fletcher, 1953).

at which overload occurs and the pressure level of noise in the system. Obviously, it is not the same for all points in the system but it is the dynamic range at the listener's ear that is important. To determine the dynamic range required of a communication system, variations in pressure level from speech sound to speech sound and condition to condition must be studied. From Tables 5–1, 5–2, and 5–3 we note that:

1. The range of fundamental speech-sound levels is from 0 to 28.2 dB. (See Table 5–1.)
2. The range (difference) from speech minima with minimum normal vocal effort to peak instantaneous pressures with maximum normal effort is 60 dB (39–99 in Table 5–2).
3. The range (difference) from speech minima to peak instantaneous pressures is about 40 dB for a given level of vocal effort (such as average normal level in Table 5–2).
4. The range of variations of talkers in normal conversation is 20 dB (see Table 5–3).

Design Recommendations for Dynamic Range

1. Very high-quality communications call for a dynamic range of 60 dB.
2. For commercial broadcast purposes, the dynamic range can be 40 to 45 dB.
3. With a mechanism that compensates for variations in average speech levels among talkers, a dynamic range of 30 dB is adequate.
4. With practiced talkers and listeners, communication can be quite effective at a dynamic range of only 20 dB.

5.2.4 Estimating Speech Level

By reading either a special voltmeter, called a volume indicator (VU meter) or a sound level

TABLE 5–2. SOUND-PRESSURE LEVELS OF SPEECH 1 M. FROM THE TALKER

Measure of sound pressure	Whisper (dB)	Normal level (dB)			Shout (dB)
		Minimum	Average	Maximum	
Peak instantaneous pressures	70	79	89	99	110
Speech peaks	58	67	79	87	98
Long-time rms pressures	46	55	65	75	86
Speech minima	30	39	49	59	70

TABLE 5–3. DISTRIBUTION OF TALKER LEVELS FOR PERSONS USING THE TELEPHONE

Percent of talkers	Level range (dB*)
7	Below 54
9	54–57
14	57–60
18	60–63
22	63–66
17	66–69
9	69–72
4	72–75
0	Above 75

* Above sound pressure of 0.0002 μbar at a point 1 m. from the talker's lips (Fletcher, 1953).

meter, personnel in commercial radio, television and recording can find the approximate overall r.m.s. pressure. Ordinarily, the sound-level meter is set for "slow" meter action and "flat" response. It can be calibrated to indicate the r.m.s. pressure at the microphone location.

If the engineer takes the averages, by eye, or the highest deflections corresponding to the individual words for several spoken sentences, his readings will be about 4 dB above long-time r.m.s. pressure level as measured by more precise techniques. As can be seen from Figures 5–2 and 5–3, the long-time r.m.s. pressure level of normal speech, measured 1 m. from the talker's lips, is about 65 dB above the standard reference level of 0.0002 μbar (i.e., 9 dB below 1 μbar or 129 dB below 1 bar). Normal speech intensity at different positions around a talker's head is shown in Figure 5–4 (Fletcher, 1953).

Long-time r.m.s. pressure levels of Figures 5–2, 5–3, and 5–4 are for speech uttered with normal effort by a talker using a telephone or in an office. When the talker is speaking to a listener only 1 m. away in a quiet place, the level of conversational speech is somewhat lower than the 65 dB considered normal—it is about 58 dB. Because most communication systems include microphones and/or background noise, however, it is appropriate to identify "normal" with 65 dB, relative to 0.0002 μbar, 1 m. in front of the talker.

5.3 Intelligibility in Speech Communication

Many design decisions must be based on measurements of speech intelligibility. This may be accomplished using one of two procedures, each having certain limitations. Usually the engineer will predict or calculate intelligibility with the Articulation Index (AI). In the face of extreme noise masking conditions, or distortion of frequency or amplitude, or the need to evaluate systems involving complex speech processing, he may decide to obtain empirical data derived from intelligibility testing. This latter procedure should not be attempted, however, unless careful laboratory control is possible.

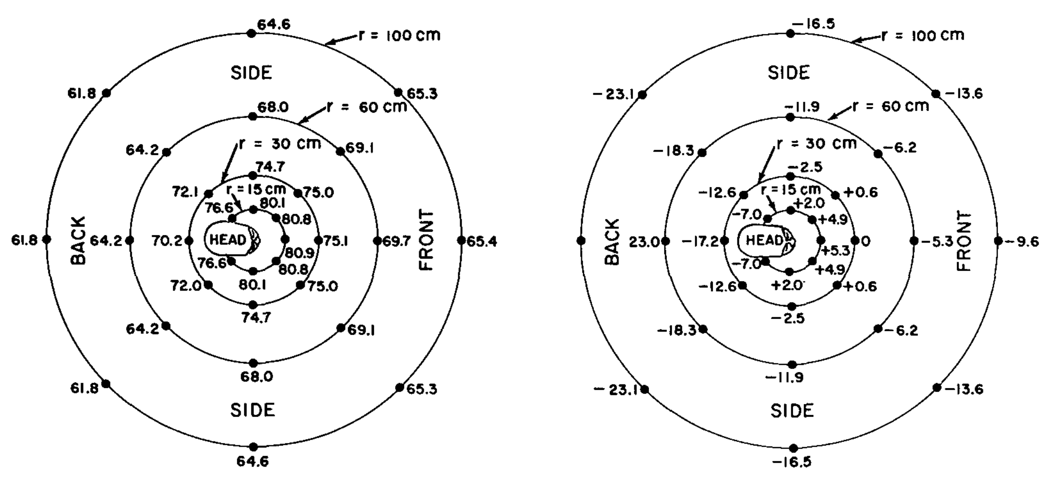


FIGURE 5–4. Speech-intensity levels around head of speaker. Numbers in left-hand diagram are for whole speech and are in decibels relative to 0.0002 μbar; numbers in right-hand diagram are for band of speech of 2800 to 4000 Hz and are in decibels relative to level 30 cm. in front of lips (Fletcher, 1953).

5.3.1 Steps in Intelligibility Testing

1. Formulate the objective explicitly. Intelligibility test scores are most useful as bases for comparison. If the objective is evaluation or comparison with other systems, testing a reference system is recommended.

2. Match the testing environment with that of actual use. The parameters of noise are especially important. If jet aircraft operate in the vicinity, or if an interior of an armored vehicle is to be the environment, noise levels and spectra conditions should be duplicated.

3. Schedule tests to counterbalance the effects of learning, fatigue, and boredom. If word lists are used repeatedly, be sure to rearrange the order of the words with each use.

4. Select a group of listeners who have normal hearing, know the language, and are motivated to do their best.

5. Select a group of talkers who are to be representative of the actual end users with respect to dialect, voice training, voice power, etc. Ordinarily, this means that they should not have unusual regional dialects or noticeable speech defects. They should be thoroughly familiar with the language to be used in the tests, and try to work without speaking too slowly, and without abnormally accentuating words.

6. Train talkers and listeners in the proper use of the equipment. Two hours a day for five days a week is not too long a period. The talker's lips should touch the microphone if a close-talking or noise-cancelling microphone is used. The talker's voice level should either be kept constant from condition to condition, using an independent microphone and meter for monitoring, or free to change as it would if the tests were being made in the operational context. The listening level should be representative of the level of actual use. Headsets, microphone noise shields, and ear protective devices should fit the wearers as in the actual system.

7. Select speech material to be used in the tests. The listeners should be familiar with the spelling and meaning of every word in the lists. Test material should score readily without ambiguity. Choose material that has been used in earlier tests or which may be used subsequently, so a basis for comparison may be established.

8. In nonsense-syllable or word tests, the talkers should insert each test syllable or word in a "carrier" sentence, such as, "You will write ———— now." The carrier sentence should be repeated with each new test item, without undue stress or emphasis.

9. Have the listeners write their best estimates of what the talker transmitted, and then compare their records with the standard list.

5.3.2 Selecting Test Material

Four types of tests have been used in measuring intelligibility: (a) nonsense syllable, (b) monosyllabic Phonetically Balanced (PB) word, (c) Modified Rhyme (MRT) (see Fairbanks, 1958, and House et al., 1965), and (d) sentence tests. Lists of items used in these tests are given in Tables 5–4 through 5–17. Testing material in use also includes the multiple-choice intelligibility tests developed by J. W. Black et al.

The items in a nonsense-syllable test are usually random combinations of fundamental speech sounds in the consonant-vowel-consonant pattern. In PB-word tests, items are usually drawn from a set of 20 lists of 50 words each in which the frequency or occurrence of types of speech sounds (plosives, fricatives, etc.) are proportional to those in everyday speech.

TABLE 5–4. NONSENSE-SYLLABLE LIST NO. 1

1. monz	26. dahf	51. zohm	76. duhm
2. nihf	27. fohf	52. gohn	77. map
3. nan	28. fook	53. pahz	78. zaf
4. ʒeef	29. kohth	54. thoop	79. puhf
5. dayth	30. thehʒ	55. dad	80. gahk
6 thayd	31. muhd	56. koof	81. pohd
7. gayf	32. kawd	57. pooth	82. nohg
8. thawf	33. zihg	58. fuhp	83. ʒuhg
9. dohp	34. kuhk	59. gehg	84. dihʒ
10. fayg	35. ʒihd	60. nood	85. pawg
11. meek	36. zehd	61. fehm	86. nawz
12. thuhn	37. ʒayp	62. dehz	87. mawʒ
13. geed	38. theez	63. mihth	88. fahd
14. kihp	39. fihn	64. faz	89. dawk
15. zahp	40. mehf	65. kaʒ	90. fawth
16. kayz	41. keem	66. nahʒ	91. gihz
17. pam	42. ʒehth	67. mahm	92. gawp
18. payʒ	43. nehk	68. kehn	93. neep
19. naym	44. ʒawm	69. goom	94. guhʒ
20. mayn	45. feeʒ	70. doon	95. zayk
21. deeg	46. peen	71. pihk	96. zeeth
22. ʒahn	47. thag	72. ʒak	97. zuhz
23. thahth	48. kahg	73. moog	98. gath
24. zawn	49. thihm	74. zooʒ	99. pehp
25. thahk	50. ʒohʒ	75. ʒooz	100. nuhth

Egan (1948).
Note: See Table 5–33 for symbols.

TABLE 5–5. NONSENSE-SYLLABLE LIST NO. 2

1. neen	26. nehp	51. theeg	76. fawf
2. nahz	27. mood	52. thuhm	77. puhk
3. maym	28. ʒoog	53. fohk	78. fayd
4. kaz	29. poof	54. zohʒ	79. muhtb
5. dehg	30. nihk	55. duhʒ	80. zeef
6. man	31. kook	56. dath	81. fehʒ
7. geeth	32. zahn	57. mahʒ	82. thawk
8. gawn	33. mawz	58. mihf	83. kahd
9. thad	34. gehd	59. ʒahm	84. meep
10. nohd	35. kehm	60. ʒayn	85. nayʒ
11. fag	36. ʒihth	61. thayth	86. gohm
12. dayf	37. mehk	62. mohg	87. payz
13. dawp	38. gihg	63. zuhg	88. zooz
14. dihz	39. fihm	64. gaf	89. kihn
15. thehz	40. guhz	65. ʒap	90. doom
16. thahf	41. nawg	66. ʒuhd	91. peem
17. kahf	42. gooʒ	67. gahp	92. kawth
18. zihd	43. fuhn	68. fahth	93. gayk
19. pawd	44. ʒohz	69. pehn	94. zawm
20. ʒeek	45. dohn	70. pahg	95. ʒehf
21. dahk	46. ʒawʒ	71. foop	96. deed
22. feez	47. nuhg	72. zehth	97. nooth
23. thihʒ	48. kayg	73. pohth	98. nam
24. keeʒ	49. thoon	74. thohp	99. kuhp
25. zayp	50. pihp	75. paʒ	100. zak

Egan (1948).
Note: See Table 5–33 for symbols.

TABLE 5–6. PHONETICALLY BALANCED (PB) WORD LIST NO. 1

1. smile	14. box	27. are	40. fuss
2. strife	15. deed	28. cleanse	41. folk
3. pest	16. feast	29. clove	42. bar
4 end	17. hunt	30. crash	43. dike
5. heap	18. grove	31. hive	44. such
6. toe	19. bad	32. bask	45. wheat
7. hid	20. mange	33. plush	46. nook
8. creed	21. rub	34. rag	47. pan
9. rat	22. slip	35. ford	48. death
10. no	23. use (yews)	36. rise	49. pants
11. there	24. is	37. dish	50. cane
12. then	25. not	38. fraud	
13. fern	26. pile	39. ride	

Egan (1948).

TABLE 5–7. PHONETICALLY BALANCED (PB) WORD LIST NO. 2

1. gill	14. dab	27. mute	40. start
2. suck	15. earl	28. rib	41. bounce
3. perk	16. bean	29. awe	42. bud
4. fate	17. nut	30. trash	43. frog
5. five	18. ways	31. corpse	44. quart
6. need	19. wish	32. bait	45. rap
7. pick	20. pit	33. job	46. charge
8. log	21. cloud	34. hit	47. sludge
9. nab	22. scythe	35. hock	48. tang
10. else	23. blush	36. niece	49. them
11. gloss	24. shoe	37. tan	50. vamp
12. hire	25. snuff	38. vast	
13. bought	26. moose	39. our	

Egan (1948).

TABLE 5–8. MODIFIED RHYME TEST (MRT) TALKER'S LIST

Form 1

List A

#———you will mark the———please.

1. rang	14. dig	27. sud	40. led
2. hark	15. book	28. fill	41. sold
3. pool	16. sad	29. dent	42. bun
4. tab	17. sun	30. pass	43. seep
5. sing	18. page	31. teach	44. pin
6. map	19. hot	32. duck	45. meat
7. pun	20. kick	33. bean	46. fame
8. top	21. kit	34. way	47. sip
9. nest	22. kith	35. ten	48. bat
10. cut	23. foil	36. raw	49. came
11. sake	24. jig	37. late	50. rave
12. bust	25. peace	38. gale	
13. hear	26. pill	39. will	

House et al. (1965).

TABLE 5-9. MODIFIED RHYME TEST (MRT) LISTENER'S ANSWER SHEET

Form 1

	1.		2.		3.		4.		5.
fang	bang	mark	bark	peel	keel	tang	tab	sick	sit
rang	hang	park	hark	feel	eel	tam	tap	sing	sin
gang	sang	lark	dark	reel	heel	tack	tan	sill	sip
	6.		7.		8.		9.		10.
mass	map	pup	pug	hop	pop	best	west	cuff	cup
mad	man	putt	puff	top	cop	nest	rest	cud	cut
mat	math	pun	pub	shop	mop	test	vest	cub	cuss
	11.		12.		13.		14.		15.
sale	sake	dust	rust	heave	heal	dim	din	took	look
safe	save	just	gust	heath	heap	did	dig	cook	hook
sane	same	bust	must	hear	heat	dip	dill	book	shook
	16.		17.		18.		19.		20.
sap	sat	gun	run	page	pale	got	hot	tick	wick
sag	sass	bun	nun	pane	pay	tot	pot	pick	sick
sack	sad	sun	fun	pave	pace	lot	not	kick	lick
	21.		22.		23.		24.		25.
wit	fit	kith	kit	foil	oil	fig	rig	peach	peas
sit	hit	kiss	kid	coil	toil	pig	wig	peal	peak
bit	kit	king	kill	soil	boil	big	jig	peat	peace
	26.		27.		28.		29.		30.
pill	pip	sup	sung	fizz	fit	bent	tent	pat	pang
pig	pin	sun	sum	fill	fib	went	dent	pass	pan
pit	pick	sud	sub	fig	fin	sent	rent	pad	path
	31.		32.		33.		34.		35.
teach	tear	dud	dun	beak	beam	way	say	then	hen
teak	team	dub	dull	beat	bead	may	day	pen	men
teal	tease	dug	duck	beach	bean	gay	pay	ten	den
	36.		37.		38.		39.		40.
paw	saw	lane	lace	pale	tale	till	bill	bed	wed
thaw	law	lake	lay	bale	gale	fill	kill	fed	led
jaw	raw	lame	late	male	sale	hill	will	red	shed
	41.		42.		43.		44.		45.
hold	gold	bun	buff	seed	seem	sin	tin	neat	heat
fold	cold	bug	buck	seep	seen	win	din	beat	meat
sold	told	but	bus	seethe	seek	fin	pin	seat	feat
	46.		47.		48.		49.		50.
fame	name	sip	rip	bath	back	cake	cape	race	rate
came	same	hip	tip	ban	bad	case	cane	rake	ray
game	tame	lip	dip	bass	bat	cave	came	raze	rave

House et al. (1965).

TABLE 5-10. MODIFIED RHYME TEST (MRT) TALKER'S LIST

Form 2

List A

\# ——You will mark the—— please.

1. top	14. pill	27. meat	40. sad
2. pin	15. ten	28. came	41. sun
3. bat	16. kick	29. seep	42. pun
4. hot	17. foil	30. hark	43. late
5. cut	18. bust	31. peel	44. sip
6. peace	19. sold	32. gale	45. dig
7. jig	20. map	33. rang	46. nest
8. sake	21. way	34. rave	47. kit
9. fame	22. book	35. will	48. kith
10. duck	23. fill	36. page	49. sing
11. sud	24. dent	37. teach	50. bun
12. tab	25. pass	38. led	
13. raw	26. bean	39. hear	

House et al. (1965).

TABLE 5–11. MODIFIED RHYME TEST (MRT) LISTENER'S ANSWER SHEET

Form 2

	1.		2.		3.		4.		5.
mop	top	din	sin	back	bath	tot	lot	cut	cuff
hop	shop	fin	pin	bass	ban	hot	got	cud	cub
cop	pop	win	tin	bad	bat	pot	not	cuss	cup
	6.		7.		8.		9.		10.
peas	peace	jig	big	safe	same	name	same	dun	dud
peach	peal	rig	pig	save	sane	game	fame	dub	dug
peak	peat	wig	fig	sale	sake	came	tame	duck	dull
	11.		12.		13.		14.		15.
sup	sud	tam	tang	law	saw	pill	pip	ten	hen
sun	sung	tap	tab	raw	paw	pin	pick	den	pen
sub	sum	tan	tack	jaw	thaw	pit	pig	then	men
	16.		17.		18.		19.		20.
sick	pick	coil	oil	bust	dust	fold	hold	mad	mass
lick	tick	toil	foil	rust	must	cold	sold	mat	map
wick	kick	soil	boil	gust	just	gold	told	math	man
	21.		22.		23.		24.		25.
say	gay	shook	look	fit	fill	tent	sent	pass	pat
pay	way	book	took	fig	fizz	bent	went	pad	pang
may	day	cook	hook	fib	fin	dent	rent	path	pan
	26.		27.		28.		29.		30.
beat	beak	neat	heat	cave	cape	seed	seethe	dark	hark
beam	bean	beat	meat	came	cane	seek	seep	bark	park
beach	bead	feat	seat	case	cake	seen	seem	mark	lark
	31.		32.		33.		34.		35.
feel	peel	pale	tale	fang	gang	rave	rate	bill	will
heel	eel	gale	male	bang	sang	raze	race	kill	hill
keel	reel	bale	sale	hang	rang	ray	rake	till	fill
	36.		37.		38.		39.		40.
pane	pay	team	teak	red	led	heat	heave	sag	sap
pale	pave	tear	teal	fed	wed	heath	heal	sass	sat
page	pace	tease	teach	bed	shed	hear	heap	sad	sack
	41.		42.		43.		44.		45.
fun	nun	pun	pup	lane	lace	tip	lip	dill	dig
gun	run	puff	putt	lame	late	sip	dip	din	dip
sun	bun	pub	pug	lake	lay	rip	hip	dim	did
	46.		47.		48.		49.		50.
vest	best	fit	wit	kid	king	sit	sin	but	bug
test	rest	kit	sit	kill	kit	sip	sing	bus	bun
nest	west	bit	hit	kiss	kith	sill	sick	buff	buck

House et al. (1965).

TABLE 5–12. MODIFIED RHYME TEST (MRT) TALKER'S LIST

Form 3

List A

#———You will mark the———please.

1. sing	7. foil	13. sold	19. fame	25. bun	31. peace	37. teach	43. bean	49. bat
2. book	8. bust	14. map	20. duck	26. hear	32. way	38. sud	44. rang	50. hot
3. nest	9. jig	15. gale	21. rave	27. sad	33. ten	39. pill	45. seep	
4. kith	10. sake	16. raw	22. will	28. sun	34. meat	40. led	46. hark	
5. pun	11. kit	17. dent	23. pass	29. kick	35. sip	41. top	47. pin	
6. fill	12. came	18. page	24. peel	30. cut	36. dig	42. late	48. tab	

House et al (1965).

TABLE 5–13. MODIFIED RHYME TEST (MRT) LISTENER'S ANSWER SHEET

Form 3

	1.		2.		3.		4.		5.
sing	sit	look	shook	vest	rest	kill	kid	putt	puff
sin	sill	cook	took	nest	test	kit	king	pub	pun
sip	sick	hook	book	best	west	kith	kiss	pup	pug
	6.		7.		8.		9.		10.
fin	fig	toil	boil	rust	must	rig	pig	sane	save
fit	fib	foil	soil	just	gust	wig	big	safe	same
fill	fizz	coil	oil	dust	bust	jig	fig	sale	sake
	11.		12.		13.		14.		15.
bit	hit	came	cape	hold	cold	mass	map	sale	pale
sit	wit	cane	cake	fold	gold	math	man	gale	bale
fit	kit	cave	case	told	sold	mad	mat	male	tale
	16.		17.		18.		19.		20.
raw	saw	rent	went	pace	pale	came	game	dub	dull
paw	thaw	dent	sent	page	pay	name	fame	dun	duck
jaw	law	tent	bent	pave	pane	same	tame	dud	dug
	21.		22.		23.		24.		25.
rake	rave	bill	hill	pan	pang	keel	peel	bus	bun
ray	raze	fill	will	pad	pass	reel	eel	buff	buck
rate	race	kill	till	pat	path	feel	heel	bug	but
	26.		27.		28.		29.		30.
heath	heat	sag	sack	gun	nun	tick	pick	cuff	cup
heave	hear	sat	sass	run	sun	sick	wick	cud	cub
heal	heap	sap	sad	bun	fun	lick	kick	cuss	cut
	31.		32.		33.		34.		35.
peace	peak	pay	way	den	pen	seat	beat	dip	hip
peach	peat	gay	may	hen	men	meat	heat	rip	sip
peal	peas	say	day	ten	then	feat	neat	lip	tip
	36.		37.		38.		39.		40.
dip	din	team	teak	sub	sun	pig	pill	fed	red
dim	did	tease	tear	sung	sup	pin	pick	shed	wed
dig	dill	teach	teal	sud	sum	pip	pit	bed	led
	41.		42.		43.		44.		45.
mop	shop	lane	lame	beach	beat	sang	hang	seep	seed
top	hop	lace	lay	bean	beak	gang	bang	seem	seethe
cop	pop	lake	late	bead	beam	rang	fang	seen	seek
	46.		47.		48.		49.		50.
park	dark	pin	din	tab	tang	bath	back	hot	not
mark	bark	sin	tin	tan	tam	bat	ban	tot	got
lark	hark	fin	win	tack	tap	bass	bad	lot	pot

House et al. (1965).

TABLE 5–14. TEST SENTENCE LIST NO. 1

Question	Answer	Question	Answer
1. What do you saw wood with?	Saw	14. What number comes before 12?	11
2. What letter comes after B?	C	15. What day comes before Wednesday?	Tuesday
3. What is the color of coal?	Black	16. How many pennies are there in a dime?	10
4. Which is smaller, a dog or a horse?	Dog	17. Does a horse eat oats or chickens?	Oats
5. What is the opposite of white?	Black	18. What do you spread butter with?	Knife
6. What comes between 2 and 4?	3	19. What number comes before 2?	1
7. How many weeks are there in a month?	4	20. What color is ketchup?	Red
8. What do you hear with?	Ears	21. What is the opposite of young?	Old
9. Does a cat eat bricks or mice?	Mice	22. What month comes after March?	April
10. What do you use to unlock a door?	Key	23. Does an eagle have wings or arms?	Wings
11. What letter comes before D?	C	24. What number comes after 20?	21
12. Do elephants have a hump or a trunk?	Trunk	25. How many wheels does a bicycle have?	2
13. What is the opposite of wet?	Dry		

Egan (1948).

TABLE 5–15. TEST SENTENCE LIST NO. 2

Question	Answer
1. What letter comes after C?	D
2. What is the opposite of narrow?	Broad, Wide
3. Which is higher, a hill or a mountain?	Mountain
4. What is the opposite of tall?	Short
5. What day comes after Tuesday?	Wednesday
6. How many months are there in a year?	12
7. What number comes after 5?	6
8. Does a man wear a hat or a table?	Hat
9. What do you chop wood with?	Axe
10. What letter comes before E?	D
11. What is the color of butter?	Yellow
12. What is the opposite of dry?	Wet
13. What day comes after Thursday?	Friday
14. What country is Moscow in?	Russia
15. What letter comes after B?	C
16. How many toes are there on each foot?	5
17. What do you tell the date by?	Calendar
18. What number comes before 3?	2
19. What is the opposite of love?	Hate
20. What color is the cloth on a pool table?	Green
21. What month comes after June?	July
22. How much is 1 and 8?	9
23. Does an owl lay books or an egg?	Egg
24. What number comes before 20?	19
25. Do palm trees grow in Alaska?	No

Egan (1948).

TABLE 5–16. TEST SENTENCE LIST NO. 3

1. *Deal* the *cards from* the *top*, you *bully*.
2. *Jerk* the *cord*, and *out tumbles* the *gold*.
3. *Slide* the *tray across* the *glass top*.
4. *Heat* the *rod* and *tap* it *against* the *skin*.
5. The *cloud moved* in a *stately way* and was *gone*.
6. *Light maple makes* a *swell room*.
7. *There* were *high jinks* on the *gala day*.
8. *Set* the *piece here* and *say nothing*.
9. *Gay lads make hay* in the *moonshine*.
10. *Drop that notion*, and *be a sport*.
11. *Stale bread* is a *waste* in *pudding too*.
12. The *gray paint clashed* with the *gaudy tints*.
13. The *stiff drink hit* like a *bolt* of *lightning*.
14. *Sell* it *short*, and *lose your money*.
15. *Dull stories make her laugh*.
16. The *clown moulded* his *pale face* into a *smirk*.
17. *Sew* and *knit*, it *helps evenings pass*.
18. *Men* are *but infants* to a *smart girl*.
19. *Dip* the *pail once* and *let* it *settle*.
20. A *stiff cord* will *do* to *fasten* your *shoe*.

Egan (1948).

TABLE 5–17. TEST SENTENCE LIST NO. 4

1. Get *the trust fund* to the *bank early*.
2. A *tight bow* is not *found* in *worn shoe-strings*.
3. A *mean child pulls wings* from *insects*.
4. There is *plenty* of *pork*, but *spring lamb* is *short*.
5. *Jean* is a *man trap* in *silk stockings*.
6. *Can such falsehoods* be *proved true*?
7. The *nude man* was *jailed* by the *angry cop*.
8. *She works* in a *boiled shirt factory*.
9. A *number* of *tunes tell* of *June nights*.
10. *Men dare* not *scorn* a *dab* of *lipstick*.
11. *Choose between* the *high road* and the *low*.
12. A *plea* for *funds seems* to *come again*.
13. He *lent* his *coat* to the *tall gaunt stranger*.
14. There is a *strong chance* it will *happen more* than *once*.
15. The *line* will *march until* they *halt* or *drop*.
16. The *duke rides* the *park* in a *silver coach*.
17. The *donkeys bray* as the *ghosts flit* in the *fields*.
18. As the *boys fought* the *cops brought* the *wagon*.
19. *Greet* the *new guests* and *leave quickly*.
20. *When* the *frost* has *come* it is *time* for *turkey*.

Egan (1948).

The MRT's, lists of which are given in Tables 5–8 through 5–13, differ from the nonsense-syllable and PB-word tests in two respects: (a) the listener must perceive correctly but one phoneme in each test word—either the initial or final consonant; (b) because the listener has access to the printed list of test words, he can check which word he thought was correct. However, the words are grouped into subsets of six words differing only with respect to initial or final consonant, one word of which was spoken by the talker. Although the lists of words in the MRT are not phonetically balanced to represent everyday speech, the MRT test is efficient and useful because it requires perception of consonantal sounds, sounds that are difficult to transmit successfully and are thus more important than vowels to intelligibility. Also, because of the multiple-choice form of the answer sheet, only minimal listener training is required.

If only a few sentence tests are to be made, the sentences are usually drawn from a set compiled for that purpose (Egan, 1948). Extensive testing calls for a large ensemble of sentences. They can be drawn from any reasonably homogenous source, but random selection from the ensemble is important. No sentences should be used twice with the same listener.

Two variations of sentences can be used. One, illustrated in Tables 5–14 and 5–15, is a question that can be answered with a single word or short phrase. The other, illustrated in Tables 5–16 and 5–17, is an ordinary sentence in which certain "key" words have been designated.

5.3.3 Scoring Tests

Score each nonsense syllable correct if all its component sounds are correct. The percentage of the consonants heard correctly is the "percent consonant articulation." The percentage of the vowels heard correctly is called the "percent vowel articulation." The percentage of the syllables heard correctly is called the "percent syllable articulation."

Nonsense syllables are difficult to understand and are, therefore, sensitive to small amounts of noise and distortion. They reveal small differences between systems. For example, the contribution to intelligibility of the speech components above 5000 Hz would be difficult to measure in quiet with sentences. With nonsense syllables, however, one can tell after a reasonable amount of testing that some intelligibility is lost when the components above 5000 Hz are removed from the signal.

If the case of PB-word tests, score each recorded word on its phonetic agreement with the corresponding word on the talker's list. The percentage of the words heard correctly is called the "percent word articulation" or the "percent word intelligibility." Though PB-word tests are less sensitive than nonsense-syllable tests, they do not require as much training time.

The MRT is scored by: (a) counting the number of words correct, and (b) determining the percent words correct according to the following formula:

$$\%\,\text{correct} = (\text{No. right} - \frac{\text{No. wrong}}{5} \times 2); \quad (5\text{–}1)$$

this formula is required to correct the word score for "chance" or guessing made possible by the multiple-choice form of the answer sheet. The less difficult MRT (Kryter and Whitman, 1965), is therefore less sensitive for system evaluation than nonsense-syllable or PB-word tests that contain a greater number of test items. (See Figure 5–5.)

There are two ways of scoring a sentence test. The first uses an experienced scorer (or group of scorers) to judge whether the listener's record indicates that he understood the sentence. The second way is to designate five "key" words in each sentence. These words are then scored.

The "percent sentence intelligibility" is the percentage of the sentences of which the essence was understood or the key words recorded correctly. The two scores usually approximate each other. Because of redundancy, sentence intelligibility tests yield rather high scores, even for noisy or distorted systems.

5.3.4 Interpreting Test Results

In interpreting the results of intelligibility tests, consider the vocabulary. Will numbers, single words, short phrases, or complete sentences actually be used? What percentage of each kind of item must be heard correctly for the system to operate? The test scores can be taken at

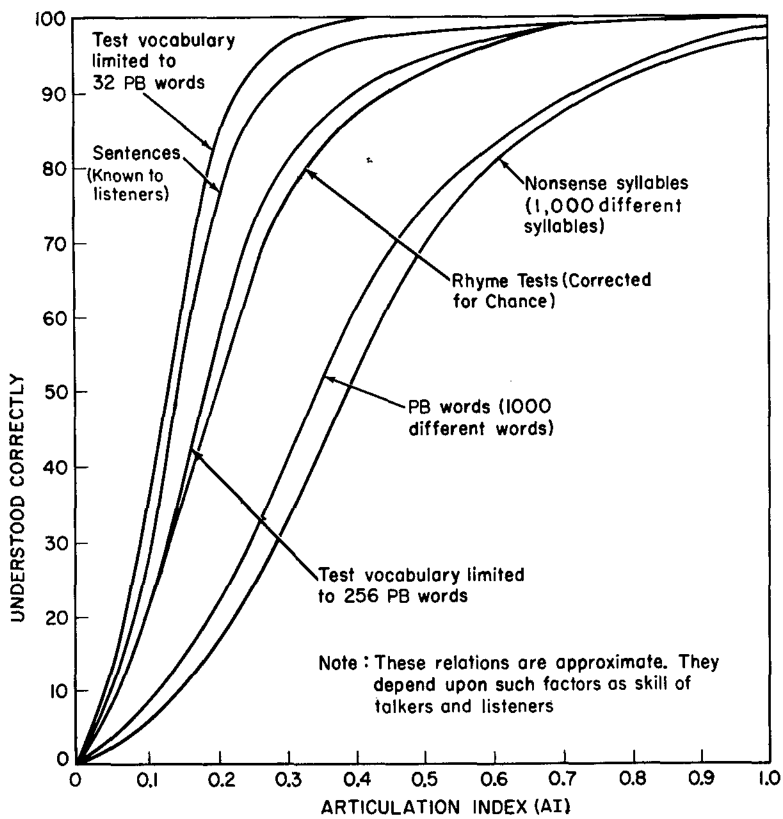


FIGURE 5–5. Relationship between the Articulation Index (AI) and the intelligibility of various types of speech-test materials (Kryter, 1966).

face value if the testing material is the same as that to be transmitted over the system. If not, inferences must be made. Figure 5–5 shows approximate relationships between the Articulation Index and the intelligibility of various types of speech-test materials.

Although simple in concept, intelligibility testing is actually an art with many ramifications. A person rarely obtains usable results immediately. Supervision by experienced persons and study of the published literature is recommended. (Fletcher and Steinberg, 1929; Egan, 1948; USASI 1960; and House et al., 1965.)

5.3.5 Predicting Intelligibility in Communication Systems

Many design decisions are based on calculated predictive measures of intelligibility. Two general methods of measuring are the Articulation Index (AI) (French and Steinberg, 1947) and the speech interference level (SIL). The (SIL) can serve as a rule-of-thumb guide in making decisions regarding face-to-face communication. The Articulation Index is a more precise way of calculating intelligibility. It has two methods of calculation, the 20-band and the weighted one-third and full-octave band. The 20-band is more detailed and accurate.

Communication systems—the articulation index formulation. For high intelligibility, a considerable fraction of total speech bandwidth must be delivered to the listener and the signal-to-noise ratio at his ear must be reasonably high. If the speech peaks are 30 dB or more above the noise throughout the frequency band from 200 to 6100 Hz, the listener will make essentially no errors (AI = 1.00). If the speech peaks are less than 30 dB above the noise in any part of the speech band, he will make some mistakes (AI = 0.50). If the speech peaks are never above the noise (ratio of speech peaks to r.m.s. noise less than 0 dB), the listener will rarely be able to understand anything (AI = 0).

Methods of Computing the Articulation Index

In order to compute AI, one may use the 20-band method, the one-third octave band, and the octave band (preferred frequency (p.f.)) methods. It is desirable to specify which method was used when reporting results in the literature. The octave band method is not as sensitive to variations in the speech and noise spectra as the 20-band or the one-third octave band methods and accordingly is not as precise. It should not be used, for example, in situations where an appreciable fraction of the energy of the masking noise is concentrated in frequency bands that are one octave or less in width; under these conditions the one-third octave band or preferably the 20-band method should be used.

The 20-band method. This method is based upon measurement or estimates of the spectrum level of the speech and noise present in each of 20 contiguous frequency bands that contribute equally to speech intelligibility when the signal level is equal to the noise level in the band.

Step 1. Plot on the worksheet shown in Figure 5–6 the spectrum level of the speech peaks reaching the listener's ear. The mid and cut-off frequencies of 20 frequency bands of equal contributions to speech intelligibility are given in Table 5–18.

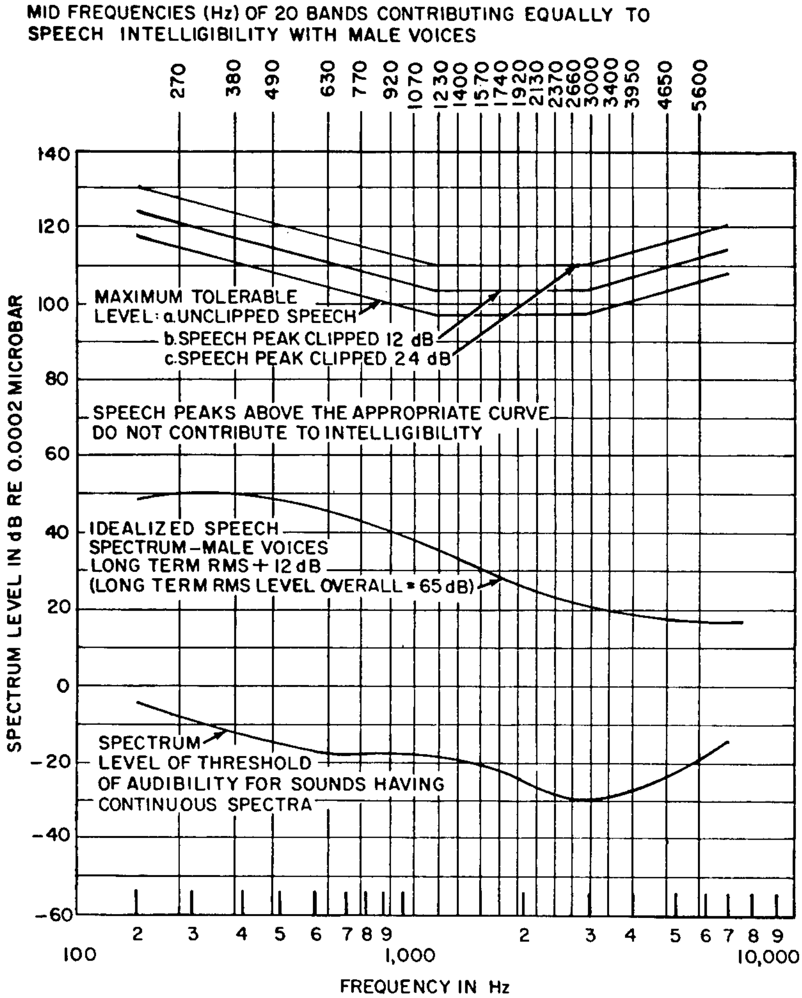


FIGURE 5–6. Worksheet for 20-band method of calculating the Articulation Index (AI) (Kryter, 1966).

TABLE 5–18. TWENTY FREQUENCY BANDS OF EQUAL CONTRIBUTION TO SPEECH INTELLIGIBILITY, MALE VOICES

Band No.	Limits (Hz)	Mid-frequency (Hz)	Band No.	Limits (Hz)	Mid-frequency (Hz)
1	200–330	270	11	1660–1830	1740
2	330–430	380	12	1830–2020	1920
3	430–560	490	13	2020–2240	2130
4	560–700	630	14	2240–2500	2370
5	700–840	770	15	2500–2820	2660
6	840–1000	920	16	2820–3200	3000
7	1000–1150	1070	17	3200–3650	3400
8	1150–1310	1230	18	3650–4250	3950
9	1310–1480	1400	19	4250–5050	4650
10	1480–1660	1570	20	5050–6100	5600

Beranek (1947).

The spectrum level of speech peaks may be approximated by adding together algebraically the following:

(*a*) The frequency response characteristic in dB of the system to be evaluated. The frequency response at each frequency is the difference between the sound pressure level at the listener's ear and that at the talker's microphone at that frequency.

(*b*) The idealized speech spectrum found on the worksheet, Figure 5–6, (1) raised by an amount equal to the difference between the long-term r.m.s. for speech and 65 dB, which is the overall long-term r.m.s. sound pressure level of the idealized speech spectrum of Figure 5–6; and (2) in the case of loudspeaker presentation of speech in a non-free field lowered by a number of decibels as indicated in Table 5–19. The correction given in Table 5–19 is not required for speech presented to listeners via earphones or by a loudspeaker operating in a free field.

For example, suppose the long-term level for speech over a public address system in a room equals 95 dB. Accordingly, adding 26 dB (95 dB minus 65 dB, minus 4 dB from Table 5–19) to the idealized speech spectrum modified by the frequency response characteristic of the system as found in Step 1a above, gives an effective spectrum plot of the speech from the public address system. This correction for effective spectrum of speech from a loudspeaker in a non-free field is based on experiments which show degradation in intelligibility of such speech when presented in the quiet at higher speech levels.

TABLE 5–19. CORRECTION FOR LEVEL OF SPEECH PRESENTED BY A LOUDSPEAKER IN A REVERBERANT OR SEMIREVERBERANT ROOM

OA SPL of speech (dB)	Amount to be subtracted from speech level (dB)
85	0
90	2
95	4
100	7
105	11
110	15
115	19
120	23
125	27
130	30

Kryter (1962).

Step 2. Plot on the worksheet the corrected spectrum level of steady-state noise reaching the ear of the listener. Combine ambient noise levels and noise from the system.

The masking effectiveness of a noise increases at a faster than normal rate when the band sensation level of the noise exceeds 80 dB. This increased masking is figured in the calculation of AI by adding a decibel correction to the sound pressure level (SPL) of the noise.

Whenever the band sensation level of the noise exceeds 80 dB at any intersection of the center frequency of a band (the vertical lines of Figure 5–6), the SPL of the noise at that point on the frequency scale should be increased by an appropriate amount obtained from Table 5–20.

TABLE 5-20. CORRECTIONS FOR NONLINEAR GROWTH OF MASKING

Band sensation level (dB)	No. of dB correction to be added to SPL of noise (dB)
80	0
85	1
90	2
95	3
100	4
105	5
110	6
115	7
120	8
125	9
130	10
135	11
140	12
145	13
150	14

French and Steinberg (1947).

The resulting curve is called the corrected noise spectrum.

Determine the band sensation level of the noise by subtracting the spectrum level of the threshold of audibility values given on Figure 5-6 from the spectrum level of the noise.

Step 3. Plot on the worksheet, Figure 5-6, the effective masking spectrum of the noise as follows:

(*a*) Locate the extreme right-hand point at which a horizontal line, 3 dB below the maximum of the noise spectrum or corrected noise spectrum, whichever is higher, intersects the noise spectrum. Call this the "starting point."

(*b*) Drop vertically 57 dB from the starting point. Then draw a line to the left and upward with a slope of 10 dB per octave. This line is the low-frequency part of the masking spectrum.

(*c*) From the starting point draw a horizontal line to the right and then downward. The length of the horizontal portion of this line and the slope of the downward portion of this line depend upon the frequency of the starting point and upon the maximum spectrum level of the noise as shown in Table 5-21. This line represents the high-frequency part of the masking spectrum.

Step 4. Determine at the mid-frequency of each of the 20 frequency bands indicated on the worksheet, Figure 5-6, the difference in decibels between the spectrum level of the speech peaks and that of the noise spectrum or the masking spectrum, whichever is higher. Whenever the difference is zero or less, assign a value for zero to that difference; whenever the speech exceeds the noise or masking spectrum by 30 or more decibels, assign a value of 30 to that difference.

NOTE 1. The threshold of audibility curve on the work sheet, Figure 5-6, is to be considered as the minimum noise spectrum whenever the threshold curve exceeds the noise or masking spectrum.

NOTE 2. If the speech peak curve exceeds the maximum tolerable level, then the maximum tolerance level itself is to be considered the maximum tolerable level.

Step 5. Add the 20 differences found in Step 4 and divide by 600. The resulting number is the AI for that particular speech system operating under the noise conditions and for the speech level specified.

An illustrative example of the calculation of an AI by the 20-band method is given in Figure 5-7.

The one-third octave band and the octave-band methods. These methods are derived from the 20-band method but require measurements or estimates of the speech and noise present either in certain one-third octave bands or in certain full-octave bands.

Step 1. Plot on the appropriate worksheets, Figures 5-8, 5-9, or 5-10, depending on the filters used, the band levels of the speech peaks reaching the listener's ear.

NOTE. The center and cut-off frequencies on the one-third octave band and the octave band filters are given in Tables 5-22, 5-23, and 5-24. (See USASI, 1953 and 1960.)

The band levels of the speech peaks may be approximated by adding together algebraically:

(*a*) The frequency response characteristics of the system under evaluation. The frequency response at each center frequency is the difference between the band sound pressure level at the listener's ear and the band sound pressure level at the microphone of the talker at that frequency.

(*b*) The idealized band levels on the worksheet, Figures 5-8, 5-9, or 5-10, adjusted in accordance with the difference between the long-term r.m.s. level for speech as measured or estimated and 65 dB which is the overall long-term r.m.s. sound pressure level of the idealized

TABLE 5–21. HIGH-FREQUENCY PART OF MASKING SPECTRUM—UPWARD SPREAD OF MASKING

A: Draw from starting point horizontal line to right for this number of Hz.

B: Draw from right-hand end of horizontal line a downward line that has this slope in dB per octave.

Maximum spectrum level or corrected spectrum level, in dB whichever is higher, of noise re 0.0002 microbar (dB).	Frequency of starting point located in step 3									
	50–800 Hz		800–1600 Hz		1600–2400 Hz		2400–3200 Hz		3200–5200 Hz	
	A (Hz)	B (dB)	A (Hz)	B (dB)	A (Hz)	B (dB)	A (Hz)	B (dB)	A (Hz)	B (dB)
96– —	250	10	500	8	1000	5	1500	3	3000	0
86– 95	200	15	500	13	1000	10	1500	5	3000	0
76– 85	200	20	400	18	800	15	1500	10	3000	0
66– 75	150	25	250	23	500	20	1000	15	2000	5
56– 65	75	35	150	30	300	25	500	25	800	20
46– 55	50	45	100	40	200	35	200	40	200	40

Carter and Kryter (1962).

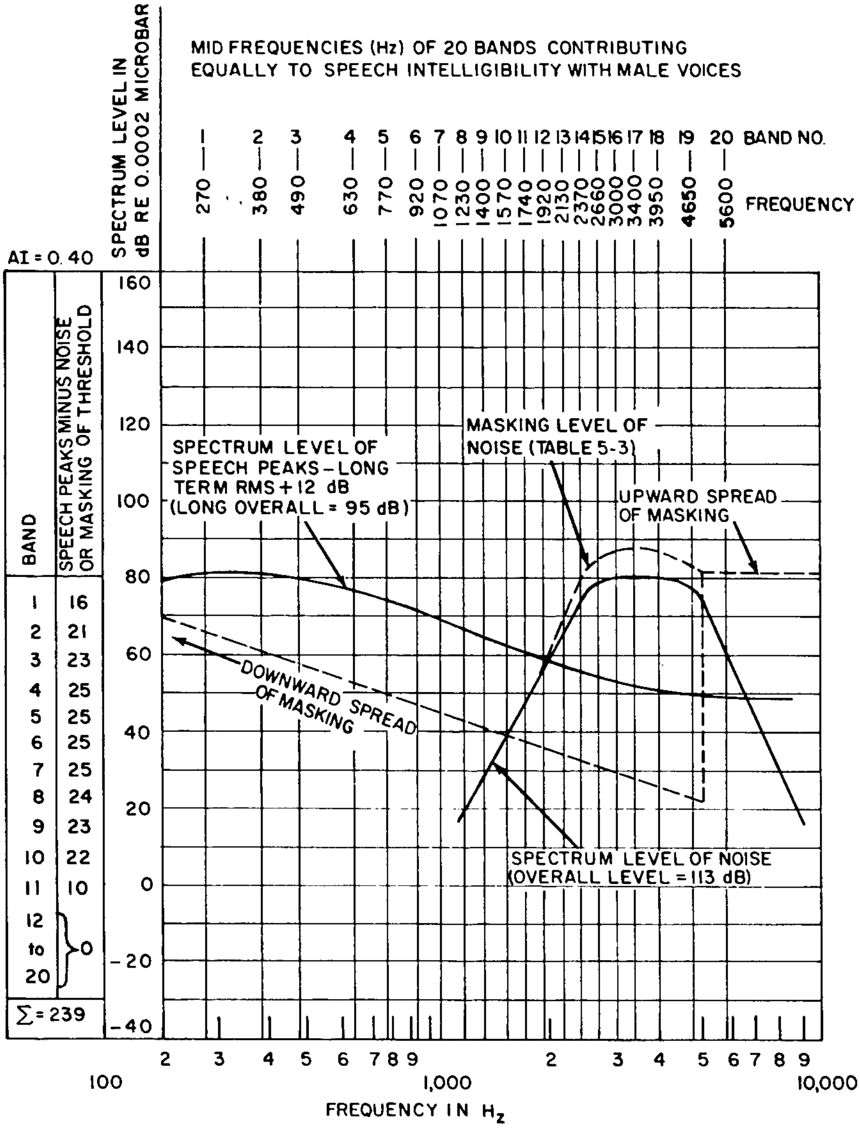


FIGURE 5–7. Example of the calculation of an Articulation Index (AI) by the 20-band method (Kryter, 1966).

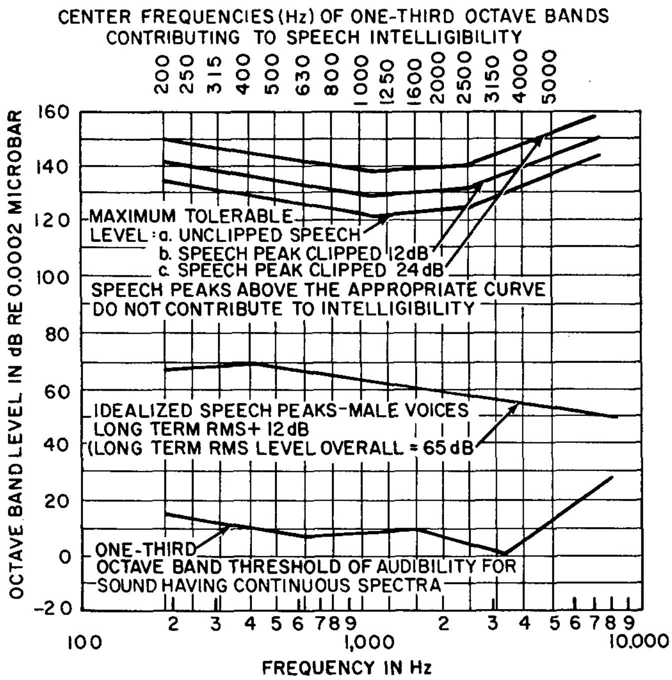


FIGURE 5–8. Worksheet for the one-third octave band method of calculating the Articulation Index (AI) (Kryter, 1966).

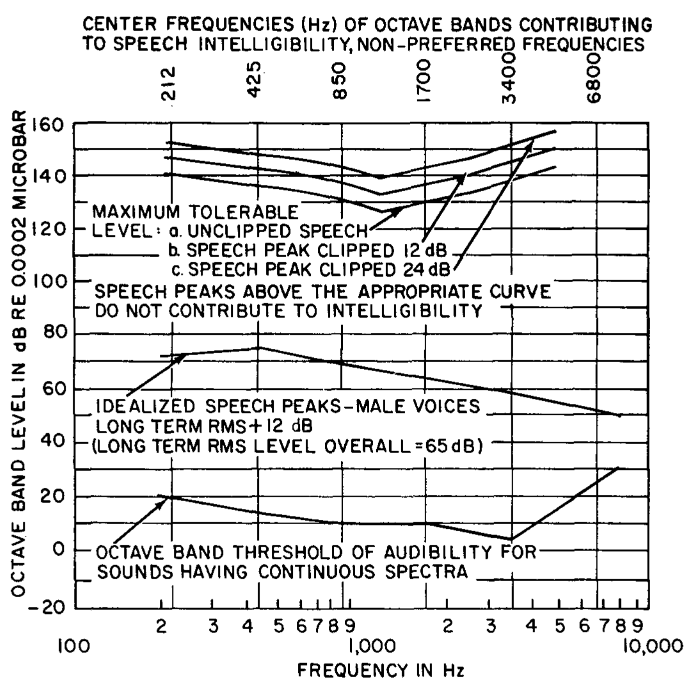


FIGURE 5–9. Worksheet for the octave band method (nonpreferred frequencies) of calculating the Articulation Index (AI) (Kryter, 1966).

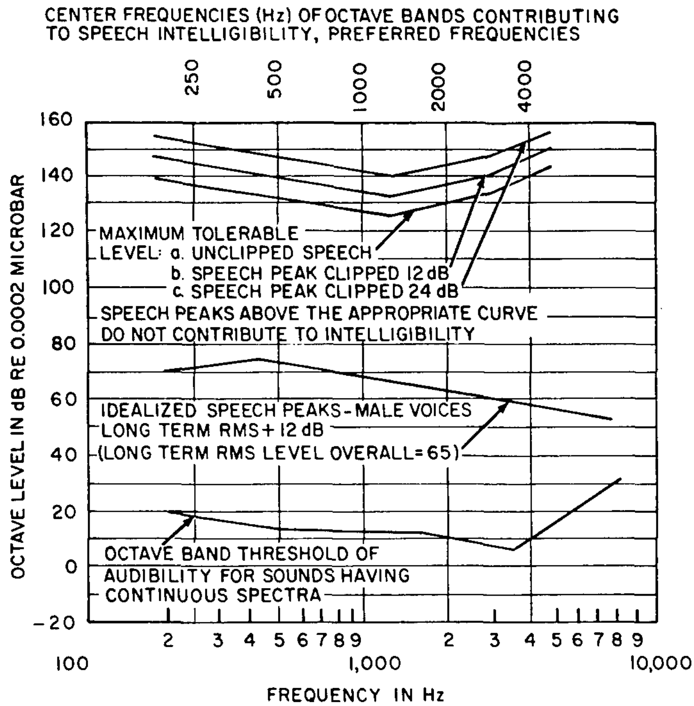


FIGURE 5–10. Worksheet for the octave band method (preferred frequencies) of calculating the Articulation Index (AI) (Kryter, 1966).

TABLE 5–22. ONE-THIRD OCTAVE BANDS

Col. 1 One-third octave band (Hz)	Center frequency (Hz)	Col. 2 Speech peak-to-noise difference in decibels (from Step 3)	Col. 3 Weight	Col. 4 Col. 2×Col. 3
180–224	200	______	0.0004	______
224–280	250	______	0.0010	______
280–355	315	______	0.0010	______
355–450	400	______	0.0014	______
450–560	500	______	0.0014	______
560–710	630	______	0.0020	______
710–900	800	______	0.0020	______
900–1120	1000	______	0.0024	______
1120–1400	1250	______	0.0030	______
1400–1790	1600	______	0.0037	______
1790–2240	2000	______	0.0038	______
2240–2800	2500	______	0.0034	______
2800–3530	3150	______	0.0034	______
3530–4480	4000	______	0.0024	______
4480–5600	5000	______	0.0020	______

Kryter (1966).

TABLE 5–23. OCTAVE BANDS—NON PREFERRED FREQUENCIES

Col. 1 Octave band (Hz)	Center frequency (Hz)	Col. 2 Speech peak-to-noise difference in decibels (from Step 3)	Col. 3 Weight	Col. 4 Col. 2 × Col. 3
150– 300	212	______	0.0017	______
300– 600	425	______	0.0040	______
600–1200	850	______	0.0065	______
1200–2400	1700	______	0.0107	______
2400–4800	3400	______	0.0084	______
4800–9600	6800	______	0.0020	______

Kryter (1966).

TABLE 5–24. OCTAVE BANDS—PREFERRED FREQUENCIES

Col. 1 Octave band (Hz)	Center frequency (Hz)	Col. 2 Speech peak-to-noise difference in decibels (from Step 3)	Col. 3 Weight	Col. 4 Col. 2 × Col. 3
180– 355	250	______	0.0024	______
355– 710	500	______	0.0048	______
710–1400	1000	______	0.0074	______
1400–2800	2000	______	0.0109	______
2800–5600	4000	______	0.0078	______

Kryter (1966).

speech band spectrum of Figures 5–8, 5–9, and 5–10.

Step 2. Plot on the worksheet, Figures 5–8, 5–9, or 5–10, the band levels of steady-state noise reaching the ear of the listener. Again, ambient noise in the listener's environment is to be combined with the noise, if any, reaching the listener via the speech transmission system.

NOTE. Whenever the band sensation level of the sound exceeds 84 dB at any intersection of the center frequency of a band contributing to intelligibility (the vertical lines on Figures 5–8, 5–9, and 5–10), the speech and noise spectra should be converted to spectrum level values and the results plotted on Figure 5–6. The AI may then be calculated in accordance with the procedures prescribed for the 20-band method above. The purposes of converting to spectrum level is to be able to consider AI calculation of nonlinear and spread-of-masking effects. These become significant only at band sensation levels above 84 dB. The band sensation level of a band of frequencies, within a broadband sound, can be determined by subtracting from the band level of the sound the threshold of audibility values given in Figures 5–8, 5–9, and 5–10.

Step 3. Determine at the center frequency of each of the bands indicated on the worksheet, Figures 5–8, 5–9, or 5–10, the difference in decibels between peak level and that of the noise band level or the masking level of the noise, whichever is higher. Whenever the difference is zero or less, assign a value of zero to that difference; whenever the speech exceeds the noise band level or the masking level of the noise by 30 or more decibels, assign a value of 30 to that difference.

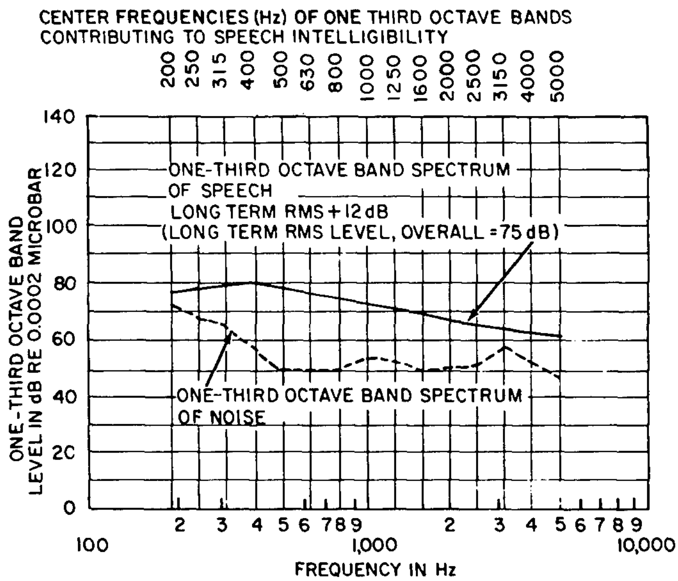
NOTE 1. The threshold of audibility curve, in Figures 5–8, 5–9, or 5–10, is the minimum noise level whenever the threshold curve exceeds the noise band level or the masking level of the noise.

NOTE 2. If the speech peak curve exceeds the maximum tolerable level, then the maximum level itself (not the higher value) is taken as the speech peak level.

Step 4. Multiply the values for the respective bands found in Step 3 according to weighting factors in Column 3, Table 5–22.

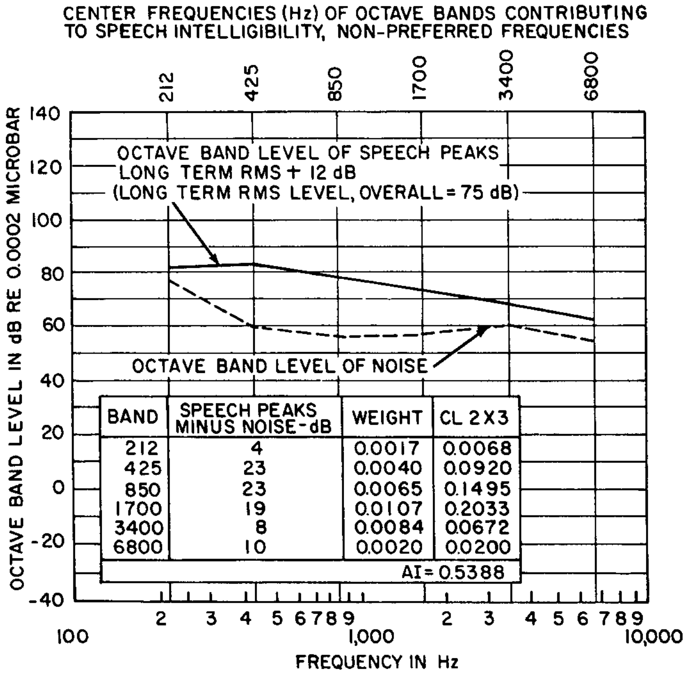
Step 5. Add Column 4 on Tables 5–22, 5–23, or 5–24. The resulting number is the AI for that particular speech system operating under the noise conditions and for the speech level specified.

Examples of the calculation of an AI by the one-third octave and octave-band methods are given in Figures 5–11, 5–12, and 5–13.



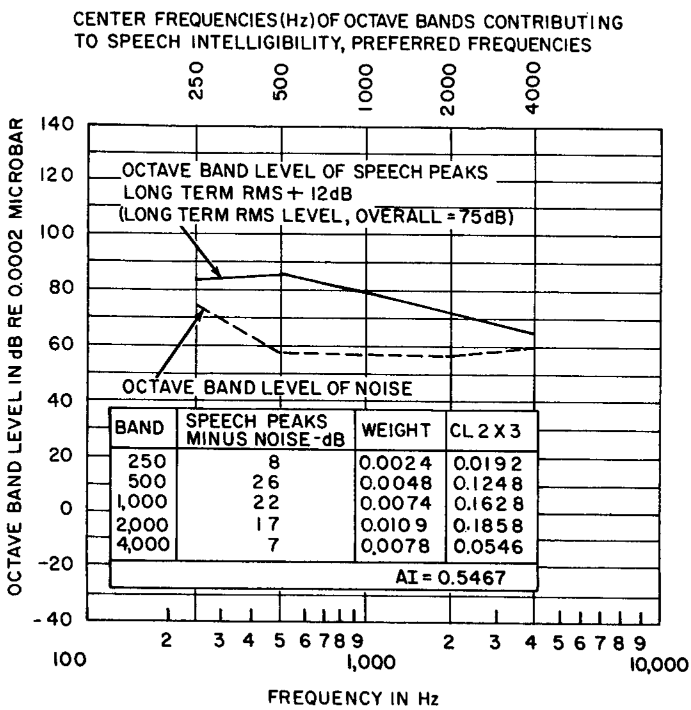
BAND	SPEECH PEAKS MINUS NOISE-dB	WEIGHT	CL 2 X 3
200	4	0.0004	0.0016
250	10	0.0010	0.0100
315	13	0.0010	0.0130
400	24	0.0014	0.0336
500	26	0.0014	0.0364
630	26	0.0020	0.0520
800	24	0.0020	0.0480
1000	21	0.0024	0.0504
1250	18	0.0030	0.0540
1600	18	0.0037	0.0666
2000	15	0.0038	0.0570
2500	15	0.0034	0.0510
3150	6	0.0034	0.0204
4000	8	0.0024	0.0192
5000	12	0.0020	0.0240
		AI = 0.5372	

FIGURE 5-11. Example of the calculation of an Articulation Index (AI) by the one-third octave band method (Kryter, 1966).



BAND	SPEECH PEAKS MINUS NOISE-dB	WEIGHT	CL 2X3
212	4	0.0017	0.0068
425	23	0.0040	0.0920
850	23	0.0065	0.1495
1700	19	0.0107	0.2033
3400	8	0.0084	0.0672
6800	10	0.0020	0.0200
			AI = 0.5388

FIGURE 5–12. Example of the calculation of an Articulation Index (AI) by the octave band method (non-preferred frequencies) (Kryter, 1966).



BAND	SPEECH PEAKS MINUS NOISE-dB	WEIGHT	CL 2X3
250	8	0.0024	0.0192
500	26	0.0048	0.1248
1,000	22	0.0074	0.1628
2,000	17	0.0109	0.1858
4,000	7	0.0078	0.0546
			AI = 0.5467

FIGURE 5–13. Example of the calculation of an Articulation Index (AI) by the octave band method (preferred frequencies) (Kryter, 1966).

NOTE. The AI's calculated for the same speech signal and masking noise as given in Figures 5–11, 5–12 and 5–13 are as follows:
One-Third Octave Band Method AI = 0.54.
Octave Band Method, Non-preferred Frequencies AI = 0.54.
Octave Band Method, Preferred Frequencies AI = 0.55

As previously mentioned, the octave band methods must be used with some caution, because of their relative insensitivity either to sharp changes in spectrum shape of the speech signal or to a masking noise.

Effects of various factors on AI. Extreme distortion and stress imposed upon speech during its transmission influence intelligibility test scores in a way not fully understood or quantifiable at present. Some factors can be evaluated using the AI principle. Among these are:

1. Masking by Steady-State Noise. AI's adequately predict the effects of wideband, continuous spectrum noise and the effects of bands of noise wider than 200 Hz in the frequency range from about 200 Hz to 6000 Hz and for sound pressure levels up to approximately 125 dB.
2. Masking by Nonsteady-State Noise. The duty cycle, or the fraction of the time that a masking noise is on, affects speech intelligibility. Whenever the noise is not steady state and the on-off duty cycle is known, calculate the AI as though the noise were steady state and then apply the correction as indicated in Figure 5–14. This procedure may be followed when the noise falls during the "off" periods to a level at least 20 dB below the level of the noise during the "on" periods.
3. Rate of Interruption of the Noise. The effective AI found in accordance with Figure 5–14 for a communications system in which a noise having a definite on-off duty cycle is present should be further adjusted in accordance with the functions shown in Figure 5–15. The vertical ordinate gives the effective AI to be expected for a given parameter when the masking noise is interrupted at the rates shown on the abscissa.
4. Frequency Distortion of the Speech Signal. Frequency distortion, i.e., transmission of the signal with unequal gain as a function of frequency, usually affects the intelligibility of speech. These effects are accounted for by the AI, provided that the unequal emphasis is applied to either high-, medium-, or low-frequency components of the speech signal.

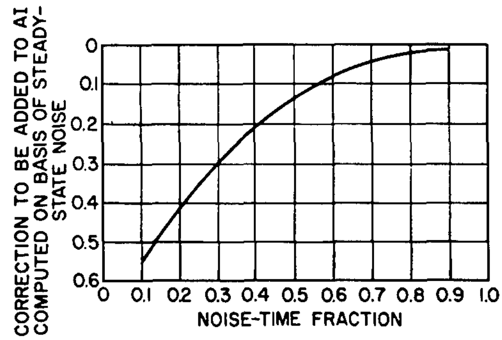


FIGURE 5–14. The ordinate shows a correction to be applied to the Articulation Index (AI) computed on the assumption that a masking noise is steady state for various noise-time fractions. The corrected AI cannot exceed 1.0 (Kryter, 1966).

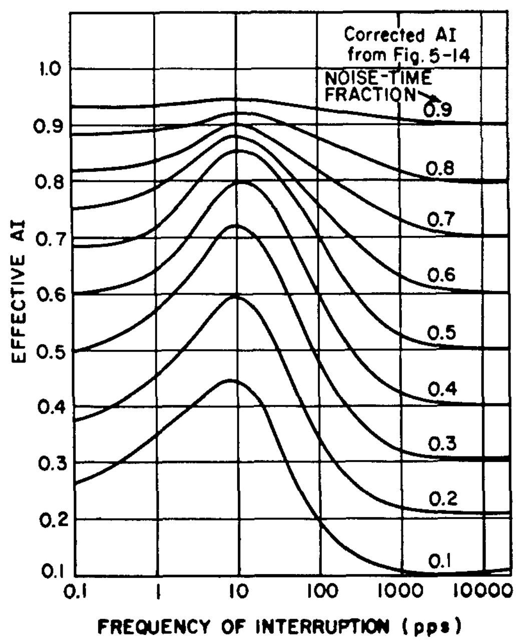


FIGURE 5–15. Showing the effective Articulation Index (AI) as a function of the frequency with which a masking noise is interrupted. The parameter of the curves is the corrected AI calculated on the assumption that the masking noise is steady state and then adjusted according to Fig. 5–14 or the fraction of the time the noise is on (Kryter, 1966).

However, the AI will not provide a valid means for estimating the intelligibility of speech that has a very irregular long-term spectrum,

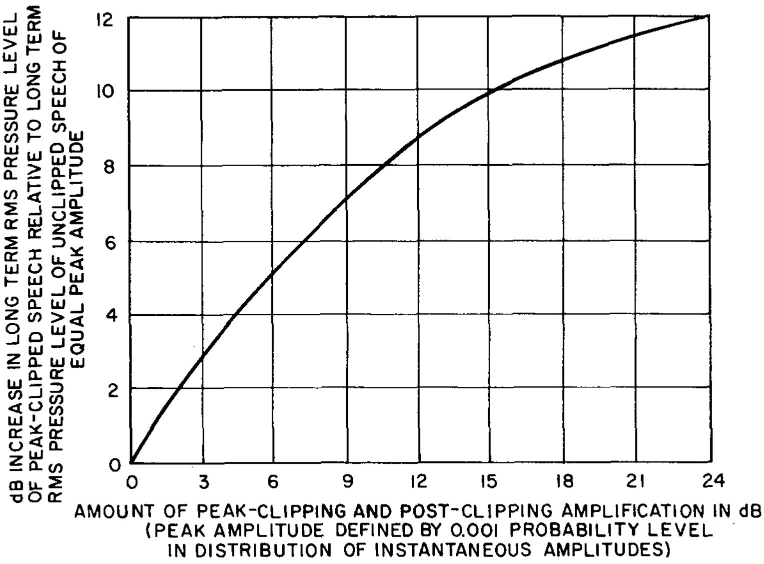


FIGURE 5–16. Showing the increase in r.m.s. speech power as a function of clipping when clipped level is raised to clipping reference level (Licklider, 1946, and Wathen-Dunn and Lipke, 1958).

i.e., a spectrum that goes through a series of peaks and valleys, the slopes of which, on the average, exceed 18 dB per octave.

5. Amplitude Distortion of the Speech Signal. The effects upon intelligibility of sharp symmetrical peak clipping can be estimated by use of a computed AI. The steps to be taken follow:

(*a*) Determine from Figure 5–16 the increase in the level of the long-term r.m.s. of speech as the result of the particular amount of peak-clipping and post-clipping amplification present in a system.

(*b*) Add the result of Step 5a above to the level of the speech peaks (unclipped speech +12 dB) that would reach the listener's ears without the peak-clipping and comparable post-clipping amplification. Post-clipping amplification is the amount added to the system to achieve peak-to-peak amplitudes equal to the peak amplitudes that would be achieved without peak clipping. If the post-clipping amplification does not equal in decibels the amount of peak clipping applied to the speech signal, the increase in the long-term r.m.s. found in Step 1 should be reduced by a factor equal to the ratio between peak-clipping and post-clipping amplification in decibels.

(*c*) Plot the result of Step b on an AI work sheet and proceed to compute AI according to previous instructions. Note that the maximum tolerable level indicated on the AI work sheets for the speech is higher for peak-clipped than for non-clipped.

NOTE. In general, peak-clipping should be used only when the speech signal is relatively free of noise prior to peak-clipping but is heard in the presence of noise or has noise mixed with it after being peak clipped.

6. Reverberation in a Room Will Cause a Decrease in Intelligibility. The amount of degradation will be a function of the reverberation time of the room. Reverberation time is the time required for a steady-state sound to decrease 60 dB after the source is stopped.

Refer to Figure 5–17 to correct the AI found for a given speech communication system when the reverberation time is known.

7. Vocal Effort. Very weak or very intense vocal efforts by a talker will tend to reduce speech intelligibility. A given AI value can be expected to be accurate when the vocal effort of the talker is maintained consistently between a measured long-term r.m.s. sound pressure level of 50 dB to 85 dB, measured 1 m. from the talker's lips. In systems where very strong or very weak vocal efforts are used, the effective level, rather than the measured speech level, should be plotted on the AI work sheets. The relation between actual and effective speech levels is shown in Figure 5–18.

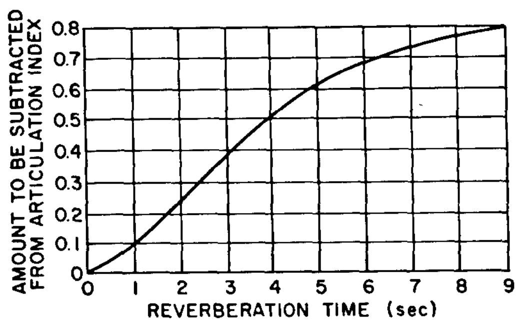


FIGURE 5–17. The ordinate shows the amount to be subtracted from the Articulation Index (AI). AI cannot be less than 0.0 Corrected AI's less than 0.0 are to be called 0.0 (Knudsen and Harris, 1950).

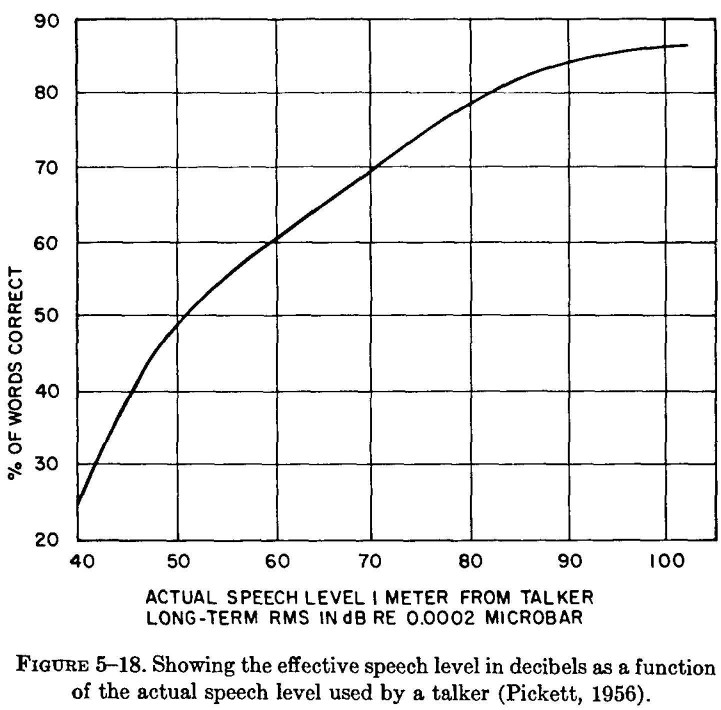


FIGURE 5–18. Showing the effective speech level in decibels as a function of the actual speech level used by a talker (Pickett, 1956).

8. Visual Cues. Visual cues obtained from observing the talker's lips or face can contribute to speech intelligibility, particularly in the presence of noise. However, an AI can be modified or adjusted in accordance with Figure 5–19 into an "effective AI" to reflect the effect of visual cues upon speech intelligibility.

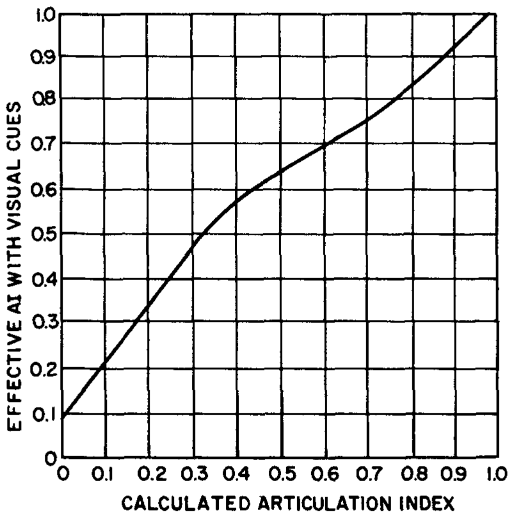


FIGURE 5–19. Relation between calculated AI and effective AI for a communication system wherein the listener can see the lips and face of the talker (Sumby and Pollack, 1954).

The increment in intelligibility contributed by visual cues is a function of the prevailing speech-to-noise ratio; if the ratio is high, the listeners hear the words clearly and therefore do not have to take advantage of the cues provided by lip reading; if it is low, they need to use the visual cues. (See Figure 5–20.)

The effective AI for a given face-to-face communication situation in which normal visual cues are available can be estimated by the procedure given above.

5.3.6 Relation of Articulation Index to Speech Intelligibility

AI's may be converted to estimated speech intelligibility scores using Figure 5–5. Note the the intelligibility score, in percentage correct, is highly dependent upon the constraints placed upon the message being communicated. Greater constraints (i.e., the smaller the average information content of each item in the total ensemble of messages) yield higher percent intelligibility scores. Grammatical structure and context in sentences, limitations on vocabulary, and syllabic length of words are also factors.

AI as presently calculated does not evaluate many factors influencing speech communication.

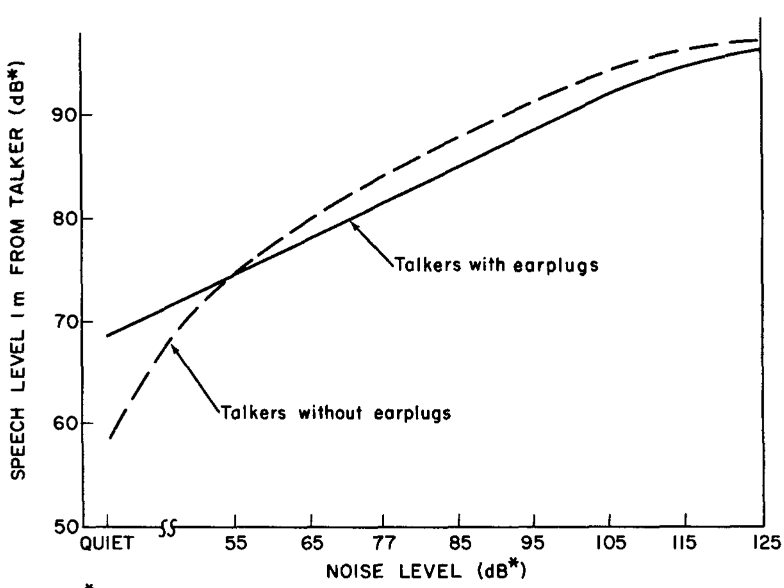


FIGURE 5–20. Speech level as a function of ambient noise level surrounding talker.

The method is designed for and has been validated against tests involving male talkers. With what degree of accuracy AI would predict the relative intelligibility of speech of female talkers over different communication systems is not known.

Quantitative effects upon speech intelligibility to be obtained from listeners receiving a mixture of the speech signal directly from a talker and also from a loudspeaker are not known. Accordingly, AI should probably not be applied to such a system.

The speech-interference level. A simple method for predicting the intelligibility of face-to-face speech communication has been devised where noise has a relatively continuous spectrum (e.g., ventilation noise in offices, engine rooms, aircraft noise, and around milling machines). The method, called the speech-interference-level (SIL) method, yields the maximum noise level that will permit correct reception of 75% of PB words or about 98% of test sentences. This criterion is equivalent to an AI of about 0.5.

To find the SIL of a given noise:

1. Measure the sound pressure level of the ambient noise in octave bands of 600 to 1200, 1200 to 2400, and 2400 to 4800 Hz; or preferably bands 710 to 1400, 1400 to 2800 and 2800 to 5600 Hz.
2. Determine the arithmetic average of the decibel levels in the three octave bands. (To convert spectrum levels to octave-band spectra or vice versa, see Figure 5–1.) This average value is the SIL.
3. Consult Table 5–20 to find the maximum distance between talker and listener at which 75% of PB words will be heard correctly.

The received-speech level. Accurate prediction of a received speech level is difficult in a face-to-face communication situation. Communicators move about, the talkers' vocal efforts vary; but reasonable estimates for engineering applications are possible.

The effects of variations in distance between talker and listener can be estimated with the aid of the "inverse distance law," which can be stated as follows:

$$\frac{P_1}{P_2} = \frac{d_2}{d_1}, \qquad (5\text{–}2)$$

where P_1 is the sound pressure at distance d_1, and P_2 is the sound pressure at distance d_2. The formula indicates that doubling the distance between talker and listener causes a halving (6-dB decrease) of the sound pressure.

The above equation is an exact formula only if the sound is propagated in a free field (e.g., outdoors). There is reverberation in a room, and hence the formula can only be approximated. It should not be used in determining the speech level at any great distance from the talker. Adjustments in vocal effort normally made to compensate for variations in noise level can be estimated from Figure 5–21.

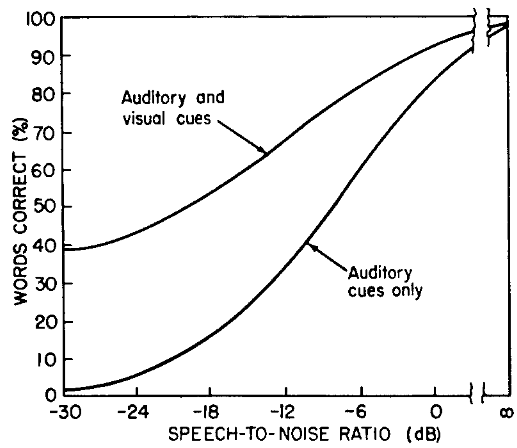


FIGURE 5–21. Intelligibility of words when perceived with and without visual cues from observing the talker (Sumby and Pollack, 1954).

For accurate values indoors at a distance from the talker, actual measurements are indispensable. A word of caution, however: when making measures of speech levels, do not use so-called carbon microphones because the amplitude response of most carbon microphones is nonlinear.

To determine the level of received speech in intense noise, use the following procedure:

1. Record with a voltmeter or VU meter the signal level produced by a magnetic throat microphone worn by the talker. (The talker should be taking with the level of effort typically used in intense noise.)
2. Turn off the source of the noise or take the talker into a relatively noise-free place.

3. Have the talker read the words or messages used in Step 1 so that the throat-microphone signal causes the voltmeter or VU meter to indicate the same levels as recorded in Step 1.

4. Place a sound-pressure-level meter at the position of the listener, and record the peak level reached on each word.

5. Subtract 4 dB from the arithmetic average of the decibel readings obtained from the sound-pressure-level meter. This gives, approximately, the long-time r.m.s. level of the speech at the position of the listener.

5.4 Noise in Speech Communication

Unwanted noise often masks speech unless appropriate design and operational procedures are adopted. The noise might come from the environment of the talker and/or listener, or from electrical and/or electromagnetic interference within the system itself. The effects of noise on speech communication can be reduced by using ear-protective devices, microphone noise shields, or noise-cancelling microphones, signal control and processing, or special headsets.

The most effective way to protect speech from noise interference is to control and isolate noise at its source—to keep it out of the communication system. Assuming that the communication engineer has applied noise control at its source, the following discussion concerns control of the noise that remains.

5.4.1 Masking of Speech by Noise

The masking spectrum of a noise is the curve that shows, as a function of frequency, how strong a sinusoidal signal must be to be just audible above the noise. As a practical matter, the masking spectrum of a wideband noise is the same as the acoustical spectrum of the noise. This is the reason why the acoustical spectrum can be used in the calculation of AI, even though it is what can be heard (and not the noise-pressure level *per se*) that is important in determining intelligibility. Accordingly, the AI for a communication system operating in the presence of wideband noise can be computed directly from a plot of the speech and noise spectra as has been described.

How well AI predicted the performance of several systems in wideband noise is shown in Figure 5–22. The solid curve shows predicted nonsense-syllable scores. The various points represent scores actually obtained. Extending from 100 to 10,000 Hz, the noise spectrum was the same for all of the systems tested; it was uniform up to 400 Hz and then fell at the rate of 10 dB per octave above that frequency.

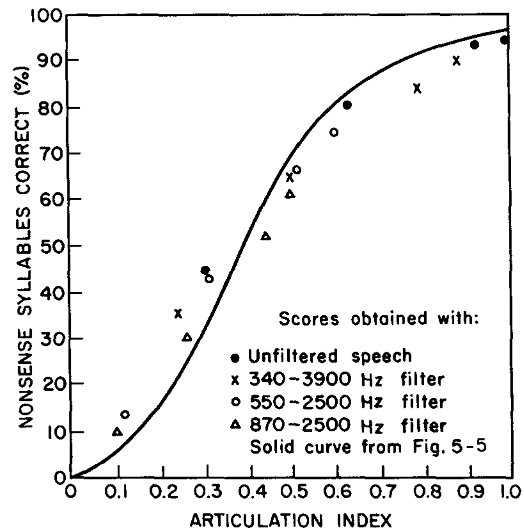


FIGURE 5–22. Relation between intelligibility of syllables and AI in the presence of wideband noise (Egan and Weiner, 1946).

5.4.2 Narrowband Noise

Most industrial and military noises have approximately continuous spectra and exist in a more or less steady state for several minutes or hours. Sometimes, however, a noise can be more or less steady, and have its energy concentrated in relatively narrow bands of frequency.

The masking effect on speech intelligibility of narrow bands of noise cannot be estimated in the same way as that of wide bands. In computing the AI for narrowband noises, one must consider the spread of masking from noise components in one frequency band to speech components in another, as well as the spread down the frequency scale which sometimes must be taken into account.

The scores predicted from the AI's calculated for various narrow bands of noise are shown in

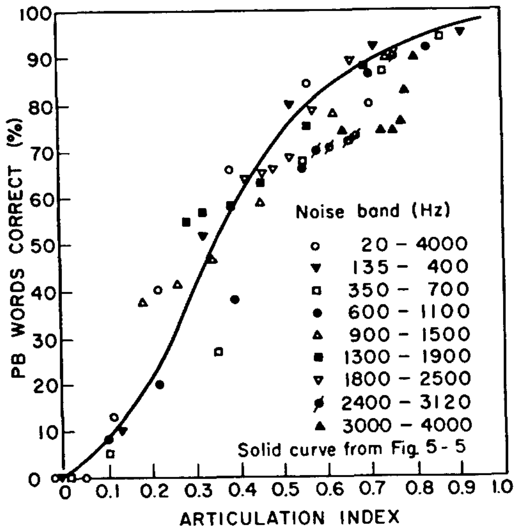


FIGURE 5–23. Relation between intelligibility of words and AI in the presence of narrowband noise (Miller, 1957b).

Figure 5–23. The scatter of the actual scores around the curve of prediction is largely attributable to the inherent variability in experimental test scores.

5.4.3 Criteria for Communication in Noise

For signals of moderate intensity, the signal-to-noise ratio, and not the absolute level of the speech or the noise, determines speech intelligibility. The absolute level is important also, because of the risk of hearing damage which can be caused by intense speech or noise. The absolute level indicates extreme levels of speech and noise in various working environments.

Damage-risk criteria. Intense speech can cause damage to the human hearing mechanism just as any other audio signal of sufficient intensity and duration. In intense noise, it can be dangerous to provide a sufficiently high signal-to-noise ratio for high intelligibility—the speech can become as great a source of danger to the ear as the noise.

If communication is to be frequent or continued, the overall r.m.s. level of the speech signal should not exceed 115 dB, relative to 0.0002 μbar, at the listener's ear. Higher speech levels can damage hearing. Intelligibility is not improved even though the signal-to-noise ratio is increased. If the listeners wear earplugs, however, speech presented by the loudspeaker or the head set (worn over the ear-plugs) can be made stronger than 115 dB by an amount equal to the actual earplug attenuation.

Intelligibility criteria. No single AI value can be a criterion for "acceptable" communications. The efficiency of communications as shown in Figure 5–19 depends on the message to be transmitted, and the proficiency of the talker and listener as well as the expectations of the end user. (See Table 5–25.)

Present-day commercial communication systems are designed for conditions that provide AI's in excess of 0.5. For communication systems used under stress conditions and by numerous randomly skilled talkers and listeners, an AI of 0.7 or higher is appropriate.

Table 5–26 presents criteria found to be useful for estimating acceptability of communication systems for which AI's have been calculated. The criteria are based, primarily, on experience with communication systems designed for military use.

Noise criteria. The effect of noise on a person's efforts to communicate clearly has an important influence on his tolerance for noise. Other factors, such as the general annoyance of a continued high-pitched "hiss" and individual dislikes for particular noises, also influence this tolerance.

Acoustical noise criteria (NCA) evaluate communication for offices, conference rooms, and factory workplaces. Figure 5–24 (Beranek, 1957) and Table 5–27 outline these criteria. These criteria apply however, only when the background noise is fairly steady with a continuous spectrum.

To determine the NCA of a given communication environment:

1. Measure the background noise in octave bands with work proceeding normally (but with no one talking).
2. Plot the octave-band spectrum of the noise on a worksheet as shown in Figure 5–24.
3. Find, in Figure 5–24, the NCA curve next above the highest measured octave-band level as determined by the plot obtained in Step 2, and assign its number to the environment under study.

TABLE 5–25. SPEECH-INTERFERENCE LEVELS THAT BARELY PERMIT RELIABLE CONVERSATION

Distance between talker and listener (ft)	Speech-interference level (dB)*			
	Normal†	Raised†	Very loud†	Shouting†
0.5	71	77	83	89
1.0	65	71	77	83
2.0	59	65	71	77
3.0	55	61	67	73
4.0	53	59	65	71
5.0	51	57	63	69
6.0	49	55	61	67
12.0	43	49	55	61

Beranek (1949).
*Correctly hearing 75% of PB words.
†Voice level.

TABLE 5–26. INTELLIGIBILITY CRITERIA

An AI of—	Provides communications—
0.7 to 1.0	Satisfactory to excellent.
0.3 to 0.7	Slightly difficult to satisfactory—up to 98% of sentences are heard correctly.
0.0 to 0.3	Impossible to difficult—special vocabularies and radio-telephone voice procedures are required.

Beranek (1949).

TABLE 5–27. NOISE CRITERIA (ACOUSTICAL) FOR OFFICES AND SHOP AREAS

NCA	Communication environment
20 to 30	Very quiet office, telephone use satisfactory, suitable for large conferences.
30 to 35	Quiet office, satisfactory for conferences at 15-ft table, normal-voice range 10–30 ft, telephone use satisfactory.
35 to 40	Satisfactory for conferences at 6–8 ft table, telephone use satisfactory, normal-voice range 6–12 ft.
40 to 50	Satisfactory for conferences at 4–5 ft table, telephone use slightly difficult, normal-voice range 3–6 ft, raised-voice range 6–12 ft.
50 to 55	Unsatisfactory for conferences of more than two or three people, telephone use slightly difficult, normal-voice range 1–2 ft, raised-voice range 3–6 ft.
55 to 60	Very noisy office, telephone use difficult.
60 to 70	Raised-voice range 1–2 ft, telephone use difficult.
70 to 80	Raised-voice range 1–2 ft, shouting range 3–6 ft, telephone use very difficult.
Above 80	Communication extremely difficult, telephone use unsatisfactory.

Beranek (1957).
Note: Measurements made to compare office noise with these criteria should be performed with the office in normal operation but with no one talking where the measurement is being made.

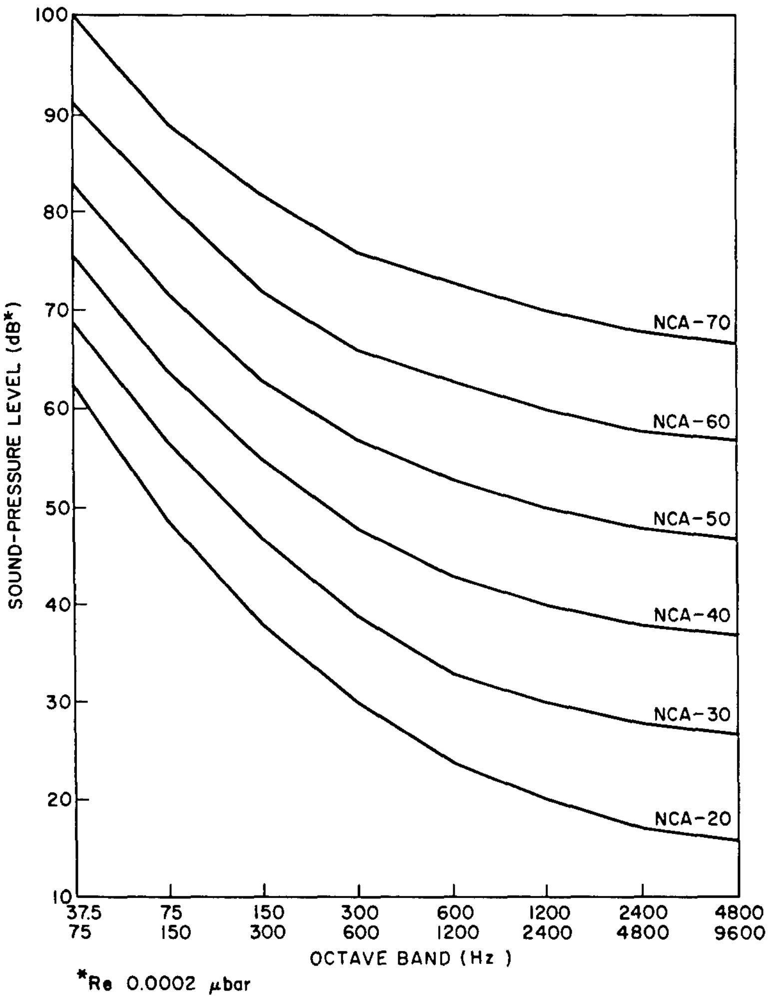


FIGURE 5–24. Acoustical-noise-criteria curves. These curves are to be used with Table 5–27 in determining permissible sound-pressure levels in eight octave bands. Curves show recommended maximum noise that is free of beats between low-frequency components (Beranek, 1957).

4. Enter Table 5–27 with the NCA number and read out a description of the communication environment.

Damage-risk criteria may be found in Kryter (1966).

5.4.4 Ear Protective Devices

Under most noise conditions, a listener can wear earplugs without reducing the intelligibility of speech (see Figure 5–25). The earplugs attenuate the speech and the noise by the same amount so that the signal-to-noise ratio at the listener's eardrums is the same. If the ambient noise is sufficiently intense to override the "physiological noise" in the listener's ear, the AI is also the same.

To override the physiological noise, even after being attenuated by earplugs, the ambient noise must be 30 to 40 dB above threshold in each of the 20 bands; otherwise, the earplugs can reduce both ambient noise and all of the speech in some bands to levels below the hearing threshold. This is what happens when the noise has an overall level of 65 dB or less; the use of earplugs in this case causes a decrease in speech intelligibility.

On the other hand, as Figure 5–25 shows, when the speech level exceeds 85 dB, the use of earplugs causes an increase in intelligibility,

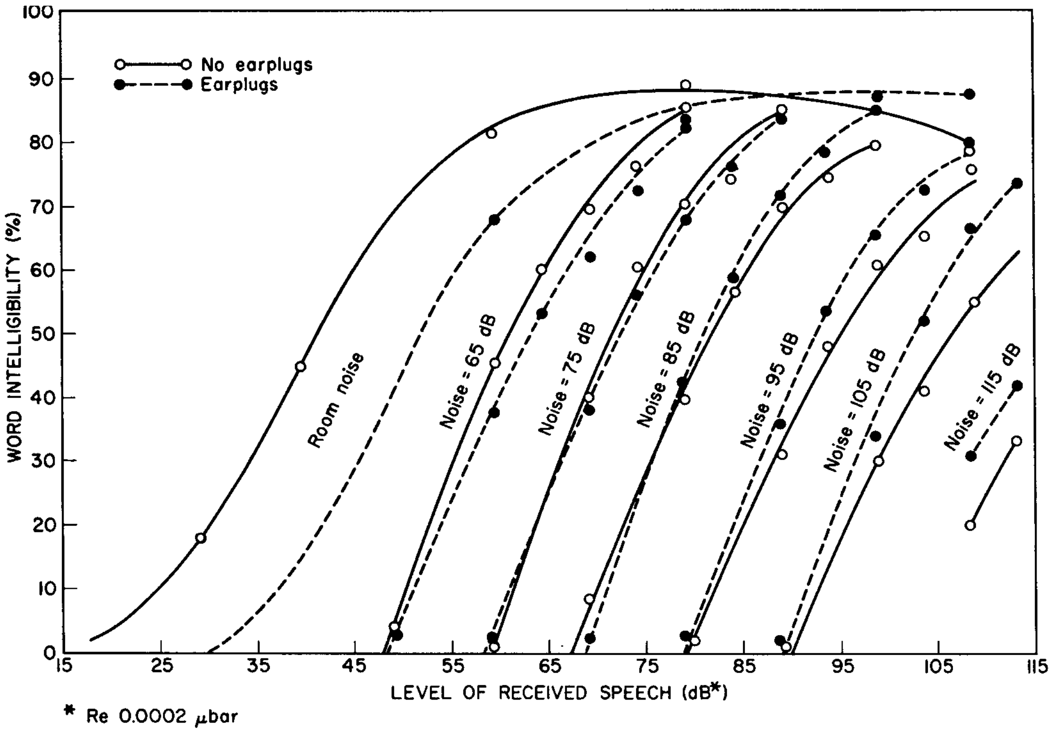


FIGURE 5–25. Relation between PB-word intelligibility and speech level in various levels of masking noise with and without earplugs. (NDRC type V-51R earplugs were used.) Data show higher intelligibility with earplugs than without in presence of intense noise (Kryter, 1946).

with or without background noise. When the long-time r.m.s. level of speech is 85 dB, the speech peaks have a spectrum level of about 97 dB. As mentioned earlier, at about 95 dB the speech peaks start to overload the ear.

A word of caution: the effect of overload on speech intelligibility can be more serious than is implied by the procedure for computing AI. While peaks above the overload level are assumed to make no further contribution to intelligibility, actually overload impairs reception; any device that attenuates both speech and noise (and there are several over-the-ear devices as well as ear plugs) will reduce the speech-peak level and the amount of overload.

Not only can ear-protective devices improve speech intelligibility in intense noise, but they can also retard a decline in intelligibility resulting from long exposure to that speech and noise.

With no background noise and with maximum-attenuation (50 dB in the mid-frequencies) ear-protective devices, speech reception will be as good with the ears protected as with them unprotected, providing the long-time r.m.s. level is at least 90 dB at the listener's head, a listener, that is, who has normal hearing. The following procedure indicates whether a given ear-protective device will improve, or interfere with, speech reception under specified noise conditions:

Step 1. Plot on a worksheet, such as Figure 5–6, the noise spectrum, the peak speech spectrum of the talker, or from a communication system at the listeners' ears, and the pure-tone threshold of hearing of the listeners.

Step 2. Subtract, from the speech and noise spectra curves plotted in Step 1 above, the attenuation afforded sound at each audible frequency by the ear-protective devices. Plot the results on a separate worksheet, such as Figure 5–6, along with the pure-tone threshold of hearing.

Step 3. Compare the area between the peak speech spectrum and the noise spectrum or pure threshold curve, whichever is the higher, for the plots obtained in Steps 1 and 2.

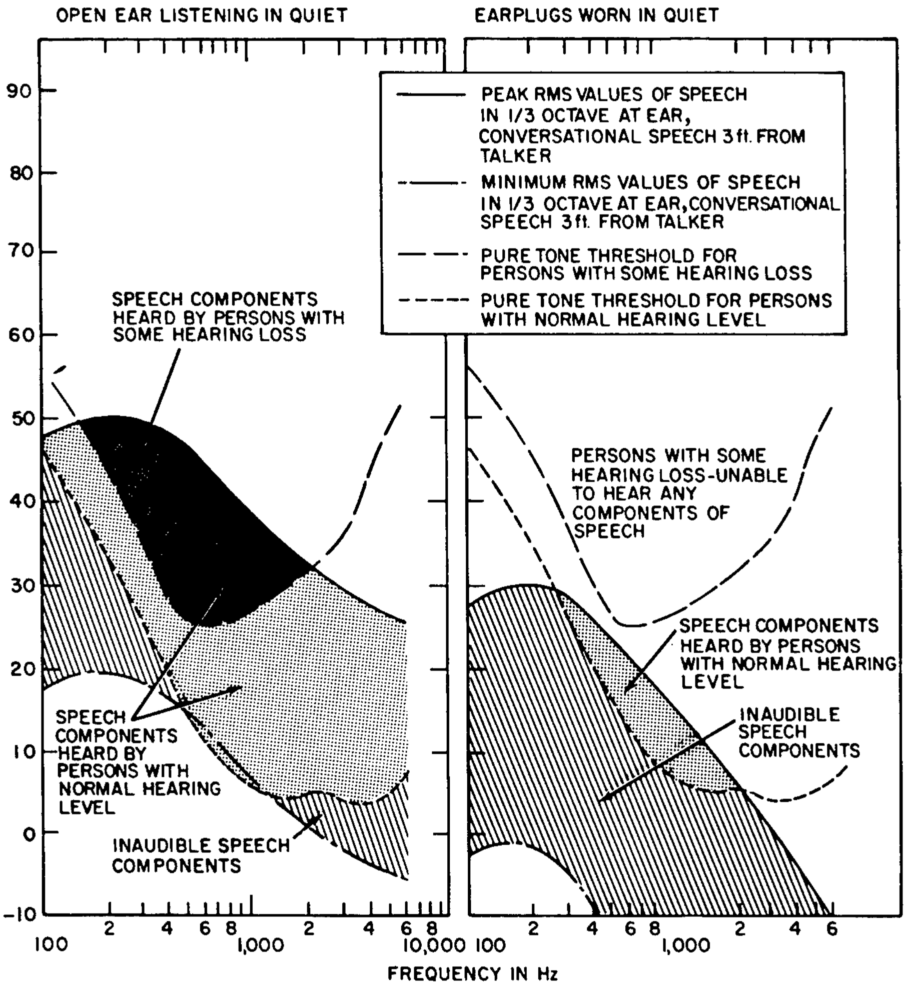


FIGURE 5–26. Effect of earplugs upon intelligibility of speech components in quiet environment (Garinther and Kryter, 1948).

If the areas requested in Step 3 are equal or greater for Step 2 than Step 1, ear-protective devices should be worn inasmuch as speech communications will be as good with ear protection as for open-ear hearing, and the listener's ears will be protected from temporary and permanent hearing loss because of noise exposure.

The application of this procedure is illustrated for persons with normal hearing and with hearing loss as in Figures 5–26, 5–27, and 5–28. It is seen that in "intense" noise, ear-protective devices should be worn by persons with normal hearing or with hearing loss as indicated. However, in "moderate" noise, ear-protective devices are indicated only for persons with normal hearing.

5.4.5 Microphone Noise Shields

When the talker is in an intense noise field, his microphone should be put in a noise shield. This applies to noise-cancelling microphones as well as to other microphones. A noise shield protects the microphone more from high- than low-frequency noise; noise cancelling does just the opposite. But, as shown in Figure 5–29 (Hawley and Kryter, 1957), a noise-cancelling microphone in a noise shield can attenuate noise by 30 dB.

Noise shields should be designed to meet the following requirements:

1. The shield should have a volume of at least 250 cm.^{3} to permit a pressure-gradient microphone to function normally, if this is the type used.

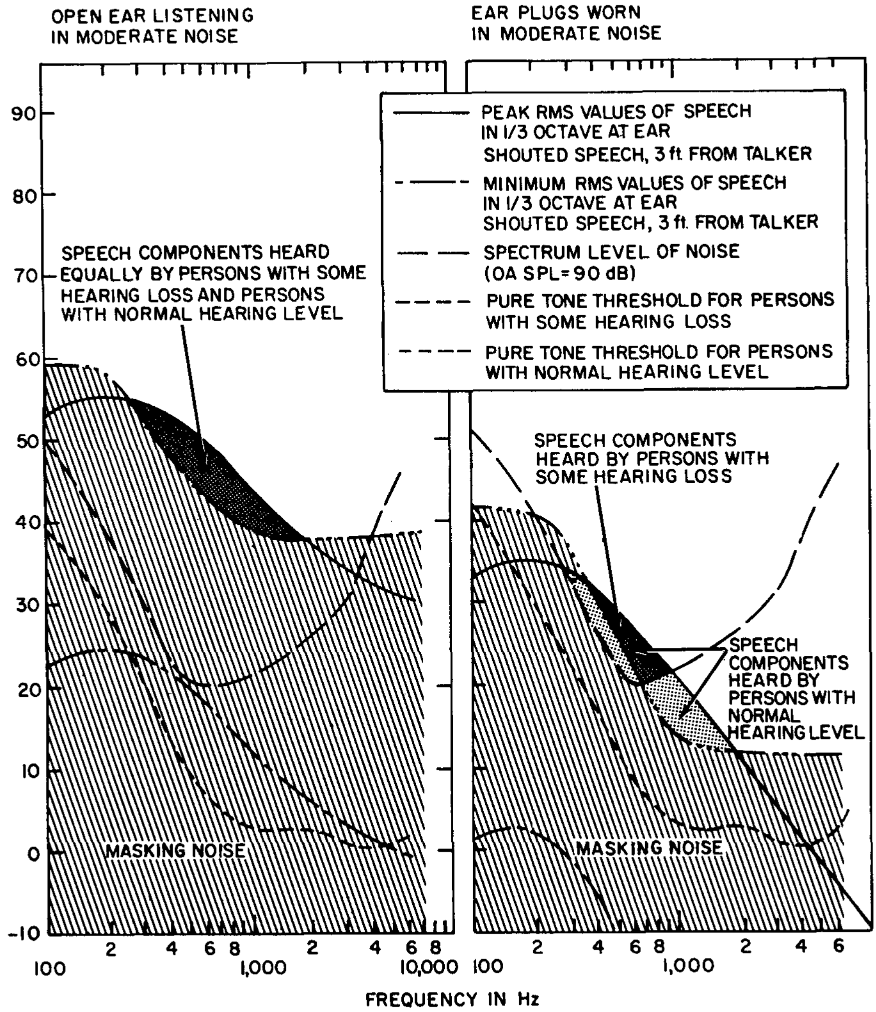


FIGURE 5–27. Effect of earplugs upon intelligibility of speech components in moderate noise (Garinther and Kryter, 1948).

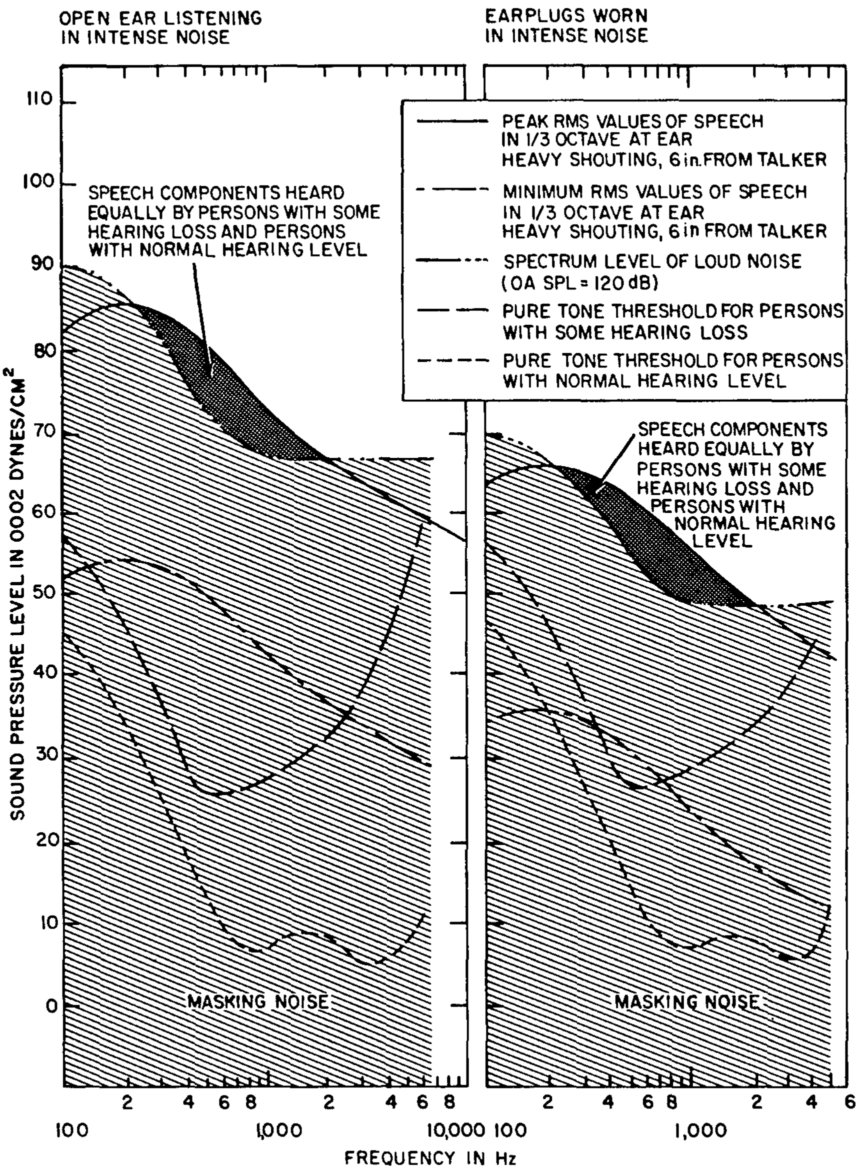


FIGURE 5–28. Effect of earplugs upon intelligibility of speech components in intense noise (Garinther and Kryter, 1948).

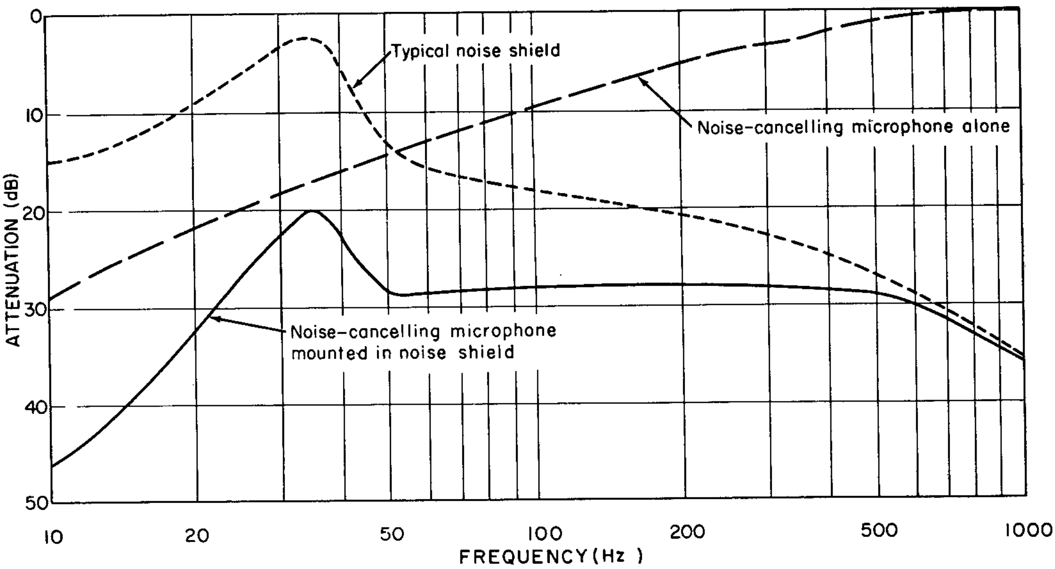


FIGURE 5–29. Attenuation of noise by noise shields around microphone and by noise-cancelling microphone (Hawley and Kryter, 1957).

2. The shield should not be so large as to be unwieldy.

3. The shield should fit tightly against the face to obtain a good seal with hand pressure or strap tension.

4. A hole, or combination of holes, having a total area of 0.1 in.^{2} should be provided in the shield to permit exhalation while talking without pressure build up in the shield. The hole or holes should be as far as possible from the microphone.

5. A standing-wave pattern inside the shield should be avoided. Sound-absorbing material should be enclosed in acoustically transparent, waterproof material (such as thin-sheet polyurethane).

6. To provide good intelligibility, the shield should not impede the talker's voice effort, mouth or jaw movement, or breathing.

7. To obtain the data necessary for the prediction of intelligibility, the designer should measure the response of the shield and the microphone together.

5.5 Component Selection and Application

Microphones, amplifiers, headsets, loudspeakers, etc. are components of a communication system and often must function in noise. Some of these components can distort the speech signal adversely, while others can be used to advantage. For instance, automatic gain, or volume control, or clipping of the peaks of speech waves, may yield improved communications.

5.5.1 Microphones

The three most important characteristics to be looked for in a microphone are:

1. High sensitivity to acoustic speech signals.
2. Faithful transduction of the acoustic speech signal into an electric signal.
3. Ability to reject other acoustic signals and noises that are present at the location of the talker.

Noise-Cancelling Microphones

Some microphones have good noise-cancelling characteristics. These microphones are so constructed that sound waves can reach the diaphragm from the back as well as from the front. A microphone placed directly in front of the lips of a person who is talking is in the spherically expanding part of the speech wave pattern, and there is a large gradient of speech pressure between the front and back of the diaphragm. Noise, on the other hand, usually comes from

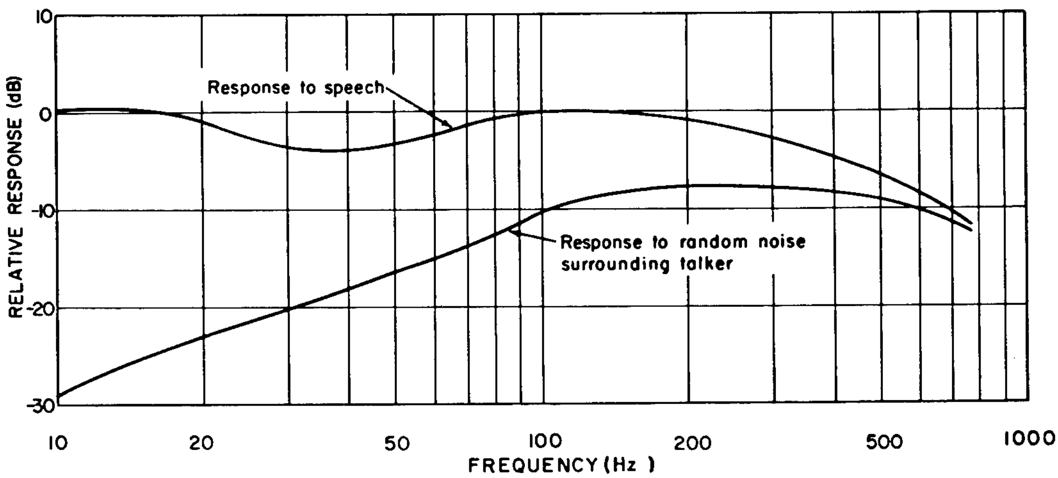


FIGURE 5–30. Amount of discrimination available from a typical noise cancelling microphone placed ½ in. from speaker's lips (Hawley and Kryter, 1957).

more distant sources. With noise-cancelling microphones, this noise has equal access to both the back and front of the diaphragm and is thus largely "cancelled," whereas the speech is not. The amount of discrimination that is available from a typical noise-cancelling microphone placed ½ in. in front of the talker's lips is shown in Figure 5–30 (Hawley and Kryter, 1957).

Design Recommendations for Noise-Cancelling Microphones

1. For close talking, do not use the large pressure-gradient microphones that are usually called "velocity microphones." They are not designed for noise cancelling, but as ordinary microphones used at a several-foot distance from the sound source.

2. Use a close-talking noise-cancelling microphone in low-frequency noise of 100-dB overall sound pressure level.

3. Design the microphone mounting or grip so that talkers *always* speak along the axis of the microphone, i.e., along a line through the holes leading to the active element.

4. Design the microphone grip to be held in such a way that the talker's hand, clothing, etc. will not cover the holes.

5. Noise-cancelling microphones must be held very close or touching the lips if their noise discrimination properties are to be realized.

Contact microphones. Contact microphones are placed directly in contact with the throat, forehead, jaw or surfaces of the head, in the ear canal, or on a tooth. Such microphones can be constructed and shielded to yield a good signal-to-noise ratio even though the talker is in an intense-noise field. Contact microphones tend to introduce distortion, however, because of the absorption of the high speech frequencies by body tissues, and because different microphone locations favor certain speech sounds.

With a talker in intense noise, and a listener in relative quiet, the best contact microphones provide as much intelligibility as some other types. When both are in noise, however, noise-cancelling microphones are preferred, even though the overall signal-to-noise ratios are comparable.

General Design Recommendations for Microphones

1. Choose a microphone with a smooth frequency-response characteristic as wide as that of the rest of the system. It should extend, at least, from 200 to 6100 Hz for highest intelligibility.

2. The dynamic range of the microphone, when working into the selected amplifier, should be great enough to admit 50-dB variations in signal input as a minimum.

3. For close talking, consider only microphones that do not overload with signals as high as 125 to 130 dB.

4. Avoid condenser microphones if the bias voltage (typically 200) would constitute a safety hazard, i.e., in a high oxygen environment.

5. Avoid carbon microphones in which "packing" of the carbon granules occurs.

6. Do not use carbon microphones if the quality criteria demand a truly linear response characteristic of very low background noise.

7. Do not use a ribbon microphone for close talking unless the microphone has been designed specifically for this use.

8. For close talking, protect a microphone against breath blast, which may damage it and will certainly make the reproduced sound quality objectionable. Protect it also against condensation of moisture, particularly saliva.

5.5.2 Amplifier, Transmitter, and Receiver Characteristics

These components should have the following characteristics:

1. Sufficient bandwidth to provide a "flat" audio-frequency response from at least 250 to 4000 Hz (preferably 200 to 6100 Hz for intelligibility and 100 to 7500 Hz for quality of reproduction).

2. Sufficient dynamic range and gain to handle the range of instantaneous pressures found in speech and to develop the necessary signal level at the headset or loudspeaker terminals.

3. They should introduce less background noise than is introduced by a microphone.

Automatic Gain Control

When both talkers and listeners are in relative quiet, linear amplification is usually desirable. In noise, it may be desirable to deliberately introduce nonlinearity. Two kinds of nonlinear amplification are of particular interest: automatic gain control (AGC), sometimes called automatic volume control (AVC), and peak clipping.

Although automatic gain control and peak clipping have different actions and effects, they can be used together. The one essential difference is in their response times; ordinary AGC operates on relatively long-time measures of the intensity of a signal, whereas a peak clipper can be thought of as an AGC that operates instantaneously.

Essentially, an AGC system acts like a linear system during any interval that is short relative to the attack time (the time required for the AGC to adjust the gain to the level of the input signal) and the release time (the time it takes the gain to return to normal after the signal stops). This linear behavior is illustrated by waveforms A and B of Figure 5–31. Note that input A is of a greater intensity than input B, but the output, following AGC gain adjustment, is the same for both A and B.

The AGC system derives a measure of the average signal strength over a period of time; and

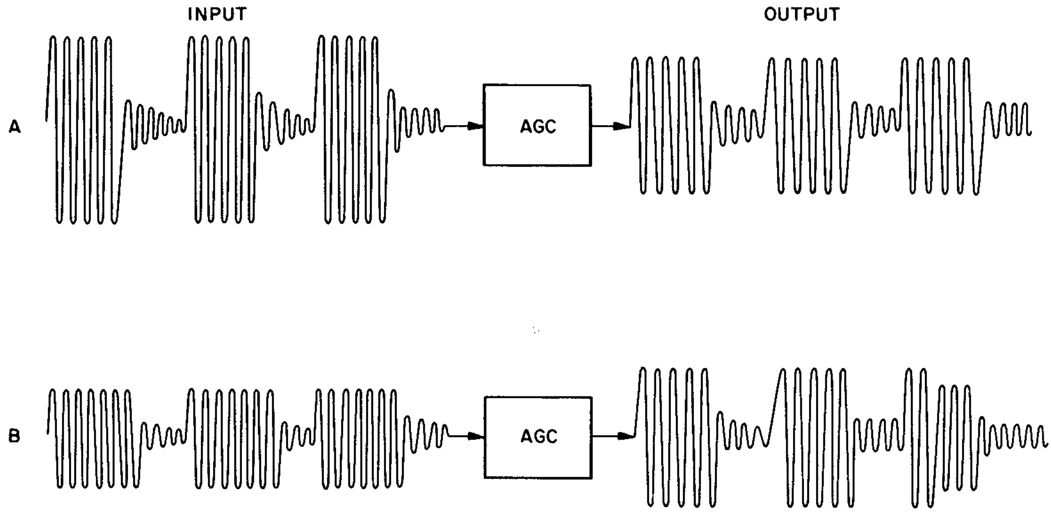


FIGURE 5–31. Effect of AGC amplifier on two signals. AGC action preserves intensive relations within short segments of signal but it automatically de-amplifies stronger signal in A, and automatically amplifies weaker signal in B, to give the same average long-term output.

this information is used to adjust the amplifier's operating characteristic. Sustained, intense signals lead to reduction of the gain; sustained, weak signals lead to increase of the gain. Therefore, the average output level is about the same, no matter what the average input level. But AGC does not eliminate variations in intensity between parts of the signal occurring together within a short interval; the consonants remain weaker than the neighboring vowels, for example, because the AGC averages over a longer interval than a single speech sound.

The uses of AGC. When AGC is used to maintain full modulation of a carrier, it is located between microphone and modulator. If the "side-tone" circuit through which the talker hears his own voice is controlled by the AGC, the AGC will tend to keep the level of the side tone (the talker's own voice) constant even though the input speech level is reduced. The natural reaction of the talker is to reduce his voice level even further. He will, of course, hear the signal-to-noise ratio decrease as he lowers his voice level, but, ordinarily, talkers do not pay much attention to the signal-to-noise ratio of the side tone. This problem can be avoided by removing the talker's side tone from the system at a point ahead of the AGC.

AGC provides a constant listening level without requiring adjustment of a manual gain control. When the strength of the received signal varies (e.g., fading in radio transmission), AGC in the receiver will tend to keep the output constant.

In addition, AGC is used to maintain a constant signal-to-noise ratio in spite of variations in the ambient-noise level at the listener's location. A noise-controlled AGC system can provide dual advantages of high speech intelligibility during intense noise and protection of hearing from intense speech during quiet periods.

For example, aboard aircraft, communication is transmitted in a noise field by a powerful loudspeaker system generating a speech signal intense enough to be heard in the noise. In quiet moments however, the speech signal sounds excessively loud, and intelligibility declines.

A noise-controlled AGC system has a "noise microphone" in the listening area so that noise intensity picked up by the noise microphone can control the speech amplifier gain. Variations in level, even at rapid rates, introduced by a noise-operated AGC do not adversely affect speech intelligibility unduly.

Attack and release times. The attack- and release-time constants employed in the "limiter" amplifiers of commercial broadcast work are 10 msec. and 600 msec., respectively. For some military communication systems designed to operate in noise, it has been found that an attack time of 0.1 sec. and a release time of 10 sec. are most satisfactory. (When the release time is made appreciably shorter, there is an objectionable fluctuation in the transmitted background noise.) Because the optional time constants depend on conditions and requirements, it is best to determine the AGC time constants empirically for each new application.

Peak clipping. Peak clipping is, as the name implies, simply clipping the peaks from the speech signal and leaving the remainder. Ordinarily, this clipping involves clipping both the positive (upward) and negative (downward) peaks. For all practical purposes, peak clippers have no attaak or release times; they operate instantaneously.

Peak clipping brings homogeneity in amplitude. Often clipping alone can reduce the amplitudes of vowels to the level of consonants. Reamplifying a clipped signal so the peak amplitude of the remnant is the same as that of the original wave (see Figure 5–32) increases the intensity of the weak consonant sounds even though the peak level of the speech (and, therefore, the peak power requirements of the amplifiers, radio transmitters, etc.) is not increased. Figure 5–33 (Licklider, 1946) shows word intelligibility as a function of peak amplitude of received speech, with peak clipping as the parameter. As can be seen, with equal peak-to-peak amplitude, clipped speech in noise is much better understood than is unclipped speech, but only if the talker is in quiet.

The optimum amount of clipping. The amount of peak clipping to be used in a communication system depends on the noise conditions. For instance, as the amount of pre-clipping noise increases relative to that of post-clipping noise, the optimum degree of peak clipping decreases. Tests have been conducted with various amounts of peak clipping and noise introduced at various

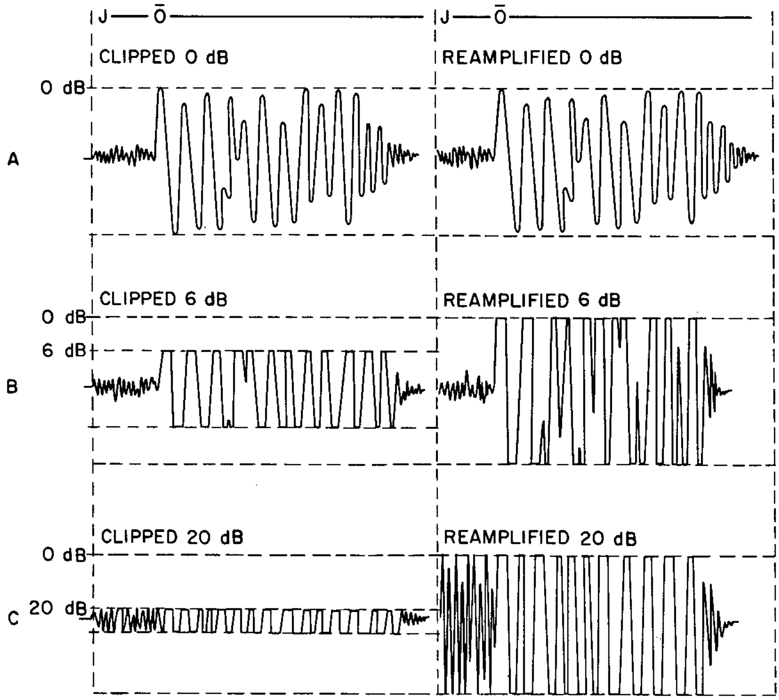


FIGURE 5–32. Schematic representations of word "Joe." A is undistorted, B is after 6-dB clipping, and C is after 20-dB clipping. Clipped signals in B and C are shown reamplified until their peak-to-peak amplitudes equal peak-to-peak amplitude of A (Licklider et al., 1948).

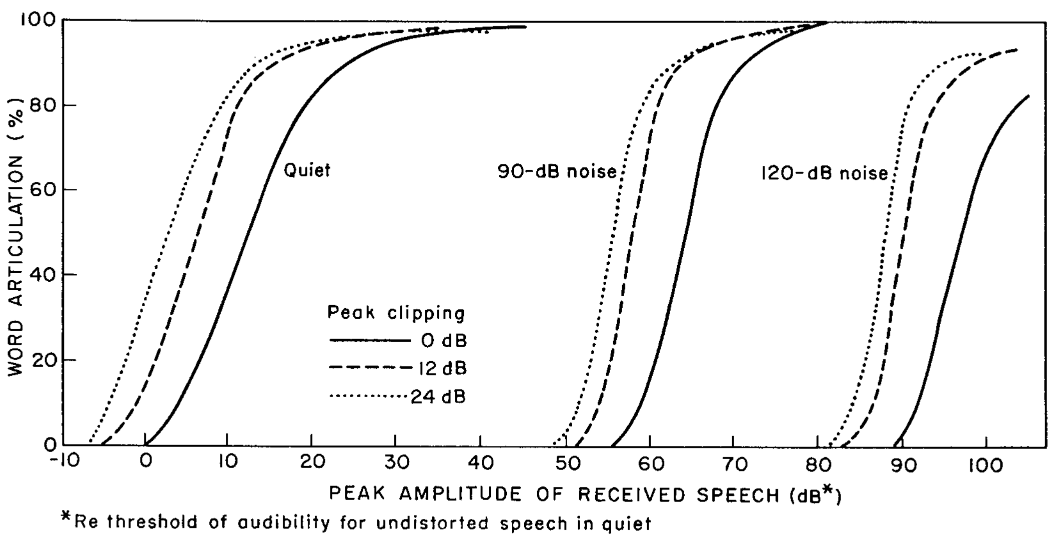


FIGURE 5–33. Effect of different levels of peak clipping on intelligibility of words in quiet and in noise (Licklider, 1946).

places into the communication system. The results are summarized in Table 5–28 and Figure 5–34.

When it is possible to have the talker in a quiet environment, or to have him use a noise-cancelling microphone and noise shield, so that the signal-to-noise ratio coming from the microphone is greater than 10 dB, Figure 5–35 can be used to estimate the optimum amount of peak clipping and reamplification. To determine the amount of clipping to be used, first calculate the AI for the system without peak clipping; then enter Figure 5–35 with the AI and read out the optimum amount of peak clipping and reamplification.

Distortion products. When listened to in quiet, severely clipped speech has a harsh, unpleasant sound because clipping introduces distortion products. When listened to in noise that enters the system at a point following the clipping, distortion products tend to be masked by noise, and the speech sounds about as unclipped speech would in the same noise. The listener might be aware that consonants and other weak sounds are louder, but the speech is more intelligible. Such subjective appraisals of the quality of clipped and unclipped speech are shown in Table 5–29.

Low-frequency-attentuating filters between microphone and clipper help reduce distortion products and the harsh sound of heavily clipped speech. High-pass filters cutting off frequencies below 300 to 400 Hz are also successful. Where cost is a consideration, coupling circuits between the stages of preclipper amplifiers can provide 6-dB-per-octave attenuation of frequencies below 1000 Hz with considerable improvement in the quality of the clipped speech.

Heterodyne clipping. Another way of avoiding distortion caused by peak clipping is by using "heterodyne clipping":

1. Use single-sideband suppressed-carrier modulation to shift the spectrum of the speech signal up the frequency scale by x Hz.
2. Peak clip and then reamplify the single-sideband-modulated carrier by the desired amount.
3. Pass the resulting signal through a band-pass filter (x to $x+5000$ Hz).
4. Use the signal in an ordinary single-sideband suppressed-carrier transmission or, if an audio signal is required, demodulate with the aid of standard single-sideband suppressed-carrier techniques.

Because the distortion-product noise introduced by peak clipping consists of harmonics and intermodulation products, it will be high and low in frequency, relative to the shifted speech frequencies and will, therefore, fall outside the band of the bandpass filter (x to $x+5000$ Hz). The transmitted signal will not contain the distortion products even though clipped.

Such a process will make the received signal sound less harsh to a listener in quiet. Eliminaation of the distortion products that lie outside the filter bands will affect the shape of the transmitted wave in such a way that less power is actually transmitted than would be transmitted by an ordinary premodulation peak-clipping system. Thus, heterodyne clipping does not improve intelligibility of speech received in noise as much as peak clipping prior to modulation.

Computing the AI for clipping speech. The increase in intelligibility of noise-free speech that has been clipped and reamplified is closely related to that produced in the long-time r.m.s. of the speech signal by clipping and reamplification, assuming that the signal is reamplified until the peak of the remnant is the same as the peak of the original signal.

Because the AI of speech heard in noise is proportional to the speech-to-noise ratio (if it is between 0 and 30 dB), computing an AI for a system that utilizes peak clipping is possible if the speech entering the clipper is relatively free of noise.

5.5.3 Loudspeakers and Headsets

Loudspeakers should be used when any of the following conditions prevail:

1. Ambient-noise levels are low, and no special equipment need be worn.
2. Listeners must move around so much that a headset cable is impractical.
3. A large number of listeners should hear the same message.
4. Warning or alerting signals might have to be sent to people who would not be wearing headsets.

Headsets should be used when any of the following conditions exist:

TABLE 5–28. OPTIMUM AMOUNT OF PEAK CLIPPING FOR VARIOUS NOISE CONDITIONS AND SYSTEM CONFIGURATIONS

Environmental noise surrounding talker	System noise	Environmental noise surrounding listener	Highpass filter* after microphone	AGC after microphone	Peak clipping before modulation (dB)	Lowpass filter† before modulation	AGC in receiver	Peak limiter in receiver
Negligible‡	Negligible	Negligible	No	Yes	0	No	No	No.
Negligible‡	Negligible	Moderate	No	Yes	12	Yes	No	No.
Negligible‡	Negligible	Intense	No	Yes	18	Yes	No	No.
Negligible‡	Moderate	Negligible	No	Yes	12	Yes	Yes	Yes.
Negligible‡	Moderate	Moderate	No	Yes	18	Yes	Yes	Yes.
Negligible†	Moderate	Intense	No	Yes	24	Yes	Yes	Yes.
Moderate§	Negligible	Negligible	No	Yes	0	No	No	No.
Moderate§	Negligible	Moderate	No	Yes	0	No	No	No.
Moderate§	Negligible	Intense	No	Yes	8	No	No	No.
Moderate§	Moderate	Negligible	No	Yes	6	Yes	Yes	Yes.
Moderate§	Moderate	Moderate	No	Yes	8	Yes	Yes	Yes.
Moderate§	Moderate	Intense	No	Yes	12	Yes	Yes	Yes.
Intense¶	Negligible	Negligible	Yes	No**	0	No	No	No.
Intense¶	Negligible	Moderate	Yes	No**	0	No	No	No.
Intense¶	Negligible	Intense	Yes	No**	6	Yes	No	No.
Intense¶	Moderate	Negligible	Yes	No**	0	No	Yes	Yes.
Intense¶	Moderate	Moderate	Yes	No**	6	Yes	Yes	Yes.
Intense¶	Moderate	Intense	Yes	No**	12	Yes	Yes	Yes.

*6 dB/octave
†5000 Hz.
‡Or moderate with noise-cancelling microphone in noise shield.
§With noise-cancelling microphone not in noise shield or intense with noise-cancelling microphone in noise shield.
¶With noise-cancelling microphone not in noise shield.
**If noise fluctuates widely in intensity, noise-operated AGC might be beneficial. If used, AGC should have release time greater than 15 sec.

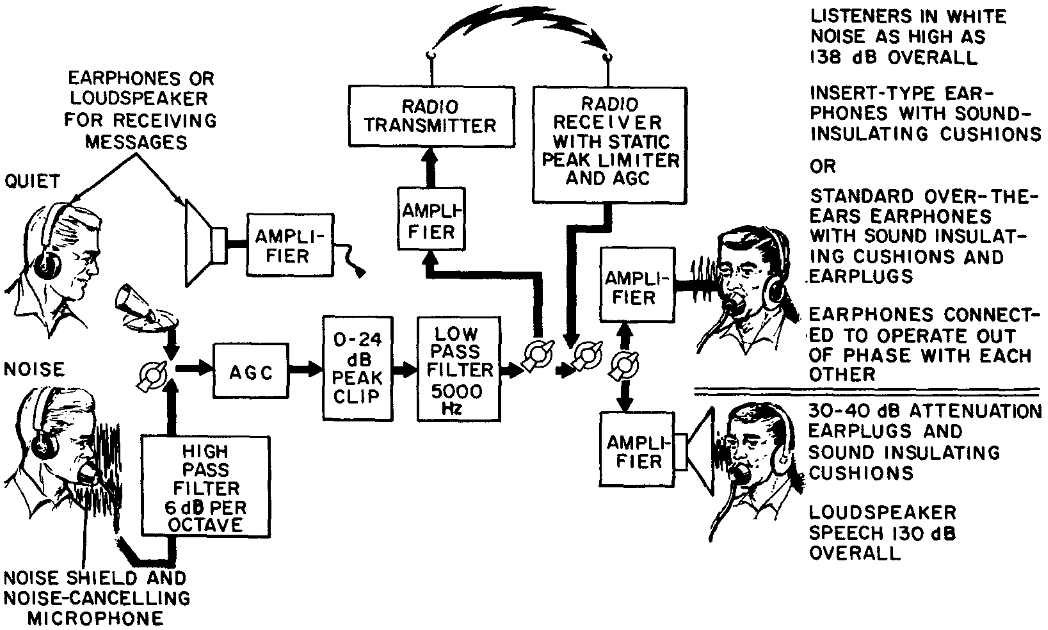


FIGURE 5–34. Proposed speech-communication systems for use in extreme-intensity noise (Kryter, 1958).

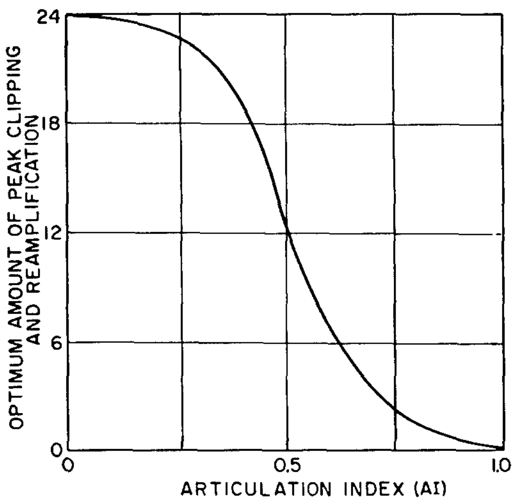


FIGURE 5–35. Optimum amount of peak clipping for systems with AI's.

TABLE 5–29. QUALITY OF CLIPPED AND UNCLIPPED SPEECH

Clipping (dB)	Sound in quiet	Quality
0	Normal	Excellent
6	Essentially normal, effect barely detectable	Probably acceptable as of broadcast quality.
12	As though talker enuciated with special care	Usable for military communication.
18	Sharp, "sandy"	Fair, usable for most military communication.
24	Coarse, "grainy"	Poor, but usable if intelligibility is of paramount importance.

Licklider (1946).

1. Ambient-noise levels are so high that ear-protective devices are required to protect the ears of the listener.
2. Different listeners must receive different messages.
3. Reverberation interferes with loudspeaker listening.
4. The listener must wear special equipment such as a protective helmet.
5. The electric power available is inadequate to operate a loudspeaker.

From a human engineering point of view, it makes little difference whether horns or direct-radiator loudspeakers are used or what kind of enclosure is provided. It is only necessary that the signal strength, sensitivity, bandwidth, and distance are such that an adequate signal-to-noise ratio exists at the listener's ears. The problem of reverberation, however, must be taken into consideration whenever loudspeakers are to be used.

Both reverberation and echoes are caused by reflection off walls or other surfaces. This indirect (and therefore delayed) signal is superposed upon the signal received over the direct path. With an echo, the delay is great enough so that the listener distinguishes two separate sounds. In the case of reverberation, the delayed signal (or, as is usually the case, the complex of variously delayed signals) fuses with the direct signal in the listener's perception.

Often the addition of the delayed components to the direct-speech waves interferes with intelligibility; the direct-speech sounds are, in effect, masked by the echoes or reverberation of preceding speech sounds. Some reverberation is desirable, to avoid the abnormal "dead" sound of a completely anechoic space.

The reverberation characteristics of a room usually are measured by determining the length of time required for a tone abruptly terminated at the source to decay 60 dB in sound pressure level. The longer the reverberation time, the greater the masking effect. Reverberation is usually greater at the rear of a large space than at the front. Much of the sound reaching the rear is reflected off of more than one surface. This reflection comes to the service of listeners in the rear of the space by distributing sound energy more or less uniformly throughout the space and keeping the signal from becoming excessively weak at great distance from the source.

Reducing reverberation. A certain amount of reverberation is desirable because it makes speech sound "alive" and natural. The computation of AI to take into account reverberation effects on intelligibility has been given.

Figure 5–36 (Farrell, 1958) gives the reverberation times (measured with a 500-Hz pure tone) generally agreed upon by acoustical experts as being optimum for various types of rooms. Complete information on the effects of reverberation on speech reception is not available, but Figure 5–37 (Fletcher, 1953) shows the main relation

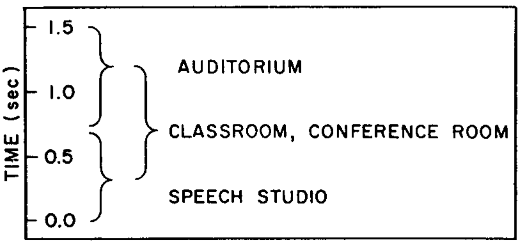


FIGURE 5–36. Optimum reverberation time allowed for different types of rooms (Farrell, 1958).

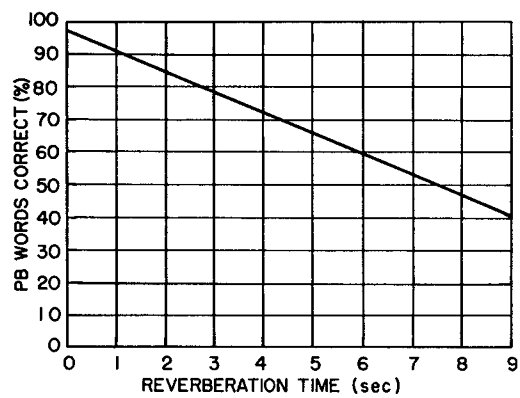


FIGURE 5–37. Effect of reverberation upon intelligibility of words.

between reverberation time and speech intelligibility.

To lessen undesired reverberation effects in large rooms, use many *low-powered* instead of a few powerful loudspeakers. They should be directional or oriented to cover but not overlap the most important parts of the room. Talkers should speak slowly, deliberately, and allow time for echoes to decay.

Preventing feedback "squeal." Public-address systems in highly reverberant rooms go into oscillation and "sing" or "squeal" because of acoustical feedback from loudspeaker to microphone. To prevent this:

1. Place talker close to microphone and reduce gain from microphone to loudspeaker.
2. Use components with uniform frequency response.
3. Use a directional microphone.
4. Provide a booth for the talker or a shield around the microphone.

Headsets. Human engineering considerations in the selection and application of headsets are:

1. Choose a headset with a smooth frequency response as broad as that of the remainder of the system.
2. Be sure that the dynamic range, without appreciable distortion, is at least 40 dB and that the power-handling capacity is adequate to receive the peaks of the amplifier output.
3. Pick a combination of earphone and socket or cushion for which the earphone sensitivity and the earcap attenuation together will provide an adequate signal-to-noise ratio.
4. Select an earphone cushion that is comfortable enough to permit the user to wear it as long as necessary.

Measuring Headset Characteristics

Frequency response. The cushions or sockets holding earphones influence their frequency response and, thereby, the fidelity of speech signal reproduction. The characteristics of the earphone-and-socket combination can be measured with the earphone and socket on an "artificial ear" that simulates acoustically the head and ear cavity of a human being. When sinusoidal signals of various frequencies are applied at constant voltage across the earphone, the pressure developed in the artificial ear is measured with the aid of a calibrated microphone.

A more meaningful way to determine the characteristics of the earphone and its socket, however, is based on the equal-loudness judgment of people actually wearing the equipment.

1. Set up a loudspeaker and associated equipment as shown, producing a pure tone of variable frequency at a sound-pressure level of 80 dB relative to 0.0002 μbar. (See Figure 5–38.)
2. Set up a "transfer-standard" earphone, and alternately apply the test signal to it and to the loudspeaker.
3. Adjust the voltage applied to the transfer-standard earphone until its tone equals the loudspeaker tone in subjective loudness. This will require separate adjustments at several different representative frequencies.
4. Put earphone x in its socket to the ear with which the listener has been listening to the loudspeaker.
5. Deliver a test signal alternately to x and to the transfer-standard earphone.

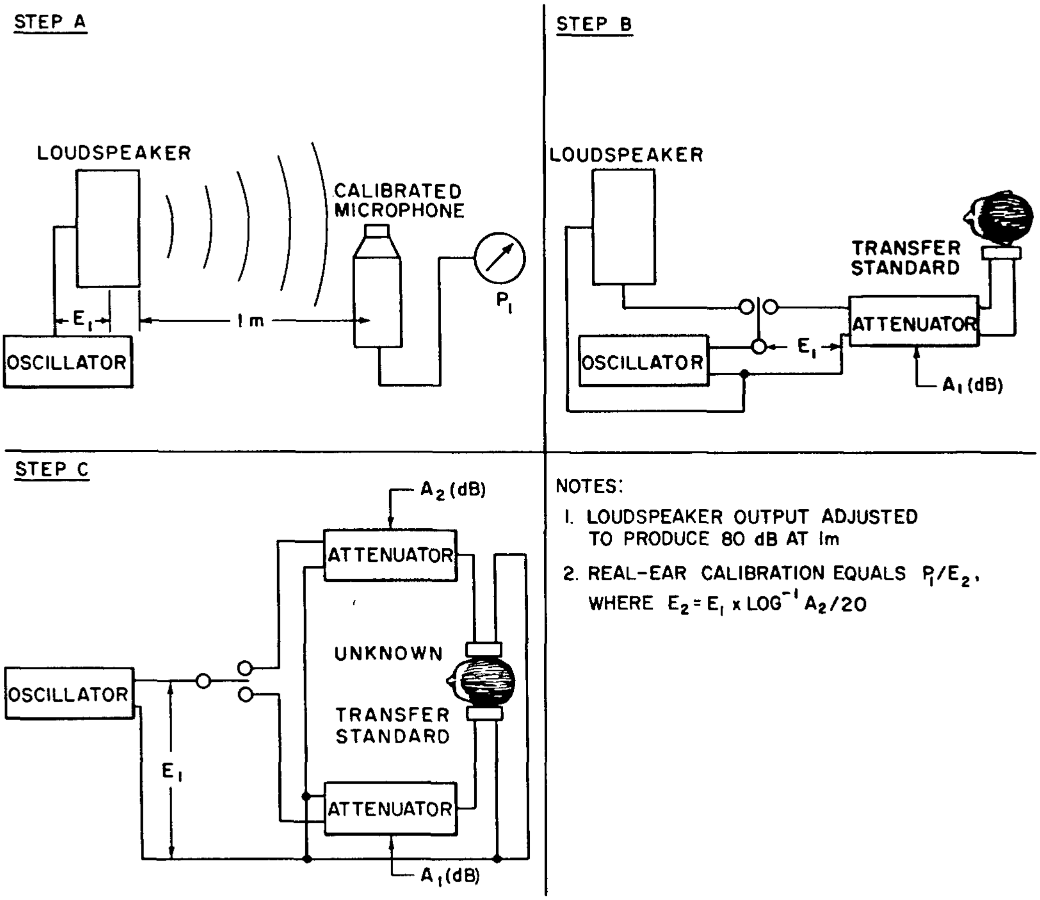


FIGURE 5–38. "Loudness-balance" method for measuring "real-ear" response characteristics of earphones (Beranek, 1949).

6. Adjust the voltage applied to x until the loudness produced by the two earphones is equal.

Insofar as the assumption of "equal sound pressure level for equal loudness" holds, earphone x is now developing the equivalent of 80 dB (re 0.0002 μbar) or 2 μbar. The "response" at each frequency tested is simply 2 μbar divided by the voltage required for equal loudness. This can be expressed in decibels relative to 1 μbar/v.

Typical results of real-ear and artificial-ear earphone measurements are shown in Figure 5–39 (Weiner and Filler, 1945). Note that the response obtained with the artificial ear differs from the real-ear response; the latter offers a better basis for engineering decisions because it is free of the unevaluated errors inherent in the simulation of so complex a system as the human ear.

Noise attenuation. These characteristics of an earphone and its socket are best measured by tests based on the "threshold of hearing" for pure tones. To perform this test, proceed as follows:

1. Place a listener in a soundproof, anechoic chamber with a loudspeaker on one side of him.
2. Plug the ear that is not oriented toward the loudspeaker.
3. Have the listener adjust the sound pressure level of a pure tone until he can just detect its presence at several frequencies.
4. Have him place the earphones (in their sockets) on his ears as they would be worn normally, and again adjust the sound pressure level until he can just detect the presence of a tone at each of the test frequencies. The differ-

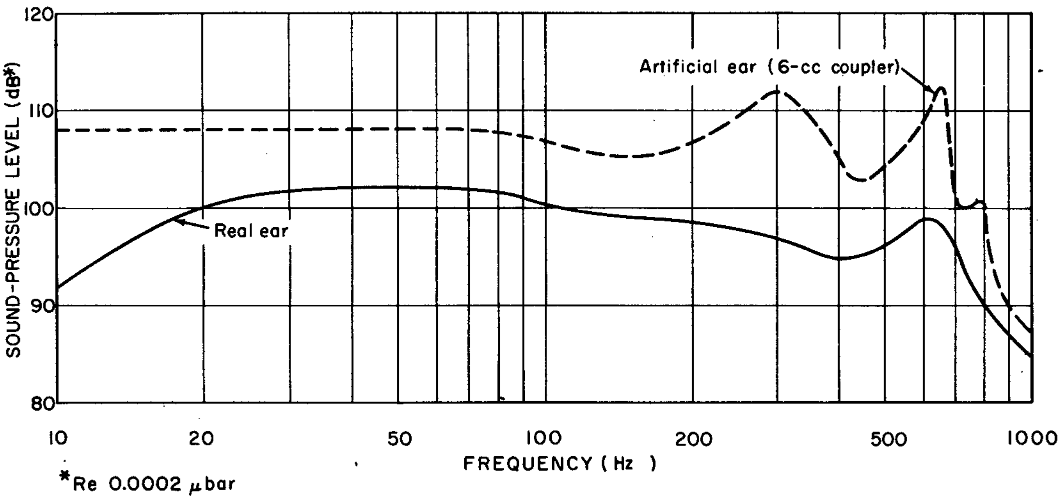


FIGURE 5–39. Typical difference between the frequency response of an ear-phone measured on an artificial ear and by the loudness-balance (real ear) method (Beranek, 1949).

ence between the open-ear and the covered-ear thresholds is the amount of attenuation or noise exclusion provided by the earphone and its socket.

Noise-attentuation characteristics also can be measured with the aid of an artificial or "dummy" head in which a calibrated microphone is located at the bottom of the "ear cavity." To perform this test, proceed as follows:

1. Place the earphone and socket on the dummy head.
2. Produce ambient sounds, preferably tones of known frequency and sound pressure levels.
3. Read the voltage developed by the microphone on the dummy head.
4. Remove the earphones and repeat Steps 2 and 3.
5. For each test frequency, determine the ratio between the voltage generated by the microphone when the earphone and socket are not on the artificial head and the voltage generated when they are.
6. Convert this ratio into decibels. The result is the amount of sound attenuation afforded by the earphone and socket.

A comparison of typical real-ear and artificial-ear measurements of earphone–socket and –cushion attenuation is shown in Figure 5–40 (Beranek, 1949). Note, again the difference between the values for real-ear and artificial-ear measurements; the latter provide only a rough approximation.

Comfort vs. performance. The sensitivity of an earphone depends on the size of the cavity formed by the earcap, earphone, ear canal, and any parts of the outer ear that are inside the earphone socket. The smaller this cavity is, the higher the sound pressure produced by a given power delivered to the earphone. Sensitivity, as well as power-handling capacity and attenuation, and comfort tend to be incompatible and compromises between them are at least partially unsatisfactory.

To achieve comfort with a headset, large lightweight earcaps that apply little pressure to the ears might be used. The combination of low pressure and light weight, however, provides little attenuation, so that a strong signal is necessary to provide a satisfactory signal-to-noise ratio. A large earphone socket means a large cavity, and to get a strong signal in a large cavity, with leaks around the socket, requires high electrical power and, therefore, high power-handling capacity. High power-handling means large, heavy earphones, and heavy earphones are uncomfortable.

The most common compromises are:

1. Large earphones and sockets mounted in any rigid helmet that has to be worn anyway (e.g., an oxygen helmet or diving helmet). The large socket surrounds the ear and presses very

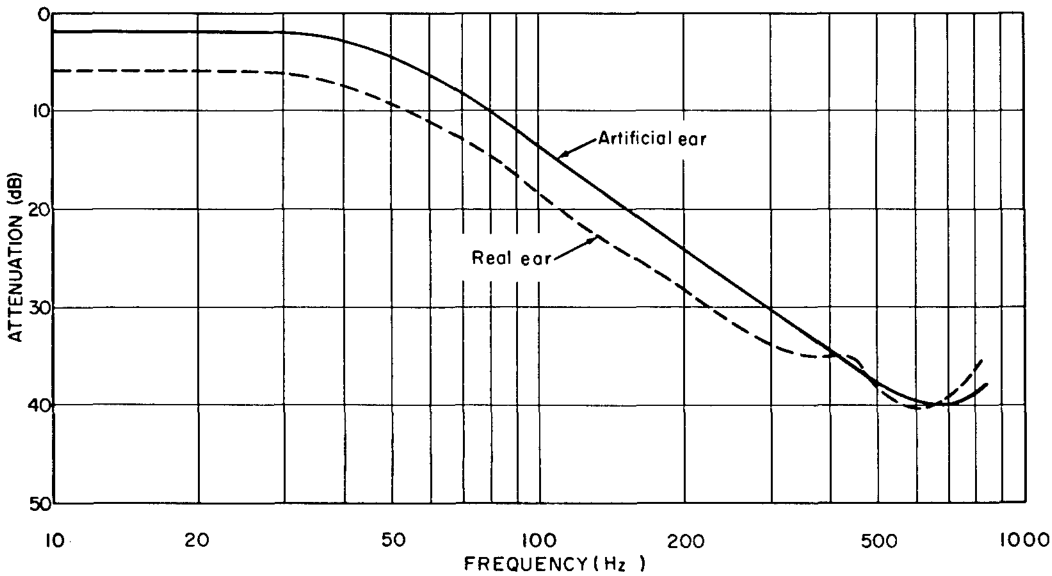


FIGURE 5–40. Typical difference in attenuation of noise furnished by an earphone–socket combination as measured by artificial ear and by the loudness-balance (real-ear) method (Wiener et al., 1945).

gently on the head bones. Such headsets can be worn for long periods of time, in some cases continuously for twenty hours. They provide adequate intelligibility if the noise level is low, or if the helmet provides good attenuation.

2. Medium-size earphones and sockets mounted in a rigid or flexible helmet or on a headband. The socket, large enough to press tightly against head bones, also rests lightly on the outer ear, because the cavity size is not great enough for complete clearance. Such headsets can be worn up to six hours and will provide high intelligibility if the sockets have good attenuation.

3. Small earphones and sockets mounted on a headband or in a telephone handset or radioset. The earcap presses tightly against the outer ear. Such headsets can be worn for less than an hour usually but are convenient for intermittent use and provide high intelligibility with small electrical power.

4. Insert earphones attached to a very light headband. Such earphones are very small and light. They are attached to small insert or semi-insert tips that partially enter the ear canal. These headsets can be worn for several hours if properly fitted. For the insert tips, this usually means custom fitting for each person because ear canals vary widely in size and shape. The tips might be unsatisfactory medically; they might start or aggravate infections in the ear canal, or be physically dangerous; a blow to the earphone might damage the ear. On the other hand, this headset gives high intelligibility with low electrical power, especially if covered with an earmuff or rigid helmet.

Design Recommendations for Headsets

1. Connect earphones to operate out of phase. Slightly better intelligibility can be obtained in ambient noise when the earphones are so connected. "Out of phase" means that the diaphragm of one earphone moves toward the center of the listener's head while the diaphragm of the other earphone moves away from it. This is just the opposite of the normal way of connecting earphones. Electrically, it is quite simple. The improvement in intelligibility, relative to that obtained with in-phase connection of the earphones, is only about 5%, but even this amount can be helpful.

2. Delaying the speech signal delivered to one ear relative to the signal delivered to the other ear by 500 μsec. results in about the same degree of improvement as out-of-phase connection.

3. Use binaural headsets if the listeners will be in intense noise. Single-ear listening (e.g., the common telephone) is quite satisfactory under good listening conditions, but somewhat less effective than binaural listening under conditions of noise stress or signal distortion. The advantage of binaural over monaural listening lies in the superior discrimination achieved with two ears. However, noise reaching the uncovered ear has little or no effect on the intelligibility of speech delivered to the covered ear, unless the noise level at the uncovered ear is 40 or 50 dB higher than the noise level at the covered ear. Even then, the effect probably is not truly an interaural interaction, but appears to be caused by transmission of the noise through the tissues of the head to the covered ear.

4. Make the talker's side tone meet certain criteria. The signal from the talker's microphone is usually returned to the talker via his earphones. This feedback signal is called the side tone. The following are some ways in which the talker's speech can be manipulated by varying the side tone:

(a) Talkers usually adjust their vocal effort (speech intensity) in such a way as to compensate, within limits, for changes in the level of the side tone. The more intense the side tone, the less intense the talker's speech. To maintain a high signal-to-noise ratio, the side tone should be kept weak. Talkers tend to speak slowly when the side tone level is low because of the increased vocal effort required to restore "normal" side tone—especially in high-noise conditions.

(b) Talkers tend to increase their vocal effort and to enunciate more precisely and at a slower rate if the side tone signal is passed through either a high-pass or a low-pass filter before the signal enters the binaural headset. Duplex binaural headsets in which only one channel is either high-pass or low-pass filtered usually yield even more intelligible speech than simple binaural headsets.

(c) Talkers tend to increase their level of effort and to talk at a slower rate if the electronic side tone is slightly delayed. A delay of 0.05 sec. can be introduced intentionally to produce this effect. However, a 0.15- to 0.21-sec. delay should be avoided.

5.6 Special System Requirements

Certain special requirements may influence the design of a communication system. Some examples of unusual situations are: (a) the listener must receive messages from several sources at the same time; (b) the talker and/or listener must be at a high altitude; (c) the talker and/or listener must be submerged in water; (d) the talker must communicate through a mask; and (e) the telephone or radio-telephone link used for the transmission of the speech signal is limited in frequency bandwidth or channel capacity so that the normal frequency bandwidth of speech must be compressed.

5.6.1 Multichannel Listening

While it is impossible to listen to two simultaneous, nonredundant messages and receive the full content of each message, some communication systems seem to pose just these conditions. An air traffic controller having to receive voice communications from several aircraft, from other controllers, and from his supervisor; a taxi dispatcher receiving messages from roving taxicabs as well as customers, are two examples of the requirement of multichannel listening.

Design Recommendations for Multichannel Listening:

If the system cannot be designed so that only one message will be heard at a time:

1. Use a separate loudspeaker for each speech channel, and locate the loudspeakers at different angles from listeners. (See Figure 5–41.)
2. If there are two speech channels, feed one channel into one ear and the other channel into the other ear so the listener can switch his attention. If this is not possible, "picket-fence" both signals at 30 to 40 Hz and interleave the two.
3. Use frequency-selective filters to give characteristic timbre to signals. If there are three channels involved, leave one channel unfiltered, use a high-pass filter (with 1000-Hz cutoff) in the second channel, and use a low-pass filter (with a 2500 Hz cutoff) in the third channel.
4. Use a visual signal to show which channel is in use.

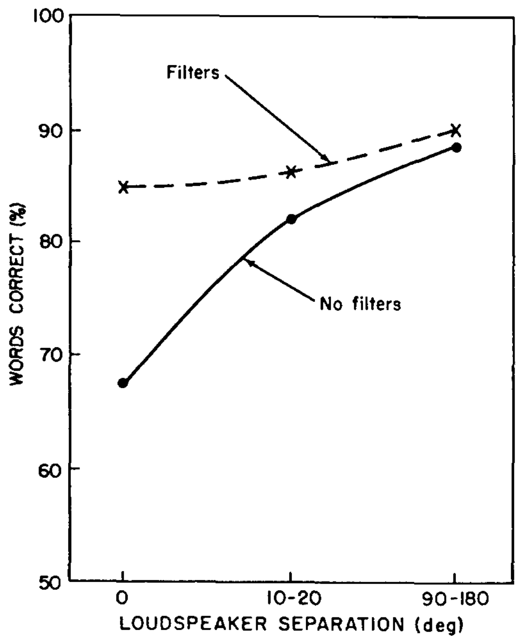


FIGURE 5-41. Effect of angular separation of loudspeakers on multichannel speech intelligibility. Comparing dashed curve with solid curve shows effect of inserting low-pass filter in channel feeding one loud-speaker and high-pass filter in channel feeding the other (Spieth et al., 1954).

5. If feasible, use a message-storage device for all incoming messages. Use a separate recording channel for each communication channel and set up switching arrangements for message review. The oldest messages could be automatically erased to make recording space available for new messages.

5.6.2 Communication at High Altitudes

Personnel and communication equipment in unpressurized or partially pressurized aircraft will be subjected to low ambient pressures when the aircraft fly at high altitudes. The human voice and earphones and loudspeakers become less efficient generators of sound, and microphones become somewhat less than sensitive at certain frequencies, as the ambient pressure is reduced. These effects become appreciable at altitudes above 10,000 ft.

The effect of reducing the ambient pressure on talker, microphone, and earphones from that sea level to that at 40,000 ft. is shown in Figure 5-42 (Kryter, 1944). The upper solid curve represents the overall acoustic pressure response of the interphone as a function of frequency when tested at sea level.

The dotted curve in Figure 5-42 is the overall acoustic response when the microphone-mask combination was tested at 40,000 ft. while the other components of the interphone remained at sea level. This curve reflects the changes in sensitivity of the microphone as a function of high altitude. The dashed curve is the overall acoustic response with the entire interphone calibrated at 40,000 ft.

The above three curves were obtained with a constant input signal. The fourth, the lower solid curve, shows the drop in voice level to be expected when the speaker, wearing an A-14 oxygen mask at 40,000 ft., talks with the same estimated effort that he would use at sea level.

It can be seen from Figure 5-42 that the sound pressure level of speech reaching the listener's ears is reduced about 25 dB if the effort of the talker and the speech amplifier gain are kept constant. The average talker cannot raise his voice enough to compensate fully for this decrease in level.

Design Recommendations for Communication at High Altitudes

1. Use a pressure-sensitive device to adjust the gain of the amplifier as the ambient pressure changes.
2. Use microphones and earphones designed to have uniform frequency-response characteristics at all altitudes in which they are expected to operate.

5.6.3 Underwater Communication

Underwater hearing. Sometimes frogmen and divers need to speak to one another when underwater. Also, a person underwater outside a vessel might need to communicate with people inside the vessel. In both cases, communication is beset by problems of listening and talking underwater.

Hearing underwater is limited by (a) reverberation of sound caused by reflections off the bottom, surface, and such gradients and discontinuities in the water as are due to thermal and salinity conditions, micro-organisms, fish, etc.

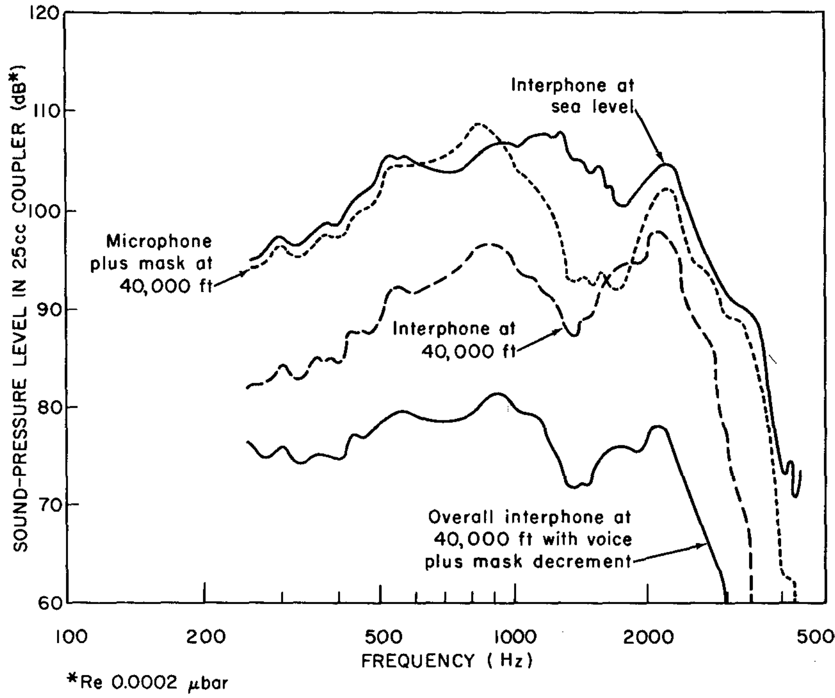


FIGURE 5–42. Frequency response characteristics of microphone and interphone at sea level and at 40,000 ft (Kryter, 1944).

(b) noises made by movement of the water, fish, vessels, and the listener himself, (c) an increased absolute threshold of hearing. The threshold is raised about 40 dB by the impedance mismatch between the water and the air in the middle ear. The following are design recommendations for listening underwater:

1. Use an amplifying system and an underwater transducer (loudspeaker) to compensate for the increased threshold of the listener's hearing.
2. Use a directional receiver to discriminate against reverberation and noise coming from other directions.

Talking underwater. Underwater, it is more difficult to talk than it is to listen. Noise accompanying emitted bubbles tends to mask speech. Consequently, the listener might hear little but vowel sounds.

If a diver has to hold the end of his air hose between his teeth he cannot move his jaw or close his lips freely when he speaks. His pronunciation of consonants is generally poor, and he cannot use some sounds e.g., "b" (as in bat) and "p" (as in pat), at all. The following recommendation will be helpful for underwater talking design:

1. The diver should be provided with an air-exhaust tube that is long enough to delay the release of the air bubble into the water until after the word or phrase is completed.
2. The diver should be instructed to wait before saying the next word until the bubble that goes along with the previous word escapes and the bubble dies away.
3. Vocabulary restriction, message simplicity, and a vowel sound code are three language controls.
4. If possible, provide the diver with a full-face mask that contains a microphone connected to an underwater amplifier and transducer.

Communication in Deep Sea Diving Bells or Diving Suits

In order for man to operate for extended periods of time without suffering so-called nitrogen bends in a deep sea diving bell or diving suit where the pressure of the atmosphere breathed by the man is comparable to the pressure of the surrounding water, it is necessary that the nitrogen content normally present in air be replaced by another gas. Nitrogen must be replaced by some inert gas which will provide proper gas tension but will not react physiologically as readily as nitrogen with tissues of the body. Helium is typically used as a replacement for nitrogen in the air used by deep sea divers or persons living in deep underwater chambers or diving bells. Research is being conducted with other inert gases such as neon. The weight of neon is similar to that of nitrogen and avoids the so-called Donald Duck or helium speech.

Helium speech is less intelligible to persons attempting to communicate within or from deep chambers of diving suits; a decrease from about 70% of special test sentences correct in normal air to about 50% of the sentences correct in a mixture of 81% helium, and 19% oxygen (Sergeant, 1963). The difficulty stems from the fact that although the fundamental pitch of the voice will remain about the same, the higher frequency formants of speech which are formed in the resonance chamber of the throat, nose, and mouth, are shifted upwards by a measured factor of about 1.51 to 2.5 times the component frequency of the formants. This upward shift follows from the fact that the resonance frequency of a cavity is, other things being equal, a function of particle velocity of a gas which is greater for helium than for the nitrogen in normal air. Increased pressure (depth) has increasing distortion effects.

A downward shift proportional to the upward shift caused by the helium, a factor of about 0.65 at each frequency, helps restore the speech signal intelligibility, although, such a processing does not completely remove all distortions. This division can be accomplished by playing back a recorded speech signal at a speed slower than that used for the recording (Holywell and Harvey, 1964). While such a system restores most of the intelligibility that is lost, it is not a procedure that allows practical person-to-person speech communications because of the time delay. Unfortunately, division in real time of a frequency spectrum appears to be rather difficult in the present state of electronic art. However, a heterodyne system, used in conjunction with a number of relatively narrow passband filters for dividing the speech spectrum into parts (Kryter, 1960), would enable one to shift each narrow band of speech downward by an amount proportional to its center frequency; such a system should provide a relatively simple procedure for achieving in real time an approximately normal speech signal from helium speech, particularly for that portion of the speech signal at and near the center frequency of each narrow passband filter. Other real-time systems also exist.

5.6.4 Communication Through Masks

Speech communication is often of great importance in situations requiring the wearing of gas masks. Yet, because of gas mask construction, communication is difficult; the wearer must talk through the mask, whether face-to-face or to a microphone. The distortion and attenuation introduced by the mask reduces both speech intelligibility and the audible distance over which it can be heard. Various masks' effects on speech intelligibility are illustrated in Figure 5–43 (Egan et al., 1943).

Most of the sound is transmitted through ordinary masks by way of the exhaust valve. To improve the sound transmission and divorce it from the valve action, masks may be equipped with diaphragms especially designed to transmit speech. Masks A and B of Figure 5–43 were equipped with such diaphragms; mask C was not.

5.6.5 Digitized Speech and Speech-Compression Systems

When designing an electronic system for speech transmission, the engineer attempts to preserve the waveform of the speech signal with as much fidelity as possible, as seen in the upper left-hand corner of Figure 5–32. Speech systems usually transmit the signal in analog form maintaining its true amplitude as a continuous function of time. However, it is sometimes advantageous to quantize the speech signal both in time and in amplitude not only to maximize the peak power capability but to permit processing for various exotic purposes such as "scrambling." Thus, the signal is unintelligible if intercepted prior to "unscrambling." This time–amplitude quantization results in a "digital signal" which in turn is typically coded into a binary form.

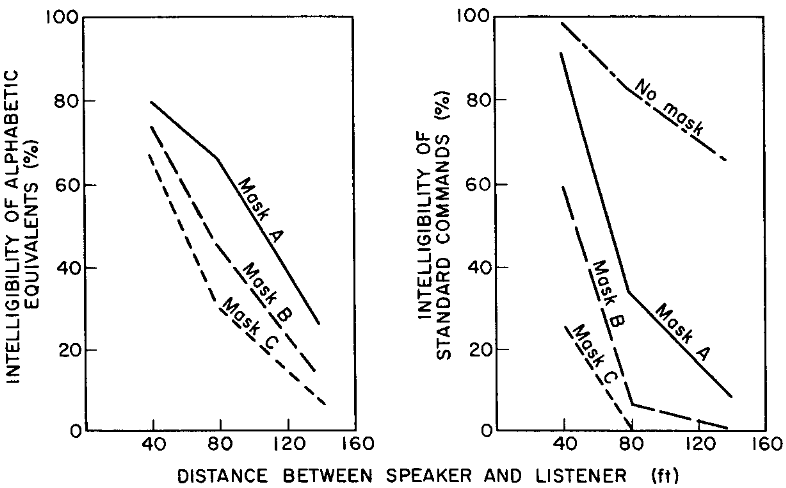


FIGURE 5–43. Effect of various gas masks on intelligibility of speech.

The number of times per second for determining the signal's waveform amplitude to reconstruct with fidelity the analog waveform at the receiving end of the transmission system is a function of: (a) the bandwidth of the signal, in the case of undistorted speech, about 5900 Hz; and (b) the range of amplitude variations in the signal which, for practical purposes, is taken to be 30 dB. The channel capacity, H, in bits per sec., required to transmit perfectly intelligible speech in time-and-amplitude quantized form is, according to the Shannon formula:

$$H = B \log_2\left(1+\frac{S}{N}\right) \text{ bits/sec.,} \qquad (5\text{–}3)$$

where B is the channel bandwidth, and S and N represent signal and noise power respectively. One "bit" is defined as the information at one moment of time, at a given level of intensity when the information is binary in form (that is, the information is no more nor less than the presence or the absence of an event, usually expressed as an impulse of energy). Table 5–30 shows the effect of band limiting the analog speech signal and Table 5–31 shows the effect of noise on AI, percent PB word intelligibility, and channel capacity in "bits per sec." To remain free from distortion due to digitization a noise-free analog speech signal must be sampled in the time domain at twice (2 bits) the rate of the highest frequency of its bandwidth (5900 Hz), and in the amplitude or intensity domain at about 32 levels (5 bits) of equal importance to intelligibility, for a total bit rate of 59,000 per sec. To achieve amplitude steps of "equal im-

TABLE 5–30. CALCULATION OF POINTS FOR BL CURVE IN FIGURE 5–26.

Frequency band (Hz)	AI	Channel capacity (bits/sec)	PB-word intelligibility (percent)
200–6100	1.0	59,000	98.5
200–4250	.9	40,500	96.5
200–3200	.8	30,000	93.5
200–2500	.7	23,000	89.0
200–2020	.6	18,200	83.0
200–1660	.5	14,600	75.0
200–1310	.4	11,100	61.0

Kryter (1960).
*A signal-to-noise ratio of 30 dB is assumed.

TABLE 5–31. CALCULATION OF POINTS FOR SN CURVE IN FIGURE 5–24.

S/N (dB)	AI	Channel capacity (bits/sec)	PB-word intelligibility (percent)
30	1.0	59,000	98.5
27	.9	53,100	96.5
24	.8	47,200	93.5
21	.7	41,300	89.0
18	.6	35,400	83.0
15	.5	29,500	75.0
12	.4	23,600	61.0

Kryter (1960).
*A bandwidth of 5900 Hz is assumed.

portance" to intelligibility requires that rather than divide the total range of amplitudes (30 dB) covered by a speech signal uttered at a constant level of effort into evenly divided levels or steps, one divides the amplitude range into 128 steps (7 bits) but allows the transmission of information about a change of amplitude between successive samples of the speech waveform only up to 5 bits. This form of amplitude-following, called "delta" coding, is possible and effective because it is found that between immediately successive samples of the speech waveform the level will not, or at least will only seldom, change from one amplitude extreme (0 dB) to the other amplitude extreme (30 dB) so that it is therefore not necessary to provide a system capable of sensing up to 7 bits in a single step, a step change of up to 5 bits between successive samples has been found to be sufficiently large for level digitalization of speech.

Bandwidth compression. Expressing information channel capacity of a speech communication system, whether it be analog or digital in form, in bits per sec. is useful for design specification. It also helps compare the efficiency of speech bandwidth compression devices.

The need for compression of normal bandwith or channel capacity stems from inadequacy of frequency space available on telephone or radio-telephone systems. This is particularly true when the speech signal has been converted from analog to digital form, inasmuch as only 2400 bits/sec. can be transmitted over a telephone channel having a frequency bandwidth of about 3000 Hz. Undistorted speech would require about 59,000 bits/sec. for completely noise-free digitalization.

Meaningful bandwidth or channel capacity compression can be achieved by one of four methods or by combinations of these methods: i.e. sampling: (a) the total speech signal in the temporal domain; (b) the speech signal in the frequency domain; (c) the speech signal in the amplitude domain; or (d) information about the amplitude-frequency pattern in time within specially chosen, relatively narrow frequency bands within the speech signal.

Temporal compression. The time-compression method works, because it is possible to remove small segment of the signal or the quiet interval in speech without reducing intelligibility. Two techniques have been used with analog speech signals. One system, (Schiesser, 1949, Fairbanks et al., 1954) involves a system of magnetic tape recording pickups:

Stage 1. Rotating magnetic tape pickup heads are passed over a speech message pre-recorded on magnetic tape that is moving in the same direction of rotation as the pickup heads. These rotating heads pick up and re-record as temporally contiguous every other segment of speech and "skip" the intermediate segments. The pickup heads are moved over the pre-recorded magnetic tape at a speed slower than the duration of the original message by an amount proportional to the total duration divided by the duration of the segments not re-recorded. At this stage, the frequency of the speech components have been proportionately reduced below their normal frequency. At the present time, there are several electronic devices which achieve the same effect.

Stage 2. The recording of compressed speech is played back at a speed faster than that used during the re-recording process so that the spectrum of the speech is returned to normal, but the total message is now of shorter duration.

A second method (Garvey, 1953a) of obtaining time-compressed speech is to mechanically cut alternate portions from a magnetic tape on which speech has been recorded, and then to splice together the remaining sections. It has been found that searching for and eliminating, with normal connected discourse, only the silent intervals or portions of the longer duration vowel sounds, does not achieve more intelligible speech than does a systematic elimination in time of segments of the speech signal.

A reduction of more than about 10% in the time segments of the speech signal causes a measurable decrease in speech intelligibility (Fairbanks and Kodman, 1957).

This temporal saving can be utilized in one of three ways:

1. In a system where a number of separate, but simultaneously occurring, messages are to be transmitted, the messages can be interleaved or multiplexed in time. In communication situations where multiplexing of a large number of speech signals is practical, 10 messages could possibly be multiplexed without too noticeable a degradation in intelligibility over a 9-channel

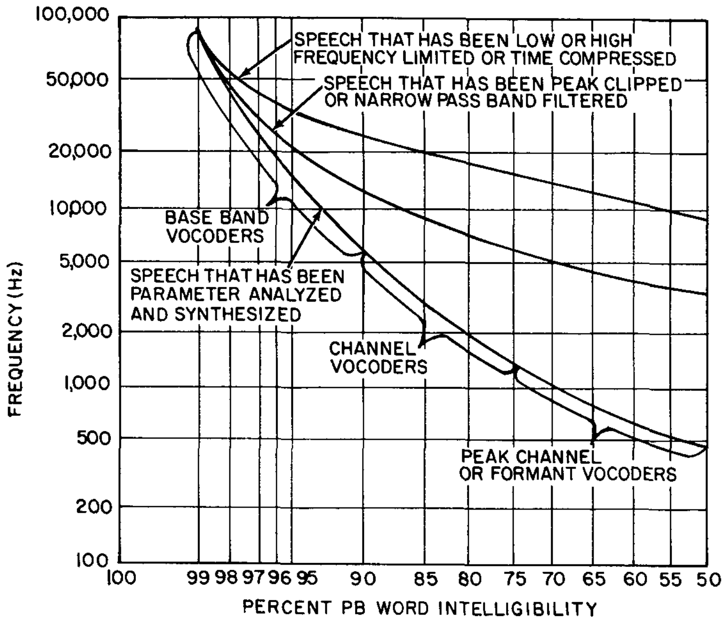


FIGURE 5–44. Speech-compression system chart showing the channel capacity required to transmit, with a given approximate level of PB Word Intelligibility, speech processed by various compression systems (after Stevens, 1960).

transmission system. (This type of time-compression multiplexing is not to be confused with a message multiplexing system used by the Bell Telephone System called "TASI" (Time Assignment Speech Interpolation). In this system an electronic device scans a number of telephone channels that are being used simultaneously and transmits only those channels where there is an actual speech signal present; it is found that because of the natural pauses that occur in telephone conversations it is possible to transmit all the words being uttered on, say, ten telephone links that are in use over a telephone cable that has a normal capacity of say, only eight channels. The TASI system, however, unless it in turn attempts to switch among too many message channels, does not time compress in any way the speech signal from any of the channels.

2. A given speech signal that has been only partially sampled in time can be "pushed" together in time so that the total elapsed time for a message is shortened by a period of time equal to the sum of the temporal intervals removed from the original signal. The latter application has been applied occasionally to the preparation of "commercials" prepared for radio and television and has been proposed as a means of increasing the rate at which recorded lectures or instructions are presented to auditors.

3. Inasmuch as the duration of the time sampled speech has been reduced its frequencies can be proportionally reduced by stretching the signal over the time occupied prior to time compression. This is done automatically in Stage 1 of the system described in 1 above and could be achieved with the tape from which segments have been cut by a deliberate slowing of a playback system. This process would, of course, transfer the saving achieved in the time domain into the frequency domain where the saving would be reflected as a reduction in the analog bandwidth of the signal. However, if bandwidth compression is the goal desired, band limiting or filtering of the speech signal is about equally effective for a given level of intelligibility and is easier to achieve (Figure 5–44).

Frequency bandwidth reduction. In terms of equipment required, the simplest way to achieve modest degrees of bandwidth compression is to pass the speech signal through electronic filters. The effect on speech intelligibility of simple high pass and/or low pass filtering is illustrated in Figures 5–22 and 5–44, and Table 5–31.

A greater effective bandwidth reduction can be achieved by narrow passband filtering of the speech signal, rejecting alternate bands of the signal (Kryter, 1960). By appropriate heterodyne and/or digitizing procedures the relatively narrow bands can be transmitted with a channel capacity suitable for a speech signal having a bandwidth equal to the sum, taken at the 30-dB downpoints, on the filter skirts, of the narrow passbands to be transmitted. Upon reception the narrow passbands of speech must be transposed to their proper position by further filtering and heterodyning before presentation. (See Figure 5–44.)

Table 5–32 gives the bandwidth and location on the frequency scale of a narrow passband filter system. The system noted in Table 5–32 has been found to provide speech intelligibility of about 80% in response to PB words.

TABLE 5–32. NARROWPASSBAND SYSTEM FOR REDUCTION OF BANDWIDTH OF SPEECH.

Upper-lower cutoff frequency (6-dB downpoints) (Hz)	Upper-lower cutoff frequency (30-dB downpoints) (Hz)	Band center frequency (Hz)
250– 350	225– 375	300
550– 650	525– 675	600
850– 950	825– 975	900
1500–1600	1475–1650	1500
2075–2175	2050–2200	2125
3125–3225	3100–3250	3175

Kryter (1960).
Note: Total bandwidth = 900 Hz (30-dB downpoint).

Amplitude compression. As previously noted, about 5 bits (or 32 steps of equal importance) in the amplitude variations of the speech signal must be maintained for the faithful transmission of speech. Allowing the speech wave to take only two amplitude conditions, either positive or negative, results in "infinitely" peak-clipped speech. The resulting signal, provided it is sampled in time and sufficiently often, retains considerable intelligibility (Licklider, 1950). Infinitely peak-clipped speech has an unpleasant sound. Decreasing the amount of peak-clipping from infinite clipping and providing the transmission of more information about intermediate amplitude levels achieved by the speech waveform provides improved intelligibility, as shown in Figure 5–44.

Bandwidth reduction by selection of information within narrow speech bands. The greatest amount of reduction, for a given level of intelligibility, in the bandwidth capacity required to transmit speech is achieved by a class of systems called vocoders. The basic vocoder (Dudley, 1936) attempts to determine the fundamental pitch of the talker as a continuous function of time, which may vary but a few hundred cycles. It also determines the slowly varying amount of total energy within about 18 narrow frequency bands covering the total speech frequency range. This information is transmitted to a special device, the speech synthesizer. A speech signal is synthesized by applying the transmitted information to a "buzz' (pitch) generator and to the gain controls of 18 or so narrow passband filters. A broadband noise generated in the synthesizer is applied continuously to inputs of the 18 passband filters.

Although vocoders have an artificial sound, due to inaccuracies in vocal pitch detection and reproduction, they have been developed to the stage where highly intelligible speech, about 85% correct of PB words, is achieved with a channel capacity of but 2400 bits per sec, see Figure 5–44. By transmitting intact a lower frequency band of speech which contains vocal pitch but requires additional channel capacity, it is possible to achieve a speech signal that sounds nearly normal. This type of vocoder is called the Base Band Vocoder (Flanagan, 1960). (See Figure 5–44.)

A number of vocoder-type systems have been developed in attempts (Fant and Stevens, 1960) to further reduce the channel capacity required to transmit intelligible speech. These systems essentially transmit, as does the vocoder, pitch information from the original speech signal and, at any one period of time, information from only a few of the narrow bands into which the original signal has been filtered. These bands, and the information to be transmitted, are selected on the basis that most speech sounds or phonemes contain energy maxima in but two or three narrow, shifting frequency bands. The logic used for the determination of the loci on the frequency scale of these two or three most important, narrow portions of the speech signal has been:

(a) to find the three or so narrow bands in a bank of narrow passband filters where the energy at a given moment is at a maximum compared to the other bands (this system is called the "peakpicker" vocoder (Peterson and Cooper, 1957)); or (b) a set of rules whereby the analyzer selects those bands which, on the basis of theory concerning the way in which the speech signal is formed in the vocal tract, should contain the most significant information about the speech signal at each moment of time. These systems are usually called formant vocoders (Fant, 1960). Unfortunately, "peakpicker" and formant vocoders, while capable of operating with transmission systems having rather small channel capacities, are not at the present state of the art able to provide highly intelligible speech, as indicated in Figure 5–44.

Combining bandwidth compression techniques. Some of the techniques described above operate in different domains or on different dimensions of the speech signal. The possibility exists that by combining these procedures an effective, efficient speech compression system would result.

Speech that is time-compressed could be band-limited or narrowband filtered, or even peak-clipped before submitting to a vocoder for analysis and resynthesis. However, past attempts at combining speech compression techniques have been unsuccessful. The intelligibility of the speech signal has been drastically reduced below the level achieved with the application of either technique alone. The intelligibility of speech appears to be maintained in several but independent ways within the time–amplitude waveform of the speech signal. Removing a dimension of information to achieve an effective reduction in channel capacity reduces that aspect of intelligibility, and makes the speech signal more vulnerable than normal to other distortions or processing.

5.7 Human Factors Affecting Communication Systems

5.7.1 Personnel Selection and Training

A combination of heavy information loads and poorly trained personnel can be as hard on system performance as are noise and distortion. Successful communication depends on vocabulary, message set size, and degree of standardization, as well as familiarity with message and equipment. These factors in turn depend on selection and training of operators. Large differences in fundamental intelligibility among individual talkers and listeners tend to persist even through practice and training, although proper training can improve performance.

Trainability of persons to understand distorted or processed speech. Personnel working in high noise fields, or with communications equipment that itself adds noise, are apparently able to hear and understand messages the novice finds completely unintelligible. They have learned to identify the slight differences that exist between speech and noise. The same applies to persons using speech compression systems where certain distortions are introduced into the speech signal.

A person with training can better comprehend severely time-compressed speech, than can an untrained person (Orr et al., 1965), and trained listeners do better in understanding "helium" speech than do untrained listeners (Sergeant, 1963). The functions in Figure 5-44 are based on intelligibility scores for listeners who were about equally trained with each of the systems tested.

Bearing in mind that if system performance is better with little or no training on the part of its personnel, it will likewise be better following a great deal of training. However, it is not wise to rely only upon training as a method for improvement unless there are severe cost, channel capacity, or time constraints. Not only are relatively untrained users required to operate the communications systems from time to time, but the "margin of safety" for satisfactory communications on the part of trained observers becomes dangerously small.

5.7.2 Language Factors

Message intelligibility is greatly affected by the language used, by the kind and set of messages that are sent. Important language factors to be considered are:

Information content of words. Other things being equal, the more frequently a word occurs in everyday usage, the more readily it is correctly identified when transmitted over a speech-communication system. As Figure 5–45 shows, the length of the word also influences its intel-

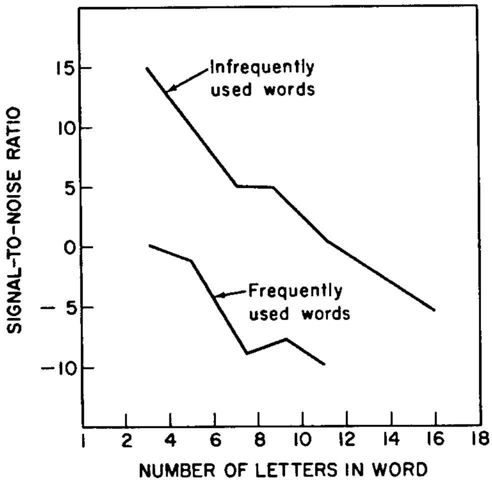


FIGURE 5–45. Relation between length of word (number of letters) and signal-to-noise ratio required for 50 percent intelligibility. Infrequently used words require higher signal-to-noise ratio than frequently used words when word length is held constant (Howes, 1957).

ligibility—the longer the word, the more readily it is correctly identified. The listener is able to identify a long word by hearing portions of it, particularly a familiar probable word, whereas missing one syllable of a short word is more likely to prevent the identification of the entire word. Both frequency and length factors can change by 10 to 15 dB the signal-to-noise ratio required for a given level of intelligibility.

Sentence or phrase structure. A similar effect occurs as words are formed into phrases and sentences. Listeners comprehend meaningful messages at signal-to-noise ratios, or with distortions, that they would be unable to understand if the message consisted of an equal number of unrelated words. This fact is reflected in Figure 5–5; under conditions that permit correct reception of about 75% of PB words, over 95% of test sentences are correctly understood.

The size of the message set. The size of the set or ensemble of messages is related to frequency and familiarity factors: the smaller the set, the easier for communication.

To take advantage of limited message-set size, both talker and listener must know the set thoroughly. Useful rules to follow are: (a) deliberately set limits to message and form, (b) include digits and (c) short, well-known highly stereotyped phrases, and (d) send what conforms to the listener expectations.

Situational constraints. Although not fully quantified, communication system performance is influenced by the "extra" information that is contained in *a priori* knowledge of various aspects of the prevailing situation. In landing, a pilot approaching a control tower is helped in the operational context to "hear" and understand messages concerned with landing instructions, and to know the particular arrangement of the air field, the wind conditions, and the time of day.

Interaction of speech sounds. Situations arise in which the noise level is so great, the time available for communication so brief, or the requirement for communication so critical, that it is necessary to specify the elemental speech sounds to be used in building a set of messages.

Fundamental speech sounds and symbols. The individual letters of the alphabet are not adequate for identification of the fundamental speech sounds because most letters of the alphabet cán be pronounced in any of several ways. (Compare the sounds of the "a" in "apply," "may," and "any."). Accordingly phoneticians have devised a special set of phonetic symbols refering to basic sounds used in spoken language. Table 5–33 (Fletcher, 1953) lists and defines, by "key" words, these phonetic symbols and relates them to the dictionary symbols (letters with diacritical marks).

Confusion among speech sounds. Figure 5–46 (Miller and Nicely, 1955) shows a "tree of confusion" characteristic of a wideband communication system that is limited only by random background noise at the *signal-to-noise* (*S/N*) *ratios* indicated at the left-hand side of the illustration. Note the following:

1. At S/N= −18 dB, all consonants are confused with one another.

2. At S/N= −12 dB, the voiced consonants (m, n, d, g, b, v, z, and ʒ) are confused with one another, and the unvoiced consinants (t, k, p, f, θ, and s) are confused with one another, but the consonants in the first group are seldom confused with those in the second group.

3. At S/N= −6 dB, although the "m" and "n" are confused with each other, they are clearly distinguishable from the other consonants.

TABLE 5–33. PHONETIC SYMBOLS

Key word	Dictionary symbol	BTL* symbol	IPA† symbol	Key word	Dictionary symbol	BTL* symbol	IPA† symbol
t oo l	o͞o	ū	u	m ate	m	m	m
t o ne	ō	ō	o or ou	n ate	n	n	n
t a lk	ä	ó	ɔ	si ng	ng	ng	ŋ
t a r	ȧ	a	a	v oice	v	v	v
t a pe	ā	ā	e or ei	f un	f	f	f
t ea m	ē	ē	i	z ero	z	z	z
t oo k	oo	u	ʊ	s it	s	s	s
t o n	u	o	ʌ	th en	th	th	ð
t a p	a	á	æ	th in	th	th	θ
t e n	e	e	ɛ	a z ure	zh	zh	ʒ
t i p	i	i	I	sh ip	sh	sh	s
t i me	ī	ī	aI	b at	b	b	b
t ow n	ou	ou	aʊ	p at	p	p	p
t oi l	oi	oi	ɔI	d en	d	d	d
f e w	ū	ew	ju	t en	t	t	t
w oo	w	w	w	j udge	j	j	dʒ
y ou	y	y	j	ch ur ch	ch	ch	ts
h ow	h	h	h	g oat	g	g	g
l ate	l	l	l	c oat	k	k	k
r ate	r	r	r				

*Bell Telephone Laboratories.
†International Phonetic Association.

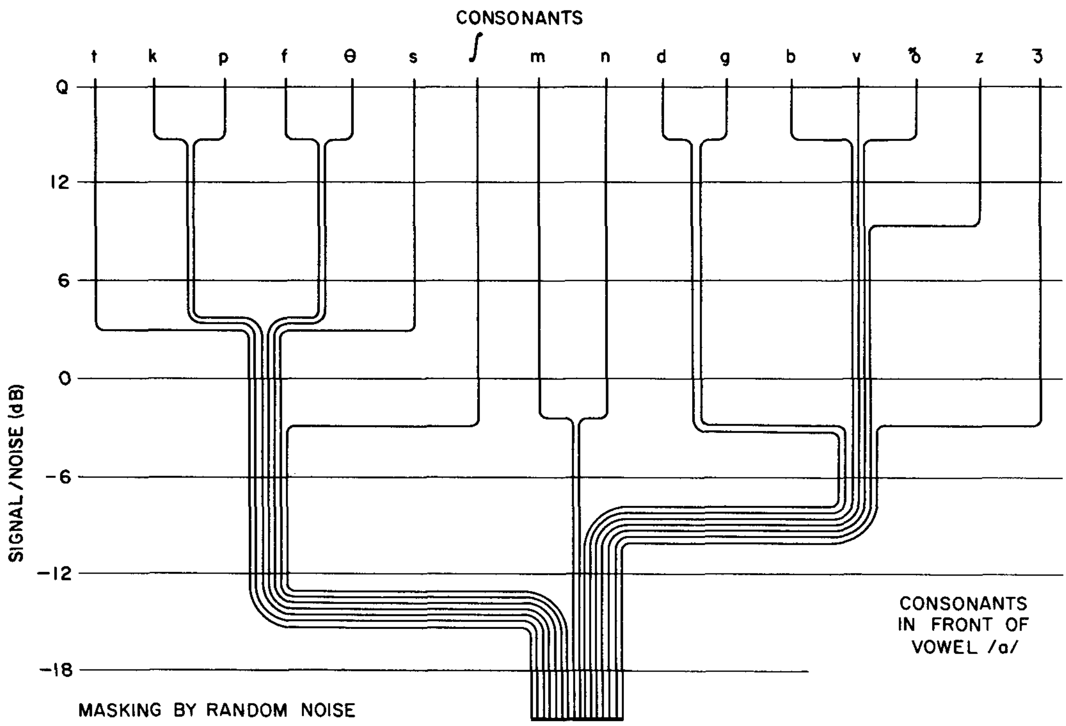


FIGURE 5–46. Confusion among consonants because of masking by random noise (Miller and Nicely 1955).

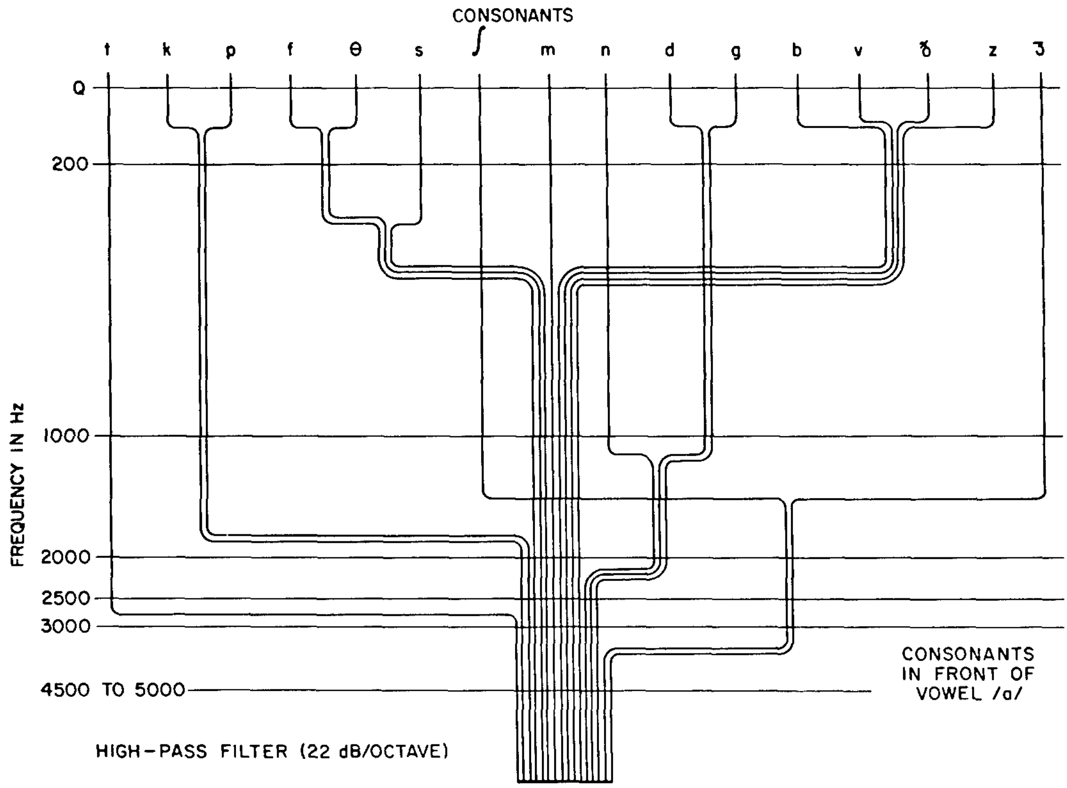


FIGURE 5–47. Confusion among consonants because of high-pass filtering of speech (Miller and Nicely, 1955).

4. At S/N = 0 dB, there are seven groups of easily recognizable consonants.
5. At S/N = 12 dB, all of the consonants are readily distinguished.

Figures 5–47 and 5–48 (Miller and Nicely, 1955) show similar trees of confusion for communication channels in which there is high-pass and low-pass filtering, respectively. In general, the wider the spectrum passed, the greater the intelligibility. Conversely, the greater the amount of filtering (either high-pass or low-pass), the more frequent the confusion.

Word-spelling alphabets. Even in extremely noisy conditions, or bad audio circuits, communication can be maintained by spelling the words out with a word-spelling alphabet, a technique that works well, but is time consuming.

The word-spelling alphabet given in Table 5–34 (Moser and Bell, 1955) is used by the International Civil Aeronautics Organization and the U.S. Armed Forces. The product of extensive research, it is designed to yield high intelligibility when used by talkers and listeners of any nationality represented in NATO (North Atlantic Treaty Organization).

Design recommendations for standard word lists. Standard phrase or word use can mean the difference between poor and good communication. If there are a sufficient number of words and phrases in the message set, standard phraseologies provide a margin of safety for emergency conditions. Nevertheless, a special language or voice procedure should be used only as a supplement to the best possible equipment design. Five rules for constructing standard word and phrase lists for communication follow, in order of importance:

1. Embed critical words in context phrases or sentences. This technique many people use in ordinary telephone conversations. One is much less likely to mishear a sentence like, "Your house is on fire." than the single word "Fire!"
2. Use familiar words rather than unfamiliar ones. The "Teacher's Word Book" (Thorndike

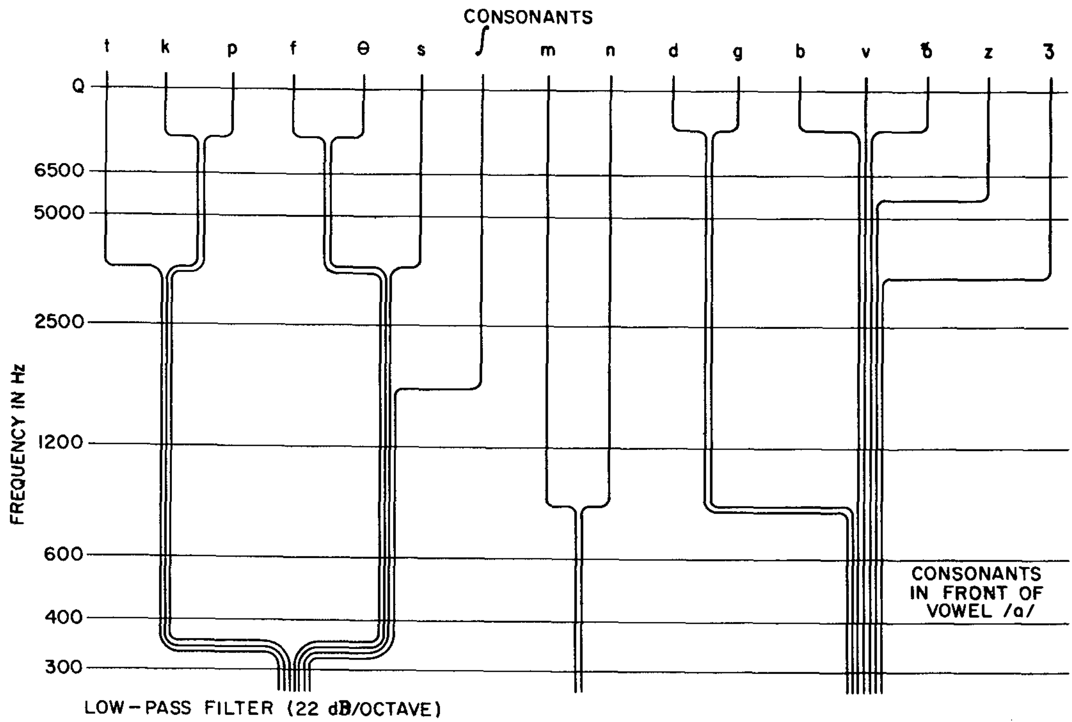


FIGURE 5–48. Confusion among consonants due to low-pass filtering of speech (Miller and Nicely et al., 1955).

TABLE 5–34. INTERNATIONAL WORD-SPELLING ALPHABET

A—Alpha	N—November
B—Bravo	O—Oscar
C—Charlie	P—Papa
D—Delta	Q—Quebec
E—Echo	R—Romeo
F—Foxtrot	S—Sierra
G—Golf	T—Tango
H—Hotel	U—Uniform
I—India	V—Victor
J—Juliet	W—Whiskey
K—Kilo	X—X-Ray
L—Lima	Y—Yankee
M—Mike	Z—Zulu

Moser and Bell (1955).

and Lorge, 1944) lists the 30,000 most common English words in order of frequency of use.

3. Use as small a total vocabulary as possible. If the listener knows that a critical word can be one of only, for example, eight possible words, he is more likely to hear it correctly.

4. To obtain words that are easily distinguished, select polysyllables. The words "negative" and "affirmative," although used less often than "no" or "yes," are much less likely to be misunderstood.

5. Avoid words that contain confusing sounds.

5.7.3 Feedback to the talker

Feedback is important in design for the following reasons:

1. The talker can be sure that the system is operating properly. An example is the telephone dial tone; it tells the caller that the system is prepared to accept his call. The ringing sound after a number has been dialed tells the caller that the system has accepted and is processing the call.

2. The talker can be sure that the listener is receiving the message.

3. The talker can be more confident that his message is correctly understood or can ask for repetition.

4. The talker can be sure that appropriate action is taken. An example is the helmsman advising the officer of the deck that he has carried out a command to change bearing.

There are other advantages in two-way communication, particularly in emergency situations; morale of both talker and listener tends to be higher. If separate pieces of information are being held at different stations, problems can be solved more efficiently.

References

Beranek, L. L. The design of speech communication systems. *Proceedings of the Institute of Radio Engineers*. 1947, **35,** 880.

Beranek, L. L. *Acoustic measurements*. New York: Wiley, 1949.

Beranek, L. L. Revised criteria for noise in buildings. *Noise Control*, 1957, **3**(1), 19.

Carter, N. L., & Kryter, K. D. Masking of pure tones and speech. *Journal of Auditory Research*, 1962, **2,** 66–98.

Dudley, H. W. The vocoder. *Bell Laboratories Record*, 1936, **18,** 122–126.

Dunn, H. K., & White, S. D. Statistical measurements on conversational speech. *Journal of the Acoustical Society of America*, 1940, **11,** 278.

Egan, J. P. Articulation testing methods. *Laryngoscope*, 1948, **58,** 955.

Egan, J. P., & Weiner, F. M. On the intelligibility of bands of speech in noise. *Journal of the Acoustical Society of America*, 1946, **18,** 435.

Egan, J. P., et al. The effects of Army gas masks on speech communication. OSRD Report No. 1816, 1943, Harvard Psycho- Acoustic Lab., Harvard University, Cambridge, Mass.

Fairbanks, G. Test of phonomeic differentiation: The rhyme test. *Journal of the Acoustical Society of America*, 1958, **30,** 596–600.

Fairbanks, G., & Kodman, F., Jr. Word intelligibility as a function of time compression. *Journal of the Acoustical Society of America*, 1957, **29,** 636–641.

Fairbanks, G., Everitt, W. L., & Jaegar, R. P. Method for time or frequency compression-expansion of speech. *IRE Transactions on Audio* **AU 2,** 1954, 7–12.

Fant, C. G. M. *Acoustic theory of speech production*. 's Gravenhage (the Hague): Mouton Co., 1960.

Fant, C. G. M., & Stevens, K. N. Systems for speech compression. *Fortschritte der Hochfrequenztechnik*, **5,** Akademische Verlagsgesellschaft m. b. H., Frankfort/a.M., 1960, 229–262.

Farrell, W. R. *Reverberation time criteria*, Cambridge Mass: Bolt Beranek & Newman, 1958.

Kryter, K. D. Study of a speech compression system (spectrum selection). Quarterly Technical Report No. 4, September 1960, for USASRDL under contract No. DA–36–039 SC–78078.

Kryter, K. D. Speech bandwidth compression through spectrum selection. *Journal of the Acoustical Society of America*, 1960, **32,** 547–556.

Kryter, K. D. Standard procedures for the calculation of the articulation index. Report No. ESD–TDR–62–1, 1962, Electronics Systems Division, Air Force Systems Command, L. G. Hanscon Field, Bedford, Mass.

Kryter, K. D. Chairman S3.5, Methods for the calculation of the articulation index, 1969. U. S. American Standards Institute. 10 East 40th Street, New York, N. Y.

Kryter, K. D., & Whitman, E. C. Some comparisons between Rhyme and PB-word intelligibility tests. *Journal of the Acoustical Soceity of America*, 1965, **37,** 1146.

Licklider, J. C. R. Effects of amplitude distortion upon the intelligibility of speech. *Journal of the Acoustical Society of America*, 1946, **18,** 429.

Licklider, J. C. R. The intelligibility of amplitude-dichotomized time quantized speech waves. *Journal of the Acoustical Society of America*, 1950, **22,** 820–823.

Licklider, J. C. R., et al. The intelligibility of rectangular speechwaves. *American Journal of Psychology*, 1948, **61,** 1.

Miller, G. A. The masking of speech. *Psychological Bulletin*, 1947(b), **44,** 105.

Miller, G. A., & Licklider, J. C. R. The intelligibility of interupted speech. *Journal of the Acoustical Society of America*, 1950, **22**, 167.

Miller, G. A., & Nicely, P. E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, **27**, 338.

Moser, H. M., & Bell, G. E. Joint United States, United Kingdom report. AFCRC–TN–55–56–1955, Air Force Cambridge Research Center, Cambridge, Mass.

Orr, D. B., Friedman, H. L. & Williams, J. C. C. Trainability listening comprehension of speeded discourse. *Journal of Educational Psychology*, 1965, **56**, 148–156.

Flanagan, J. L. A resonance vocoder and baseband complement: A hybrid system for speech transmission. *IRE Transactions on Audio* **AU 8**, 1960, 95–102.

Fletcher, H. *Speech and hearing in communication*. New York: D. Van Nostrand Co., Inc., 1953.

Fletcher, H., & Steinberg, J. C. Articulation testing methods, *Bell System Technical Journal* 1929, **8**, 806.

French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 1947, **19**, 90.

Frick, F. C., & Sumby, W. H. Control tower language. *Journal of the Acoustical Society of America*, 1952, **24**, 595.

Garinther, G. R., & Kryter, K. D. Auditory and acoustical evaluation of several shoulder rifles. Technical Memo No. 1–65, 1965, U. S. Army Human Engineering Laboratories, Aberdeen Proving Grounds, Md.

Garvey, W. D. The intelligibility of abbreviated speech patterns. *Quarterly Journal of Speech*, 1953(a), **39**, 296–306.

Hawley, M. E., & Kryter, K. D. Effects of noise on speech. In C. M. Harris (Ed.), *Handbook of noise control*, New York: McGraw-Hill, 1957.

Holywell, K., & Harvey, G. Helium speech. *Journal of the Acoustical Society of America*, 1964, 210–211.

House, A. S., Williams, C. E., Hecker, M. H. L., & Kryter, K. D. Articulation-testing methods: Consonantal differentiation with a closed-response set. *Journal of the Acoustical Society of America*, 1965, **37**, 158–166.

Howes, D. H. On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America*, 1957, **29**, 296.

Kryter, K. D. Articulation test comparisons of six Signal Corps aircraft interphones at low and high altitudes. OSRD Report No. 1974, 1944, Harvard Psycho-Acoustic Lab., Harvard University, Cambridge, Mass.

Kryter, K. D. Effects of ear protective devices on the intelligibility of speech in noise. *Journal of the Acoustical Society of America*, 1946, **18**, 413.

Kryter, K. D. Human engineering principles for the design of speech communication systems. Report No. AFCRC–TR–58–62, 1958, Air Force Cambridge Research Center, Cambridge, Mass.

Peterson, E., & Cooper, F. S. Peakpicker: A band-width compression device. *Journal of the Acoustical Society of America*, 1957, **29**, 777(A).

Pickett, J. M. Effects of vocal force on the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 1956, **28**, 902–905.

Schiesser, H. Device for time expansion used in sound recording. *Funk Und Ton*, Vol. 3, No. 5, 1949: and *IRE Transactions on Audio* **AU–2**, 1954, 12–15.

Sergeant, R. L. Speech during respiration of a misture of helium and oxygen. *Aerospace Medicine*, 1963, **34**, 9.

Spieth, W., et al. Cues that aid listening to one of two simultaneous voice messages. Report No. 446, 1954, U. S. Navy Electronics Lab., San Diego, Calif.

Stevens, K. N. Review of existing speech compression systems. Report No. TN 60–167, October 1960, Rome Air Development Center, Rome, New York.

Stevens, S. S., et al. Methods of measuring speech spectra, *Journal of the Acoustical Society of America.* 1947, **19**, 771.

Sumby, W. H., & Pollack, I. Visual contribution to speech intelligibility in noise. *Journal of the Acoustical Society of America*, 1954, **26**, 212–215.

Thorndike, E. L., & Lorge, I. *Teachers word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

USASI. Octave band filter set for the analysis of noise and other sounds. Report No. Z24.10–1953, U. S. American Standards Institute, 10 East 40th Street, New York, N. Y.

USASI Method for measurement of monosyllabic word intelligibility. Report No. S3.2–1960, U. S. American Standards Institute, 10 East 40th Street, New York, N. Y.

USASI. Preferred frequencies for acoustical measurements. Report No. S1.6–1960, U. S. American Standards Institute, 10 East 40th Street, New York, N. Y.

Wathen-Dunn, W., & Lipke, D. W. On the power gained by clipping speech in the audio band. *Journal of the Acoustical Society of America*, 1958, **30**, 36.

Weiner, F. W., & Filler, A. S. The response of certain earphones on the ear and on closed couplers. Report No. PNR–2, 1945, Harvard Psycho-Acoustics Lab., Harvard University, Cambridge, Mass.

Selected Readings

Fant, C. G. M., & Stevens, K. N. Systems for speech compression. Fortschritte der Hochfrequenztechnik, 5, Akademische Verlagsgesellschaft m. b. H., Frankfort/a.M., 1960, 229–262.

French, N. R., & Steinberg, J. C. Factors governing the intelligibility of speech sounds. *Journal of the Acoustical Society of America*, 1947, **19**, 90.

Hawley, M. E., & Kryter, K. D. Effects of noise on speech. In C. M. Harris (Ed.), *Handbook of noise control.* New York: McGraw-Hill, 1957.

Kryter, K. D. Chairman. Methods for the calculation of the articulation index, S3.5, 1969. U. S. American Standards Institute, 10 East 40th Street, New York, N. Y.

Chapter 6

Man—Machine Dynamics

George Frost

Human Engineering Division
Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio

This chapter identifies and discusses the factors affecting human performance in tracking and watchkeeping (vigilance) tasks and presents design techniques for optimizing the performance of man-machine systems where these tasks are required. Direct analytical techniques are presented wherever possible; otherwise, general design recommendations are made and the available research results are referenced.

CONTENTS

Contributions to this chapter were made by John W. Senders of Brandeis University.

Symbols Used in Chapter 6

- a — Amplitude factor invariant with time
- b — Amplitude factor invariant with time
- c — Damping constant (force/velocity)
- K — Spring constant (force/displacement); gain constant
- m — Mass
- K_p — Human operator's gain
- e — Naperian base 2.71828...; general instantaneous voltage
- e_i — Input voltage
- e_o — Output voltage
- Y_c, $Y_c(s)$ — Transfer function of controlled element
- Y_d, $Y_d(s)$ — Transfer function of display device
- Y_p, $Y_p(j\omega)$ — Describing function of human operator
- T — Time constant; sampling interval
- T_I — Lag time constant adopted by the human operator
- t — Time
- T_L — Lead time constant adopted by the human operator
- x — General random variable
- $R_{xx}(\tau)$ — Autocorrelation function of a random variable, x, evaluated on the interval, τ.
- F_s — Control stick force
- f — Frequency, Hertz
- j — $\sqrt{-1}$; index
- s — Dependent variable for the Laplace transform
- δ — Control deflection
- ϵ, $\epsilon(j\omega)$ — Error signal
- θ_c — Command signal
- θ_o — Output signal
- α — Mid-frequency phase angle coefficient in the extended crossover model of the human operator
- τ — Time interval of correlation; time constant
- τ_e — Effective low-frequency time delay of human operator
- $\Phi_{xx}(\omega)$ — Power spectral density of x
- π — 3.14159 . . .
- ζ — Damping ratio of second order dynamic system
- ω — Circular frequency—radians per second
- ω_c — Unity gain crossover frequency
- ω_i — Input frequency bandwidth
- ω_{ic} — Effective input bandwidth
- ω_n — Undamped natural frequency
- ω_b — Frequency bandwidth in general; "break" frequency

6. Man–Machine Dynamics

6.1 The Closed-Loop Manual-Tracking System

A *closed*-loop system is one in which information about an output is fed back to an earlier stage of the system to stablize or control the output. Thus, closed-loop systems are frequently referred to as *feedback* control systems. In an *open*-loop system, no such feedback is present. Virtually all manual control systems and most real-world monitoring systems are closed-loop systems. Man senses deviant outputs and closes the loop by appling corrective inputs.

The similarity between manual control systems and servomechanisms—feedback controllers—makes the choice of servo terminology natural for discussiong manual control—tracking—systems. Figure 6–1 shows the similarity between a typical autopilot-airframe system and a corresponding manual flight-control system. This similarity is more than superficial. For many non-trivial applications, the techniques of servoanalysis and synthesis are directly applicable to the design of *manual* control loops, whether they be for aircraft, submarines, or radar trackers. These techniques encompass control system *aiding*, dispaly *quickening*, and operator *unburdening* and allow analytical prediction of the effects of proposed system changes.

The foregoing should not be construed to imply that the man in the loop is a servomechanism. The man in the loop is a thinking being and he (usually) behaves in a rational manner. The important point here is that the techniques of servoanalysis do serve to describe many facets of the behavior of a well-trained human operator, and these techniques are the appropriate techniques for manual control system design.

In the following sections of this chapter the emphasis is heavily servo-oriented. The attempt has been made to provide clear examples of each point as it is made and to minimize the number of formulae introduced. This has not, in all cases, been possible. The mathematical tools of servoanalysis are sufficiently powerful that their utility justifies the effort involved in becoming familiar with them. The reader who has not had an introduction to servoanalysis may find the material difficult. It is strongly urged that an introductory text in servoanalysis be used as collateral reading. (Savant, 1958; Milsum, 1966; Thaler and Brown, 1953.)

Figure 6–2 shows a generalized tracking loop. The basic elements are: the *Display,* which is generally visual but may utilize other modalities such as the auditory or tactile senses; the *Man,* who functions as an adaptive error sensor and controller; the *Control,* which serves as a transducer to convert the man's force or displacement outputs into controlled element inputs; and the *Controlled Element,* which lumps together the dynamics of the external element to be controlled and any coordinate transformations for the display. Each of these elements will be treated separately. The basics of control system analysis are discussed; some of the simpler techniques of analysis and synthesis for single-loop systems are presented; also considered are the problems of manual control of multiple-loop systems and of discrete control systems.

In many control systems, man is used as a backup system, as a system monitor, and as an adaptive, redundant element. While training and skill retention are outside the province of this chapter, the designer is cautioned not to neglect these extremely important areas. If a pilot is to be expected to control an aircraft in the event of a stability augmentation failure, means must be provided for him to practice controlling the unaugmented vehicle. Further, information must be presented to him in a manner that allows him to sense rate without depending upon the rate sensor of the failed stability augmentation system.

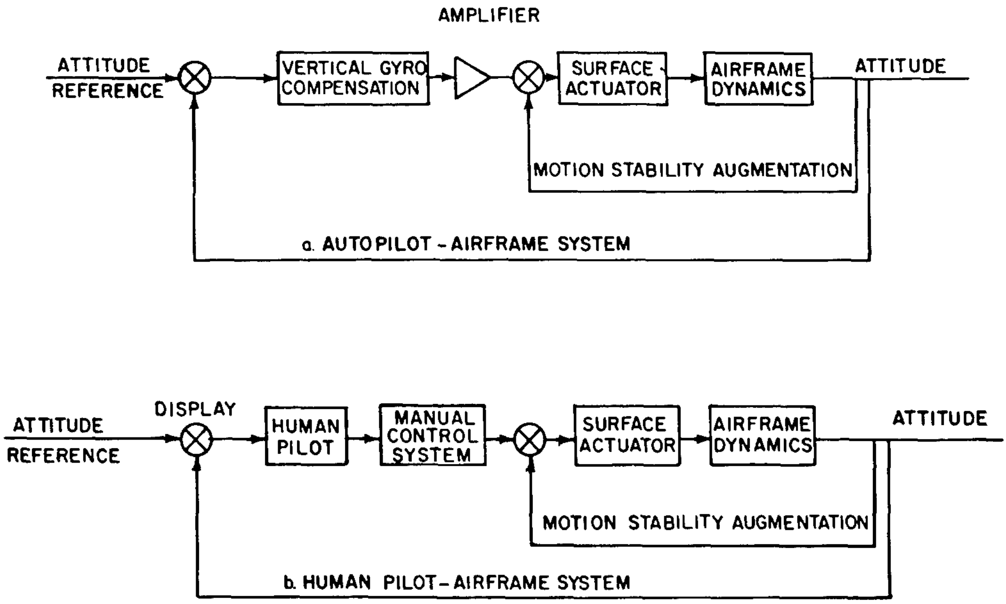


FIGURE 6–1. Comparison between autopilot-airframe systems and manual flight control systems (Westbrook, 1959).

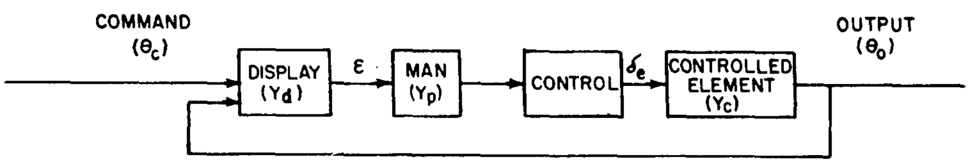


FIGURE 6–2. Generalized tracking loop.

6.2 Elements of the Manual Control System

6.2.1 Types of Input Functions

Typically a manual control system has two types of inputs: command signals, and disturbance inputs due to the external environment. If the operator has a pursuit display, he can separate the command input from the disturbance inputs and the vehicle response. If he has a compensatory display, he can see only the difference or error signal. For example, a pilot when making an Instrument Landing System (ILS) approach makes use of glide-slope and localizer beams. These beams constitute, in this case, the *command signal inputs*. During the approach, additional *disturbance inputs* exist in the form of atmospheric turbulence. It is the "effective" input, i.e., the summation of command and vehicle response to the disturbance, that the pilot actually sees on his display. (In the remainder of the chapter, unless specifically noted, reference to "input function" is to be understood as "effective" input function.) These input signals may be either periodic or nonperiodic nature, and they may be either continuous or transient. Examples of the four cases are given in Figure 6–3 below.

This classification of inputs is somewhat arbitrary since periodic transients can be approximated to any desired degree by properly weighted sums of harmonically related sinusoids. (Harmonically related frequencies are those related by integer multiples. Frequencies of 120 Hz and 180 Hz are the second and third harmonics of 60 Hz.) Conversely, the sum of four or more non-harmonically related sinusoids

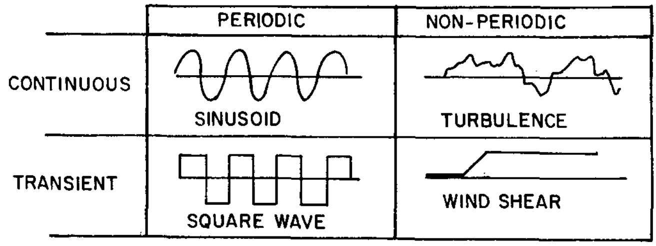


FIGURE 6–3. Types of input functions.

appears quite random to the operator in a tracking system, and a stationary random process can be considered a segment of a periodic signal of infinite period.

Periodic Inputs

Within the classification of periodic inputs are not only simple sine waves but also sums of harmonically related sine waves since these repeat once for each cycle of the fundamental frequency (See Figure 6–4). Sums of non-harmonically related sine waves are not included since, although they are periodic, their period is much longer than that of the lowest frequency. Actually their period is equal to the period of a wave whose frequency is the largest common denominator of the frequencies included. Square waves, triangular waves, sawtooth waves, and recurrent impulses are also classed as periodic inputs although, with the exception of square waves, they are seldom used for inputs to tracking tasks.

In general, an experienced operator will recognize simple periodicity in an input function and adapt his behavior to make use of this additional information. Specifically, the operator learns to anticipate the input so that he can track with virtually no error. He operates in an essentially open-loop mode and depends upon his display only to maintain synchronism with the input. A more thorough treatment of this type of behavior is given in Krendel and McRuer (1960).

The fact that operators do recognize periodicity in input functions and modify their behavior accordingly creates a problem for the investigator of tracking behavior. Specifically, it is impossible to use simple periodic inputs (e.g., sine waves) to determine how an operator will track in an operational situation where the input is in fact random. Many early tracking studies used either sinusoidal inputs or inputs made up of two or three harmonically related sine waves. The results of these studies must be considered with extreme caution since they are generally overly optimistic of the operator's ability to track random inputs.

Aperiodic Inputs (Noise)

Within the classification of aperiodic (non-periodic) inputs, *continuous* inputs must be distinguished from *transient* inputs. Continuous aperiodic inputs can only be described in terms of their statistics. Probably the most useful description of continuous aperiodic or complex periodic inputs is the *autocorrelation* function or its Fourier transform, the *power spectral density*.

The autocorreclation function of a time function $x(t)$ is given by the following formula, where $x^*(t)$ denotes the conjugate of $x(t)$ (See Blackman and Tukey, 1959.)

$$R_{xx}(\tau)=\lim_{T\to\infty}\frac{1}{2T}\int_{-T}^{T}x^*(t)x(t+\tau)dt \qquad 6\text{–}1$$

and the power spectral density is given by

$$\Phi_{xx}(\omega)=4\int_0^\infty R_{xx}(\tau)\cos\omega\tau\,d\tau \qquad (6\text{–}2)$$

Alternatively, the power spectral density is given by

$$\Phi_{xx}(\omega)=\lim_{T\to\infty}\frac{1}{T}[I^*(j\omega)I(j\omega)]=\lim_{T\to\infty}\frac{1}{T}\mid I(j\omega)\mid \qquad (6\text{–}3)$$

where $I(j\omega)$ is the Fourier transform of $x(\tau)$.

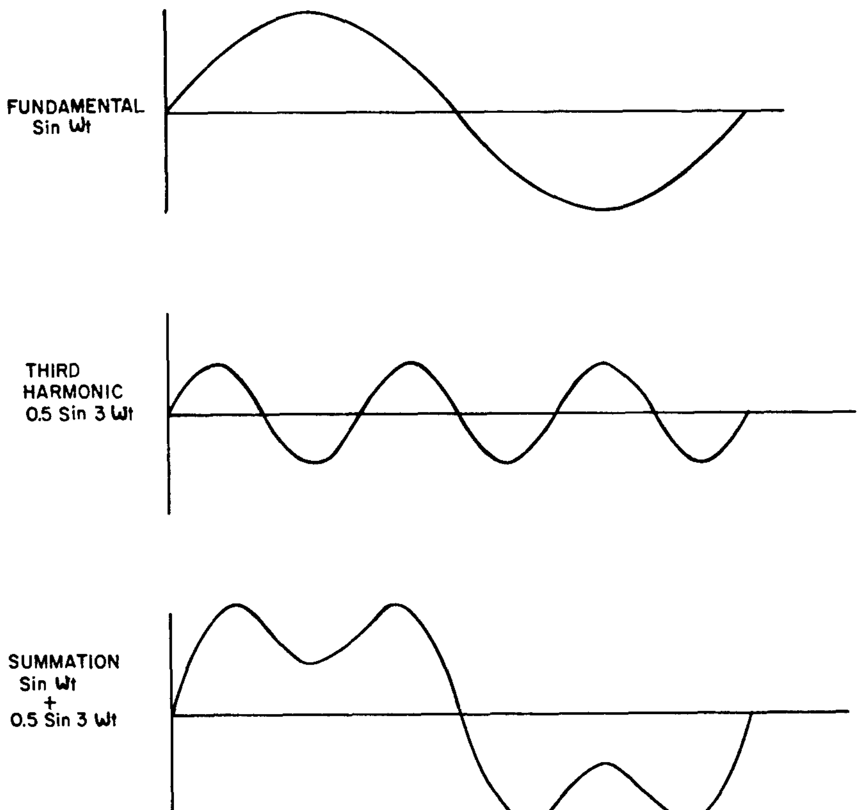


FIGURE 6–4. Summation of harmonically related sinusoids.

To gain an intuitive understanding of power spectral density, consider a random-appearing waveform made up of the sum of non-harmonically related sinusoids. A plot of power spectral density of such a random-appearing waveform appears as a comb, with tooth areas proportional to the power at each frequency, and with teeth knocked out in the regions where no signal power is present. This is shown in Figure 6–5. The magnitudes of the spectral lines in Φ_{ii} are π times the square of the peak amplitudes of the sinusoidal components. In the plot, these are shown in power *dB*, i.e., $10 \log_{10}(\pi a^2)$ where a equals the peak amplitude of each component frequency.

A true random function contains all frequencies rather than just a few as shown above. In most systems, we are not concerned with true random or white noise but rather with noise in a specific range of frequencies—band-limited noise. The line spectrum in Figure 6–5 is an approximation to such band-limited white noise. The power spectral density for a true band-limited noise would appear as the envelope enclosing the line spectrum above. Important characteristics of such band-limited noise are:

1. The attenuation rate of the upper frequencies, usually expressed in *dB*/octave or *dB*/decade. This is simply the slope of the high frequency cut-off line. (See Figure 6–5.)

2. The "corner" frequency; i.e., the frequency at which a line representing the upper frequency slope intercepts a horizontal line along the flat portion of the spectrum. This is identified as ω_b in Figure 6–5. Note that this is *not* the lowest frequency that is attenuated.

3. The average power in the spectrum below the corner frequency; or alternatively, the r.m.s. value of the function at the display, assuming no operator input.

For predicting the performance of a manual control system, the concept of a rectangular

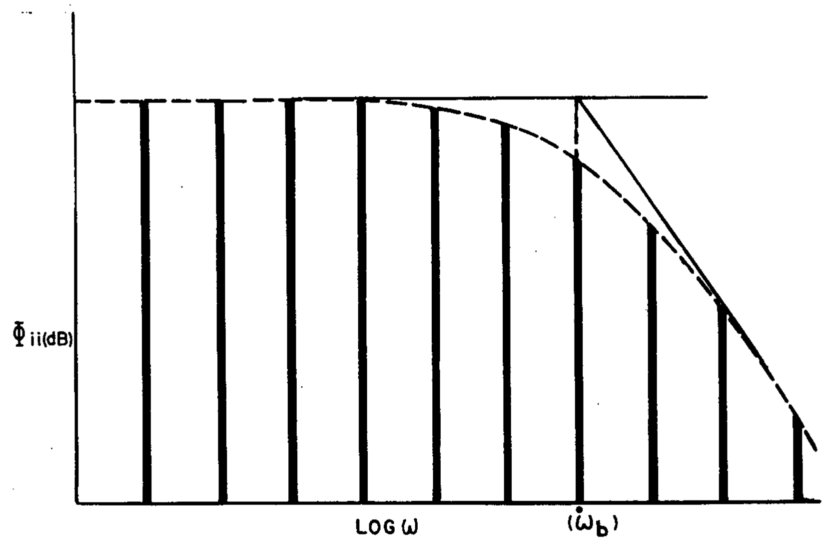


FIGURE 6–5. Power spectral density of a random-appearing function.

spectrum (which never occurs in nature) having an "equivalent bandwidth" to the actual input is useful. Several possible theoretical bases exist for deriving "equivalent" bandwidths, but recent work by Elkind (1964) supports the selection of "effective degrees of freedom" as the basis for defining an "equivalent" bandwidth. He gives the equivalent bandwidth ω_{i_e} as:

$$\omega_{i_e}=\frac{\left[\int_0^\infty \Phi_{ii}\,d\omega\right]^2}{\int_0^\infty (\Phi_{ii})^2\,d\omega}. \tag{6–4}$$

This equivalent bandwidth is the highest frequency occuring in a rectangular power spectrum which would allow the same system performance as the real non-rectangular spectrum. By providing a common spectral shape to which various spectra can be equated, the equivalent bandwidth concept reduces the number of variables that must be considered in predicting system performance.

Although *continuous* aperiodic inputs can be described rather well, very little can be said about the use of aperiodic *transients* as inputs to manual control systems. The amplitude, duration, shape and speed of the leading and trailing edges, and the intersignal interval result in so many possible input functions that they cannot be reasonably classified.

While the response of mechanical systems to transient inputs has frequently been used to determine their parameters, little success has been obtained applying this technique to manual control systems because of the human operator's basic adaptability. The problem is analogous to that of determining the transient response of a fast, adaptive, automatic, control system. It is probably this adaptive characteristic which accounts for the so called "range effect" usually found when an operator is required to track a random series of step inputs. (Ellson and Wheeler, 1949; Ellson and Coppock, 1951) In correcting a series of step inputs of varying amplitude, operators tend to undershoot the larger inputs and to overshoot the smaller ones. This range effect is a function of the *relative* amplitude of inputs and is independent of their absolute magnitudes. Thus, the operator is responding to a total situation, and any mathematical description of his input-output relationship or of the system's performance must take into account the entire situation rather than a single input. The complexity of the problem is well illustrated by Trumbo et al. (1965). When correcting a rapid series of step inputs, the operator's performance is affected by the rate of presentation of the inputs. Depending on the time interval between adjacent inputs, the operator might do any one of the following:

1. Respond to each input individually and at the proper time;
2. Respond to several inputs as if they comprised a single input;
3. Respond to each input individually but spread out in time; or
4. Fail to respond to some inputs.

While the foregoing generalities are of no direct use to the designer, they do point up the adaptive nature of the operator and the similarity of the operator measurement problem to the problem of measuring and specifying the performance of an adaptive control system. Human operator response to step inputs is treated more fully later under "Response to Transient Inputs." (See Section 6.5.2.)

6.2.2 Types of Displays

Of necessity, the choice of "dimension" for classifying types of displays is somewhat arbitrary. Displays for control of dynamic man–machine systems span the gamut of sensory input modalities and the range of complexity from full sensation of the real world to the simple presence or absence of a single input.

For the purposes of this section, we shall restrict ourselves to the primary modality, vision. While the other modalities are not trivial, and, particularly for vehicle control, the motion "feel" effects are quite important, even a full treatment of visual displays is beyond the scope of this section (See Chapter 3.)

One possible categorization of types of displays is as shown in Figure 6–6 below. This categorization is for control of a single variable. If more than one variable is to be controlled, number of axes, combined versus separate display, and perhaps perspective must be added to the categorization scheme. Within the categorization given, it must be remembered that the three axes of the figure represent three orthogonal continua, and that the "pure" end points are only laboratory abstractions. In the discussion that follows, we will treat first the nonpredictive pursuit and compensatory displays and then the predictive displays. The discussion will, for the most part, be limited to symbolic displays.

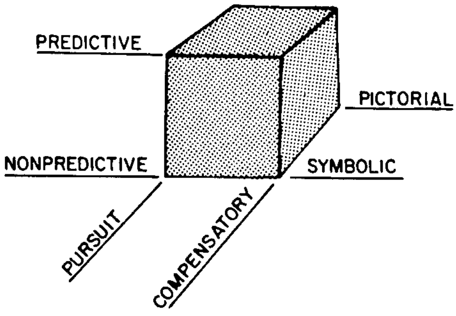


FIGURE 6–6. Categorization of display.

Pursuit Dispays

In a pursuit display, there is a fixed element and two moving elements. The fixed element may be either explicit, such as a reference point on a cathode ray tube (CRT) display, or it may be implicit such as a textured display background. The moving elements represent the "target" and the "follower." The distance from the target to the reference represents the system input or forcing function. The distance from the follower to the reference represents the system output, and the distance between target and follower represents the system error as diagrammed in Figure 6–7. The pursuit display's distinguishing feature is that the operator can see both the input and output of the system. This provides him with additional information which may or may not be helpful depending upon the dynamics of the control system and upon the type of input forcing function. Obermeyer et al (1960) provide an excellent example of the complex three-way interaction which exists between forcing function, system dynamics and display.

The most serious disadvantage of the pursuit display is that of size. The display must be large enough to accommodate the maximum range of the input signal without making the scale so small that it compromises system resolution. For example, if an altimeter is to cover a range of 50,000 feet, but only five inches of panel space are available for a linear meter, it would be unreasonable to expect the pilot to maintain an altitude tolerance of ±100 feet.

Compensatory Displays

In a compensatory display there is one fixed element and one moving element. The

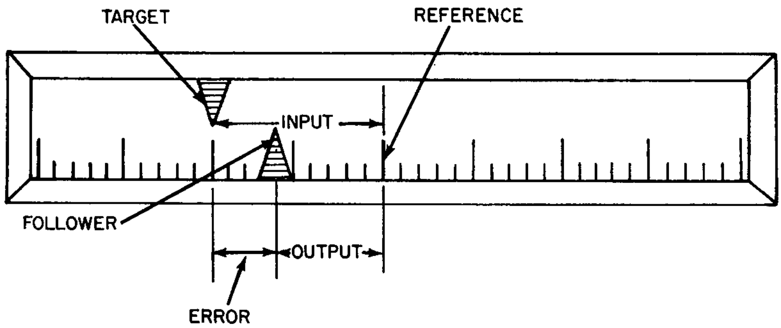


FIGURE 6–7. Pursuit display.

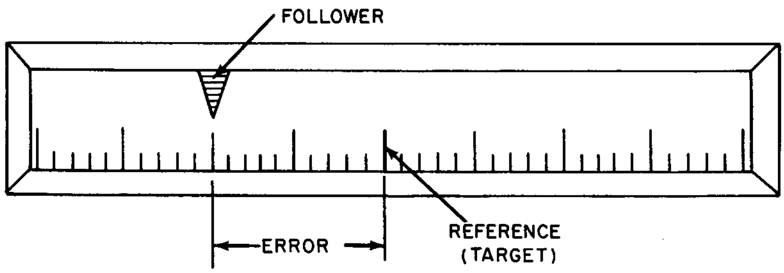


FIGURE 6–8. Compensatory display.

fixed element is the reference or "target," and the moving element is the "follower." The situation is diagrammed in Figure 6–8 below.

The distinguishing feature of the compensatory display is that the operator sees only the *difference* between the system input and the system output. This has the advantage that the display scale factor can be chosen to magnify the system error without regard to the total range of input; but it carries with it the possible disadvantage that the operator does not know the system output and can only determine the fluctuations of the input if the ouput is held stationary. This may or may not be a disadvantage depending upon the particular application. Certainly a compensatory display would not be satisfactory for an altimeter, but it is satisfactory for a glide-slope indicator.

Predictive Displays

Predictive displays show the operator some aspect of what will happen to the system in the future. Such displays may be pursuit or compensatory, symbolic or pictorial; and they may predict the vehicle output, the system input, or both. The distinguishing feature is simply that they provide advanced information that allows the operator to anticipate future requirements for control movements. Many real-world situations provide such information. Consider driving an autombile on a flat, winding road. The ability to see ahead to each curve allows the operator to "set up" for it. The same road, driven in fog, presents a much more difficult task although neither the forcing function (road) nor the dynamics (automobile) have changed. The difference is that the predicitive information has been denied by the fog. Currently available computer techniques allow the mechanization of predictive displays to give the operator a similar look at what is coming.

Wierwille (1964) developed an input-predictor display which presented mathematically optimum prediction of a random input signal. Laboratory results using this display demonstrated the concept's feasibility and utility. An average reduction in tracking error of about 25% was found over a comparable nonpredictive pursuit display with a one-radian/sec. random input. This input prediction display is based on

Wiener's optimum linear prediction of stationary random processes and should not be confused with the fast-time predicition of system output as developed by Kelley (1960) and Frost and McCoy (1965).

The system output predictor display is based on the use of a fast-time computer model of the physical system. This model continuously recomputes the predicted output of the controlled system, based on some assumed operator action and the current output, and higher derivatives of the actual system. Output predicition has been shown to be effective for acceleration control systems and for systems of higher order with low-frequency inputs. The technique has been proposed for such diverse applications as submarine depth control, orbital rendevous, high-speed terrain-following flight, attitude control of a space booster.

In theory, at least, it would be possible to combine the two techniques and predict both the forcing function and the system output, and thus give the operator complete foreknowledge of the systems status. Wierwille proposed such a concept for a pursuit display in his 1964 paper, but did not test it experimentally. Figure 6–9 shows how such a combined input and output predictor might be configured.

6.2.3 Types of Controls

This section discusses various types of continuous controls for dynamic systems. These controls are esentially transducers to convert the operator's output—force and/or displacement—into useful machine inputs. The design or selection of controls for specific operations is considered in Chapter 8.

The output of a control stick may be a physical movement of some other part of the system such as a hydraulic valve or a control surface of an aircraft, or it may be an electrical signal into a power amplifier which controls some other part of the system. The nature of this output is specified by the requirements of the system "down-stream" of the control. Considering the type of output as a given or fixed requirement, our concern will be with the input to the control and with input-output relationships.

When designing or selecting a transducer for use as an element in a system, it is useful to know what must be converted into what. When dealing with man, this information is not readily available since man's output may be a pure force into an isometric (stiff-stick) control, a pure displacement into an isotonic (unrestrained)

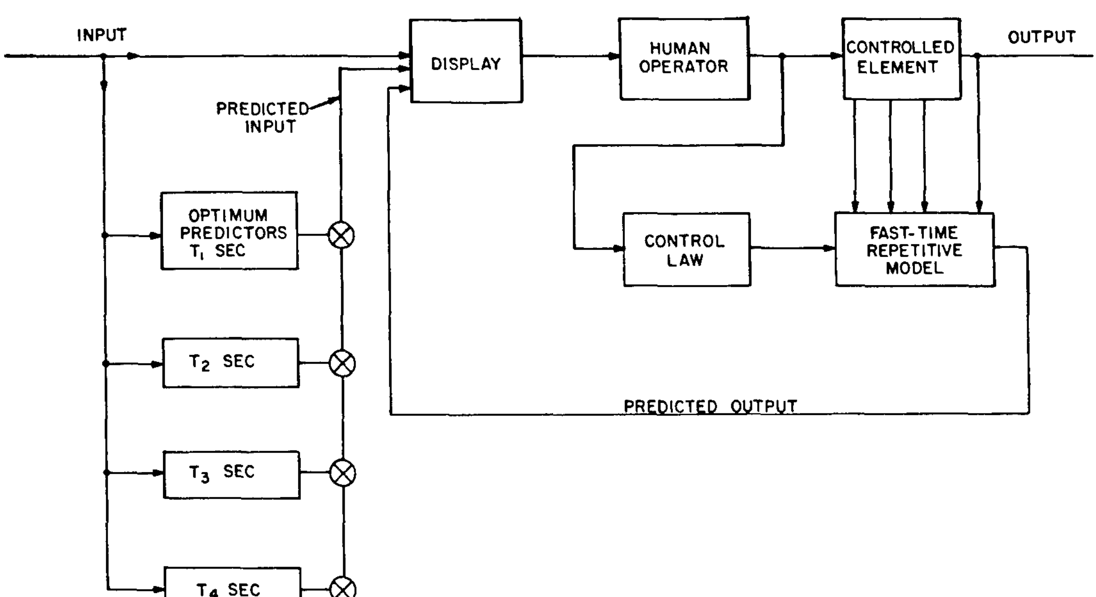


FIGURE 6–9. Control system using both predicted input and predicted output display.

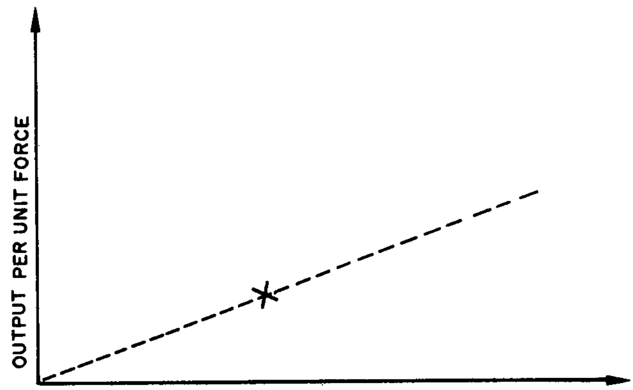


FIGURE 6–10. Force-displacement plane.

control, or some combination of force and position output into a loaded moving control. When designing or selecting a control stick, it is useful to consider the output per unit force and output per unit displacement as independent design variables. For a simple spring-loaded control stick, these variables are related by the spring constant. This is shown in Figure 6–10. The point x on the plane represents a given spring-loaded control stick. An applied force results in a given displacement and a corresponding output. The slope of the line passing through the origin and point x is the spring constant. If the gain of the system downstream of the stick is changed, the point x moves along the line. If the gain is held constant and the spring is changed, the slope of the line is changed and point x moves either vertically or laterally depending on whether the stick output is sensed by a position or a force sensor.

From the above discussion, it can be seen that the isometric and isotonic control sticks are simply limiting cases of spring-restrained controls. The isometric stick has an infinite spring constant while the isotonic stick has a zero spring constant.

The following sections treat isometric and isotonic controls as limiting cases of the general control design problem. Simple spring-loaded controls are covered next, and finally controls having appreciable mass and damping are discussed. The proper setting of "stick gain" is covered in Section 6.3.4.)

Isometric (Force) Controls

The isometric or "stiff-stick" controller is essentially a force transducer. The output of an isometric control is generally an electrical signal to a power amplifier in the controlled system. Its advantages and disadvantages derive from the fact that no appreciable motion is involved.

An isometric control's advantages are:

1. Single-axis tracking, particularly of high-frequency inputs, is better with isometric controls than with isotonic controls. This may involve as much as a 2:1 reduction in error with rate controls (North and Lomnicki, 1961).
2. No space is required for control movement.
3. Output returns to zero when force is removed.
4. There is less problem of inadvertent input under vibration and G loading if good forearm support is provided.

The principal disadvantages of isometric controls are:

1. Deadzone must be introduced in multi-axis tracking to prevent inadvertent cross-coupling between axes (Vreuls and Rheinlander 1965).
2. Continuous force application is required to maintain non-zero output with position control system or when tracking very low-frequency input signals. This in turn may lead to undue fatigue during prolonged operation. However, no fatigue was experienced with optimum gain

during five-minute tracking runs (Burke and Gibbs, 1965).

Isotonic (Position) Controls

A true isotonic control is only an engineering abstraction since any real control has some mass and some friction. Proper design can reduce the resulting inertial and frictional forces to negligible values, but some force must still be developed in the controlling muscle groups to overcome the effective mass and spring of the controlling limb. For this reason, there is always some lag associated with operation of position controls. This lag is in addition to any reaction time delay that also exists.

The advantages of an isotonic control are:

1. No force is required for constant output.
2. Visual feedback of control position is available.

Disadvantages of isotonic controls are:

1. They are more subject to inadvertent inputs from operator and environment than isometric controls.
2. They do not return to zero when released.
3. They do not provide clearly defined zero output information.
4. Tracking error is generally higher, especially with high-frequency inputs. This is due primarily to phase lag at high frequencies (McRuer and Magdaleno, 1966).

Loaded Moving Controls (Real Controls)

Most practical controls lie somewhere between the two extremes described above. That is, they can be moved by application of sufficient force. There are three types of *linear* forces which act on moving controls.

1. Elastic (spring) forces proportional only to the *displacement* (x) of the control.
2. Viscous friction (damping) forces proportional only to the *velocity* ($\dot{x}$) of the control.
3. Inertial forces proportional only to the *acceleration* ($\ddot{x}$) of the control.

In addition to these linear forces, certain nonlinear forces act on the moving control.

1. Friction forces, which are of constant magnitude and oppose the direction of motion.
2. Stiction forces, which account for the difference between starting friction and sliding friction. No adequate analytic description of stiction exists.
3. Preload forces, which are of constant magnitude and always in a direction to center the control.
4. Other nonlinearities such as backlash and nonlinear gearing. These will be treated in Section 6.3.6.

It was shown previously (Section 6.2.3) that the simple spring-loaded control lies on a plane defined by its output per unit force and its output per unit displacement. Numerous studies have been conducted by various investigators to define the optimum gain or optimum control-display ratio for different types of systems. Most of these studies are not generalizable beyond the context in which they were conducted. Two notable exceptions are the studies by Gibbs (1962) and by North and Lomnicki (1961). Gibbs showed that for a variety of controlling limbs, the optimum sensitivity of a very lightly spring-loaded (almost isotonic) control is a constant *when measured in terms of angular movement of the controlling joint.* That is, if the stick output is expressed as volts per radian of thumb, wrist, or elbow movement, the same control sensitivity is optimum for all three joints. North and Lomnicki defined the optimum control-display ratio for an isotonic controller in a rate-control tracking task and in the same task with an isometric controller. Noting that any spring-loaded controller lies on a plane defined by the isometric and isotonic axes, he suggested that the locus of optimum gains for any spring-loaded control lies along an ellipse in the plane and is anchored by the optimum sensitivities of the isometric and iotonic cases. Unfortunately, if a validation study was conducted, it was not reported in the open literature. The consensus of the open literature is that optimum gain or control-display ratio is more important than optimum spring constant. Both the tracking literature and the psychophysical studies of "blind positioning" responses (e.g., Briggs et al., 1957; Weiss, 1954) indicate slightly superior performance as the spring is made stiffer *if* the sensitivity is optimized in both cases.

The problem of control stick design is even more difficult if the control system reflects forces back to the operator through the control stick. This situation occurs in direct manual flight control systems where the control stick is mechanically linked to the control surface, in nonpowered gun aiming systems, and in a number of other systems, where, intentionally or otherwise, large inertial or viscuous friction forces affect the control stick. The most comprehensive set of studies investigating the effects of these forces is reported by Notterman and and Page (1962). In their study, it was assumed that man was a force output device. Experimentally they compared tracking performance using a mechanical tracking system having various mass, spring, and damping characteristics against tracking performance using isometric and isotonic control sticks with electronic simulation of the same mass, spring, damping chacteristics. The experimental conditions are shown schematically in Figure 6–11a through 6–11c. The natural frequency (ω_n) of the mechanical system is given by $\omega_n = \sqrt{k/m}$, while the damping ratio (ζ) is given by $c/2\sqrt{kM}$, where k, c, and m are the spring constant, damping constant, and mass, respectively. The results of this study are summarized in Figures 6–12a through 6–12d. The results of tracking with the same isometric and isotonic sticks without the simulated dynamics have been added to the figures. From these results, it is evident that:

1. Man operates more as a position output device than as a force output device.
2. An isometric (force) stick is superior to an isotonic (position) stick when the control system has oscillatory position control dynamics, but- -
3. The differences become smaller with increasing natural frequency and with increasing damping of the control system.

A more recent study (DiFranco, 1968) has shown that the dynamics of the control stick can seriously alter the controllability of a vehicle. In an in-flight study using a variable-stability fighter aircraft with a variable artificial feel system, he showed that as the natural frequency and damping of the feel system were reduced, pilot opinion of the system degraded and the system became unflyable. This situation was more severe in the flight evaluation than it was in the comparable fixed-base simulation.

In general, friction in the control stick only slightly degrades tracking performance, but

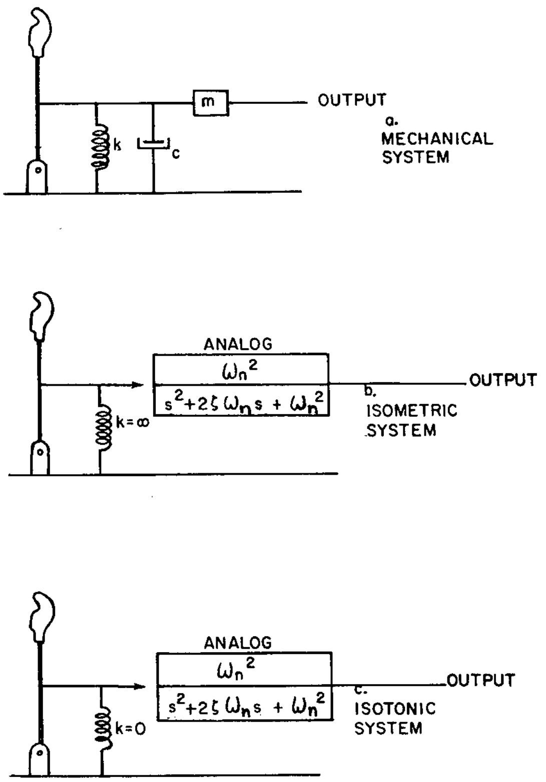


FIGURE 6–11. Mechanical and simulated stick dynamics.

frequently produces large decrements in operator opinion of the system's "feel." Thus, some stick friction is tolerable in a "jolting" environment to prevent inadvertent inputs, but the level should be kept as small as possible. The fact that the operator can close a "tight" position feedback loop around the control stick enables him to overcome the degrading effects of most force-displacement nonlinearities in the stick, but the fact that it requires "effort" to make this tight loop closure accounts for the degraded opinion.)

Valve friction in a hydraulic control system is an entirely different situation. Valve friction always degrades performance and can frequently lead to operator-induced oscillations. The addition of friction and/or preload springs to the control stick will partially negate the effects of valve friction if the total stick force required does not become excessive (Brown, 1957). A good review of the existing data on the effects of friction and preload is given in Wasicko and Magdaleno (1965).

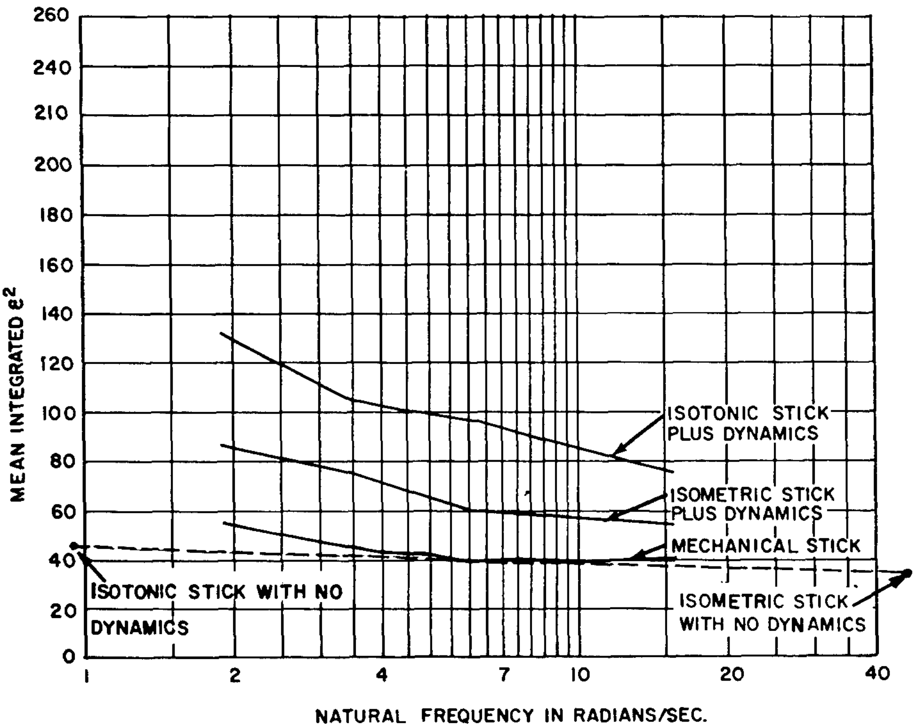


FIGURE 6–12(a). Integrated error squared vs. natural frequency. (Damping ratio 1.2) (After Notterman and Page, 1962.)

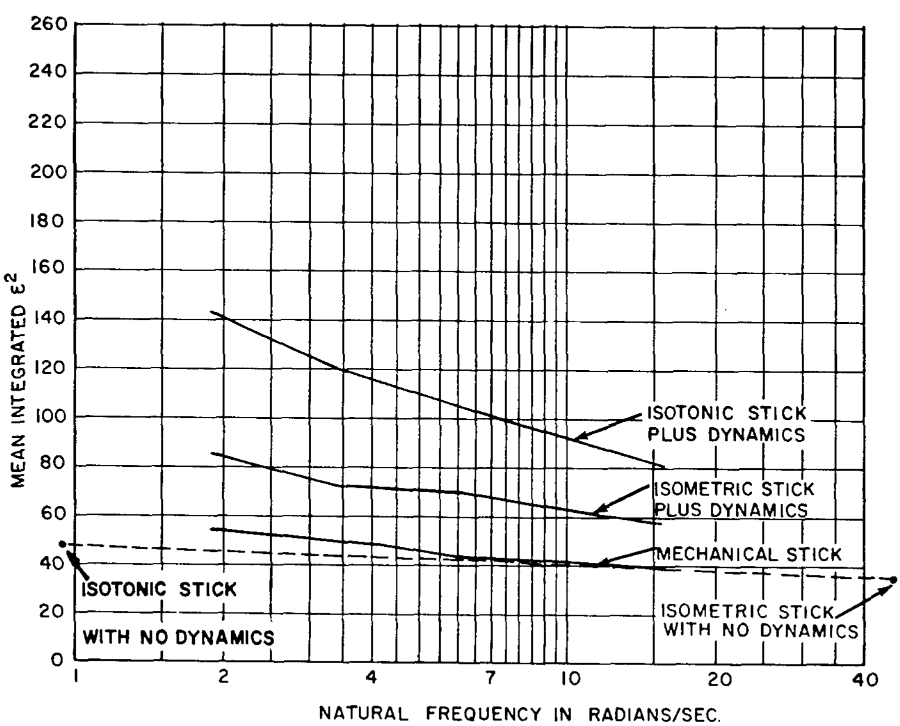


FIGURE 6–12(b). Integrated error squared vs. natural frequency. (Damping ratio 0.8) (After Notterman and Page, 1962.)

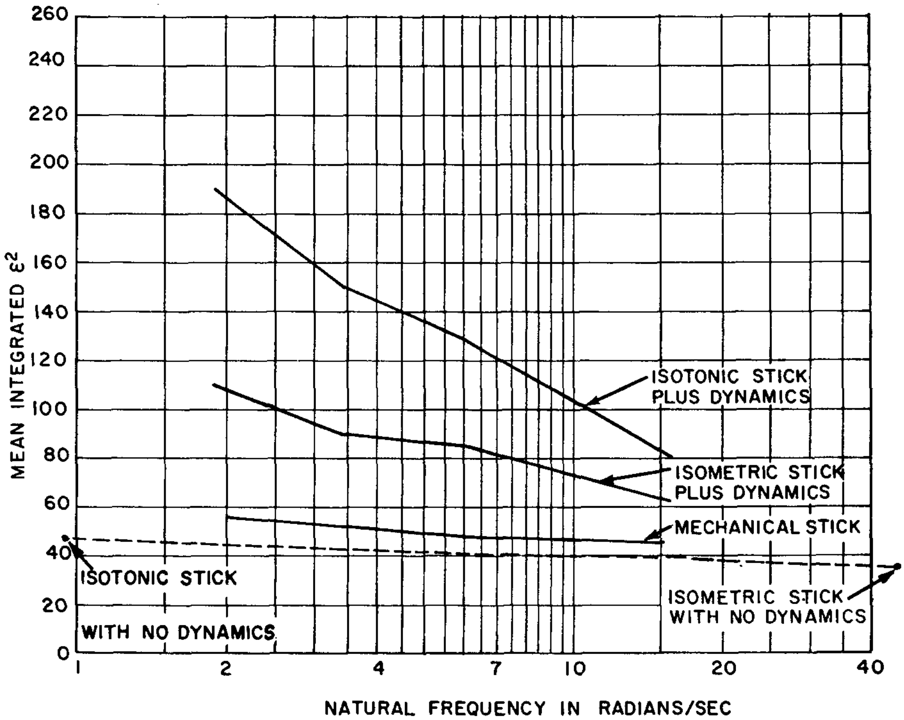


FIGURE 6–12(c). Integrated error squared vs. natural frequency. (Damping ratio 0.4) (After Notterman and Page, 1962.)

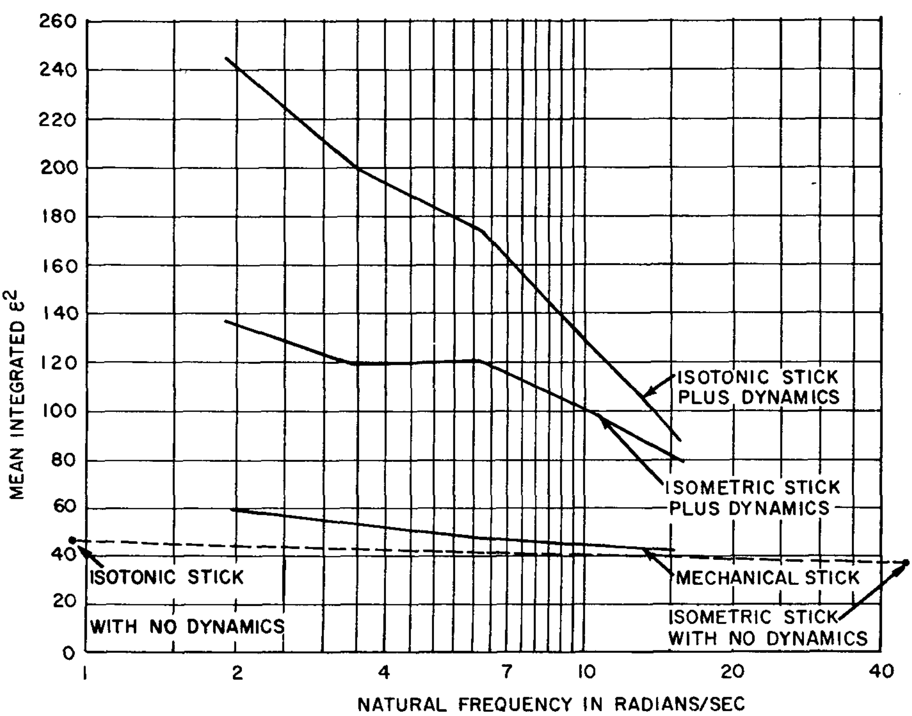


FIGURE 6–12(d). Integrated error squared vs. natural frequency. (Damping ratio 0.2) (After Notterman and Page, 1962.)

6.2.4 Controlled Element Dynamics

The terminology and concepts of control systems and displays used in the various areas—psychology, engineering, piloting, etc.—are frequently different and not always well understood by the other diciplines. An attempt will be made here to provide a common viewpoint.

Differential Equations and Transfer Functions

When dealing with control systems it is usually easier to consider the system as being made up of interconnected discrete blocks. Thus, Figure 6–2 shows the *Display*, the *Man*, the *Control*, and the *Controlled Element* interconnected to form a control system. The input-output relationships of each block can, at least in theory, be represented by one or more differential equations. For simplicity the following discussion is restricted to the case where the input-output relations can be described by linear, constant-coefficient, differential equations. Further, the Laplace transform of the differential equations is used rather that the differential equations themselves. The Laplace domain simplifies the analysis of the serial effect of several differential equations. Combining differential equations requires integral transformation, such as convolution, but in the Laplace domain this becomes simple algebraic transformation. Thus higher order equations are more simply solved by Laplace transformation which limits the mathematics to algebraic manipulations of the transformed equations. The Laplace operator s can, for the present discussion, be considered to be equivalent to the differential operator d/dt.

The *transfer function* of a block in our block diagram is simply the ratio of the output to the input. Thus, the transfer function for the block in Figure 6–2 labeled Y_c is θ_o/δ_e. This transfer function is usually expressed as a ratio of polynomials in s, the Laplace variable.

Conceptually, the transfer function is the mathematical model for representation of the tranformation that is made between input signal and output signal. Thus, if a particular element provides amplification only, then its transfer function is represented by a constant multiplier.

The Type of a Control System

The simplest transfer function is a constant, K, a simple gain. The simplest nonconstant transfer functions are those for integration and differentiation, $1/s$ and s, respectively. Double integration is represented by $1/s^2$, and so on. Figure 6–13 shows the results of one, two, and three integrations of a simple step input. Such pure integrations rarely occur alone in real physical systems, but they serve to illustrate the *type* of control system. The *type* of system is characterized by the number of pure integrations between the input and the output. Thus, a system with a transfer function K (pure gain) is a position or type O system. A system with a transfer function $1/s^2$ is an acceleration or type 2 system. Important generalizations of certain performance characteristics can be made merely by knowing the type of system.

Step Response

The *step response* of a system is simply the response of the system to a unit step change in the input. Expressed mathematically, this response is the system's "transfer function." Step response is most useful in gaining a "feel" for the response of the system to simple transient inputs. For complex physical systems, the transfer functions usually become quite complex. However, most physically realizable linear systems can be represented by combinations of simple transfer functions. In addition to the simple gain, integration, and differentiation functions mentioned previously, the following four transfer functions occur with sufficient frequency to be noted.

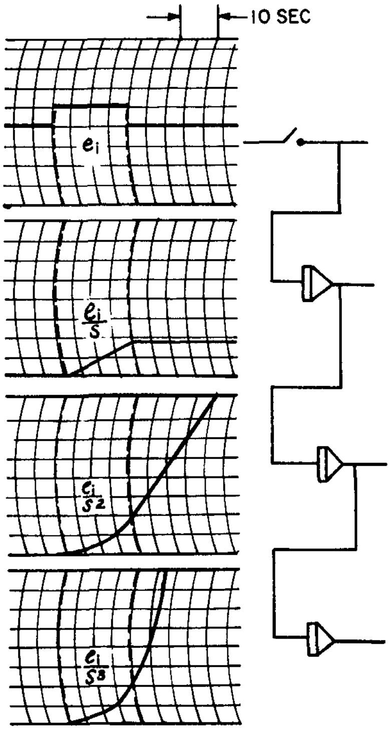
$$Y_1(s)=\frac{1}{\tau s+1} \qquad (6\text{–}5)$$

$$Y_2(s)=\frac{1}{(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1} \qquad (6\text{–}6)$$

$$Y_3(s)=(\tau s+1) \qquad (6\text{–}7)$$

$$Y_4(s)=(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1 \qquad (6\text{–}8)$$

$Y_1(s)=\frac{1}{\tau s+1}$ (Eq. 6-5) is commonly called a *lag*, and τ is the lag *time constant*. Figure 6–14



Input Step Function

e_i

Input Integrated Once (Type 1)

$e_{o_1} = (1/s)e_i$

Note linear ramp when input is present, constant position when input returns to zero. This corresponds to a constant rate while the input is present, and a constant position when the input is removed.

Input Integrated Twice (Type 2)

$e_{o_2} = (1/s^2)e_i$

There is a quadratic increase while the input is present, a linear increase when the input returns to zero. This corresponds to a constant acceleration while the input is present and a constant rate when the input is removed.

Input Integrated Three Times (Type 3)

$e_{o_3} = (1/s^3)e_i$

There is a cubic increase while the input is present, a quadratic increase when the input is removed. This corresponds to a constant rate of change of acceleration (constant jerk) while the input is present and a constant acceleration when the input is removed.

FIGURE 6-13. Types of control systems.

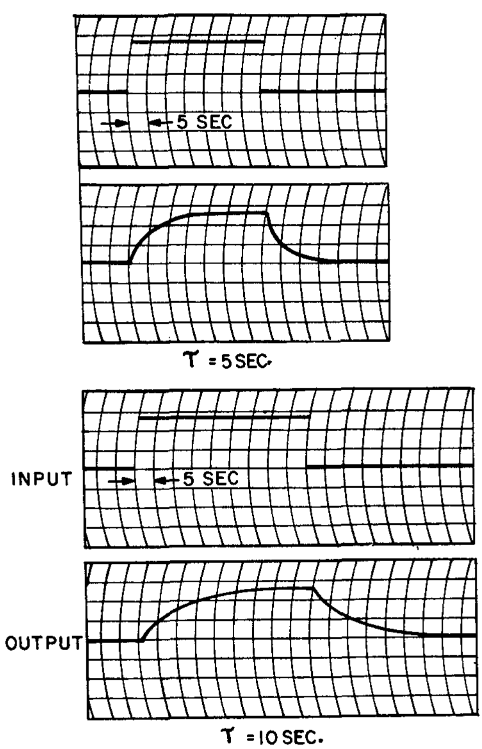


FIGURE 6-14. Step response of simple lags for several time constants.

shows the step response of a simple lag for two time constants. During a time equal to a time constant, the response always reaches 63% of its final value. Note that initially the response appears like that of a pure integrator but that it reaches a final value along an exponential curve.

$$Y_2(s)=\frac{1}{(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1} \qquad (6\text{–}6)$$

is a simple oscillatory system. Its response to a step is an oscillation which decays to a final value equal to the input step. ω_n is the *natural frequency* of the oscillation and ζ is the *damping ratio.* Natural frequency and damping are functions of the physical parameters of the system. Figure 6–15 shows the step response of such a second order system for several damping ratios with a constant natural frequency. Note that when the damping ratio equals 1.00 there is no "overshoot" or oscillatory response. Rather, the response appears similar to that of the simple lag. It is apparent that when $\zeta \geq 1.00$, $Y_2(s)$ can be factored into two terms of the form of $Y_1(s)$.

$$Y_3(s)=\tau s+1 \qquad (6\text{–}7)$$

is commonly called a *lead* and

$$Y_4(s)=\left[(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1\right] \qquad (6\text{–}8)$$

is a second order lead. They bear the same relation to pure single and double differentiation as the lag and second order lag bear to pure integration. No step response figures are given since the derivative of an instantaneous step is undefined, i.e., the instantaneous slope is infinitely large.

The Order of a Control System

An important characteristic of any transfer function is its *order.* Order is defined as the highest power of s which appears in the denominator of the transfer function. Thus, if we multiply $Y_1(s)$ by $Y_2(s)$, we have a third order system.

$$[Y_1(s)]\times[Y_2(s)]=\frac{1}{(\tau s+1)\left[(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1\right]} \qquad (6\text{–}9)$$

A generalization of this expression is a ratio of polynomials in s, as:

$$Y(s)=\frac{A(s)}{B(s)}=\frac{s^n+a_1s^{n-1}+\cdots+a_n}{s^m+b_1s^{m-1}+\cdots+b_m} \qquad (6\text{–}10)$$

The above is an "mth order system."

The *order* of the system should not be confused with the *type* of system mentioned earlier. If a factor s^k can be factored out of $B(s)$, such that $B(s)=s^k(s^j+b_1s^{j-1}+\cdots b^j)$, then the system is said to be a k type system of order $k+j$. When $k=0$, the system is a type O or position system *regardless* of the order of the system. Figure 6–16 shows several orders and types of systems. Note that a system with three or more integrators (third or higher order) may still be a position (type O) system. Note also that Figure 6–13 is the diagonal of Figure 6–16, Figure 6–14 is an expansion of the first order, type zero, cell of Figure 6–16, and Figure 6–15 is an expansion of the second order, type zero, cell of Figure 6–16.

Steady-State Response

The preceding section dealt only with the response of open-loop elements to step inputs. The purpose of this section is to examine the steady-state performance of such elements in a closed-loop system. Figure 6–17 shows the basic relationship between open- and closed-loop transfer functions. The final relationship,

$$\frac{\theta_o(s)}{\theta_i(s)}=\frac{Y(s)}{1+Y(s)} \qquad (6\text{–}11)$$

is fundamental to all closed-loop system theory.

Consider the case where $Y(s)=1/s$, a pure integrator or type 1 system. Substituting $1/s$ into the closed-loop equation yields $\theta_o/\theta_i=1/(s+1)$. This is a type O or position system of first order. Thus, closing a *position feedback loop* around a rate control element yields a position control system. Similarly, if $Y(s)=1/s(\tau s+1)$, closing a position feedback loop yields $\theta_o/\theta_i=1/(\tau s^2+s+1)$, a second order position control system. The order

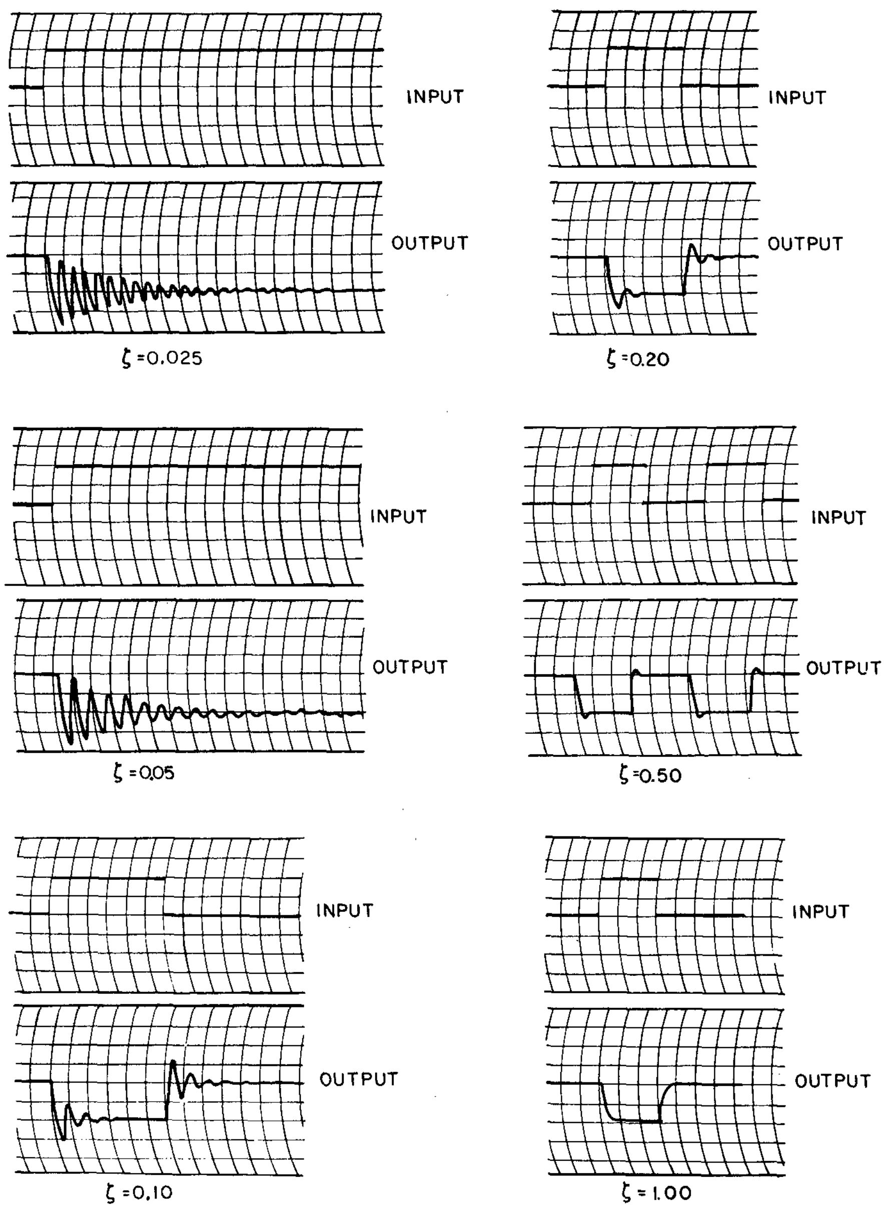


FIGURE 6–15. Step response of simple second order systems ($\omega_n = 2$ rad/sec).

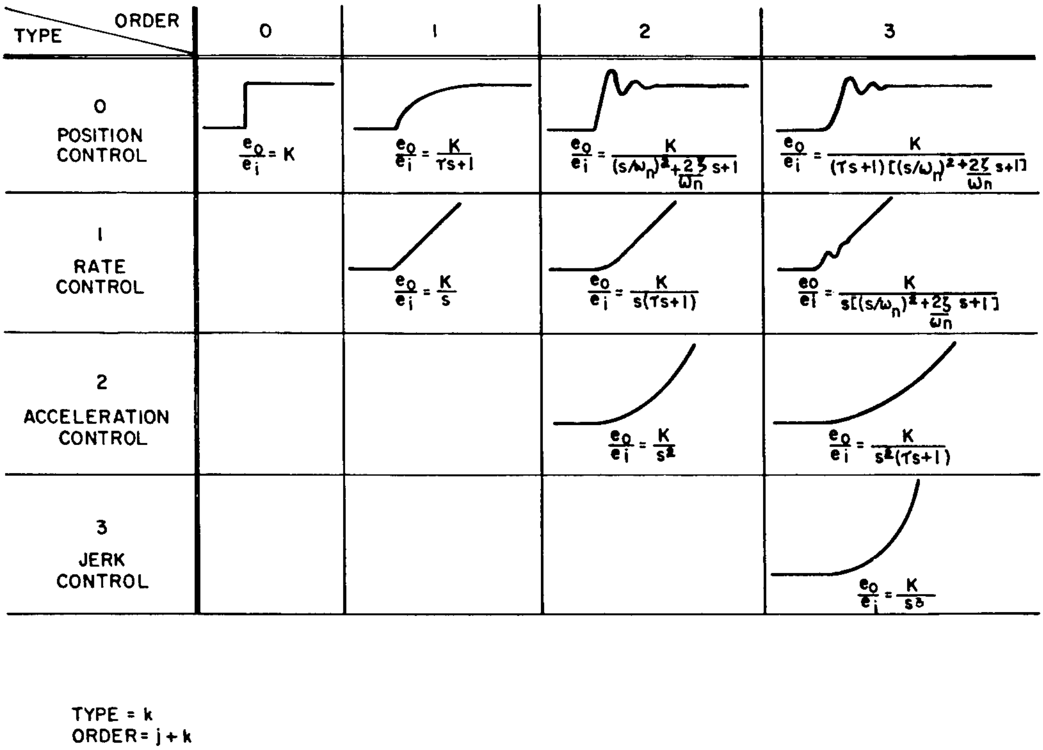
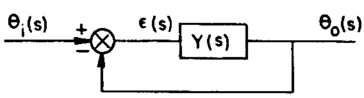


FIGURE 6–16. Step response of various controlled elements illustrating differences due to type and order

$$\frac{e_o}{e_i} = \frac{A(s)}{B(s)} = \frac{s^n + a_1 s^{n-1} + a_2 s^{n-2} + \text{----} + a_n}{s^k(s^j + b_1 s^{j-1} + \text{----} + b_j)}.$$



$\theta_o(s) = \epsilon(s)Y(s)$ OR $\frac{\theta_o(s)}{\epsilon(s)} = Y(s)$

BUT $\epsilon(s) = \theta_i(s) - \theta_o(s)$

THEREFORE $\theta_o(s) = Y(s)\,[\theta_i(s) - \theta_o(s)]$

OR $\theta_o(s) + [Y(s) \cdot \theta_o(s)] = Y(s) \cdot \theta_i(s)$

THUS $\frac{\theta_o(s)}{\theta_i(s)} = \frac{Y(s)}{1 + Y(s)}$

FIGURE 6–17. Closed-loop transfer function.

of the system is unchanged, but the type of system has been lowered.

The question of steady-state response implies a steady-state input. This is not necessarily a constant position, but may be a constant velocity or constant acceleration input. This section deals only with the existence of a steady-state error. Estimation of the magnitude of error is treated in most standard servo texts.

In general, the relationship $Y(s) = \theta_o(s)/\epsilon(s)$ relates the output *position* of the system to the error between input and output. If the input signal or command is a position, the steady-state output is also a position and the steady-state error is obviously a position. However, if the command is a constant velocity or a constant acceleration, the output is also a constant velocity or a constant acceleration; yet, the error is still a *position* error. As an example, if a gunner is aiming at an aircraft which is circling his position, the command is a constant angular velocity, the steady-state output is a constant velocity matching the command, but the error is a position error lagging the command. Table 6–1 summarizes the steady-state error characteristics for various types of systems.

Frequency Response

The Laplace operator s has been introduced as being equivalent to the differential operator d/dt. For analyzing the frequency response of a

TABLE 6–1. SUMMARY OF STEADY STATE ERROR CHARACTERISTICS

Type system	Error characteristic of closed-loop system
0	Position error at all times
1	No static position error—Lag error when operated at constant velocity. Constantly increasing error when command is a constant acceleration.
2	No static position error—No position error at constant velocity. Constant error when command has constant acceleration.
3	No static position error—No constant error at constant velocity or constant acceleration.

Thaler and Brown (1953). From *Servomechanism Analysis* by G. J. Thaler and R. G. Brown. Copyright 1953 by McGraw-Hill Book Co. Reprinted with permission of McGraw-Hill Book Co.

system or of a controlled element, it is legitimate to replace s by $j\omega$ where $j=\sqrt{-1}$ and represents a 90° phase shift and ω is the input frequency. Thus, if a signal is differentiated it would be written as $e_o(s)=se_i(s)$ or $e_o(s)/e_i(s)=s$. Consider e_i to be a sine wave. The derivative of a sine wave is a cosine wave, or a sine wave shifted ahead in phase by 90°. To exemplify the above operation, consider the following:

$$\frac{\theta_o(s)}{\theta_i(s)}=Y(s)=\frac{1}{\tau s+1}. \qquad (6\text{–}12)$$

Substituting,

$$\frac{\theta_o(j\omega)}{\theta_i(j\omega)}=Y(j\omega)=\frac{1}{(1+\tau j\omega)}. \qquad (6\text{–}13)$$

Note that $(1+j\omega\tau)$ is a vector quantity having both magnitude and direction. Evaluating $Y(j\omega)$ as a function of ω yields

for $\omega \ll 1$

$$Y(j\omega)\simeq\frac{1}{1+j0}=1\measuredangle 0^\circ \qquad (6\text{–}14)$$

for $\omega=1/\tau$

$$Y(j\omega)=\frac{1}{1+\text{j}1}=\frac{1}{\sqrt{2}\measuredangle 45^\circ}=\sqrt{2}\measuredangle -45^\circ \qquad (6\text{–}15)$$

for $\omega \gg 1$

$$Y(j\omega)\simeq\frac{1}{1+j\infty}=\frac{1}{\infty\measuredangle 90^\circ}=0\measuredangle -90^\circ. \qquad (6\text{–}16)$$

If specific values of τ and ω are chosen, the complete frequency response can be plotted. For $\tau=1$, $\omega=0.5$, $Y(j\omega)=\frac{1}{1+j.5}=\frac{1}{1.25\measuredangle 26.5^\circ}=0.8\measuredangle -26.5^\circ$. For more complex systems, direct calculation of frequency response becomes tedious. Fortunately, shortcut procedures are available. By taking the logarithm of both sides of the equation, linear asymptotic approximations can be made. For pure gain terms, no approximation is necessary, the plot being simply $20\log_{10}K$, the gain in decibels (dB). For pure differentiation τs or pure integration $1/\tau s$, substituting $j\omega$ for s and plotting $20\log_{10}(\tau j\omega)$ or $20\log_{10}(1/\tau j\omega)$ versus $\log\omega$ yields straight lines with a slope of 6 dB/octave and a 0-dB intercept at $\tau j\omega=1$. This is easily seen by evaluating a few points:

$$\omega\tau=1 \quad 20\log_{10}1=0$$

$$\omega\tau=2 \quad 20\log_{10}2=6.02$$

$$\omega\tau=4 \quad 20\log_{10}4=12.05.$$

Note that for each doubling of $\omega\tau$, $20\log_{10}(\omega\tau)$ increased 6 *dB*.

The phase angle is a constant +90° for $\tau j\omega$ and a constant −90° for $1/\tau j\omega$.

For terms of the form $(j\omega\tau+1)$, if $j\omega\tau \ll 1$, then $20\log_{10}|j\omega\tau+1|\simeq 20\log_{10}1.0=0$ dB and $\tan^{-1}(0/1)=0°$; and if $j\omega\tau \gg 1$, then $20\log_{10}|j\omega\tau+1|\simeq 20\log_{10}|j\omega\tau|=6$ dB/octave. The two lines intersect at a frequency ω_b such that $\omega_b\tau=1$. $\omega_b=1/\tau$ is known as the *break frequency.* At the break frequency, $20\log_{10}|1+j1|=3$ dB. By analogy it can be seen that for terms of the form $1/(j\omega\tau+1)$, the frequency response is asymptotic to a horizontal line of 0 dB/octave and a line of −6 dB/octave intersecting at break frequency of $1/\tau$, and the true value of the function is −3 dB at the break frequency. It can be shown that the true value of the function is 1 dB away from the asymptotes at $\omega_b/2$ and at $2\omega_b$. The asymptotic representations and the exact frequency response is shown in Figure 6–18a and b. Note that in this plot the abscissa is $\omega\tau$, not simply ω. This normalizes the curve so that it is applicable for all values of τ. Thus, if $\tau=1$ sec., the response is up 7 dB at 2 radians/sec., while if $\tau=2$ sec. the response is up 3 dB at 1 radian/sec. ($\omega\tau=2$). For the terms of the form $(\tau s+1)^N$ or $[1/(\tau s+1)]^N$, the slope becomes $6N$ dB/octave and the deviation at the breakpoint becomes $3N$ dB. The break frequency remains unchanged.

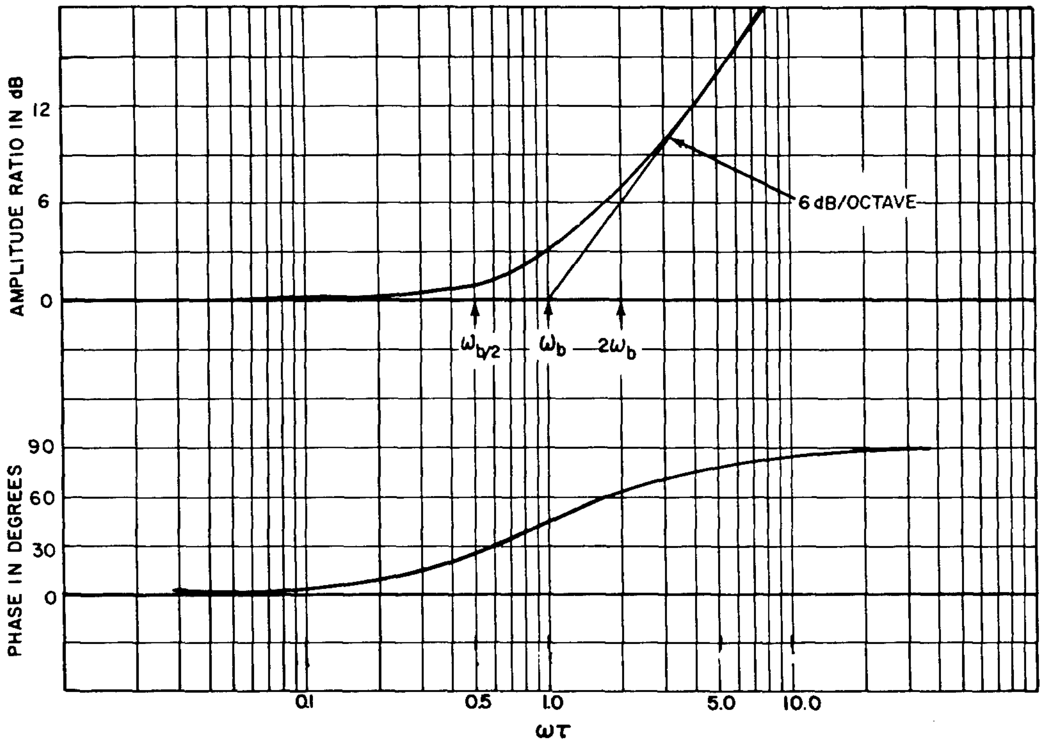


FIGURE 6–18(a). Asymptotic representation and exact frequency response $Y(j\omega)=j\omega\tau+1$.

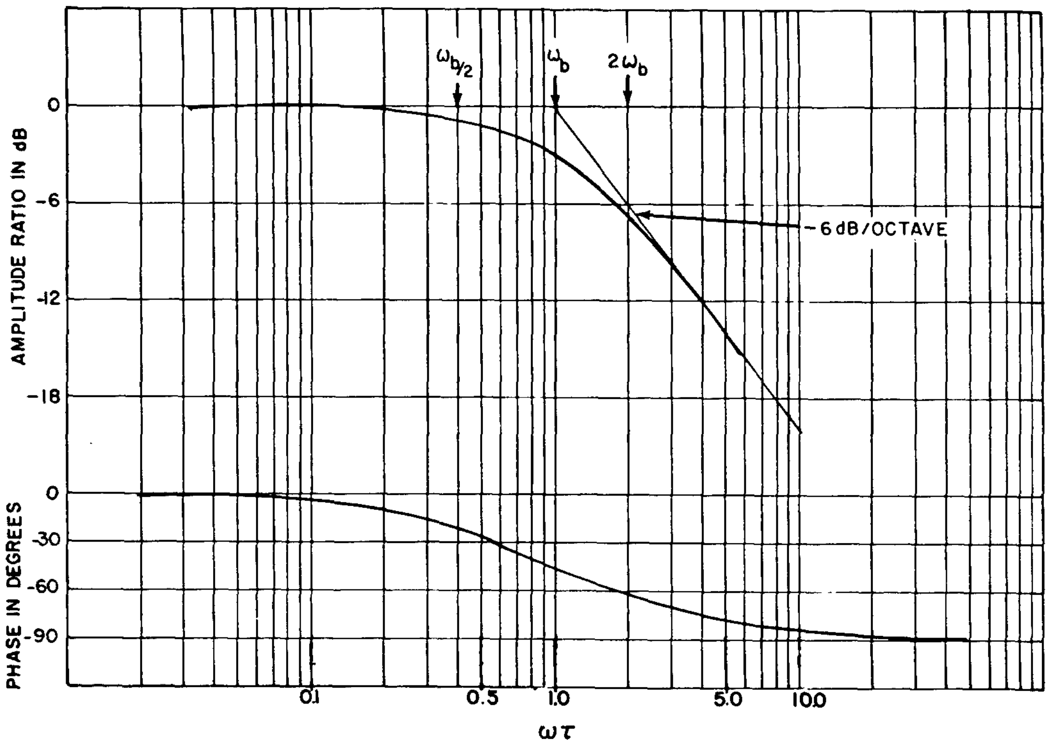


FIGURE 6–18(b). Asymptotic representation and exact frequency response $Y(j\omega)=\frac{1}{j\omega\tau+1}$.

For terms of the form

$$\frac{1}{(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1},$$

the slope of the high frequency asymptote is −12 dB/octave, the low-frequency asymptote is the 0-dB line, and the break frequency is ω_n. No simple rule can be given for the true curve since it is a function of both ζ and ω_n. The maximum deviation from the asymptote is very near the break frequency and at that point the deviation, Δ, is given by $\Delta = 20\log_{10}2\zeta$. Figure 6–19 presents the amplitude ratio and phase shift versus normalized frequency for several damping ratios and demonstrates the general shape of the curves. Again, this curve is normalized; the abscissa is ω/ω_n, not ω. This allows the curves to be used for any natural frequency. For example, if a particular system had a natural frequency of 5 radians/sec. and a damping ratio of 0.5, its response would be attenuated 10 dB at $\omega/\omega_n=2$ or $\omega=10$ radians/sec.

One of the major advantages of the logarithmic representations is the ease with which frequency responses of complex systems can be plotted. As an example, consider the system represented by the transfer function

$$Y=\frac{K(\tau_1 s+1)}{s[(s/\omega_n)^2+\frac{2\zeta}{\omega_n}s+1]}. \tag{6–17}$$

This system is the product of several terms for which the logarithmic plots are easily made. Thus, taking the logarithm of each side of the equation,

$$20\log Y=20\log K+20\log|\tau j\omega+1| -20\log|j\omega| -20\log\left|(j\omega/\omega_n)^2+\frac{2\zeta}{\omega_n}j\omega+1\right| \tag{6–18}$$

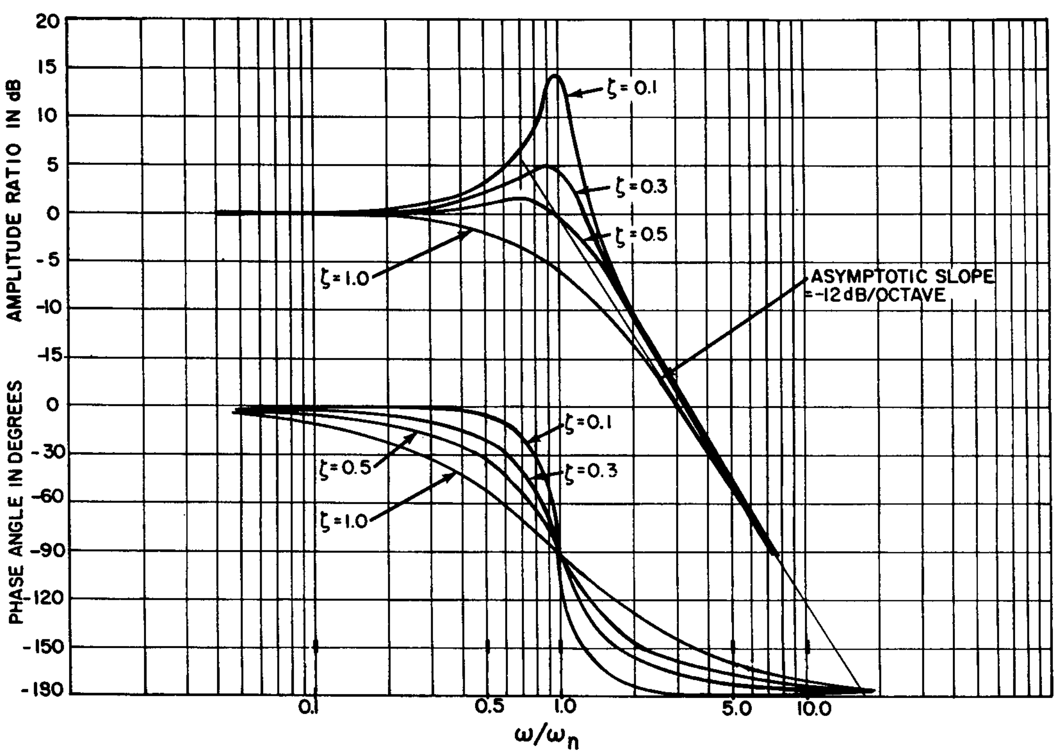


FIGURE 6–19. Amplitude ratio and phase shift vs. log frequency $Y(j\omega)=\frac{1}{(s/\omega_n)^2+\left(\frac{2\zeta}{\omega_n}\right)s+1}$.

If each term is plotted separately on the same sheet, the logarithmic response of the system—20 log Y—can be found by simple graphical addition.

Aiding, Unburdening, and Quickening

Psychologists have used the terms *aiding*, *quickening*, and *unburdening* to describe various modifications of control systems which improve human performance, improve system perfromance, or make the operator's job easier. Control system designers use various techniques to "compensate" the performance of automatic or manual control systems to meet specific performance criteria. In many instances the control system designer and the human factors specialist or psychologist have had difficulty understanding each other's language. It is the intent of this section to attempt to resolve the linguistic barrier.

"Aiding" is one technique of "compensating" the response of a control system. Specifically, aiding refers to parallel feed-forward compensation of the machine portion of the control system. Such compensation alters the frequency response and transient response of the system in a completely predictable manner. Figure 6–20 shows a simple unaided system. Figure 6–21 shows how such a system could be aided.

If the original machine consists of a gain, a_1, and three integrations, as in Figure 6–20, the machine transfer function is given by output/$\delta = a_1/s^3$. Assuming that the man and the display act as a simple gain K, the error-to-output transfer function is output/$\epsilon = Ka_1/s^3$, which is an unstable system.

If the system is aided by adding feed-forward loops around the integrators, as shown in Figure 6–21, the machine transfer function becomes

$$\frac{\text{output}}{\delta} = \frac{a_4s^3 + a_3s^2 + a_2s + a_1}{s^3}, \qquad (6\text{–}19)$$

and the error-to-output transfer function becomes

$$\frac{\text{output}}{\epsilon} = K\frac{a_4s^3 + a_3s^2 + a_2s + a_1}{s^3}. \qquad (6\text{–}20)$$

The output for a given display error signal is thus modified with an aided system. The closed-loop system transfer function becomes

$$\frac{\text{output}}{\text{input}} = \frac{a_1 + a_2s + a_3s^2 + a_4s^3}{a_1 + a_2s + a_3s^2 + \dfrac{Ka_4 + 1}{K} s^3}, \qquad (6\text{–}21)$$

which is a stable system if appropriate constants are chosen.

"Quickening" is another means of stabilizing the same type of system. Whereas aiding required adding feed-forward loops around the integrators, quickening takes the same signals and adds them into the display. Thus, quickening is feedback compensation of the control system. Figure 6–22 shows a quickened system.

Note that with the quickened system the machine transfer function is the same as for the unquickened system, output/$\delta = a_1/s^3$. That is, with a quickened system, the machine dynamics are *unchanged*. Making the same assumption as before, that the man and the display act as a simple amplifier of gain K, the display to output transfer function is unchanged. output/$\delta = Ka_1/s^3$. It is important to note, however, that the error signal displayed to the man (ϵ_D) is not equal to the actual system error (ϵ_s). In many practical situations, the operator needs both ϵ_D for control and ϵ_s for guidance. The system transfer function is

$$\frac{\text{output}}{\text{input}} = \frac{a_1}{\dfrac{(1 + Kb_3)}{K} s^3 + b_2s^2 + b_1s + a_1}. \qquad (6\text{–}22)$$

The system is stable if the values of the coefficients are properly chosen.

In many real systems the outputs of the integrators shown in Figures 6–21 and 6–22 are not explicitly available. In such cases, it is necessary to obtain these quantities by differentiating the system output. Since differentiation is inherently a noise-amplifying process, approximate differentiation is usually required. Techniques for determining these approximate derivatives are beyond the scope of this chapter. In some cases it may be possible to appromimately quicken a display with an available signal. For example, the derivative of angle of attack is not readily available in most systems, but pitch rate is frequently available. Analysis

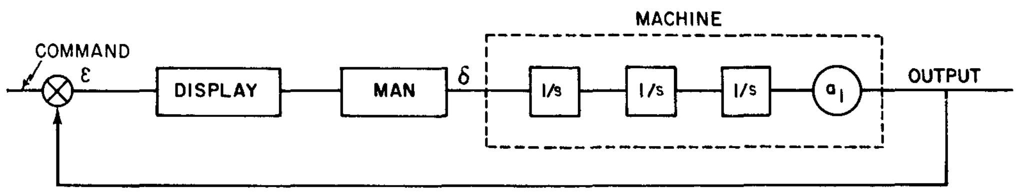


FIGURE 6–20. Uncompensated system.

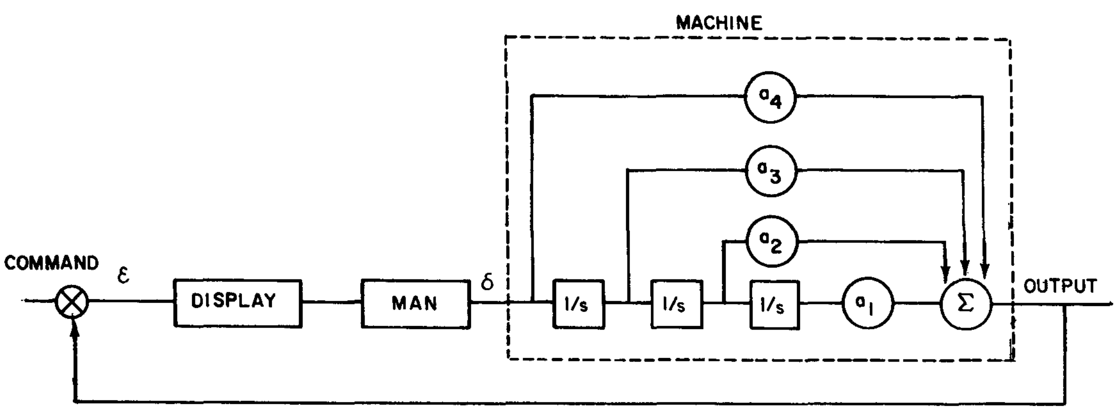


FIGURE 6–21. Aided control system.

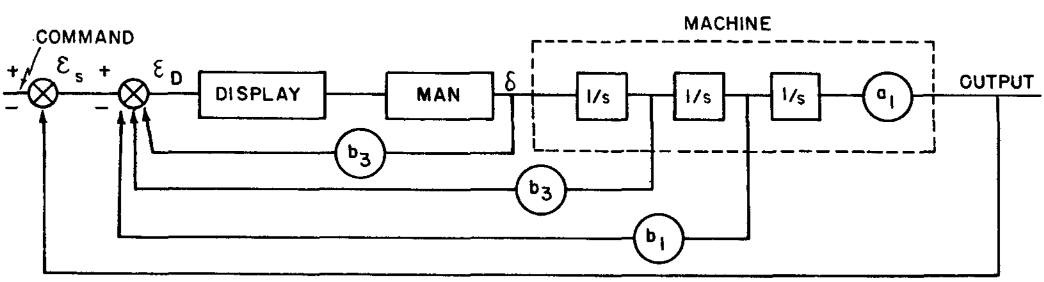


FIGURE 6–22. Quickened control system.

reveals that pitch rate is a good approximation of the derivative of angle of attack if flight path angle is not changing rapidly.

The term *unburdening* denotes a change in the control system to unburden the operator of certain compensation requirements. Specifically, unburdening is raising the *type* of the control system to relieve the requirement of the operator acting as an integrator. In Table 6–1 it was noted that a pure position control system, $Y(s) = K$, has a position error for all inputs. Since it is easily demonstrable that an operator using a position control system can reduce error to zero, the operator must behave like an integrator. By increasing the type of the control from position to rate, the operator is "unburdened" of the requirements to behave like an integrator. He can behave like a simple amplifier. The operator's behavior with various types of controlled element dynamics will be treated in greater detail in subsequent sections.

6.2.5 Man and the Human Describing Function

The human operator in a manned system serves many functions. In this section we are concerned only with his role in the control loop of the system. That is, we are concerned only with man as a controller.

As a controller, man is probably the most versatile element the control system designer has at his disposal. Man is an *adaptive* element in the control system. That is, he can vary his response to a displayed signal, in some fashion, to *optimize* the system performance. Although the criterion that the man uses to judge the optimal performance of the system is not completely known, it appears to be similar to minimization of r.m.s. error or a weighted sum of r.m.s. error and control stick movement. In addition to being an optimizing adaptive element, the operator is also able to express an opinion, occasionally quite vocally, about the quality or "feel" of the system. This "vocal adaptive controller," then, is the object of our discussion.

McRuer and Krendel (1962) summarized the advantages and attendant problems when they said:

> Those human contributions to a control system which make manned control systems most desirable are unfortunately the very contributions most difficult to describe in conventional engineering terms. Such human component attributes as judgment, multimode capability and adaptability make for a remarkable control system. Of these, human adaptability is the attribute which can most nearly be described in engineering terms. Within the confines of the physical constraints present in a given application, human adaptability includes:
>
> (1) Organization of the system structure.
> (*a*) Selection and use, as feedback quantities, of those sensible machine outputs best suited for control purposes.
> (*b*) Setup of an internal organization (equivalent to the construction of several signal-processing paths within the human) which effects highly efficient use of any coherence in the presented stimuli (system forcing function and feedbacks).
>
> (2) Adjustment, within the system structure as organized, of transfer characteristics to forms appropriate for control.
>
> The attributes of judgment, multimode behavior, and adaptability in a control system component allow the design of systems of great reliability and versatility. By using judgment, the man can evolve and modify performance criteria, select relevant inputs, decide between programmed and impromptu responses, and so forth. A multimode capacity enables the man to implement different modes of behavior and thus to effect the diverse requirements which judgment may dictate. To design such control systems, engineering descriptions of the critical component, man, are neceed.

Why Use a Human Describing Function?

For many control system design problems the engineer has available analytical or frequency domain representations for the physical elements of the control system. What in the past has been lacking is an equivalent specification of the performance of a human controller when he is embedded in a feedback control system. The human describing function provides such an engineering model. Its advantage is that it allows analytical prediction of the stability and accuracy of proposed manual control systems. These analyses may be paper-and-pencil analyses, computer simulations, or a combination of both. Frost (1961) provides a simplified example of this use of the human describing function. Other reasons for using the human describing function are to predict critical regions for experimental studies or to rationalize the results of experimental studies. As an example, analyzing manual control of a second order position control system with the aid of the human describing function allows us to predict in what regions the system will seem, to the operator, to be a zero order position control and in what regions it will seem to be a type 2 or acceleration control system. As another example, analysis with the human describing function allows us to predict the interaction effect observed by Rockway (1954) between gain and lag in a simple manual control system. Both of these examples will be treated in more detail in "Performance Prediction." (See Section 6.3.5.)

What it is—and isn't. The human describing function, like ancient Gaul, is divided into three parts: (a) a linear transfer function in general form; (b) a remnant term which accounts for the power in the operator's output that is not accounted for by the transfer function operating on the input; and (c) a set of "adjustment rules" which define how the parameters of the transfer function are adjusted to stabilize and optimize the system performance. There are several levels of detail at which the describing function can be specified. The next three subsections will describe these levels.

The human describing function is not a universal panacea for all the problems of the manual control system designer. It is, however, one of the most powerful tools available to him for analysis and synthesis of manual control systems. The describing function is most thoroughly defined for single axis compensatory systems. This single-axis model works quite well for multi-axis systems where there is no significant cross-coupling between control axes. In cases where there is significant cross-coupling of the control axes, the describing function approach permits qualitative predictions although quantitative prediction is still beyond the state-of-the-art. Research along these lines is in progress. Development of an adequate model for the pursuit display case is in progress, but at present, is not sufficiently refined to allow accurate quantitative predictions. In general, predictions made on the basis of the current model assuming a compensatory display will be slightly conservative if a pursuit display is used in the system.

The frequency variable, $j\omega$, is used in the remainder of this section rather than the Laplace variable, s, to emphasize the fact that these describing functions are valid only in the frequency domain. They should not be used to imply a system's transient response.

Simplest Form

The simplest general form of the linear transfer function part of the human describing function (Y_p) is given by the following:

$$Y_p(j\omega)=K_p\,\frac{(T_Lj\omega+1)}{(T_Ij\omega+1)}\,e^{-j\omega T_e}, \qquad (6\text{–}23)$$

where $e^{-j\omega T_e}$ represents the operator's effective reaction time as a pure transport delay. A transport delay introduces a phase shift that is linearly proportional to frequency. Unlike a simple lag, a transport delay does not affect the amplitude of the signal. This form of the operator describing function includes the main adaptive features of the operator and is sufficiently flexible that it can approximate his behavior in most random-input compensatory tracking systems. It is also of sufficient accuracy, when coupled with the appropriate adjustment rules, to allow predictions of system stability and in many cases system accuracy. The adjustment rules are given in the section on "Human Operator Adaptation". (See Section 6.3.3.) Performance prediction is considered in Section 6.3.5. When more refined estimates of system performance are required than can be made with the above representation, more detailed describing function forms are required.

Additional Terms

The describing function given is a good "mid-frequency" representation of the human operator. Its major limitation is that it does not include an adequate representation of the phase shift introduced by the operator at low frequencies. This large, lagging phase shift is believed to be due to operator neuromuscular system dynamics at frequencies much below those normally of concern in tracking studies. These phase shifts may, however, be of considerable importance in analyzing marginally stable or conditionally stable systems. McRuer et al. (1965) shows that this low-frequency phase shift can be approximated by a term $e^{-j\alpha/\omega}$, where α is an experimentally determined con-

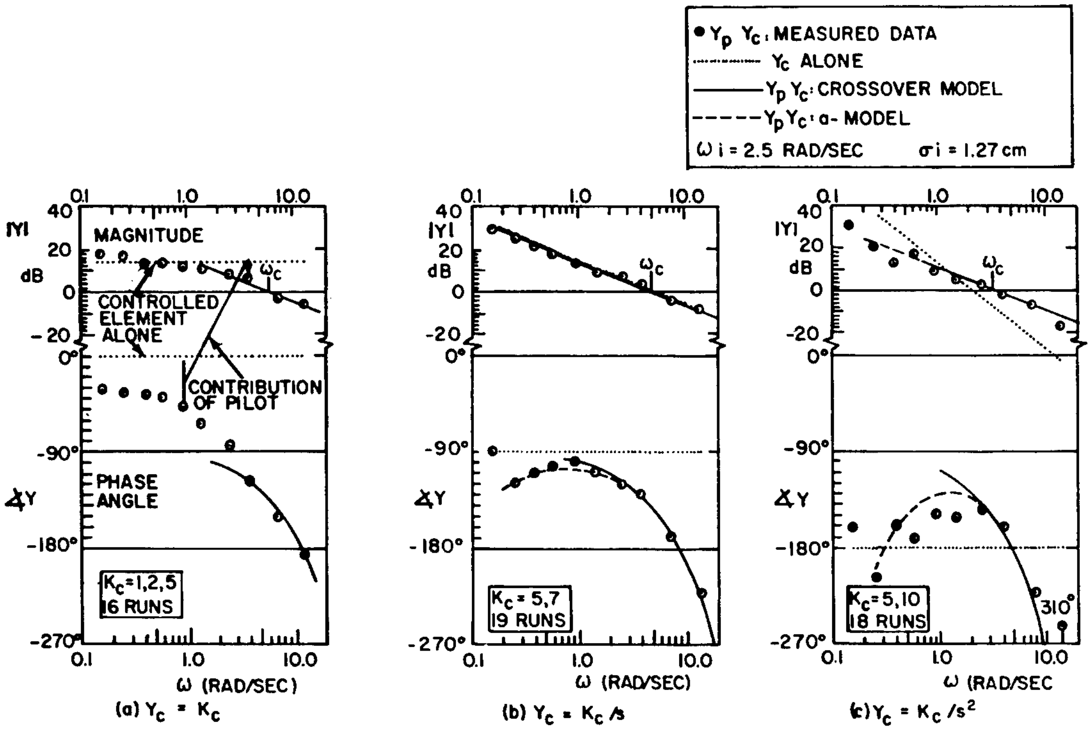


FIGURE 6–23. Measured open-loop describing functions and comparisons with crossover models for $Y_c=K$, K/s, and K/s^2 (McRuer and Jex, 1966).

stant. When this term is combined with the earlier version of the describing function, we have the "extended crossover model" for Y_p:

$$Y_p=K_p\,\frac{T_Lj\omega+1}{T_Lj\omega+1}\,e^{-j(T\omega+\alpha/\omega)}. \qquad (6\text{–}24)$$

The ability of this extended crossover model to fit actual experimental data is shown in Figure 6–23 taken from McRuer and Jex (1966).

The Complete Model (1965)

The extended crossover model given above is adequate for most engineering applications. It does not, however, fit all of the existing data. It does not include terms which, from a physiological standpoint, must be included to reasonably model the neuromuscular system. It does not explain, although it describes, the variations in the effective reaction-time delay of the operator. It is, in other words, a useful engineering approximation rather than a complete description. The most complete form of the model to date is the "precision model" developed in McRuer et al (1965). The precision model (Y_p), exclusive of remnant, is given below:

$$Y_p=K_pK_T\left[\frac{a_T}{\sigma_T}\right]e^{-j\omega\tau}\cdot\left(\frac{T_Lj\omega+1}{T_Ij\omega+1}\right)\left\{\left(\frac{T_Kj\omega+1}{T_{K'}j\omega+1}\right)\left[\frac{1}{(T_{N_1}j\omega+1)\left[\left(\frac{j\omega}{\omega_N}\right)^2+\frac{2\zeta_N}{\omega_N}j\omega+1\right]}\right]\right\} \qquad (6\text{–}25)$$

K_p is the operator's zero frequency gain, adjustable for optimizing system performance. The operator has definite gain preferences, but can adapt over a range of at least 100:1.

The term

$$K_T\left[\frac{a_T}{\sigma_T}\right]$$

is a representation of the operator's "indifference" threshold. It is a nonlinearity of relatively minor effect, but it models the observed data which show that an operator does not respond to very small tracking errors.

The term

$$e^{-j\omega\tau}$$

is a transport delay representing the minimum reaction time of the operator due to neural conduction times. For eye-hand tracking, τ is between 0.06 and 0.10 sec.

The term

$$\left[\frac{T_L j\omega+1}{T_I j\omega+1}\right]$$

is a lead and a lag which account for most of the operator's adaptability to various system dynamics. The operator may adopt pure lead, pure lag or some combination of both as required to optimize the particular system being controlled.

The term

$$\left[\frac{T_K j\omega+1}{T_{K'} j\omega+1}\right]$$

is a very low-frequency lead and lag combination which approximates even more complex neuromuscular loops which are not well understood. The mid-frequency effect of this term is the α term of the extended crossover model. It is the simplest term which can fit all of the low-frequency data.

The complex term

$$\frac{1}{(T_{N_1} j\omega+1)\left[\left(\frac{j\omega}{\omega_N}\right)^2+\frac{2\zeta_N}{\omega_N}j\omega+1\right]}$$

includes all of the high-frequency neuromuscular effects and is based only in part on the frequency response data. The other basis for this term is the known physiology of the arm. This term is the minimal analytical description of the muscle impulse response. For a more thorough treatment of this aspect of the model see McRuer and Magdaleno (1966).

Model for Periodic Forcing Functions

Sinusoidal inputs to the pilot yield different behavior than unpredictable inputs. Such sinusoidal inputs may exist alone in the laboratory or as a dominant feature in real-world inputs. In either case, they require a different model to describe the pilot's in the system. McRuer and Jex (1966) present a model which describes and summarizes virtually all the existing sinusoidal tracking data. It must be emphasized that although the operator's response to single sinusoids is presented on a frequency response plot different psychomotor mechanisms may be involved in the low-, medium-, and high-frequency regions. The ensemble of responses to single sinusoids must *not* be interpreted as a quasi-linear describing function for random appearing inputs.

Typical data for sinusoidal tracking are shown in Figure 6–24 in terms of amplitude ratio and phase shifts (McRuer and Jex, 1966). The "free wheeling" limit, shown as a shaded boundary, is the maximum oscillation frequency achievable at specified amplitudes without any input.

For sinusoidal inputs up to about 2 Hz, the operator forms an additional feed-forward path in parallel with the compensatory tracking model. This is shown in Figure 6–25.

Initially, the model is as shown. The compensatory model (Y_p) is required to track the remnant noise injected at the operator's output. With practice, the operator acts progressively more like a synchronous generator, phase-locked to the zero-crossings of the input. He is, in effect, tracking in an open-loop mode and the switches are reversed from the positions shown.

For inputs between 2 and 5 Hz, the operator can no longer close the compensatory loop. The adaptation of the compensatory loop changes, or that loop is opened completely, and the synchronous generator is no longer phase locked to the input. This results in the phase of the output gradually shifting with respect to the input. Actual frequency drifting also sometimes occurs.

6.3 Analyzing The Control Loop

The preceding sections have described the various elements of a manual control loop and defined certain basic characteristics of these elements. This section discusses how these

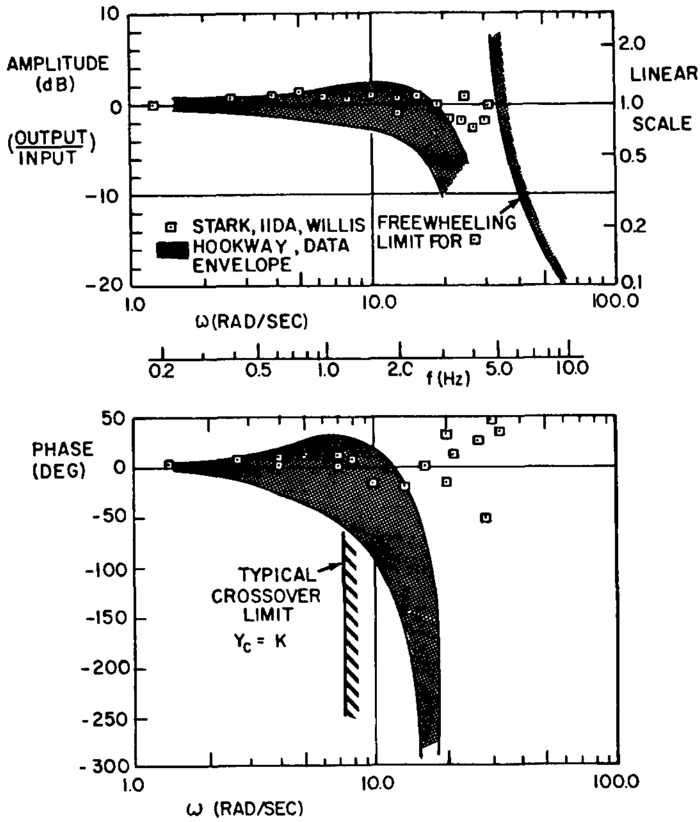


FIGURE 6–24. Frequency response data for a single sine wave input at various frequencies (McRuer and Jex, 1966).

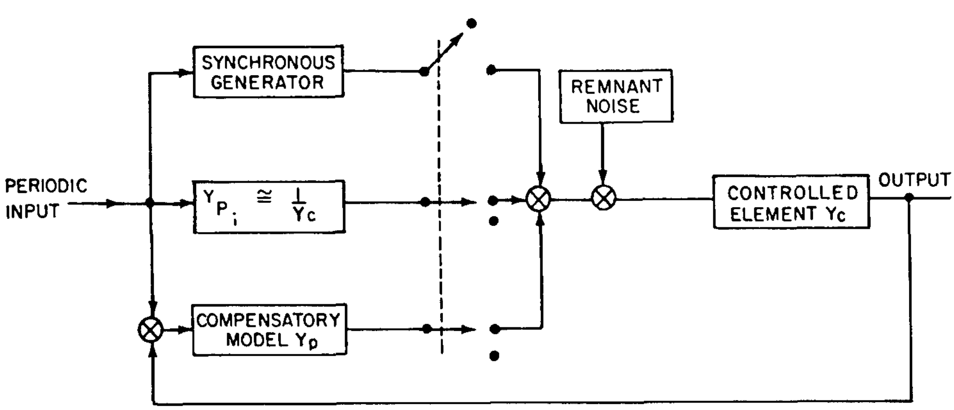


FIGURE 6–25. Block diagram for periodic waveform tracking model (McRuer and Jex, 1966).

elements are interconnected to form useful and versatile control systems. The operator in a control system may serve to stabilize a basically unstable controlled element. On the other hand, he may induce instability into a stable system. He may track very well and reduce the system error a hundred fold, or he may introduce more error than would be present if he did nothing. The intent of this section is to present sufficient information to:

1. Predict system stability;
2. Describe how and over what ranges the operator can adapt to improve system performance;
3. Indicate how the controlled element can be modified to improve system performance; and to
4. Present techniques for predicting system performance and the effects of changes in the controlled element.

6.3.1 Criteria of Useful Control Systems

The following general criteria can be applied to any control system:

1. Stability,
2. Accuracy,
3. Speed of response,
4. Reliability, and
5. Cost.

This section is concerned only with the first three criteria. They are predictable by control system analysis techniques. The other two criteria are beyond the scope of this chapter although not entirely out of the domain of the human factors specialist, since both reliability and cost are heavily influenced by whether or not there is a man in the control loop, and what controls and displays are provided for him.

Stability is the *sine qua non* of any control system. For the purposes of this chapter, the following qualitative definition will be used:

> *If a small temporary input applied to a system in equilibrium casuses only a temporary change in the output, the system is stable.*

Note that this definition, unlike the usual mathematician's definition of stability, excludes the special case of "marginal stability" in which the system response neither grows nor decays after the disturbance is removed. For most control purposes, such a system is considered unstable. Criteria for stability of linear systems are given in Section 6.3.2.

Accuracy of a control system must be defined in terms of what the system is intended to do, i.e., in terms of system criteria. For positioning systems accuracy might be defined in terms of steady-state errors. For regulating systems with external disturbances, accuracy is usually defined in terms of average absolute error or r.m.s. error. For gunnery systems time on target may be the most meaningful measure, while for landing an aircraft some time- or distance-weighted function of error may be most appropriate.

Each of the above measures is a summary statistic, and must be interpreted with the same "due respect and caution" that should be accorded all summary statistics. That is to say, each score from an individual tracking run is a summary statistic describing something about the continuous error signal during that run. Thus, the set of scores at the end of a tracking experiment is an ensemble of estimates about the errors which occurred during each of the individual runs.

Consider an arbitrary time function, $\epsilon(t)$ as shown in Figure 6–26. This time function can be quantized by periodic sampling, yielding an ordered series of discrete values. If it is sampled frequently enough the discrete points completely define the function as in Figure 6–26b. Note that nothing has been said about the mean or distribution of these points, so they must be considered analogous to "raw score" data in conventional statistics. This time function can be summarized by a plot of the frequency of occurrence of each magnitude versus that magnitude. This essentially "collapses" the time axis on itself and is shown in Figure 6–26c. In general, this distribution will be a more or less "bell-shaped" curve.

Various experimental conditions can be expected to change the center of the distribution and/or its spread. Thus, the scores must be measures of the central tendency and the variability of the error signal. Table 6–2 shows the most common tracking scores, their exact analog in discrete statistics, and the common corresponding discrete statistic. It must be

TABLE 6–2. COMMON TRACKING SCORES AND ACCOMPANYING DISCRETE STATISTICS

$1/T \int_0^T \epsilon \, dt$	Integrated error	$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i$	Mean error	$\frac{\Sigma X}{N} = \bar{X}$	Mean of distribution
$1/T \int_0^T \|\epsilon\| dt$	Integrated absolute error	$\frac{1}{N} \sum_{i=1}^{N} \|\epsilon_i\|$	Average absolute error	$\frac{\Sigma \|X\|}{N}$	No Standard Equivalent. This is equivalent to the average deviation only if the mean ($\bar{X}$) is identically zero so that $X = x$.
$1/T \left[\int_0^T \|\epsilon\| dt - \int_0^T \epsilon \, dt \right]$	Integrated absolute error corrected for non-zero mean error.	$\frac{1}{N} \left[\sum_{i=1}^{N} \|\epsilon_i\| - \sum_{i=1}^{N} \epsilon_i \right]$	Average absolute error corrected for non-zero mean.	$\frac{\Sigma \|X\|}{N} - \bar{X}$	No standard equivalent. It is *not* the average deviation.
Not computable "on-line" since the mean error is not known until end of run.		$\frac{1}{N} \left[\sum_{i=1}^{N} \left\| \epsilon_i - \frac{\Sigma \epsilon_i}{N} \right\| \right]$	Average deviation of the error	$\frac{\Sigma \|x\|}{N} = \frac{\Sigma \|X - \bar{X}\|}{N}$	Average deviation.
$1/T \int_0^T \epsilon^2 \, dt$	Integrated squared error.	$\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2$	Average squared error.	$\frac{\Sigma X^2}{N}$	No standard equivalent. This is equivalent to the variance only if the mean ($\bar{X}$) is identically zero so that $X = x$.
$\left(\frac{1}{T} \int_0^T \epsilon^2 \, dt \right) - \frac{1}{T} \left(\int_0^T \epsilon \, dt \right)^2$	Integrated squared error corrected for non-zero mean error.	$\left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \right) - \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2 \right)$	Average squared error corrected for non-zero mean equal to variance.	$\sigma^2 = \frac{\Sigma x^2}{N} = \frac{\Sigma (X - \bar{X})^2}{N}$	Variance.
$\sqrt{1/T \int_0^T \epsilon^2 \, dt}$	Root mean square error.	$\sqrt{\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^2}$		$\sqrt{\frac{1}{N} \Sigma X^2}$	No standard equivalent. This is equivalent to the standard deviation only if the mean ($\bar{X}$) is identically zero so that $X = x$.
$\frac{100}{T} \int_0^T f(\epsilon) dt$ $f(\epsilon) = 1$ if $a < \epsilon < b$ $f(\epsilon) = 0$ elsewhere	Percent time on target.	$\frac{100}{N} \sum_{i=1}^{N} F_i$ $F_i = 1$ if $a < \epsilon < b$ $F_i = 0$ elsewhere	100 times the probability that ϵ is greater than a but less than b		

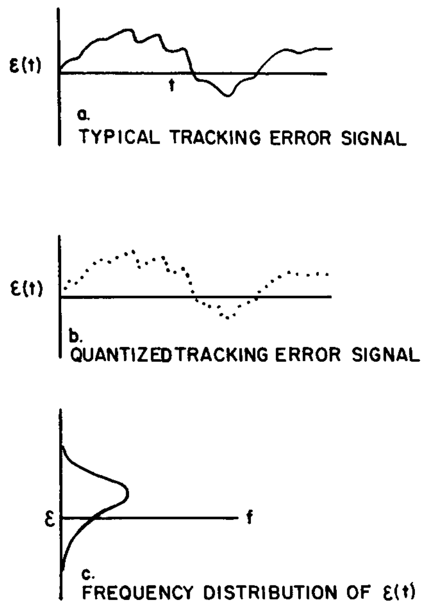


FIGURE 6–26. Tracking error function.

emphasized that neither the entire distribution nor its true shape is known. Since the underlying distribution of tracking errors is generally not known, the distribution of the scores cannot be predicted. For this reason, serious consideration should be given to nonparametric statistical analysis of the results of tracking experiments. Bradley (1960) provides an excellent review of the strengths and weaknesses of the various nonparametric tests as compared with similar parametric tests. If it is known that the error is normally distributed about a zero mean, there are little grounds for choice between Integrated Absolute Error (IAE), and Integrated Squared Error (ISE) since Magdaleno and Wolkovitch (1963) have shown that

$$1/T \int_0^T |\epsilon| \, dt = \sqrt{\frac{2}{\pi} \frac{1}{T} \int_0^T \epsilon^2 \, dt}. \quad (6\text{–}26)$$

For most simple manual control systems, Elkind and Darby (1963) have shown that if the input signal is Gaussian, the error and the system output are Gaussian. On the other hand, altitude errors in terrain following and landing approaches tend to be bimodal and biased to the high side of the command signal. Senders and Ward (1965) have shown that for step function tracking, the choice of metric can make great differences in the resulting system. In one instance they showed that if the system gain was optimized for minimum settling time, it was off optimum by 3:1 for minimum integrated error. Thus, when speaking of optimum systems, the question, "Optimum for what?" must always be asked. Because of the variety of manual control tasks, no simple set of rules to choose the appropriate metric can be formulated.

6.3.2 System Stability Criteria

To provide the reader with an intuitive "feel" for what makes a system stable or unstable and to present certain simplified techniques for predicting stability, the discussion in this section will be based on machine systems. In the next section the application of these rules to manual control systems will be treated.

A system is stable if, when the input is removed, the output eventually comes to rest at a finite value. Thus, an integrator is a stable element, while two or more cascaded integrators constitute an unstable element. (See Figure 6–13.) Any closed-loop system provides a possibility for instability. This is because of the possibility that the output may be fed back in such a way that it adds to the input and causes the output to grow. If a portion of the output of an integrator (a stable element) is fed back with the same sign as the input, the output will continue to grow even after the original input is removed. This is known as positive feedback. The feedback caused the instability. On the other hand, if the output of the integrator were fed back with the opposite sign, the output would gradually rise to the level of the input and then decay to zero when the input was removed. Thus the sign of the feedback is one of the criteria for stability.

For a unity feedback system, the sign of the feedback signal is the same as the sign of the output. Actually, the output is seldom related to the input by a pure gain. Usually the output lags the input slightly. Thus the feedback signal can be thought of as a vector quantity. The length of the vector is proportional to the gain of the system, and its angle is the phase angle between input and output. As discussed earlier under "Frequency Response," in Section 6.2.4., the amount of phase shift, as well as the gain of the system, is usually a function of frequency. As shown there, the amount of lag introduced between input and output is also a function of the order of the system. In some control systems there is sufficient lag in the system that too much gain will cause instability. Such systems are stable in normal operation, but when the gain is increased to get "tighter" control, the system output begins to oscillate and the oscillations grow until either the gain is reduced or the system destroys itself. The same thing can happen in manual control systems when the operator tries too hard to chase the input, and operator-induced oscillations result.

The differential equation of a control system contains all of the necessary information concerning system performance. *The true test for system stability is to check for the existence of positive real roots or complex roots with positive real parts.* If any such roots exists, the system is unstable; if they do not exist, the system is stable. The various tests of system stability are simply tests for the existence of positive real roots of the differential equation. One simple test is that of inspection of the differential equation. If any derivative of the equation is missing, the system is unstable. Similarly, all terms must be positive or all terms must be negative for stability. If any sign is reversed, the system is unstable. Unfortunately, although these simple tests provide necessary conditions for stability, they are not sufficient to guarantee stability. That is, it is possible for a positive real root to exist without affecting the differential equation sufficiently to indicate instability by inspection.

Routh's Criterion

In general, a linear differential equation has the form

$$A_0S^n+A_1S^{n-1}+\cdots+A_{n-1}S+A_n=0. \quad (6\text{–}27)$$

The coefficients may be arranged in a triangular array as follows

$$A_0A_2A_4A_6A_8\ldots$$

$$A_1A_3A_5A_7\ldots$$

$$b_1b_3b_5\ldots$$

$$C_1C_3\ldots$$

$$d_1\ldots$$

where

$$b_1=\frac{A_1A_2-A_0A_3}{A_1}$$

$$b_3=\frac{A_1A_4-A_0A_5}{A_1}$$

etc.

$$C_1=\frac{b_1A_3-A_1b_3}{b_1}$$

$$C_3=\frac{b_1A_5-A_1b_5}{b_1}$$

$$d_1=\frac{C_1b_3-b_1C_3}{C_1}$$

etc.

Each succeeding horizontal row will have fewer terms than the preceding row, and thus the array is triangular. The procedure for forming additional rows must be carried out until no more rows can be formed. Once the array is complete, it is necessary only to inspect the *signs* of the first term in each row. If all terms have the same sign, there are no positive real roots. If there are changes in sign, the number of changes corresponds to the number of positive real roots. A useful relationship derived from Routh's criterion is that for cubic equations of the form

$$As^3+Bs^2+Cs+D=O. \quad (6\text{–}28)$$

To be stable, AD must be greater than BC. For systems of higher than third order, the labor involved in applying Routh's criterion becomes considerable and simpler techniques become desirable.

Nyquist and Bode Criteria

When designing or analyzing a control system, it is usually more convenient to deal with the transfer function than with the differential equation of the system. It is therefore desirable to have a test for stability which can be directly applied to the transfer function equations.

It was shown in Section 6.2.3 that for a unity feedback control system with a forward loop transfer function $Y(s)$, the closed-loop transfer function must be

$$\frac{e_o}{e_i} = \frac{Y(s)}{1+Y(s)}. \qquad (6\text{–}29)$$

If the system is unstable, then $e_o/e_i = \infty$ for some value of s. This can happen only if $1 + Y(s) = 0$, which requires that $Y(s) = -1$.

Since $Y(s)$ is complex, $(s = \Delta + j\omega)$, this really says $Y(s) = -1 + j0$. It can be shown (Thaler and Brown, 1953) that if the system transfer function, $Y(s)$, is plotted in polar form for $s = j\omega$, $0 < \omega < +\infty$, the resultant plot can be used to predict system stability. This polar plot is known as a Nyquist plot.

If the plot encircles the $-1+j0$ point in a *clockwise* direction, the system is unstable. If it does not encircle the $-1+j0$ point, the system is, in general, stable. Nyquist plots illustrating stable and unstable systems are shown in Figure 6–27. The system represented by the solid line is stable, while the one represented by the broken line is unstable. Note that the encirclement requires that the magnitude of $Y(j\omega) > 1$ when the angle of $Y(s) = -180°$.

The calculation and plotting in polar coordinates $Y(j\omega)$ for many values of ω becomes quite laborious. Fortunately an easier technique exists. The logarithmic plots (Bode plots) of frequency response of various elements have been introduced under "Frequency Response" (Section 6.2.4.) It was noted that the asymptotic response of various elements was easily sketched and that the response of a complex system could be sketched by graphically adding the responses of the various elements. The amplitude and phase relationships at each frequency for the composite $Y(j\omega)$ could be read off those plots and replotted in polar form to test for stability. This is not necessary since all of the required information already exists in the logarithmic plots.

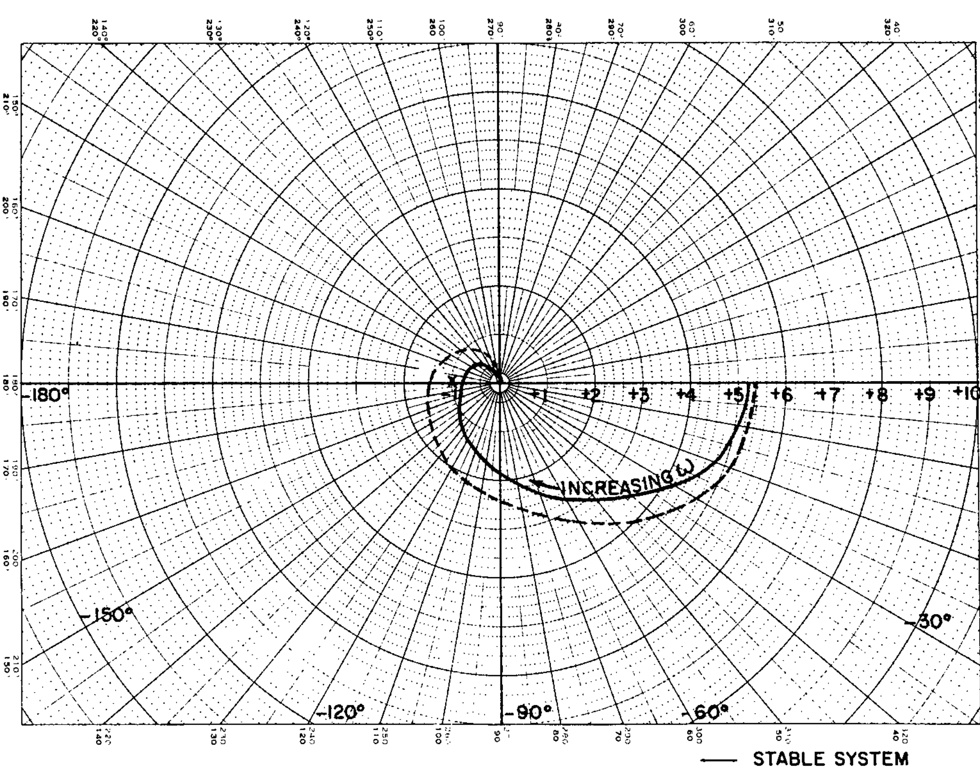


FIGURE 6–27. Typical Nyquist polar plots.

Figure 6–28a and b shows a Bode plot and the corresponding Nyquist plot for a specific system. The solid line shows a stable system while the dotted line shows an unstable system. The only difference between the two systems is in the gain term K; consequently, the phase curve is the same for both systems. Note in the Bode plot those frequencies at which the amplitude ratio lines cross the 0-dB (unity gain) line. This is the unity gain crossover frequency ω_c. It corresponds to the frequency at which the Nyquist plot crosses the unity gain circle. If ω_c on the Bode plot is at a higher frequency than that at which the Bode phase plot crosses the $-180°$ line, the Nyquist plot will encircle the -1 point, and the system will be unstable. Conversely, if ω_c is less than the frequency at which the phase curve crosses the $-180°$ line, the Nyquist plot will not encircle the -1 point, and the system will be stable.

Once it has been shown that the system is stable, the next point of interest is transient performance. *How* stable is the system? If the system overshoots wildly and continues to oscillate for a long time it is usually not useful and must be considered only relatively stable. Consider Figure 6–28. It can be seen that gradually increasing the gain of the stable system will eventually result in instability. This transition from stability to instability is not abrupt. As the gain increases, the system becomes more oscillatory and the transients take longer to damp out until the point is reached that the system is unstable.

Two convenient measures of relative stability are the *gain margin* and the *phase margin* of the system. These are most easily seen in the Bode plots. Gain margin is defined as that amount by which the system gain would have to be increased to make the unity gain crossover frequency (ω_c) correspond to the frequency at which the phase curve crosses the $-180°$ line. Thus, in Figure 6–28b the gain margin is 3 dB. Similarly the phase margin is defined as the amount of negative phase shift that would be required to make the phase curve cross the $-180°$ line at ω_c. In Figure 6–28b the stable system has a phase margin of 27°. The unstable system has a negative phase margin of 18°. That is, a positive phase shift of 18° would be required to make the system stable if the gain were not changed. For satisfactory performance, phase margins between 30° and 45° are generally required. A phase margin of about 60° is generally required for critical damping.

6.3.3 Human Operator Adaptation

The human operator in a manual control loop adapts his behavior to the requirements of the situation. In this section we will describe what changes the operator can make in his behavior and what features of the situation require these changes. The reader who desires a more complete treatment of this subject is referred to McRuer et al. (1965).

The Adaptive Terms

The principal adaptive terms of the human describing function are:

K_p	the operator's zero-frequency gain;
$[T_L j\omega+1]$	the operator's lead time constant (relative rate-to-displacement sensitivity); and
$[T_I j\omega+1]^{-1}$	the operator's lag time constant (relative integral-to-displacement sensitivity).

These are the terms which the operator varies to (a) achieve system stability, and (b) to optimize system performance. They are the terms which allow the operator to compensate for changes in the system dynamics to maintain reasonably constant system performance over a wide range of conditions. T_L can vary from 0 to about 5 sec. while T_I can vary from 0 to about 20 sec. (McRuer and Krendel, 1962). K_p can vary over a range of at least 100:1. The actual values for a particular system depend primarily upon the dynamics of the system and to a lesser extent upon the bandwidth of the forcing function. The actual values adopted by a *trained* operator are quite predictable using the adjustment rules in the next section.

It should be noted at this point that the operator describing function shows little intersubject and intrasubject variability *at or near the system unity gain crossover frequency*, although the variability outside of that region may be larger. It is these variations, well away from the crossover frequency, that account for variations

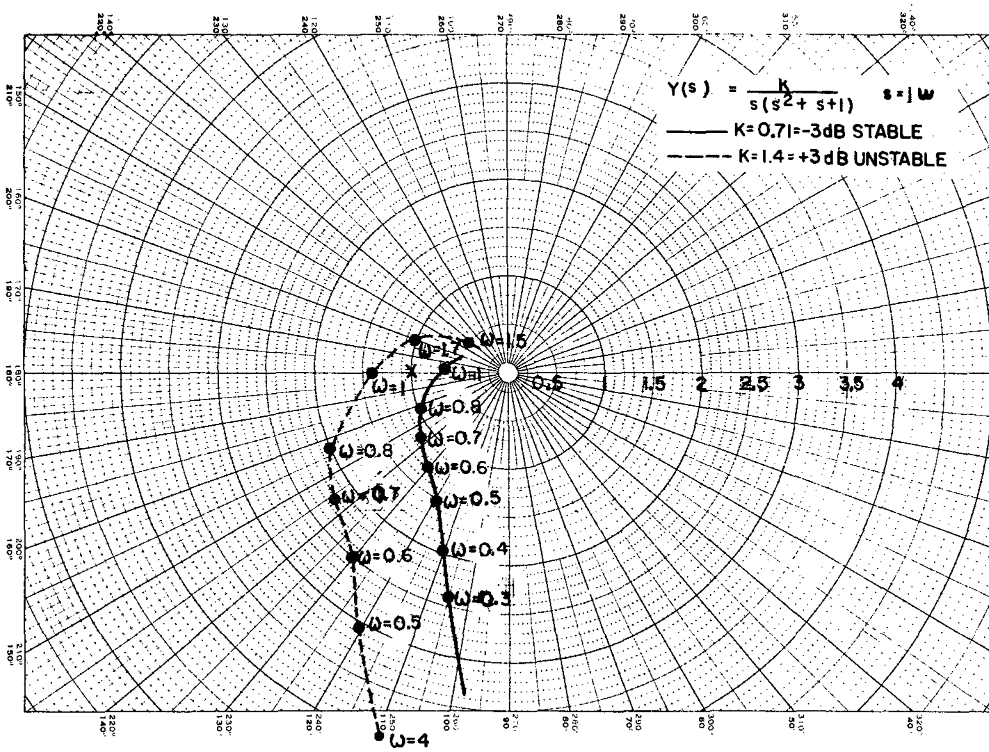


FIGURE 6–28(a). Nyquist plot showing effect of gain change.

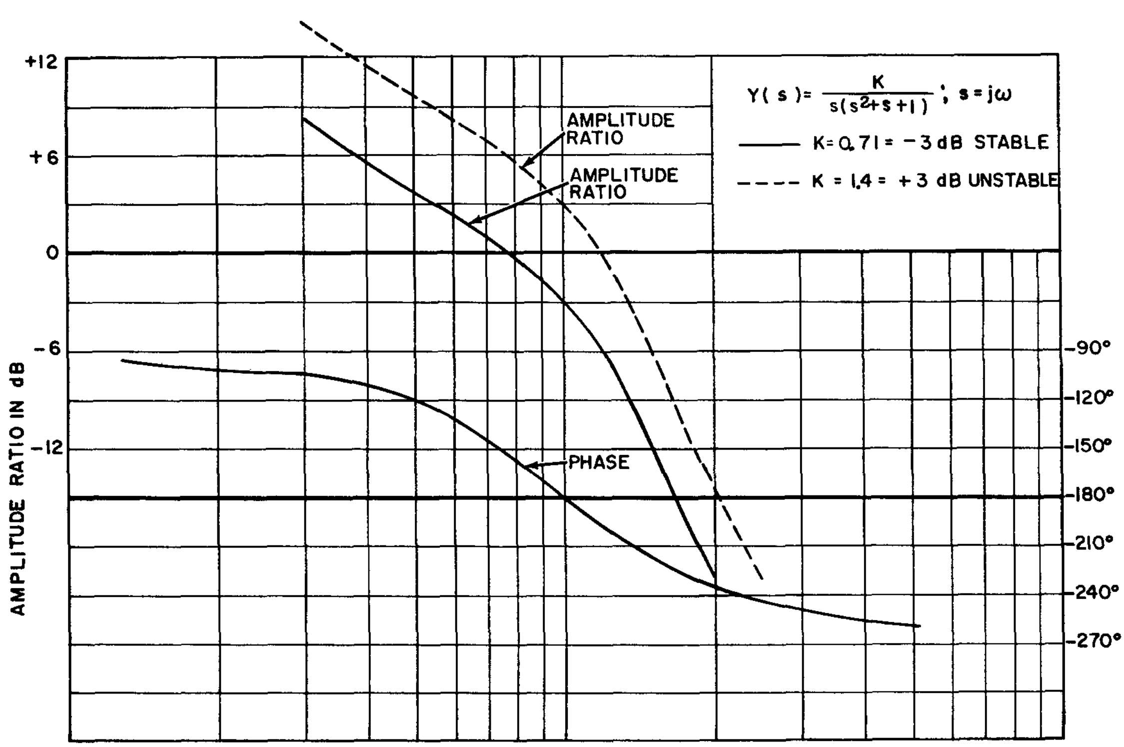


FIGURE 6–28(b). Bode plot showing effect of gain change.

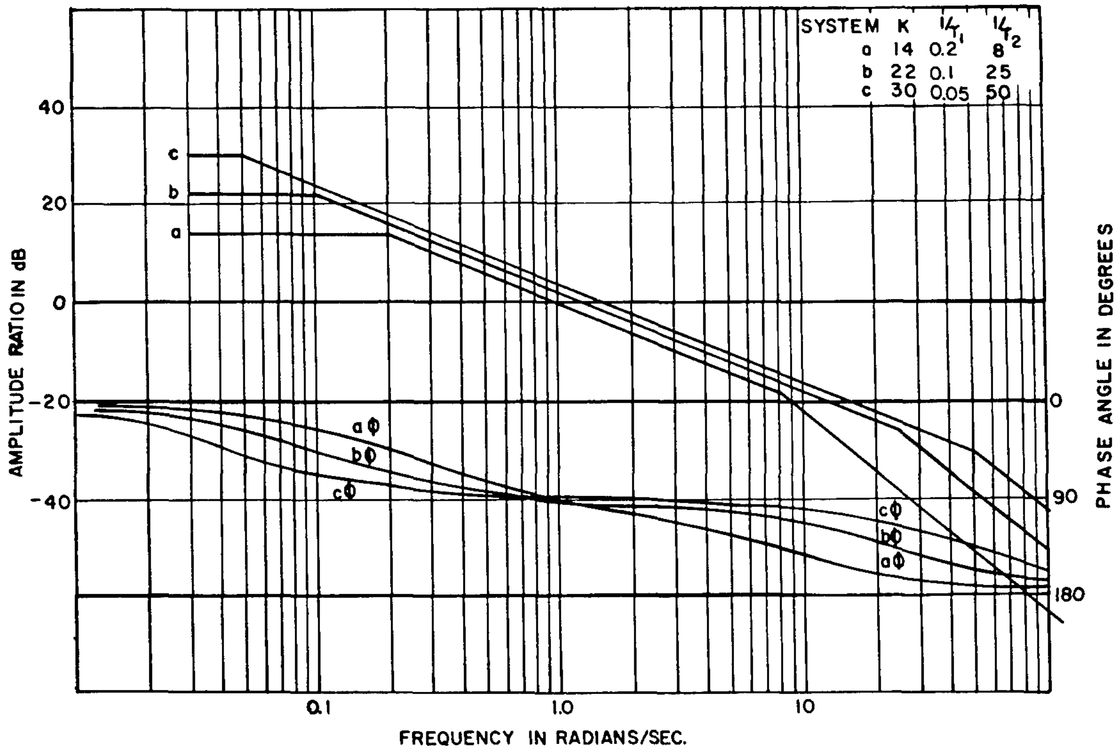


FIGURE 6–29. Bode plot showing effect of simultaneous changes in gain and time constants $\frac{K}{(T_1 j\omega + 1)\ (T_2 j\omega + 1)}$.

in operator style. Figure 6–29 shows the Bode plots for three illustrative systems to demonstrate this point. These are not typical man-machine systems. Although the zero frequency gains are more than 15 dB apart and the time constants are very different, the systems are similar in the crossover region. The systems would behave similarly since the crossover frequencies are very close, the slopes near crossover are identical, and the phase margins are almost identical. This ability to maintain nearly identical dynamics near crossover while letting the system vary away from that region has caused some confusion in the past. It must be emphasized that it is *system* behavior near crossover that is the primary determinant of system tracking performance.

The Adaptation Rules

The human operator in a control system has the ability to adjust his behavior so that the system closed-loop response fulfills the basic requirements of any good feedback control system:

1. Provide specified command response relationships;
2. Suppress unwanted inputs and disturbances;
3. Provide adequate closed-loop stability margins; and
4. Reduce the effects of variations and uncertainties in the components of the loop.

The operator appears to follow the *Primary Rule of Thumb* for synthesis in the frequency domain (McRuer and Graham, 1963):

> At frequencies just beyond the input bandwidth, seek or create (by equalization) a fair stretch of –20 dB/decade slope for the amplitude ratio and adjust the loop gain so as to put the unity-amplitude cross-over frequency near the center of this region, while maintaining adequate stability margins.

Following this rule of thumb, the operator adopts sufficient lead or lag equalization so that the slope of the *system* open-loop describing function lies very close to −6 dB/octave in the region of crossover. The price of this adaptability is an accumulation of additional phase lags due to neural transport delays and high frequency neuromuscular dynamics, all of which can be represented (near crossover) by an effective time delay τ_e. Based on the foregoing, a very simple model can be used to account for most of the significant trends in the region of unity gain crossover. The "crossover model" describes the open loop *system* behavior as follows:

$$Y(j\omega) = Y_p(j\omega) \cdot Y_c(j\omega) \cong \frac{\omega_c e^{-j\omega\tau_e}}{j\omega}; \text{ near } \omega_c. \quad (6\text{–}30)$$

Y_p is the operator's describing function and Y_c is the machine transfer function. The crossover frequency is equal to the loop gain, which is the product of machine and operator gains ($K_p \cdot K_c$). This simple crossover model is summarized in the first two "adjustment rules" (McRuer and Jex, 1966):

Rule 1. The particular equalization is selected from the general form $K_p(T_L j\omega+1)/(T_I j\omega+1)$ such that the following properties obtain:

1. The system can be stabilized by proper selection of gain, K_p, preferably over a very broad range of K_p.
2. Over a wide frequency range, near the crossover region the magnitude ratio $|Y_p \cdot Y_c|$, has approximately a −20 dB/decade slope.
3. $|Y_p \cdot Y_c| \gg 1$ at low frequencies to provide good low frequency closed-loop response to system commands and suppression of disturbances.

Rule 2. Within the operator's intrinsic limitations, and once the −20 dB/decade slope has been achieved, the adjustments of crossover frequency and effective time delay are such as to minimize the mean-square error.

The operator cannot change his basic reaction time or neural conduction delays, but he can change his effective time delay. This is true because τ_e is an accumulation of phase shifts including not only basic delays but also high frequency neuromuscular lags which can be controlled to a limited extent.

Reduction of the neuromuscular lag can be accomplished by "tightening up" the neuromuscular loop by tensing the muscles, but this requires effort. It is interesting to note that if the muscles of the arm are "tightenened" sufficiently in an isometric contraction, tremor results. This is closely analogous to increasing the gain of a servo system until the point of instability is reached. Thus, "tightening up" the neuromuscular subsystem is analogous to increasing the subsystem loop gain. If lead equalization is not required to achieve the desired open-loop describing function, the operator can generate high-frequency lead to partially cancel his effective time delay. Lead generation requires effort. Consequently, low values of τ_e are generated by the operator only when required to achieve good system performance. τ_o, the value of τ_e when the neuromuscular subsystem is relaxed and no high-frequency lead is used to cancel the lag, is a function of the form of the controlled element Y_c in the region of *system* unity gain crossover. Table 6–3 gives values for τ_o for various forms of Y_c.

TABLE 6–3. VALUES FOR τ_0 FOR VARIOUS FORMS OF Y_c

Y_c in crossover region	Amplitude ratio slope	τ_0 (sec.)
K	0	0.30
K/s	−20 dB/decade	.33
K/s^2	−40 dB/decade	.48

McRuer et al. (1965).

The approximate crossover frequency, ω_{c_i}, for very low-frequency inputs can be estimated from consideration of the system phase margin. It has often been noted that when the input to a control system is removed, the operator continues to track his own error for an indefinitely long period. This implies essentially neutral stability or zero phase margin. For a system of the form assumed in the "crossover model,"

$$Y(j\omega) = \frac{\omega_c e^{-\tau_e j\omega}}{j\omega}, \quad (6\text{–}31)$$

the phase margin is given as:

$$\phi_m = \frac{\pi}{2} - \tau_e \omega_c. \quad (6\text{–}32)$$

Letting the phase margin, ϕ_m, equal zero and substituting τ_o for τ_e yields the very low input frequency crossover frequency.

Rule 3.

$$\omega_{c_o} \cong \pi/2\tau_o. \quad (6\text{–}33)$$

Since the value of ω_{c_o} given by Rule 3 is based on phase margin, it is essential that if the system includes other time delays, they must be added to τ_o in estimating ω_{c_o}.

The value of ω_{c_o} is generally a good estimate of the actual system crossover frequency for most random-input conditions. The actual values of ω_c determined experimentally for a variety of input bandwidths and system types are given in Figure 6–30. For most preliminary design, the values estimated by Rule 3 together with the "ω_c invariance properties" of Rule 4 are quite adequate.

Rule 4.

1. $\omega_c - K_c$ *Independence.* After initial adjustment, changes in controlled element gain, K_c, are offset by changes in pilot gain, Kp; i.e., system crossover frequency, ω_c, is invarient with K_c.
2. $\omega_c - \omega_i$ *Independence.* System crossover frequency depends only slightly on forcing function bandwidth for $\omega_i < 0.8\,\omega_c$.
3. ω_c *Regression.* When ω_i nears or becomes greater than 0.8 ω_{c_0}, the crossover frequency regresses to values much lower than ω_{c_0}.

No simple rule yet exists for quantitatively predicting the degree of ω_c regression. Some empirical data are presented in McRuer et al (1965).

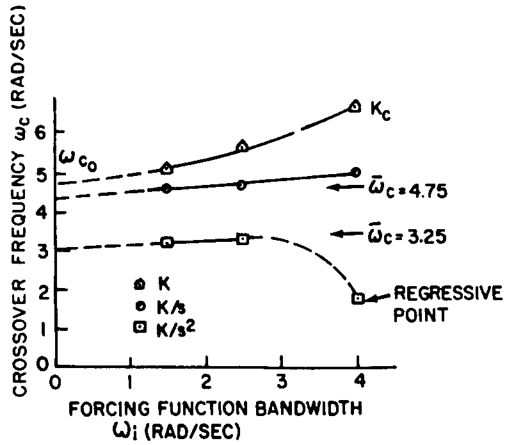


FIGURE 6–30. Dependence of crossover frequency on type of controlled element and input bandwidth (McRuer and Jex, 1966).

Once the crossover frequency is fixed, the operator apparently reduces his effective reaction time to increase the system closed-loop damping ratio and reduce the resonant peak that would otherwise result in large tracking errors. The experimental data in McRuer et al. (1965) indicate that the actual τ_e adopted by the operator is a strong function of the actual system bandtor width.

Rule 5.

$$\tau_e(Y_c, \omega_i) \cong \tau_o(Y_c) - .06\,\omega_i \qquad (6\text{–}34)$$

This relationship is entirely empirical and is based on data taken in a fixed-base simulator using a spring-loaded control stick and a single-axis compensatory CRT display. It has not been validated for multi-axis, moving base or pursuit display situations. It does, however, provide a useful approximation of how the operator adapts his behavior to higher frequency inputs.

The relatively simple crossover model, presented above with the operator adjustment rules, provides a useful model for many design problems. Near crossover it fits the available data quite well, but unfortunately, at the lower frequencies it does not fit the phase data. It is to fit these low-frequency phase data that the "extended crossover model" introduced in Section 6.2.4 is required. This model adds a new term to the simple model to describe the low-frequency phase "droop." This additional term, $e^{-\alpha/\omega}$, is a simple approximation to account for phase shifts due to extremely low (below the available measurement band) frequency leads and lags in the neuromuscular subsystem. These low frequency lags are especially important in analyzing conditionally stable systems (e.g., when a pilot is stabilizing an unstable vehicle) because such situations result in low phase margins. These low phase margins limit the stable crossover region and require a fairly accurate phase representation at low and mid-band frequencies. There is currently no simple adjustment rule to predict the value of α to be used in modeling the operator, although a considerable body of data exists. These data are summarized in Figure 6–31 and Figure 6–32 (McRuer et al., 1965). For pure gain-controlled elements, the data do not show any α effect. While there are no simple rules for predicting the value of α, the empirical data are sufficiently consistent that the α effect should not be overlooked when predicting system performance.

Handling Qualities

The operator in a control system has been referred to as "the vocal adaptive controller."

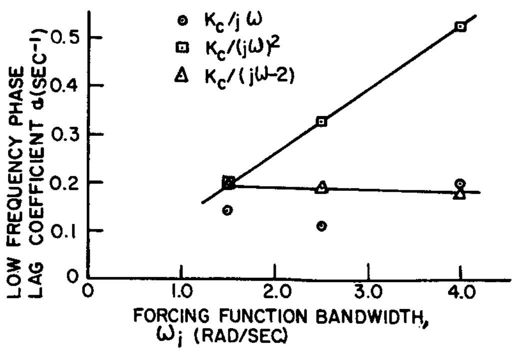


FIGURE 6–31. Variations of low-frequency incremental phase lag with forcing function bandwidth (McRuer et al., 1965).

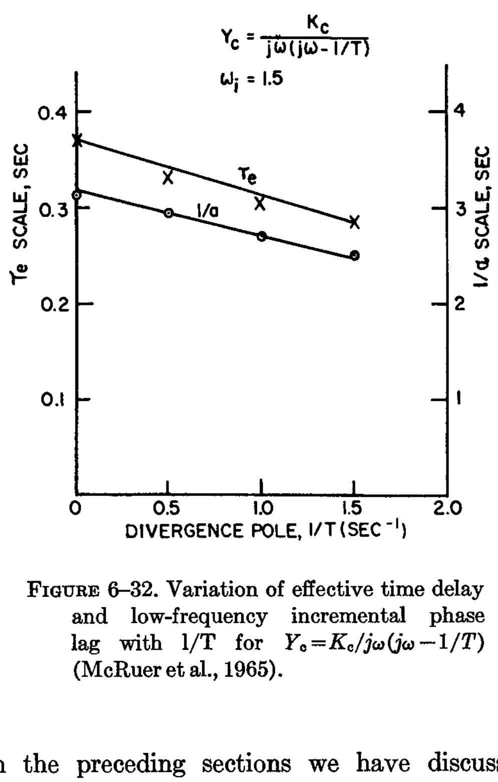


FIGURE 6–32. Variation of effective time delay and low-frequency incremental phase lag with 1/T for $Y_c = K_c/j\omega(j\omega - 1/T)$ (McRuer et al., 1965).

In the preceding sections we have discussed the adaptive controller (Man) but we have ignored the vocal part of the problem. Frequently, operators become quite vocal about the systems they are asked to control, particularly when the task is difficult. This ability of the operator to form and report an opinion can be very useful to the designer. Usually, operator opinion is more sensitive to system changes than is objective performance, although it is not always as consistent. Fortunately, operators can be trained to give consistent ratings, and these ratings do correlate with objective measures of both system variables and operator variables. The technique of opinion rating has been used extensively in aircraft handling, but is also a useful technique for other manual control systems.

Systems having widely different dynamics may feel equally "good" or "bad" to the operator. This goodness or badness is apparently a measure of how easy or difficult the system is to control, is definitely related to the mission to be performed, and is responsive to a number of different system parameters. Pilot opinion ratings of aircraft flying qualities have been used for design guidance since the early 1930's. In 1957 George Cooper, of the Ames Research Laboratory, pointed out the need for a standardized system of ratings and emphasized that such ratings were highly mission related. The original Cooper Rating Scale is shown in Table 6–4. Unfortunately, many of Cooper's cautions about using it, or any such rating scale, have been ignored. Nonetheless, it is the most widely used pilot opinion scale. Figure 6–33, adapted from Chalk (1958), shows the general shape of the "iso-opinion" plots for fighter type aircraft. This plot shows lines of equally good opinion or equally good handling qualities plotted on a scale of aircraft short period natural frequency, and short period damping ratio. The data points show the "best tested" configurations. The comments are summaries of prevalent pilot remarks in various regions of natural frequency and damping. Similar plots can be constructed for various other system parameters.

Since the original Cooper Scale was developed, a number of modifications and revisions have been published by various organizations to overcome certain problems with the original scale. Two of these, the Cornell Aeronautical Laboratory Scale (Harper and Cooper, 1966) and the Systems Technology Scale (McDonnell, 1968) are shown for comparison in Tables 6–5 and 6–6. The intent of these revisions has been to reduce the variability in pilot rating by making the mission dependence more explicit and to spread the ratings more uniformly on a psychological scale of acceptability.

Because variations in many different system parameters can result in equivalent ratings, it is often more useful to attempt to predict operator

TABLE 6–4. THE ORIGINAL COOPER SCALE

COOPER					PR
DESCRIPTION	ADJECTIVE RATING	MISSION	PRIMARY MISSION ACCOMPLISHED?	CAN BE LANDED?	
Excellent, includes optimum	Satisfactory	Normal operation	Yes	Yes	1
Good, pleasant to fly			Yes	Yes	2
Satisfactory, but with some mildly unpleasant characteristics.			Yes	Yes	3
Acceptable, but with unpleasant characteristics	Unsatisfactory	Emergency operation	Yes	Yes	4
Unacceptable for normal operation			Doubtful	Yes	5
Acceptable for emergency operation (stab. aug. failure) only.			Doubtful	Yes	6
Unacceptable even for emergency condition (stab. aug. failure)	Unacceptable	No operation	No	Doubtful	7
Unacceptable-dangerous			No	No	8
Unacceptable-uncontrollable			No	No	9
$Q=*! Did not get back to report	Unprintable	What mission?			10

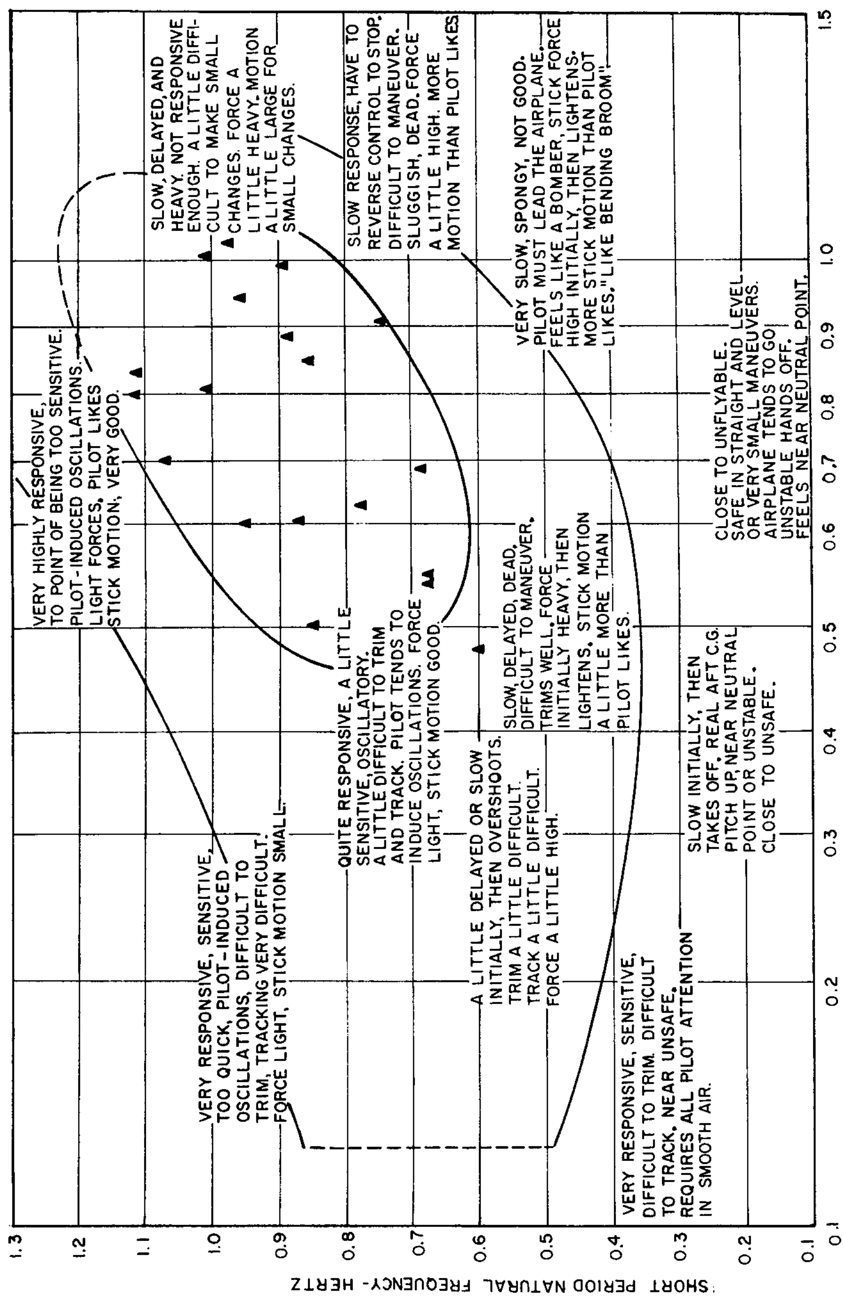


FIGURE 6–33. Pilot opinion contours and typical comments (Webb, 1964; after Chalk).

TABLE 6–5. THE CORNELL AERONAUTICAL LABORATORY SCALE

<table>
<tr><th colspan="2">MISSION SUITABILITY (CAL'S "CATEGORY")</th><th rowspan="2">PILOT ATTENTION OR EFFORT REQUIRED</th><th rowspan="2">ADJECTIVE DESCRIPTION WITHIN CATEGORY</th></tr>
<tr><th>FLYING QUALITIES</th><th>AIRCRAFT ACCEPTABILITY</th></tr>
<tr><td rowspan="3">SATISFACTORY
Criterion: Mission performance is not seriously affected by any flying quality deficiencies which may be present.
Definition: "Seriously affected" = pilot would ask that the deficient characteristics be improved.</td><td rowspan="3">ACCEPTABLE</td><td rowspan="3"></td><td>Excellent</td></tr>
<tr><td>Good</td></tr>
<tr><td>Fair</td></tr>
<tr><td rowspan="3">UNSATISFACTORY
Criterion: Mission performance is sufficiently affected by flying quality deficiencies that pilot asks that characteristics be fixed.</td><td rowspan="3">"RELUCTANTLY" ACCEPTABLE
Criterion: Mission performance deficiencies cannot be improved without a serious compromise of the other factors which influence the mission capability of the airplane.</td><td rowspan="3"></td><td>Fair</td></tr>
<tr><td>Poor</td></tr>
<tr><td>Bad</td></tr>
<tr><td rowspan="3"></td><td rowspan="3">UNACCEPTABLE</td><td>Requires major portion of pilot's attention.</td><td>Bad</td></tr>
<tr><td>Controllable only with a minimum of cockpit duties.</td><td>Very bad</td></tr>
<tr><td>Aircraft just controllable with complete attention.</td><td>Dangerous</td></tr>
<tr><td></td><td>UNFLYABLE</td><td>Control will be lost sometime during mission.</td><td>Unflyable</td></tr>
</table>

TABLE 6-6. THE SYSTEMS TECHNOLOGY SCALE

CONTROLLABLE Capable of being controlled or managed in context of mission, with available pilot attention.	ACCEPTABLE May have deficiencies which warrant improvement, but adequate for mission. Pilot compensation, if required to achieve acceptable performance, is feasible.	SATISFACTORY Meets all requirements and expectations, good enough without improvement Clearly adequate for mission.	Excellent, highly desirable.	A1
			Good, pleasant, well behaved.	A2
			Fair. Some mildly unpleasant characteristics. Good enough for mission without improvement.	A3
		UNSATISFACTORY Reluctantly acceptable. Deficiencies which warrant improvement. Performance adequate for mission with feasible pilot compensation.	Some minor but annoying deficiencies. Improvement is requested. Effect on performance is easily compensated for by pilot.	A4
			Moderately objectionable deficiencies. Improvement is needed. Reasonable performance requires considerable pilot compensation.	A5
			Very objectionable deficiencies. Major improvements are needed. Requires best available pilot compensation to achieve acceptable performance.	A6
	UNACCEPTABLE Deficiencies which require mandatory improvement. Inadequate performance for mission even with maximum feasible pilot compensation.		Major deficiencies which require mandatory improvement for acceptance. Controllable. Performance inadequate for mission, or pilot compensation required for minimum acceptable performance in mission is too high.	U7
			Controllable with difficulty. Requires substantial pilot skill and attention to retain and continue mission.	U8
			Marginally controllable in mission. Requires maximum available pilot skill and attention to retain control.	U9
Uncontrollable. Control will be lost during some portion of mission.			Uncontrollable in mission.	10

McDonnell (1968).

ratings on a limited set of operator variables. Birmingham and Taylor, (1954) postulated that optimum system performance would be obtained when the operator in the system was not required to act in any more complex manner than a simple amplifier. Ashkenas and McRuer (1960) showed that opinion, as measured on the Cooper Scale, varied as a function of both operator gain, K_p, and operator lead time constant, T_L. This variation is shown in Figures 6–34a and 6–34b. It should be noted that the *level* of pilot gain shown as best in Figure 6–34a is *not* a universal optimum, but depends upon the device being controlled. Although the curve is specific for longitudinal control of an aircraft with a column and wheel, the *form* of the curve is typical. In Figure 6–34b the ordinate T_L (dA/d_u) is pilot lead time constant multiplied by the slope of the amplitude ratio.

Ashkenas and McRuer (1962) listed the following conditions necessary for "Good" ratings:

(1) Pilot equalization essentially nil.
(2) Pilot gain adjusted to near-optimum values.
(3) $\omega_i < \omega_c$ and, in addition, for closed-loop systems requiring a third order approximation, $\zeta_{CL} > 0.35$.
Lag Indroduction—slight degradation in opinion.
Lead Introduction—degradation in ratings which increases to maximum values as required lead increases.
Nonoptimum Gain—rating degradation which increases to maximum values as gain varies on either side of optimum.

There is some evidence that the more tightly the system constrains the operator's adaptation, the poorer the opinion rating will be, but this effect has not been quantified.

These requirements for good pilot ratings will be used in the following section, in conjunction with the operator adjustment rules, to show how manual control loops can be designed and modified for good performance and good operator opinion.

6.3.4 Modifying the Controlled Element

When a control system is designed to perform a specific function there is always, explicitly or implicitly, a system performance criterion to be met. Frequently, for a variety of reasons, the "first cut" design will not meet the performance criterion, and the system must be modified. If the system is a manual control system, the modification may take the form of additional operator training. Unfortunately, no amount of operator training will make up for poor system design. This was rather graphically demonstrated by Birmingham et al (1954) when they compared the tracking performance of an operator controlling one or two systems as originally designed with the performance of operator controlling four systems simultaneously after the system had been modified by "quickening" the display. The results are presented in Figure 6–35.

There are numerous ways of modifying a control system to improve system performance. They range from simple changes in control or display sensitivity, through design of compensating networks, to changes in the order and type of system. A few of the more common ways of modifying manual control systems and their relationship to standard control system compensation techniques are presented in the following sections.

Gain Adjustment

Gain adjustment is usually one of the easiest ways to achieve system stability. Most systems are stable if the gain is sufficiently low, although they may not track very well. Conversely, most systems can be driven to instability if the gain is sufficiently high. Conditionally stable systems, which are stable for only a narrow range of gain adjustments, are beyond the scope of this chapter.

In a manual control system, the loop gain is equal to the product of the display, operator, controller, and controlled element gains. Changes in any of these have similar effects. As noted previously, the operator adjusts his gain over a fairly wide range (100:1) to compensate for nonoptimum system gain. This compensation by the operator requires additional effort on his part and results in degraded opinion ratings of the system. It also results in fatigue and frequently results in poorer system performance. Thus, although the operator can adjust the loop gain, it is desirable to adjust the gain in other parts of the system so that the operator functions at or near his preferred level.

Display gain is bounded on its upper limit by the necessity to keep the displayed quantity on the display device without saturation. It is

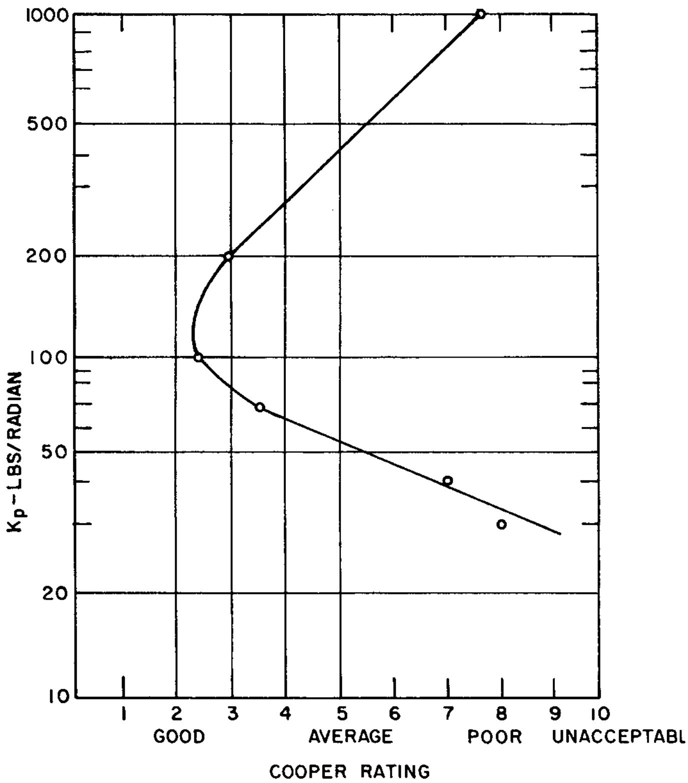


FIGURE 6–34(a). Y_p parameters vs. pilot opinion (Webb, 1964 after Ashkenas and McRuer, 1960).

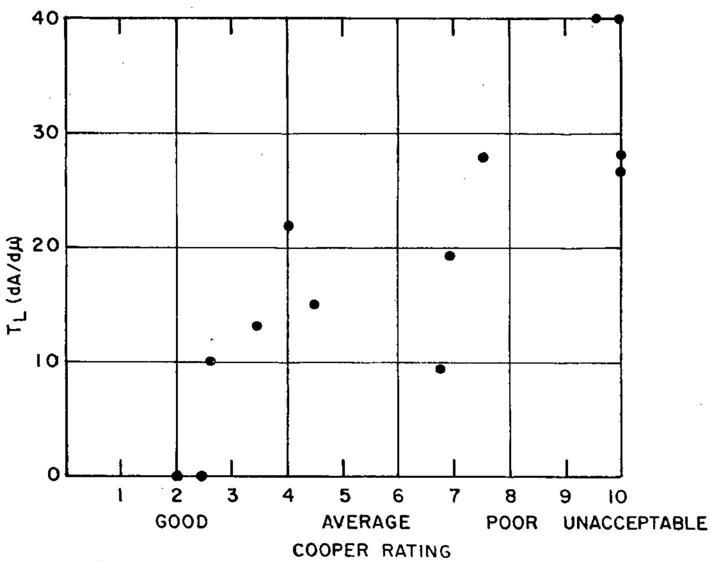


FIGURE 6–34(b). Y_p parameters vs. pilot opinion (Webb, 1964; after Ashkenas and McRuer, 1960).

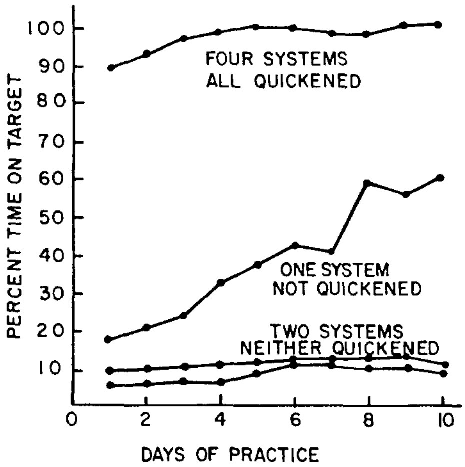


FIGURE 6–35. Effect of display quickening (Webb, 1964; after Birmingham, Kahn and Taylor, 1954).

bounded on its lower limit by the necessity for the operator to see the smallest significant error. For tracking random or random appearing inputs, it is usually adequate to scale the display so that full scale equals $\pm 5\sigma$ of the expected input. A deviation of a factor of two either way from this recommendation will generally not seriously affect system performance.

Control gain is bounded on its upper limit by the inability of the opeator to make very precise changes in force or position. It is bounded on its lower limit by the maximum displacement or maximum force the operator can achieve. The control must have sufficient power or authority to respond to the extreme expected command within the system criterion time. Care must be taken to insure that the gain is high enough to prevent limiting due to the controller hitting its stops and/or rate-limiting downstream of the operator in the control loop.

Within the practical limits outlined above, it is the loop gain or display/control ratio rather than the individual control or display gain that is important. For isotonic controls, if the gain of the system is expressed in terms of the angular movement of the display (see Chapter 3 for discussion of visual angle) as seen by the operator at his normal viewing distance divided by the angular movement of the controlling limb, as suggested by Gibbs (1962), the data of numerous investigators will agree quite well on the optimum gain. For a pure position control with no lag, the optimum gain is about 0.15. If there is some lag in the system, the optimum gain is somewhat higher. For a lag time constant of 0.5 sec., the optimum gain is about 0.4. If the lag time constant is of the order of 2.0 sec., the optimum gain is about 1.0. This gain is somewhat higher than that determined by Gibbs but lower than found by interpolation from Rockway (1954). In any event, both sets of data are very nearly flat in this region, indicating little change in performance for relatively large changes in gain. Systems having a lag time constant of greater than 2 to 3 sec. appear to the operator to act like rate control systems.

Gibbs found the optimum gain for rate control to lie in the vicinity of 1.5 sec.$^{-1}$. This value was optimum for a pure rate control and for rate control cascaded with various first-order lags up to the maximum lag tested, 2.0 sec. This is a rather broad optimum, and a 2:1 change in either direction has very little effect.

There are fewer data available on optimum gains for force or isometric control sticks. The data of North and Lomnicki, (1961) and of Gibbs and Baker (1952) indicate that for continuous tracking with a rate control system, the optimum gain lies in the vicinity of 3°/sec per pound of applied force. This gain refers to tracking with a large (approximately 20-in.) controller and obviously may be too low for tracking with a pencil force stick. Burke and Gibbs (1965) found the optimum gain for a position control force stick to be between 1.5° and 3° per pound. These data were obtained for pursuit tracking with a small (8.5-cm. long) force stick.

Feed-Forward Compensation—Aiding

Aiding, previously discussed under "Controlled Element Dynamics" in Section 6.2.4, is one form of forward loop compensation. Its first use, in the 1940's, was to aid anti-aircraft gunners tracking incoming aircraft. Unaided azimuth and elevation controls were hand cranks—position controls. To track at a steady rate, the operator had to crank continuously at a steady rate. The aided system introduced an electric motor driving through a differential so that the gun position was the sum of the motor shaft position and the crank position. The motor speed was also controlled by the crank position. This is shown schematically in Figure 6–36.

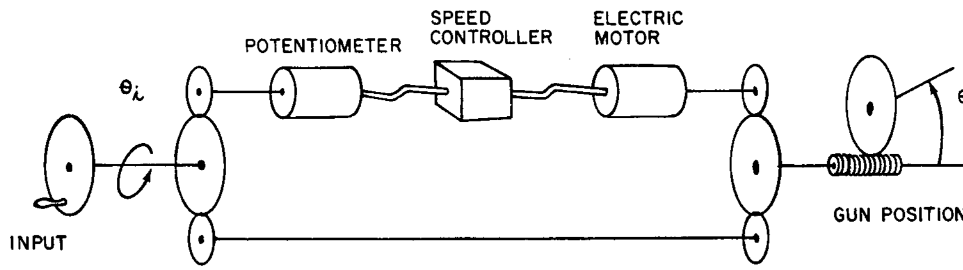
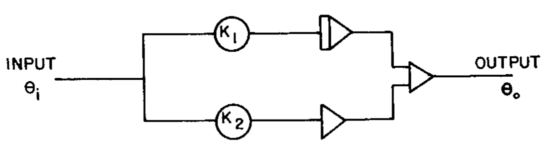


FIGURE 6–36. Aided anti-aircraft control.

The electric motor in the system acts as an integrator. The mechanical differential acts as a summing unit to algebraically add the two shaft positions. The gear ratios set the gain of the system. Figure 6–37 shows the equivalent system and the transfer function derivation. As can be seen by sketching the Bode plot of the system, the aiding results in very high gain at low frequencies while maintaining reasonable high-frequency response for initial acquisition of the target. The lead introduced helps to compensate for the operator's reaction time delay. The gain and lead time constant can be adjusted by changing the gear ratios to achieve the desired system response.



$$\theta_o = \frac{K_1 \theta_i}{s} + K_2 \theta_i$$

$$\theta_o = \left(\frac{K_1}{s} + K_2\right) \theta_i$$

$$\frac{\theta_o}{\theta_i} = \frac{K_1 + K_2 s}{s} = K_1 \left[\frac{\frac{K_2}{K_1} s + 1}{s}\right]$$

FIGURE 6–37. Aided tracking system and transfer function derivation.

The term "aiding" has come to be used to refer to any forward loop compensation. Thus, a rate control system to which a position feed-forward has been added is also referred to as an aided system. The objective of aiding, or forward loop compensation, is to modify the closed-loop behavior of the system. The proper aiding ratios to achieve the desired man–machine system response can be determined analytically using standard control system design techniques in conjunction with the describing function model of the operator.

Figure 6–38 shows the Bode plots for two sets of assumptions: (a) an operator reaction time of 0.25 sec. and an aiding ratio of 0.2 sec. and (b) an operator reaction time of 0.35 sec. and an aiding ratio of 0.5 sec. The similarity of the systems is obvious. Experimental data by a number of investigators show a broad optimum for aiding ratios between 0.2 and 0.8 sec. over a wide range of experimental conditions.

Raising the Order of the System—Unburdening

The term "unburdening," in the control system sense, refers to raising the order of the system. The idea is to remove the "burden" of performing as an integrator to meet the requirements of system performance. Previously, Table 6–1 gave the *type* of system to meet various steady-state error requirements. In the example from the preceding section, the addition of "aiding" also "unburdened" the operator. The addition of the electric motor, which served as an integrator, raised the order of the system. This relieved the operator of the "burden" of performing one integration. For other systems further unburdening might be desirable, but when increasing the order of the system, care must be taken not to require additional compensation by the operator.

The term unburdening has also been used in reference to changes in the control system which remove the requirement for the operator to act as a differentiator. While this does reduce the operator's task load, or burden, it is perhaps

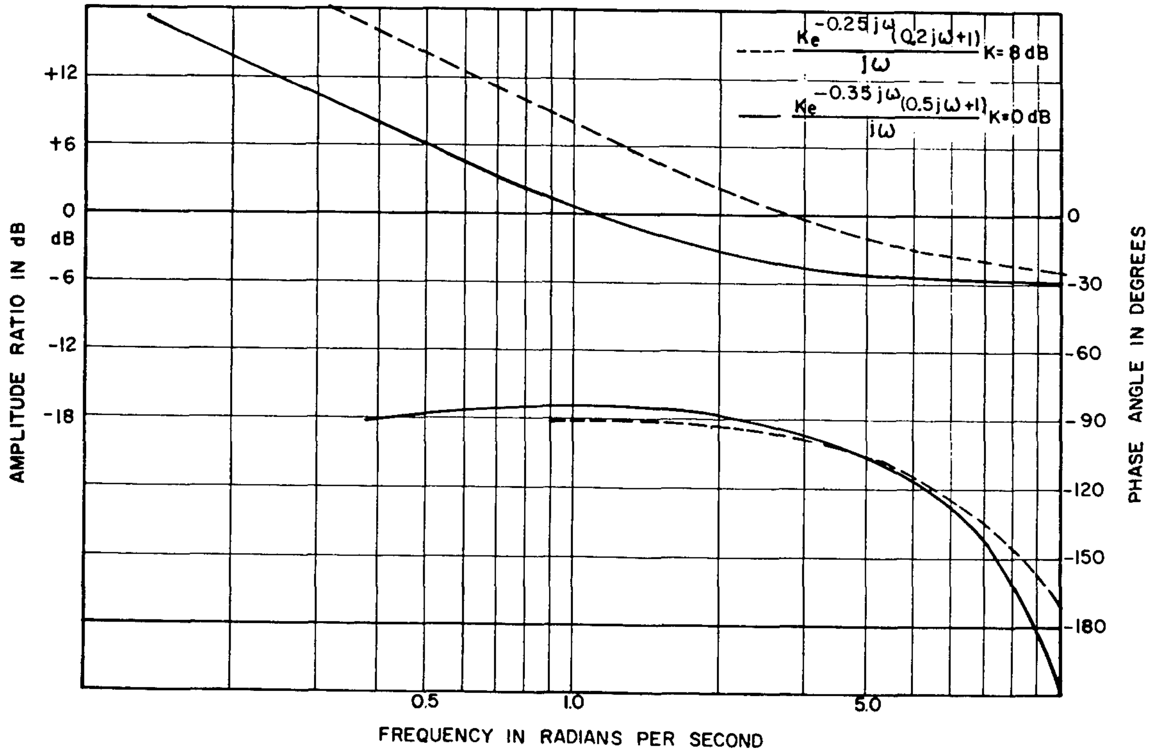


FIGURE 6–38. Bode plots of aided man–machine systems.

more properly referred to as aiding if it is done by adding feed-forward loops. If feedback loops are closed around various portions of the loop it is generally referred to simply as stability augmentation. Figure 6–39 shows the various configurations along with their transfer functions and the amplitude vs. frequency response sketches.

Quickening

In some systems it is not possible, for physical or financial reasons, to modify the hardware portion of the system. In such systems it may be possible to compensate the system by modifying the signal fed back to the operator's display. This is known as "quickening" the operator's display. Quickening is usually applied to higher order slow-responding systems in the form of derivative signals summed with the system output signal back to the display. This is shown in Figure 6–40.

From the figure it is apparent that the system appears as an aided system *to the operator.* Note, however, that the actual *system output has not been modified,* i.e., although the display responds more quickly to the operator's inputs, the sluggishness of the actual system is unaffected. For this reason it is almost always necessary to provide the operator with an indication of the actual system output in addition to the quickened system output display. The optimum quickening coefficients for a given system can be derived by relatively straightforward control system synthesis techniques so long as it is remembered that the operator describing function also changes. Verdi et al. (1965) show that as quickening is increased, the operator's transfer function changes from a lead form to a pure gain, and then to a lag form. This is consonant with the operator adjustment rules given earlier, since, as the quickening coefficients are increased, the *apparent* systems dynamics change from acceleration to rate to position dynamics.

Quickening can be an extremely useful technique when it is necessary for an operator to stabilize an unstable or marginally stable system. It can, however, result in poor response to time-varying command signals unless the

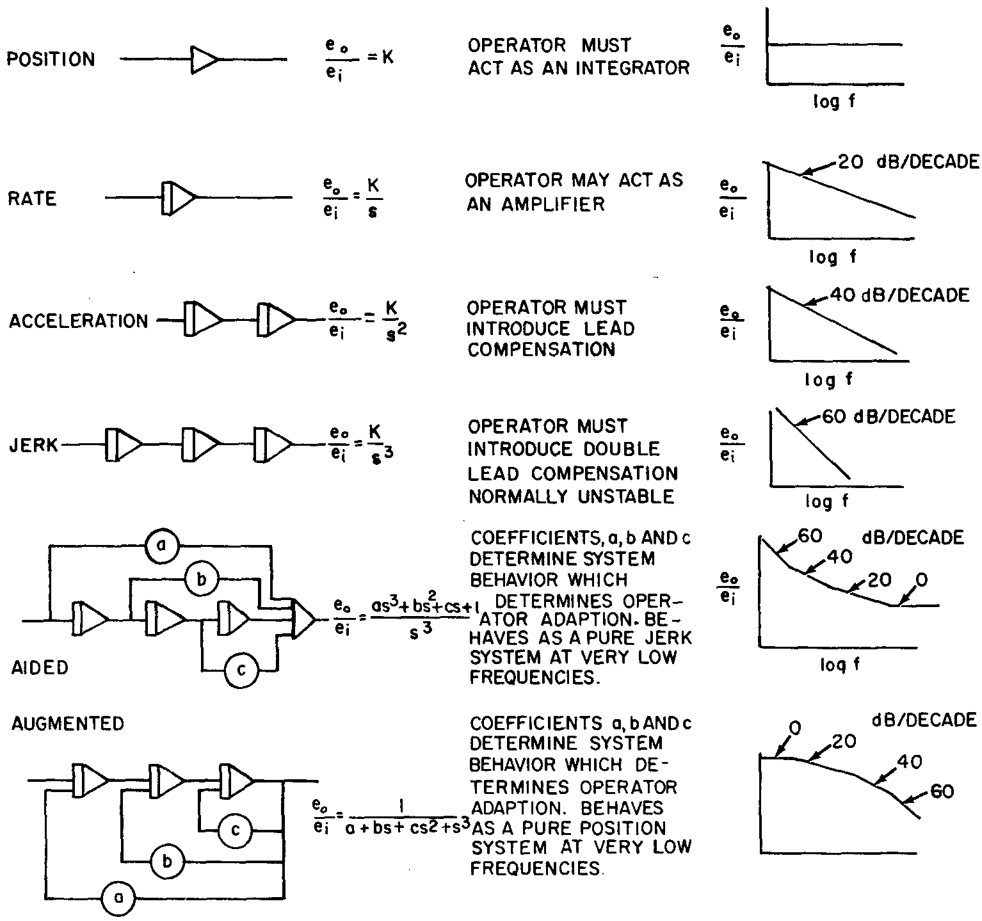


FIGURE 6–39. Unburdening, aiding, and augmenting of control systems.

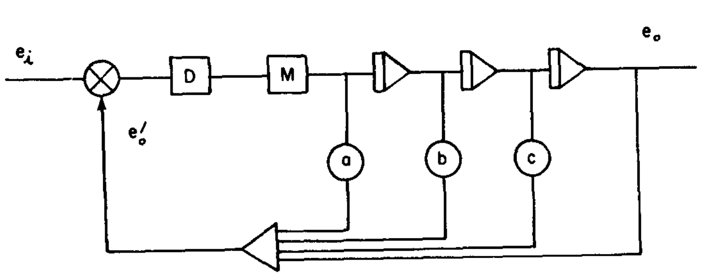


FIGURE 6–40. Quickened display system.

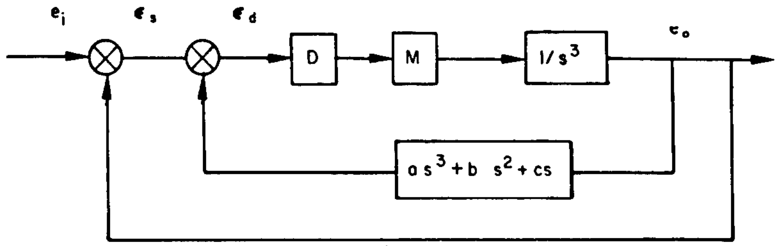


FIGURE 6-41. Quickened system showing both system and displayed errors.

command signal is passed through an anti-bias network. The need for such a network is shown in Figure 6-41. This system is identical to that shown in Figure 6-40 except that the system error $(e_o - e_i)$ is shown explicitly. Obviously, unless e_o is constant, $\epsilon_s \neq \epsilon_d$. A prefilter of the form

$$[(a+1)s^3+bs^2+cs+1]\ [s^3+1]$$

ahead of the first summing point would be required to make the system respond to all inputs (e_i) as it would without any quickening.

6.3.5 Performance Prediction

One of the major objectives in developing the describing function model of the operator, together with a set of adjustment rules, has been to allow quantitative predictions of system performance. To date this objective has been reasonably well met for single-axis compensatory tracking systems with random or pseudorandom inputs. Consider the general closed-loop system shown in Figure 6-42.

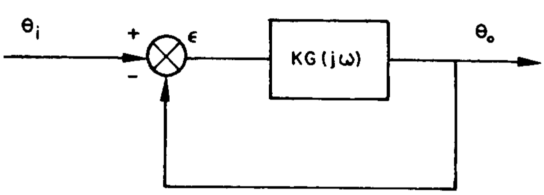


FIGURE 6-42. General closed-loop system.

The mean squared error $(\bar{e}_i)^2$ due to the input in such a system is given by

$$(\bar{e}_i)^2 = \frac{1}{2\pi}\int_0^\infty \Phi_{ee}(\omega)\, d\omega, \qquad (6\text{-}35)$$

where $\Phi_{xx}(\omega)$ denotes the power spectral density of $x(t)$. It can be shown that

$$\Phi_{ee} = \left|\frac{E(j\omega)}{I(j\omega)}\right|^2 \Phi_{ii}(\omega), \qquad (6\text{-}36)$$

but

$$\frac{E(j\omega)}{I(j\omega)} = \frac{1}{1+KG(\omega)}. \qquad (6\text{-}37)$$

Therefore,

$$\bar{e}_i^{\,2} = \frac{1}{2\pi}\int_0^{\omega_i} \Phi_{ii}(\omega)\left|\frac{1}{1+KG(\omega)}\right|^2 d\omega. \qquad (6\text{-}38)$$

If $\Phi_{ii}(\omega)$ has a rectangular power spectral density with a corner frequency of ω_i—see also Section 6.2.1, "Types of Input Functions"—this basic relationship can be used to predict the error due to the input of a man-machine system. It does not account for that portion of the system error due to operator-induced noise or remnant but for many systems this additional power is relatively low. If the operator-induced noise is considered as noise added at the operator's output, denoted as $\Phi_{nn}(\omega)$, then $(e_n)^2$, the error due to operator injected noise, is:

$$(\bar{e}_n)^2 = \frac{1}{2\pi}\int_0^\infty \Phi_{nn}(\omega)\left|\frac{Y_c(\omega)}{1+Y_pY_c(\omega)}\right|^2 d\omega. \qquad (6\text{-}39)$$

The "1/3" Law

If the "crossover model" for the man-machine system is used to estimate system performance, a very simple relationship can be developed. From Section 6.3.3, "Adaptation Rules", the crossover model is:

$$Y_p \cdot Y_c = \frac{\omega_c e^{-\tau_e j\omega}}{j\omega}. \qquad (6\text{-}40)$$

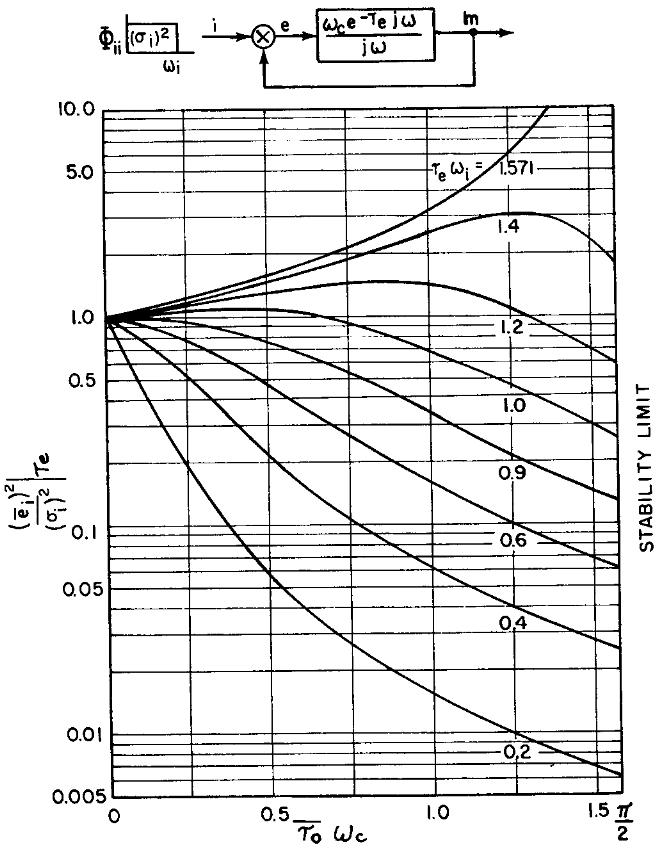


FIGURE 6–43. Mean-squared error based on crossover model (McRuer et al., 1965).

Substituting this for $KG(\omega)$,

$$(\bar{e}_i)^2=\frac{1}{2\pi}\int_0^{\omega_i}\Phi_{ii}(\omega)\left|\frac{j\omega}{j\omega+\omega_c e^{-\tau_e j\omega}}\right|^2 d\omega \quad (6\text{–}41)$$

or

$$(\bar{e}_i)^2=\frac{1}{2\pi}\int_0^{\omega_i}\Phi_{ii}\frac{\omega^2}{\omega^2-\omega\omega_c\sin\omega\tau_e+\omega^2}\,d\omega. \quad (6\text{–}42)$$

This equation has been integrated numerically and the result is given in Figure 6–43 (McRuer et al. 1965). If sin $\omega\tau_e$ is replaced by its argument (a good approximation for systems where $\omega_i\tau_e \ll 1$), the integral is readily evaluated and the resultant reduces to

$$\frac{(\bar{e}_i)^2}{\sigma_i^2}=\tfrac{1}{3}\left(\frac{\omega_i}{\omega_c}\right)^2. \quad (6\text{–}43)$$

This extremely simple result provides good predictions of system performance so long as:

1. The crossover model is appropriate,
2. The input bandwidth does not produce ω_c regression, and
3. The subjects are well trained.

Figure 6–44 shows predictions of system performance based on the ⅓ Law vs. actual system performance in experiments reported by Kelley (1967) and McRuer et al (1965). The Kelley data represent one extremely well trained subject working at his maximum capacity. The data were taken on an adaptive simulator which varied the system forcing function to keep the subject operating at his limit. Thus, they probably indicated the best possible tracking. The McRuer data represent well-trained subjects in a more conventional tracking system and are thus probably much more typical of the results to be expected in most systems. The results are predicted quite accurately. Two values of predicted performance are shown for each of Kel-

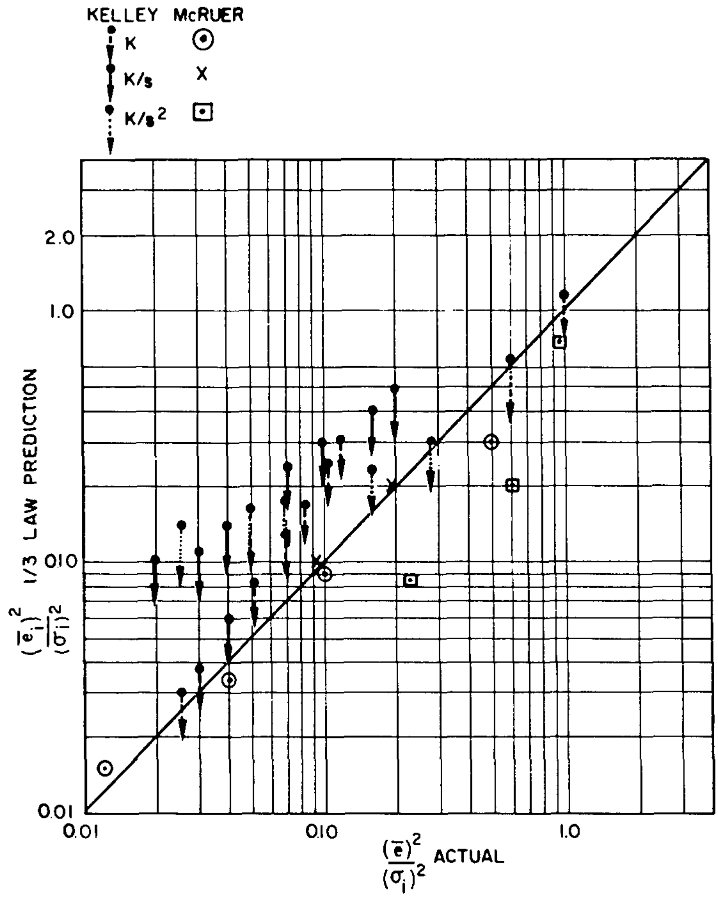


FIGURE 6–44. One-third law performance prediction vs. actual system performance (adapted from Kelley, 1967, and McRuer, et al., 1965).

ley's data points. The dot represents the prediction using the expected values of ω_c taken from Figure 6–30. The arrowheads show the predictions based on ω_c values approximately one radian higher than those of Figure 6–30 since Kelley's subject was extremely well trained and was using a force stick controller rather than an isotonic controller. For many purposes, results based on the ⅓ Law are quite adequate. More refined estimates are, however, possible.

Refined Estimates

The ⅓ Law was based upon the crossover model of the man–machine system. More refined estimates of system performance can be obtained by substituting more refined models of the man and the machine for $KG(\omega)$ in the equations developed in Section 6.3.5, "Performance Predicition." These models will not, in general, yield simple analytic solutions. They are, however, capable of being solved by numerical integration on digital computers. Alternatively, the system, including the operator model, can be simulated on an analog computer. If either case, fairly accurate representations of the operator must be included. As an example, consider the operator controlling a pure acceleration system. The first approximation considered

$$Y_p \cdot Y_c = \frac{\omega_c e^{-j\omega\tau_e}}{j\omega}. \qquad (6\text{–}44)$$

A much better fit to the actual data and a better performance prediction is obtained using

$$Y_p Y_c = \frac{(\omega_{c}/T_L)\ (T_L j\omega + 1) e^{-j(\omega\tau_e + \alpha/\omega)}}{(j\omega)^2} \qquad (6\text{–}45)$$

with $T_L \geq 5$ sec. Appropriate values for α can be taken from Figure 6–31.

Worked Example

In Section 6.2.3, "Loaded Moving Controls (Real Controls)," some of the findings of Notterman and Page (1962) are discussed. In that report, they give their results in terms of mean $\int \epsilon^2\, dt$ rather than in terms of the ratio of error to input. However, in section II A3 of the reference, they state, "the input was a random Gaussian noise with a 'double corner' at ⅓ Hz . . . . The gain of the error display was such that the untracked error never exceeded ± two inches." Their "double corner" is a filter of the form $1/(\tau s+1)^2$, which, when inserted in Elkind's equation for equivalent rectangular bandwidth (See Section 6.2.1, Aperiodic Inputs) yields a corner frequency of $2\pi/5\tau$ where τ is the time constant of the original filter. Thus, the equivalent bandwidth, in radians per second, is $\frac{1}{3}\times\frac{2\pi}{5}\times 2\pi$ radians/sec., or 2.63 radians/sec.

From Appendix A.1 of the cited report, it is found that the input amplitude distribution was Gaussian, but truncated at $\pm 2\sigma$ from the mean. Thus, $\pm 2\sigma = \pm 2$ in. at the display or $\sigma = 1$ in. Therefore, $(\bar{\sigma}_i)^2 = 1$ in. From Appendix A.5

$$S_i = K \int_0^T \epsilon^2\, dt$$

where $K = 8.05$ and the scoring interval $= 27.5$ sec. From this we see that $(\bar{\sigma}_i)^2 = 1 \times 27.5 \times 8.05 = 222$ in terms of the reported scores.

The pure isometric, pure isotonic, and the mechanically restrained control sticks can, to a first approximation, be considered essentially position tracking systems. From Figure 6–30 we see that ω_c for a position tracking system is about 5.5 radians/sec. Substituting this and the previously derived numbers into the "⅓ Law" formula

$$\frac{(\bar{e}_i)^2}{(\bar{\sigma}_i)^2} = \frac{1}{3}\left(\frac{\omega_i}{\omega_c}\right)^2 \qquad (6\text{–}46)$$

$$\frac{(\bar{e}_i)^2}{222} = \frac{1}{3}\left(\frac{2.63}{5.5}\right)^2 \qquad (6\text{–}47)$$

$$(\bar{e}_i)^2 = \frac{222}{3}\left(\frac{2.63}{5.5}\right)^2 = 16.9 \qquad (6\text{–}48)$$

It would be predicted that this error magnitude would be essentially constant for a very wide range of natural frequencies and damping ratios for the mechanically restrained stick.

Let us now consider the isotonic stick with the "equivalent" dynamics. For $\zeta \doteq 1$, we will map the variation in $(\bar{e}_i)^2$ as a function of ω_n. If $\omega_n \ll \omega_c$, the system appears to the operator as a K/s^2 system and he adapts his describing function accordingly. The lowest natural frequency tested was $\omega_n \doteq 2$. For this natural frequency the slope of the controlled-element dynamics in the likely region of crossover lies between -6 dB/octave and -12 dB/octave. From Figure 6–30 we would estimate the crossover frequency to be about 3.5 radians/sec.

$$(\bar{e}_i)^2 = \frac{222}{3}\left(\frac{2.63}{3.5}\right)^2 = 41.8 \qquad (6\text{–}49)$$

For $\omega_n = 6$ the system looks more like a rate control in the region of crossover. From Figure 6–30 we can estimate $\omega_c \doteq 4.5$ radians/sec.

$$(\bar{e}_i)^2 = \frac{222}{3}\left(\frac{2.63}{4.5}\right)^2 = 25.3 \qquad (6\text{–}50)$$

At the highest natural frequency tested, $\omega_n = 15$ radians/sec., the system looks almost like a pure position control system. Again, from Figure 6–30, we estimate $\omega_c = 5.25$ radians/sec.

$$(\bar{e}_i)^2 = \frac{222}{3}\left(\frac{2.63}{5.25}\right)^2 = 18.6 \qquad (6\text{–}51)$$

These predictions are plotted against the actual results in Figure 6–45. Point A is the average from the isometric, isotonic, and mechanically restrained sticks. Points B, C, and D are for the 2, 6, and 15 radian/sec natural frequencies. The experimental data agree quite well with the predictions on a qualitative basis but not quantitatively. The explanation for this lies in the fact that naive subjects were used and were given relatively little practice. This is borne out by points E and F on the figure which represent experimental data given in Experimental Note I at the end of the Notterman and Page report. These points show one subject's performance after extensive practice. The report

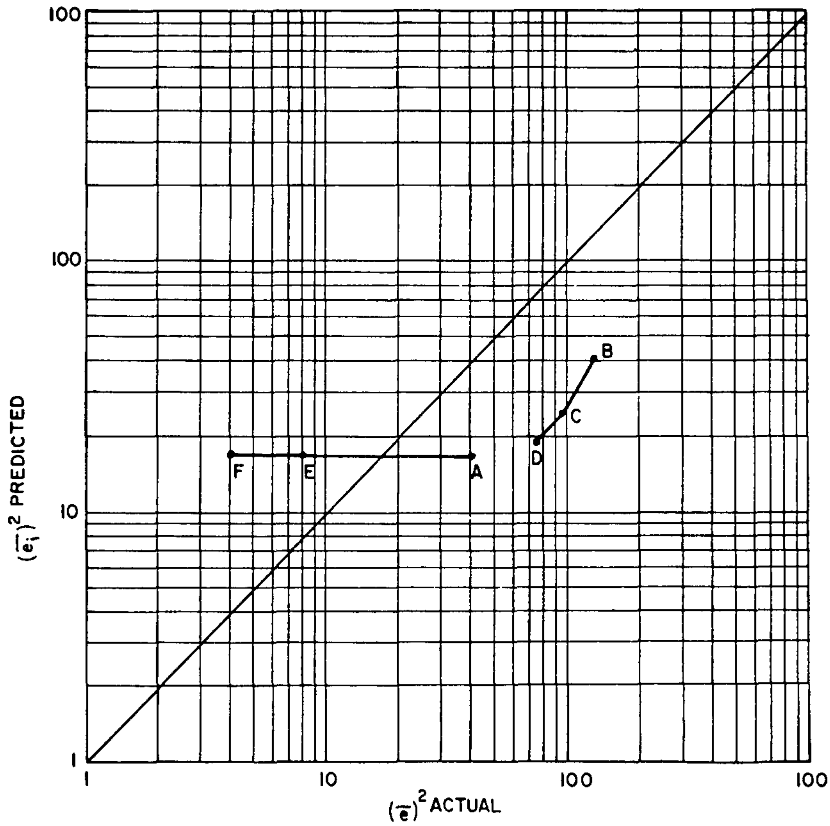


FIGURE 6–45. Predicted vs actual error scores from Notterman and Page experiments (adapted from Notterman and Page, 1962).

does not indicate how much practice is represented by point E, but points E and F are separated by 2400 runs! Thus, it can be seen that predictions based on the "⅓ Law" show the relative effects of various changes made in the control system but they do not take into account the effects of practice. The predictions based on the ⅓ law assume a moderately well-trained operator, not a naive subject and not a robot who is "locked in" on the system after extremely intensive practice.

6.3.6 Effects of Small Nonlinearities

In the previous sections, the assumption has been implicit that the control system was linear. No electromechanical system is entirely linear, but for most design work the assumption of linearity is required, at least initially. If the nonlinearities are sufficiently small, they can be ignored. If they cannot be ignored, the task of predicting their effect on system performance becomes extremely difficult.

The major source of difficulty in analyzing nonlinear systems arises from the fact that the superposition theorem does not hold for these systems. Thus, the effect of a nonlinearity in the control system is determined by its location. The magnitude of the signal input in relation to the magnitude of the nonlinearity also determines the effect of the nonlinearity, as may the frequency content of the signal. In light of these problems, it is not at all surprising that the literature on the effects of nonlinearities on manual control systems is confusing and the results frequently contradictory. The problem is made more acute by the number of different types of nonlinearities and their numerous locations in practical control systems. Figure 6–46 shows some of the common nonlinearities and their locations in a simplified aircraft longitudinal control system (Graham, 1967). As an example,

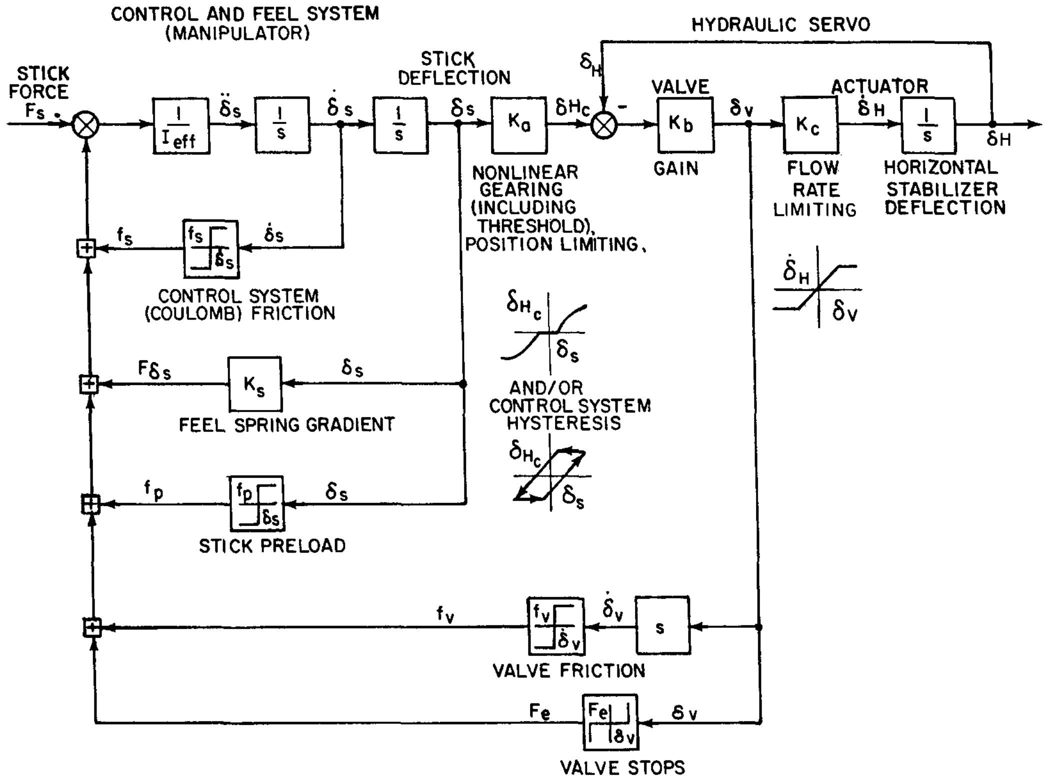


FIGURE 6–46. Simplified aircraft longitudinal control system block diagram showing common non-linearities (Graham, 1967).

friction in the hydraulic servo valve produces a markedly different result, both in terms of system performance and in terms of control stick feel, than does friction in the linkage and bearings between the stick and the valve.

Fortunately, the human operator can effectively linearize many systems having force-displacement ("feel") nonlinearities. This linearization is apparently accomplished by enclosing the nonlinearity within a high-gain position feedback loop. That is, the operator positions the control *ignoring* the force feedback. This activity requires effort which is fatiguing and results in severely degraded opinion ratings of the system. The operator's ability to maintain this effort over prolonged time periods, when fatigued, or under stress, is not well known. Control system nonlinearities have long been suspect in cases of pilot-induced oscillations (PIO), but recent research (Graham, 1967) indicates that force-displacement nonlinearity alone is not enough to cause oscillation even with PIO-prone system dynamics.

In cases where the nonlinearity is not reflected back to the operator as a force-displacement nonlinearity, the operator has virtually no cue as to the nature of the nonlinearity and therefore cannot adopt a suitable compensation. For this reason, experimental studies in which nonlinearities are electronically simulated and do not reflect back to the operator *cannot* be generalized to predict the results of a similar nonlinearity which does reflect back to the operator. In the following sections, the effects of certain common nonlinearities will be discussed.

Control Stick Friction

The principle effect of control stick friction is to degrade operator opinion of the system. Small amounts of friction—about one pound—occasionally result in slight improvement in operator opinion when no restraining springs are used. Reasonable amounts of friction seldom result in any significant degradation in tracking accuracy, but may result in problems where very precise positioning is required. Thus, for example, excessive control stick friction produces difficulty in making small corrections about a trim position. For discrete settings of rotary controls, a small amount of friction is desirable to prevent inadvertent moving of the control and to prevent vibration from altering the setting. A frictional torque of 3 to 6 in.-oz. seems to be satisfactory, with the lower value applicable to knobs down to 0.5 in. in diameter.

Friction in a control system frequently combines with other linear and nonlinear elements of the system to produce results unlike those of either the friction or the other element alone. Figure 6–47 shows the effect of spring restraint and preload on the force-displacement hysteresis loop created by coulomb friction in a control stick. Figure 6–48 shows how friction can combine with other elements to produce displacement-displacement hysteresis. It should be noted that backlash in a gear train, a loose ball end-fitting, or a loose clevis pin fitting produce the same result as the fork and pin arrangement shown in the figure.

The several studies dealing with control stick friction and spring gradient, as reported by Wasickó and Magdaleno (1965) show little or no performance decrement due to stick friction. In those cases where stick friction did produce a slight decrement, increased spring gradient appeared to help. Graham (1967) found no performance decrement due to stick friction up to 10 lb. at the stick-head, although operator opinion was severely degraded. Increased spring gradient did not improve opinion ratings.

Control Valve Friction

Friction in the control valve of a hydraulic power actuator reflects itself back to the pilot by trying to move the control stick. The reason for this can be seen by examining Figure 6–49. With no friction in the system, pulling back on the control stick pushes the valve forward because the valve spool moves more easily than the surface and the actuator piston.

As the spool moves forward, it allows fluid into the actuator, pushing the piston forward and moving the surface up. As the piston moves forward, it moves the bottom end of the vertical link between the actuator and the valve forward. Since the pilot is holding the stick, the top of the vertical link moves back, which moves the spool back and closes the valve. If there is excessive friction in the valve, it attempts to remain stationary and the piston tries to move the stick back. The stick moves back until the pilot and the centering springs (not shown) exert sufficient forward force to close the valve. If the valve overshoots the center position, the cycle repeats and the system oscillates. The addition of friction in the control stick tends to alleviate the problem, but this increases required pilot effort. There is currently no specification on maximum allowable valve friction. The "feel" of the particular system must be determined experimentally. In general, stick friction should equal or slightly exceed valve friction as measured at the stick.

Backlash

Backlash is a displacement-displacement hysteresis caused by looseness in mechanical linkage. A certain minimum of backlash is almost inevitable due to necessary running clearances to prevent binding in the linkages and gear trains. Special anti-backlash gears and fittings, which are spring-loaded to prevent backlash, are available but these are expensive and not always effective when force loading is high. The effect of backlash is always to degrade tracking performance. Small corrective movements that don't exceed the magnitude of the backlash never get transmitted to the rest of the system and reversals of inputs are delayed by the time needed to cross the zero output phase of the backlash. Figure 6–50 shows the relative increase in tracking error as backlash is increased. These data, adapted from Rockway and Franks (1959), are for an electronically simulated hysteresis rather than true mechanical backlash between the control stick and the controlled element. If the backlash boundaries can be felt mechanically, systems performance will be better than if they are not felt. Even so, the operator's opinion of the system will still be degraded. Because this hysteresis did not take into account the mass of the driving and driven elements, the results may be conservative. An excellent discussion of true backlash is given in Graham and McRuer (1961).

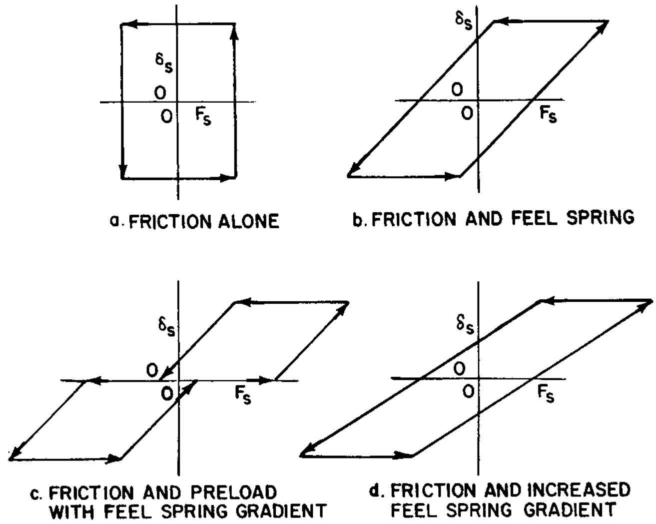


FIGURE 6–47. Functional relationships between force and displacement (approximately sinusoidal input) with common nonlinearities (Graham, 1967).

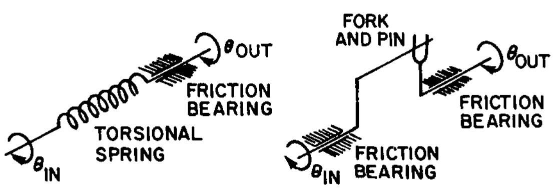


FIGURE 6–48. Two mechanisms giving rise to displacement-displacement hysteresis (Graham, 1967).

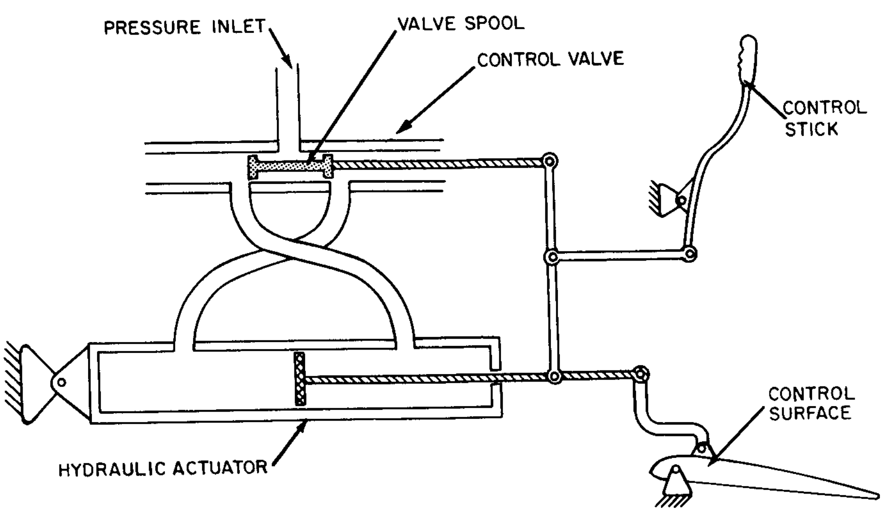


FIGURE 6–49. Simplified hydraulic control system.

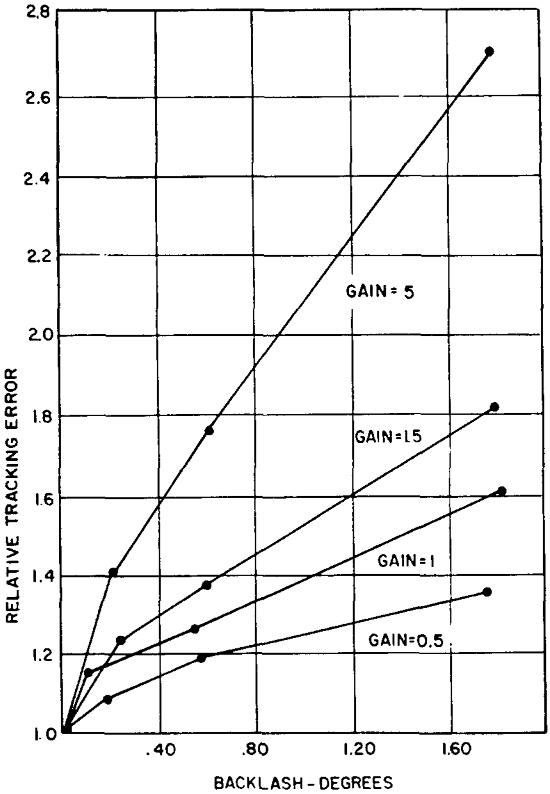


FIGURE 6–50. Effect of backlash on tracking performance (Webb, 1964).

Display Quantization

For certain simple control tasks it is sufficient to tell the operator simply that he is high, low, or on course. In more difficult situations, or where great accuracy is required, the operator requires more complete information about the error signal. Hunt (1959 and 1963) investigated the effect of display quantization on tracking performance for a single-axis acceleration control task. It would be expected that this task should be quite sensitive to display quantization since the quantized display restricts the operator's ability to perceive rate information. The results are shown in Figure 6–51. In Hunt's 1959 report, the "Difficult" forcing function was a summation of three sinusoids of 2, 4, and 10 cycles per minute and the "Easy" forcing function was the summation of three sinusoids of ⅔, ⁴⁄₃, and ⅓ cycles per minute. In the 1963 report the forcing function was the summation of three sinusoids of 1, 2, and 3 cycles per minute. Unfortunately, forcing function amplitude was not reported in either study. For this reason no quantitative predicitions can be based on these results. They are, however, indicative of the trend to be expected with quantized displays in systems requiring operators to sense rate or generate large lead terms.

Intermittency

In some control situations, the operator cannot or does not watch a display continuously as he may in a laboratory tracking situation. This may be by choice or it may be forced by the necessity to scan many instruments. Alternatively, the system may force periodic sampling as in a PPI radar display. In systems which do not force periodic sampling, operators do not sample periodically.

Recent data by Elkind and Levison (1966) indicate that when an operator is controlling two systems with widely separated displays, he continues to control the system he is not looking at by using peripheral vision. The operator's gain may be greatly reduced (10 to 20 dB), but he is still actively tracking. For two rate control systems with independent 2-radian/sec. random inputs, increasing the display separation from 0.8° to 30.0° resulted in about a 2.5:1 increase in mean squared error. Eye movement data showed no sign of periodic sampling. Measured describing functions showed no change in operator adaptation other than a gain decrease.

Earlier work by Senders (1955 and 1956) examined the forced sampling situation. Two adjacent displays were intermittently illuminated either alternately or simultaneously. Various illumination rates and light-time fractions were employed. The results are shown in Figure 6–52, along with the results of Humphrey et al. (1953) for a single-axis tracking task. All three studies were conducted using position control systems and compensatory displays. Unfortunately none of the reports gave details of the forcing function amplitude or the width of the "target" band. Only the Humphrey report gave the frequency content of the input; 2, 6, and 15 cycles per minute sinusoids.

Each of the three sets of curves shows a fairly well defined "knee" above which increasing frequency has very little effect on performance. Because the experimental situation is so

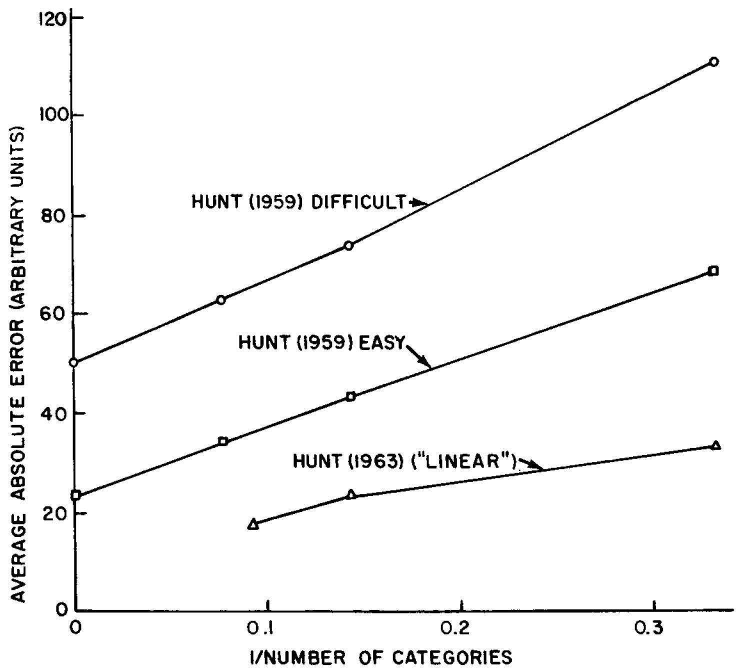


FIGURE 6–51. Effect of display quantization on control of an acceleration control system (Hunt, 1959, 1963).

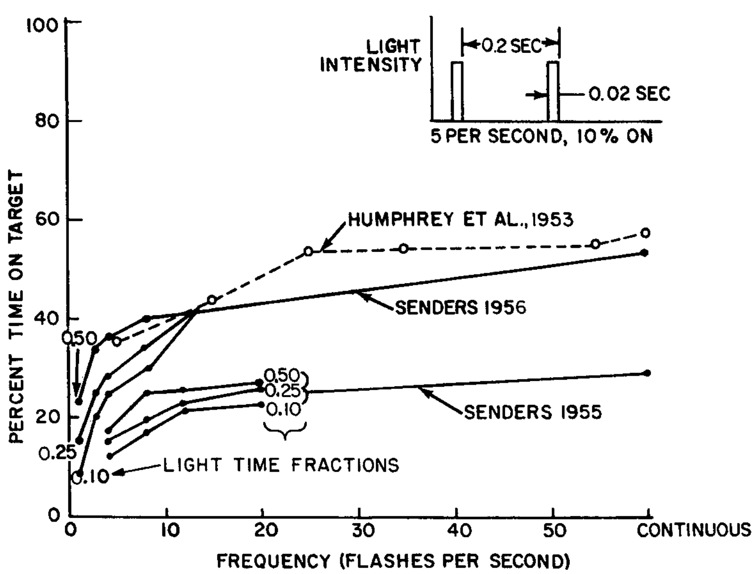


FIGURE 6–52. Effect of display intermittency (Senders, 1955), simultaneous illumination, simultaneous time on target; (Senders, 1956) alternate illumination, simultaneous time on target; (Humphrey et al., 1953) single axis, "on" time constant at 0.008 sec.

poorly defined in the reports, it is impossible to determine how this "knee" is related to the input function.

6.4 Control of Multiple Loops

In the preceding sections of this chapter, it has been assumed that the operator was dealing with only a single control variable. In the real world, this is seldom true. In most reàl situations, the operator is controlling several variables more or less simultaneously. The complexity varies from relatively simple two-axis tracking such as keeping a set of cross-hairs aligned on a radar target, to extremely complex situations such as landing a modern jet transport under adverse weather conditions.

In the simple two-axis tracking situation, the two control loops are usually not coupled. That is, a control movement in the lateral axis has no effect on the longitudinal axis and vice versa. With aircraft, this is not true. A control input which produces a bank angle will cause a loss of altitude unless an increase in pitch attitude is also commanded. An additional complication occurs in the aircraft case because there are many control loops nested one within another and all controlled by the same control variable. Pitch rate, pitch attitude, angle of attack, rate of climb, and altitude are all controlled by fore and aft movement of the control stick or column. In dealing with multi-loop systems of this sort, the operator (pilot) must scan his instruments to determine the appropriate control action. This scanning means that the operator is operating as some form of sampled-data system. The effect of the sampling must be included in any analysis of the manual control loops and in any prediction of system performance.

A critical question is, "What information should be displayed to the operator to optimize system performance?" What displayed parameters will minimize the operator's scanning workload and the destablizing effects of scanning delays? Unfortunately, complete answers to these questions are not yet available, but partial answers have been developed which offer significant assistance to the experienced designer. The remainder of this section is devoted to presenting these partial answers.

6.4.1 Display Sampling

Consider on display embedded within a complex of displays, all of which require control. This display must compete with all the others for the operator's attention. If there are only a few other displays, and all of the displays are quite stable, and well within system tolerance levels, there is no problem for the operator. But what of the situation in which the tolerances are tight and the displays can change rapidly? How frequently must the operator look at each of the displays? How long must he look at each display? How does this sampling affect the dynamic behavior of the system? The following paragraphs discuss how to predict display sampling frequency and eye dwell time for each instrument, and the effects on system performance of the additional delays introduced by this sampling.

Sampling Frequency

The Nyquist sampling theorem states, in effect, that a signal must be sampled at least twice per cycle of the highest frequency component of that signal. Thus, a random signal having frequency components from 0 to 5 Hz would have to be sampled at least 10 times per sec. Senders (1964) showed that the sampling theorem provided a very good estimate of the operator's sampling rate as a function of displayed signal band width for a monitoring task. (See Figure 6-53.) Other, more elaborate, models of scanning behavior involving decision theory and hypothesized queues of information demand are being developed, e.g., Senders et al. (1965). Currently, the improvement in precision yielded by these models does not warrant their use in preliminary system design. The simple approximation that the operator will sample the display at twice the highest frequency of the *closed-loop displayed signal*, i.e., the signal actually displayed when the operator is controlling the system, is adequate for preliminary design. It is necessary, of course, to determine the bandwidth of this signal including the operator's dynamics and remnant. Since these signals do not, in general, have rectangular spectra, it is necessary to calculate the equivalent rectangular bandwidth of the signals. (See Section 6.2.1.)

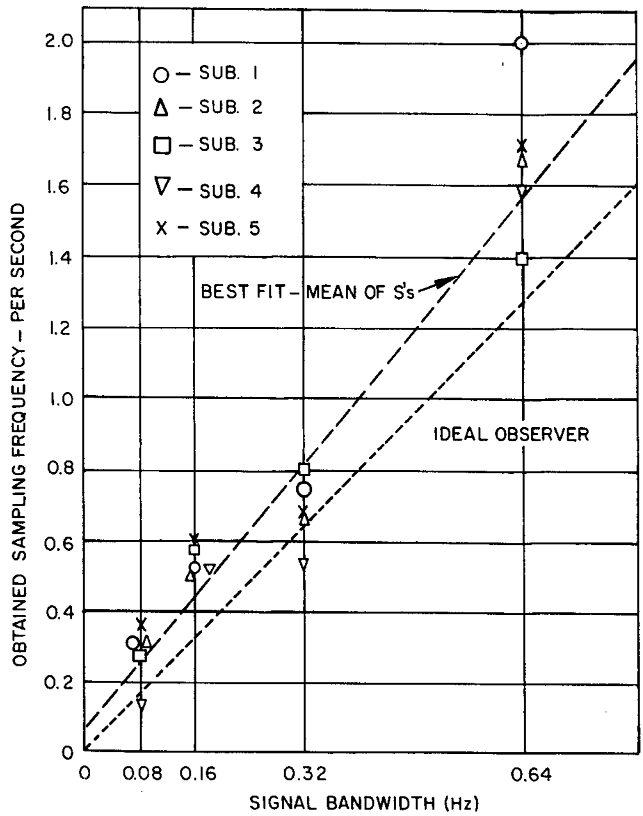


FIGURE 6–53. Regression of obtained sampling frequencies on bandwidth (Senders, 1964).

Eye Fixation Duration

The classic data on visual sampling of displays and on pilot eye movements are, of course, the *Eye Fixations of Aircraft Pilots* series of Air Force reports by Fitts, Jones, and Milton (1950) and subsequent reports by Milton et al. (1949, 1950, 1951, 1952). These data, taken in a C-45 under various flight conditions, are summarized in Table 6-7 and Table 6-8. Unfortunately, no data on the dynamic behavior of the aircraft were recorded. Thus, we know where the pilot was looking, but not what he saw. The fixation and dwell times reported are not in conflict with estimates of what would be expected, given the aircraft dynamics and assuming reasonable pilot loop closures.

Because of the paucity of data in which both describing function and eye movement data have been taken, it is difficult to provide a functional relationship between eye fixation or dwell time and the design of the instrument or the information the operator gains from it. McRuer et al. (1967) state:

> One coarsely quantized mean duration of fixations, 0.5 sec., appears to be a sufficient summary of results for all flight instruments in all maneuvers. The mean value, 0.5 sec., is termed an average 'reconstruction' dwell time, since the pilot must reconstruct his control signal output(s) from samples of flight instrument signals during fixations under instrument conditions.
>
> A second coarsely quantized mean duration of fixations, 1 sec., is termed a 'group-monitor' dwell time, since it represents an average for a twin-engine group of instruments.
>
> Results in landing approaches support more finely quantized mean values of dwell time as follows:

Duration seconds):	*Displayed signal*
0.8	glideslope/localizer deviation combined on cross-pointer in approach and tightly controlled "outer-loop" signals
0.6	primary "inner-loop" signals such as pitch and roll attitude combined in artificial horizon, heading, and airspeed

Table 6–7. Duration of eye fixations of aircraft pilots

		Fixation duration (sec.)							
		Airspeed	Directional gyro	Gyro horizon	Altitude	Turn and bank	Vertical speed	Engine instruments	Cross pointer
AFTR 5839 ILAS		0.38 ±.09	0.56 ±.08	0.52 ±.13	0.38 ±.11	0.34 ±.13	0.39 ±.12	0.79 ±.34	0.86 ±.31
AFTR 5967 GCA		.57 ±.17	.90 ±.26	.56 ±.21	.39 ±.11	.36 ±.16	.47 ±.12	.88 ±.31	
AFTR 5975 Routine maneuvers	Straight descent	.67 ±.21	.54 ±.13	.54 ±.15	.48 ±.12	.34 ±.19	.49 ±.14	1.25 ±.76	
	Descending turn	.61 ±.14	.59 ±.12	.63 ±.18	.45 ±.12	.46 ±.13	.46 ±.11	1.13 ±.42	
	Level turn	.52 ±.11	.60 ±.15	.70 ±.20	.50 ±.15	.44 ±.14	.39 ±.12	.93 ±.54	
AFTR 6018 Selected maneuvers	Straight and level	.65 ±.23	.91 ±.36	.70 ±.18	.60 ±.14	.44 ±.22	.46 ±.12	1.49 ±.20	
	Standard rate turns (timed)	.52 ±.26	1.03 ±.57	.82 ±.17	.47 ±.10	.85 ±.31	.58 ±.11	0 ±0	clock .80 ±.23
AFTR 6570 ILAS Experimental panel	Day	.49 ±.12	.54 ±.12	.37 ±.12	.38 ±.10	0 ±0	.39 ±.07	.89 ±.38	.76 ±.23
	Night	.55 ±.09	.66 ±.11	.40 ±.08	.44 ±.12	0 ±0	.53 ±.11	.89 ±.32	1.23 ±.89

TABLE 6–8. FREQUENCY OF EYE FIXATIONS OF AIRCRAFT PILOTS

		Fixation frequency (1/min.)							
		Airspeed	Directional gyro	Gyro horizon	Altitude	Turn and bank	Vertical speed	Engine instruments	Cross pointer
AFTR 5839 ILAS		16.0 ±5.0	28.0 ±8.9	16.9 ±8.9	2.7 ±1.8	0.9 ±1.1	3.5 ±3.0	1.7 ±1.5	29.6 ±6.3
AFTR 5967 GCA		18.2 ±8.0	33.3 ±7.9	20.6 ±9.5	4.7 ±3.4	3.2 ±4.7	6.2 ±5.7	2.6 ±2.3	
AFTR 5975 Routine maneuvers	Straight descent	22.0 ±7.5	24.5 ±10.4	25.0 ±10.2	8.7 ±5.6	3.8 ±4.1	10.7 ±5.8	6.4 ±3.3	
	Descending turn	18.8 ±8.6	25.4 ±7.1	24.7 ±7.5	8.9 ±5.4	6.7 ±6.6	9.0 ±5.8	5.5 ±3.1	
	Level turn	14.8 ±9.2	26.7 ±8.4	26.3 ±6.5	15.1 ±6.3	8.0 ±5.9	9.4 ±7.4	2.5 ±3.1	
AFTR 6018 Selected maneuvers	Straight and level	6.8 ±4.1	25.9 ±10.6	22.9 ±10.8	13.7 ±7.0	4.4 ±2.4	6.6 ±5.6	.9 ±1.2	
Day vs night	Standard rate turns (timed)	2.0 ±1.2	17.7 ±6.9	14.8 ±9.7	7.9 ±4.4	12.5 ±7.5	1.8 ±2.0	0 ±0	: clock 12.3 ±5.1
AFTR 6570 ILAS experimental panel	Day	9.3 ±4.2	23.4 ±6.7	18.8 ±11.0	3.8 ±3.6	0 ±0	8.2 ±5.0	2.5 ±2.3	37.3 ±7.7
	Night	9.8 ±5.7	17.6 ±7.1	10.3 ±8.5	2.9 ±3.0	0 ±0	6.1 ±3.7	1.6 ±2.2	29.9 ±7.4

0.4......... loosely controlled "outer-loop" and monitored signals, such as pressure, altitude, vertical speed, turn rate, lateral aceleration, and sometimes, airspeed

A threshold or refractory interval for dwell time appears in the range 0.2 to 0.25 sec.

Senders (1968) said that there was no indication of clear-cut differences of this sort in his data (Senders et al., 1965) or in subsequent work they did in the C-11 Link trainer. He said his data on fixation duration all fit fairly well within the following:

$$T_d = 0.45N + 0.1,$$

where T_d = fixation duration in seconds, and

N = number of discrete indicators on a single instrument, $1 \leq N \leq 6$.

While this model is undoubtedly an oversimplification of the phenomena, it does have the advantage of being easy to use, and the times it predicts are not grossly different from the early data by Fitts et al. (1950). All of the data show slight variations dependent upon the flight phase. This would indicate that additional sources of variance do exist, but these sources are not readily apparent. Additional data on pilot eye fixations *with corresponding records of the displayed parameters* are needed to resolve this question.

Delays Due to Sampling

Given estimates of the sampling frequency and the duration of each sample, it is possible to estimate the effective delay (phase shift) due to sampling. Actually, this must be done iteratively, since the delays change the signals on which the estimates of the sampling frequency are based. To calculate the loop signals, it is necessary to include the adaptive pilot dynamics and the effects of his display sampling. The amount of delay introduced is a function of the scanning frequency and the dwell time. It has been shown (McRuer et al., 1967) that the amount of delay introduced is also a function of whether or not the operator makes use of rate information. Two different classes of signal reconstruction from sampled data have been investigated by Clement (1967) as models for human operator sampling behavior. The incremental time delay due to sampling and reconstruction for both classes of reconstruction are shown in Table 6-9.

Both Linear and Truncated Cardinal reconstruction yield similar numerical answers and both are similarly affected by derivative weighting. Because they yield similar effects, and because there is currently very little human describing function data taken in the multiloop situation, it is impossible to determine which is the more accurate description of the operator.

6.4.2 Display Selection—Loop Closures

In the preceding section, it was assumed that the operator's displays were given. In many design situations this is true, but frequently they are given only in the sense that there is a traditional set of instruments, evolved over time by trial and error, that is always used in a particular class of systems. When there is not a backlog of experience from which to draw, as when designing a manually controlled booster or a new V/STOL aircraft, the choice of displays must still be made. In general, a conventional fixed-wing aircraft panel will not be optimal for these types of systems and may not even be flyable. Here again, new techniques are available.

An iterative design approach, considering various loop closures through the pilot and including the pilot's dynamics and the incremental scanning delay, will yield a reasonable set of parameters to be displayed. This procedure is not unlike that involved in determining loop closures for an autopilot, with the operator's dynamics being analogous to the designer's choice of loop compensation. Good pilot-opinion ratings and good system performance usually correlate with minimum pilot compensation. Thus, consideration should be given to selecting quantities for "inner-loop" displays which will reduce pilot compensation requirements even at the expense of additional scanning delays.

As a simple example, consider an acceleration control (K/s^2) system. As noted earlier, the operator must adopt a large lead to achieve satisfactory control. If the operator is given a rate display in addition to the position display, then the system is reconfigured as a multiloop

TABLE 6–9. EFFECTIVE TIME DELAY INCREMENT DUE TO SAMPLING BEHAVIOR

Reconstruction class	Without derivative weighting	With first derivative weighting coefficient R $0 \leqslant R \leqslant 1$
Linear	$\frac{T_s}{2}\left(1 - \frac{T_d}{T_s}\right)$	$(1 - R)\,\frac{T_s}{2}\left(1 - \frac{T_d}{T_s}\right)$
Truncated cardinal	$\frac{4T_s}{3\pi}\left(1 - \frac{T_d}{T_s}\right)$	$(1 - R)\,\frac{4T_s}{3\pi}\left(1 - \frac{T_d}{T_s}\right)$

Note: T_s = Mean sampling interval; T_d = Mean dwell (fixation) time

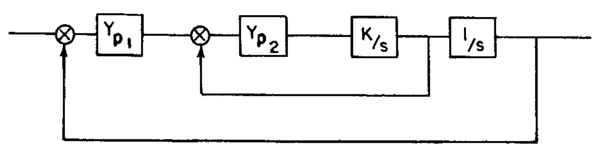


FIGURE 6–54. Multi-loop control of second order system.

system as shown in Figure 6-54. The operator can close the inner loop using the rate display with a pure gain adaptation. If the inner-loop gain is made sufficiently high, then the inner loop becomes equivalent to a simple gain at low frequencies and the outer loop can also be closed by a simple gain. The operator has thus been relieved of a lead generation task at the expense of sampling two displays.

McRuer et al. (1967) have generalized this notion to describe how the operator decides which displays to sample from a complex array. This is called the "adaptive feedback selection hypothesis" and is stated below:

Given: The task variables and evaluation criteria, the controlled element, and the potential feedbacks (quantities perceived);

Then: The pilot evolves a preferred multiloop structure similar to the one which would be chosen by a skilled automatic control system designer and having the following properties:

- satisfaction of task criteria
- wide tolerance for pilot variability
- highest pilot rating among practical closure possibilities
- least incremental delays arising from scanning and sampling; (the information used will come from the following in order of preference: the general visual field, integrated displays, and separately displayed quantities).

The adjustment rules for the multiloop pilot model depend on the function of the feedback loop under consideration.

The outer loops are equalized in accordance with the single loop rules, with some additions to account for the closure of other (inner or parallel) loops.

- *Equalization.* The equalization form and parameters are based on the 'effective controlled element' with the inner loop closed, i.e., Y_c includes the augmenter-like action of the pilot in the inner loop. In general, feedback selections which involve much equalization, or different equalization in each loop, seem to be avoided by the pilot. A wide range of acceptable outer loop gains is desirable.
- *Crossover* frequency. The same crossover frequency considerations apply as for the single-loop case. Often the outer loop has a lower crossover frequency than the inner loop or single loop cases because of a larger accumulation of effective time delays.

The inner loops often act as parallel equalization for subsequent loops, or provide feedbacks or cross feeds which suppress subsidiary degrees of freedom which have undesirable effects or subsequent loops. Because the role of the inner loops is so dependent on outer loop requirements, the rigid rules given above for the outer loop are not generally applicable to the inner loops, e.g., even stability may not be required. The types of inner loops closed and the equalization selected will generally be compatible with one or all of the following considerations:

- outer loop adjustments per the outer loop adjustment rules become more feasible, e.g., Y_pY_c can be made approximately −20 dB/decade with less outer loop equalization
- the sensitivity of the closed-loop characteristics to changes in either inner or outer loop pilot characteristics is reduced from that in an outer-loop-

only situation. This includes the improvement of gain and phase margins
- the loop structure and equalization selected are those for which total pilot rating is the best obtainable
- time delays due to sampling are minimized

These same rules can serve as a guide to the designer in determining what information he *must* present to the operator. Although the foregoing is not completely validated and must be regarded as a working hypothesis, it does provide a starting point for design. It will: (a) define *possible* sets of displays for empirical validation, (b) provide a tentative ranking of alternative candidates, and (c) provide a rationale for the interpretation of experimental results.

6.4.3 Display Arrangement

Early aircraft, with few instruments and a relatively large panel, did not place a strong constraint on instrument placement. However, as speed increased and more sophisticated instruemntation evolved in response to more stringent system requirements, panel arrangement and pilot instrument scanning became more critical. It was this problem that led to the now classic studies of pilot eye movement by Fitts, Jones, and Milton, referred to earlier. The hope of those studies was that certain associative characteristics among instruments could be found that would provide a basis for instrument arrangement. They determined the "link values," e.g., the probability of looking from instrument A to instrument B, or vice versa, and used this information to recommend modification of the instrument panel. The basic idea was that the most frequently fixated instruments should be mounted in the center of the panel. Instruments having high "link values" or transitional probabilities should be located peripherally adjacent to the central instruments, and instruments having the lowest link values should be the most remote. It was largely on the basis of these data, collected in a C-45, that the current "T" panel arrangmeent was developed.

Senders (1964) showed that the link values, or transitional probabilities, were predictable on the basis of the individual instrument fixation probabilities. The fixation probability is estimated by the relative sampling frequency of the display, as explained earlier under "Display Sampling," in Section 6.4.1.

Given the fixation probability P_i of the ith instrument, the transitional probability or link value value between any two instruments is

$$P_{ab}=\frac{2P_aP_b}{1-\sum_{i=1}^{N}[P_i^2]}. \qquad (6\text{–}52)$$

Using this model and measured fixation probabilities, Senders was able to achieve a correlation of better than 0.9 between observed and predicted link values on a monitoring task. Somewhat lower correlations were found when the model was used to predict the link values found by Fitts et al. in flight (Senders, 1966). The results were still quite good, with most of the variability being in the low-probability fixations and consequently the low link values. Since it is the high link values that are of prime importance in laying out an instrument panel, the predicted links should suffice.

McRuer et al. (1967) used the above equation to predict link values for the ILS landing phase of a Boeing 707 using *predicted* fixation probabilities. These predicted link values were then used to lay out an optimized panel. This was done for two alternative instrumentation schemes. One produced an exact correspondence to the configuration actually adopted by an airline for FAA Category II certification. The other alternative produced an almost exact correspondence to the configuration actually adopted by another airline for their Category II certification. While by no means conclusive proof of the validity of the technique, these results do suggest the utility of the technique for preliminary design. Although this technique was developed and discussed in terms of aircraft systems, there is nothing in the development that restricts its generality. It is a general technique for handling the design and layout of instrument panels for dynamic man–machine systems. Although it does not provide a complete answer automatically, it is the most powerful tool yet available for this purpose.

6.5 Discrete Control Systems

The preceding sections have dealt with continuous control or tracking systems. Many real control tasks do not require continuous tracking. These range from "bang-bang" discrete controllers used for tracking, through step input tracking, to essentially monitoring tasks where a well-defined control action is taken whenever a change of input is detected. The first of these, "bang-bang" control, is essentially a gross nonlinearity, while step input tracking and monitoring deal with discontinuous inputs and may or may not involve linear machines, although the operator's response is highly nonlinear.

6.5.1 "Almost Continuous" Control Systems

Many systems can be designed less expensively with "on-off" or "bang-bang" controllers. Motor controllers that provide a motor speed linearly proportional to control displacement are much more difficult to design and expensive to build than simple on-off controls or than controls which allow several fixed speeds. Reaction control systems with throttlable jets are very complicated compared to on-off systems. Any decision in this area must be tempered by the relative performance of the linear vs. on-off system as well as by the relative costs. Seidenstein and Berbert (1966) showed no difference between on-off and proportional rate control for a simple two-axis positioning task when the rate was optimized, although differences were quite significant when the rates were not optimum. When the task involved four degrees of freedom, Williams and Berbert (1968) showed that the proportional rate control reduced task time about 25% compared to either on-off rate control or rate-limited position control. Both of the above studies used relatively simple positioning tasks. The author knows of no studies comparing on-off controls with proportional controls for continuous tracking with rate command systems.

Bauerschmidt and Bescoe (1962) compared on-off and proportional controllers in an acceleration command system. The task was three-axis attitude stabilization of an undamped inertial system (K/s^2). The disturbance was a summation of five sinusoids from 0.02 to 0.20 Hz. The results are shown in Figure 6–55. The mean-square error at optimum gain was approximately twice as large with the on-off control as it was with the proportional control.

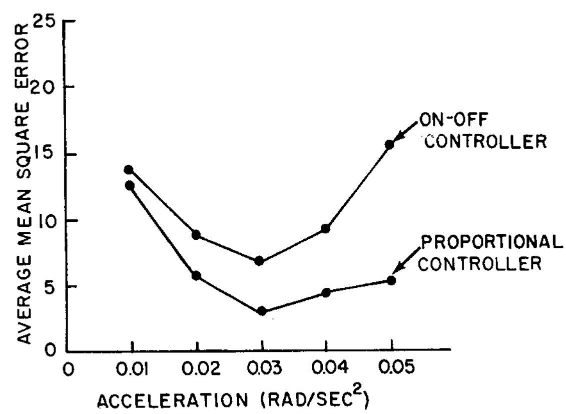


FIGURE 6–55. Comparison of on-off and proportional controllers for acceleration control (Bauerschmidt and Bescoe, 1962).

Discrete positioning or nulling an acceleration control system is quite difficult with an on-off type of controller unless explicit rate information is displayed to the operator. Pew (1965) compared performance on a single axis K/s^2 positioning task with three different types of displays. The "displacement" display showed only target displacement, a simple compensatory display. The velocity vector display was the same except that a line whose length and direction were directly proportional to instantaneous target velocity was superimposed on the target dot. The phase plane display was two dimensional with target position along the horizontal axis and target velocity on the vertical axis. Three acceleration levels, 2, 10, and 30 cm./sec.[2] on the display were used. The results are shown in Figure 6–56. Yasui and Young (1967) demonstrated that adding the optimum switching line to the phase plane display allowed subjects to match the performance of a time-optimal controller. They also generated and displayed the optimal switching surface for time-optimal control of a K/s^3 system and demonstrated that operators could use such a display effectively. They proposed that such a switching

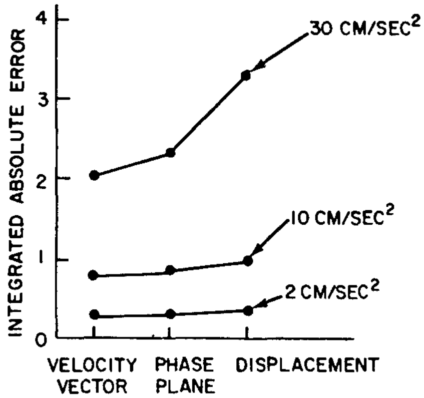


FIGURE 6–56. Positioning performance for three displays using on-off acceleration control (Pew, 1965).

surface or hypersurface could be generated for higher order dynamic systems, but no experimental data were obtained. Miller (1967) demonstrated that operators could do almost as well controlling a K/s^2 system using a phase plane predictor display ts they could using a phase plane display with the optimal switching line. Additional research is needed in this area.

6.5.2 Response to Transient Inputs

There is currently no satisfactory model of human operator response to transient inputs. When and if such a model is developed it will undoubtedly be an adaptive model, as is the current quasi-linear describing function. There will, of necessity, be a set of "adjustment rules" which account for the type of input—random or recurrent step, ramp, impulse, etc.—and the spacing between inputs, as well as system dynamics, gain and man–machine stability requirements. Because there is no such model now, this section will simply attempt to describe some of the factors which influence operator response to transient inputs.

Accuracy vs. Time

Given a good control system and enough time, the limit of a man's ability to accurately position a marker is defined by his visual acuity. That is, if the control–display ratio is appropriate to the task, there is little friction and no backlash, and there is sufficient time, the operator can null any error that he can see. Traditionally, setting time has been arbitrarily divided into slewing time and adjustment time, although there is no well-defined break between them. Figure 6–57 shows how the two components of setting time vary as a function of control gain. The slewing time curve shifts up and to the right as slewing distance increases. The whole adjustment-time curve shifts up and to the left as the accuracy requirements of the task are increased. The absolute widths of the cursor and the target also affect adjustment time, although this does not appear to interact with system gain. As an example, Jenkins and Olsen (1951) found that it took about 0.15 sec. longer to position a 0.016-in. pointer over a 0.032-in. marker than to position a 0.109-in. pointer over a 0.125-in. marker, although the tolerances were identical. The difference was highly reliable and was independent of both gain and size of slewing movement.

There are no good analytic techniques for predicting the interactions among gain, target size, adjustment accuracy, slewing distance, and task time. The data given in Figure 6–58 are typical of the effect of gain. Note that as the required tolerance is relaxed, the optimum gain increases. Figure 6–59 shows the effect of tolerance on performance time for two gains. Note that for the higher gain the performance "break" occurs much more sharply. Very similar data exist for step tracking with level controls (Jenkins and Olson, 1951). In the case of levers, the optimal gain (inches of pointer movement/inches of movement of lever tip) is between 0.1 and 0.5 for required accuracies on the order of 0.15 in.

Speed Stress

As noted above, given a good control system *and enough time*, the limits of accuracy of a manual positioning system are determined by the operator's ability to see the error. In some situations, there is not sufficient time for the operator to respond completely to one step before another occurs. In other cases the amount of time available between steps is only slightly longer, on the average, than the time required to respond. In these cases variability, both within operators and between operators, increases greatly.

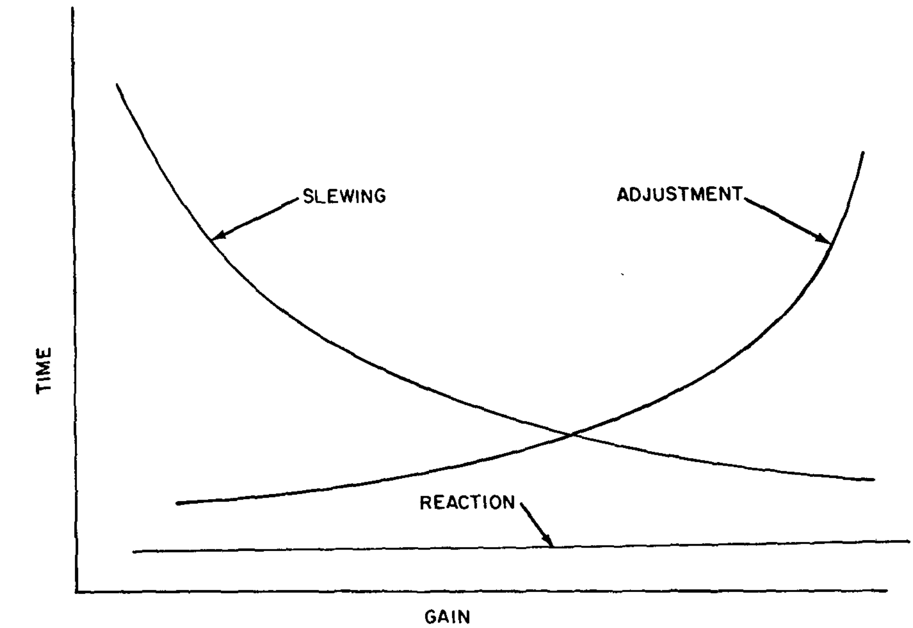


FIGURE 6–57. Adjustment and slewing time vs. control gain.

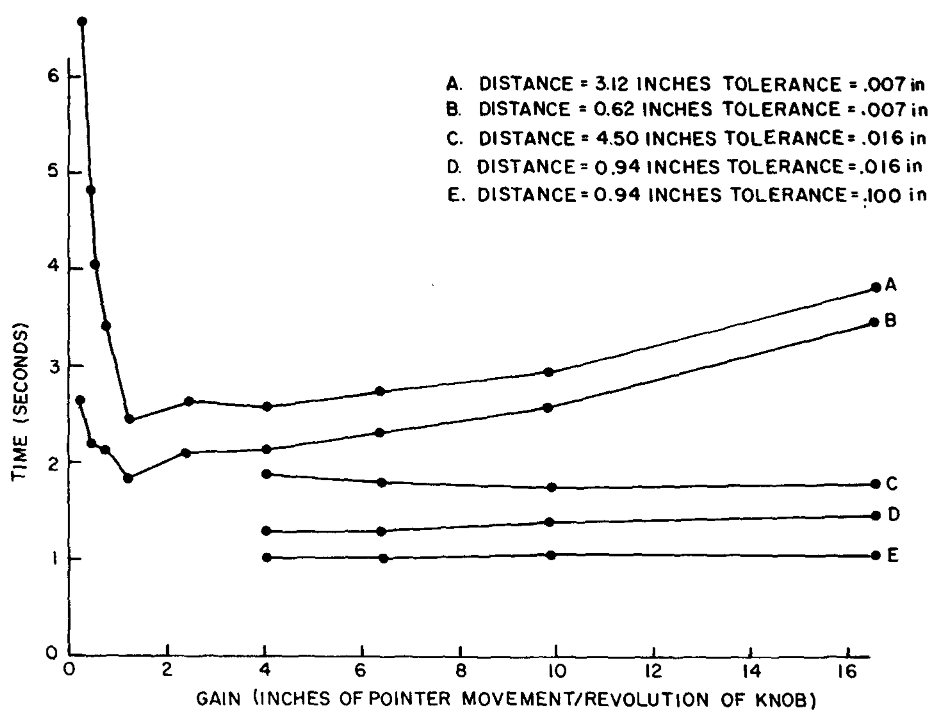


FIGURE 6–58. Response time to step inputs vs gain (Jenkins and Olson, 1951).

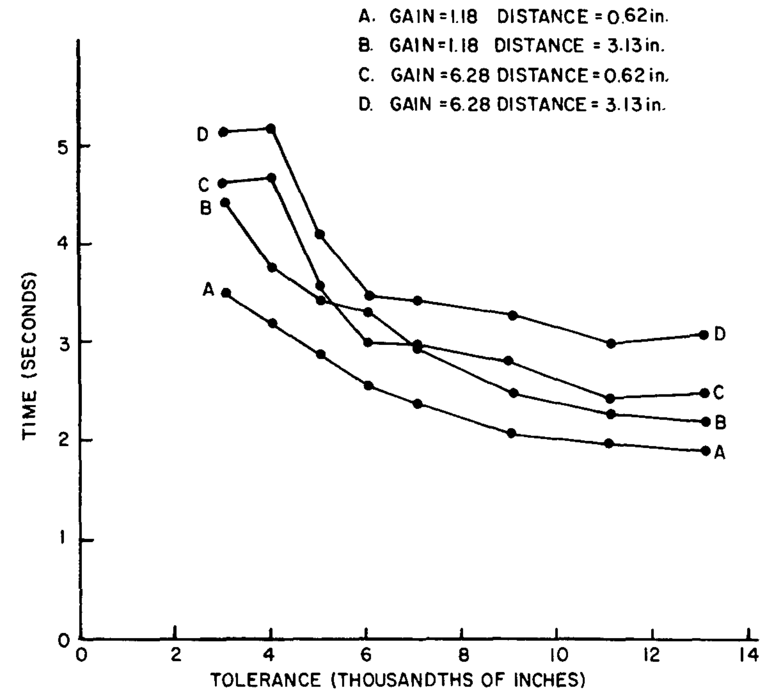


FIGURE 6–59. Response time to step inputs vs required adjustment tolerance (Jenkins and Olson, 1951).

If there is enough time to respond between steps, operators tend to overshoot the small steps and undershoot the large ones. That is, they appear to respond to an average expected step size and then apply a secondary correction if there is enough time. This is the so-called range effect and is well documented in the literature: c.f. Ellson and Wheeler (1949). When the steps are of constant amplitude and of random direction and duration, there is no "range effect." Ellson and Coppock (1951) concludes that under these conditions,

> Response amplitude is essentially proportional to input amplitude. This evidence favors Craig's hypothesis that the human operator's response to a step input is basically linear, that the nonlinearity known as 'range effect' is produced by conditions which allow the response to be based in part on differences between the momentary input and previous inputs. When these differences are not present, the linear requirement of proportionality between response amplitude and input amplitude is fulfilled.

If the signal rate is increased still further, operators may respond to several steps as if they comprised a single input, respond to each input individually but spread out in time (queueing), and/or fail to respond to some steps at all. Because the operator's behavior is so heavily determined by the precise time pattern of the input, it is impossible to make more specific predictions about his response. For systems where this type of input may occur, the input and the system must be simulated and the results evaluated.

Digital Displays

For many tasks requiring precise settings over a wide dynamic range, digital displays are advantageous. These advantages are not, however, without accompanying disadvantages. Most digital displays employ some form of detent mechanism, mechanical or electronic, so that the digits snap from one numeral to the next with no ambiguity between numerals. While this is desirable for static reading, it does not allow the operator to obtain good rate information from the display. It also limits the resolution of the display to plus or minus one-half of the value of the least significant digit. Digital displays are not good in situations where frequent "check reading" is required. It is not possible to tell at a glance whether or not the displayed variable has changed slightly. Where "check reading" and/or rate sensing is required, along with high accuracy over a wide dynamic range, hybrid displays are necessary. The "counterdrum-pointer" altimeter is an excellent example of such a hybrid. (See Chapter 3.)

To take advantage of both the accuracy and the dynamic range of the digital display requires a well-designed control system. Usually some form of slewing control is needed to avoid problems in setting system gain. That is, if the gain is low enough to take advantage of the resolution or accuracy of the display it penalizes the speed of response for large changes. Conversely, if the gain is high enough to swing rapidly through the dynamic range, it is too high for very accurate setting. For some applications a knob with a flip-out crank or a high-inertia, low-friction system which can be spun rapidly is a satisfactory solution. In other situations more sophisticated slewing systems are required.

6.5.3 Factors Affecting Reaction Time

Reaction time to a discrete stimulus is one of the most studied and least understood facets of human behavior. Virtually every variable affecting human behavior affects reaction time. A simple bibliographic listing of the literature on human reaction time would exceed the space available for this chapter. Consequently this section will be limited to those factors over which the designer has some degree of control. Motivation, fatigue, and most environmental variables will not be treated.

Reaction time can be (and has been) defined in a number of different ways. In discussing operator dynamics in the continuous tracking context, reaction time can be treated as a pure transport delay ($e_o/e_i=e^{-\tau s}$). This has been shown to vary from 0.06 to 0.09 sec. and is the shortest of all measured reaction times. Another "reaction time" used in the tracking context is the "effective" reaction time which may be on the order of 0.4 to 0.6 sec. under adverse conditions. Similarly, in responding to a single, strong, anticipated signal, the time from signal initiation until the operator starts his response may be as little as 0.1 to 0.15 sec., while response to an unanticipated or weak signal that is partially masked may take upwards of a minute, if it is detected at all. Even strong signals may not be noticed if the operator is attending very closely to his other duties.

The reaction times most frequently quoted are simple laboratory reaction times. Typical ranges of reaction time for the various sense modalities are shown in Figure 6–60. These data are from Postman and Egan (1949). They are for *laboratory conditions* and do not include any movement time. The subjects were highly practiced on the task and were forewarned when the signal was about to be presented. They had only one unambiguous and easily detectable signal and one unambiguous response. Consequently, these data represent the best discrete response times to be expected under the most favorable conditions. They are not applicable to unalerted responses or situations in which an operator must make a decision before responding.

Reaction time, or more correctly, response time, to weak signals or to unexpected signals may be much longer than the times shown above. If the operator must make a decision about which of several signals were presented, again the response time is greatly increased. When an operator is closely attending to a task, his response to a warning signal not directly associated with that task is extremely variable and frequently very long. Unfortunately, there are not yet any good models relating these various factors to predict response time in a "real" situation. In the following sections some of the factors affecting reaction time are discussed. The designer is warned, however, that these data must be treated with extreme caution due to the wide variety of factors which can affect reaction time.

Sense Modality

Figure 6–60 shows typical laboratory reaction time data for various sense modalities. These data were obtained for "moderate" intensity levels. Wundt (1874), as cited in Woodworth (1938), showed that reaction times for light, sound, and touch all averaged about 330 msec. at barely perceptible signal levels. Above these levels, it is impossible to say that a light and a

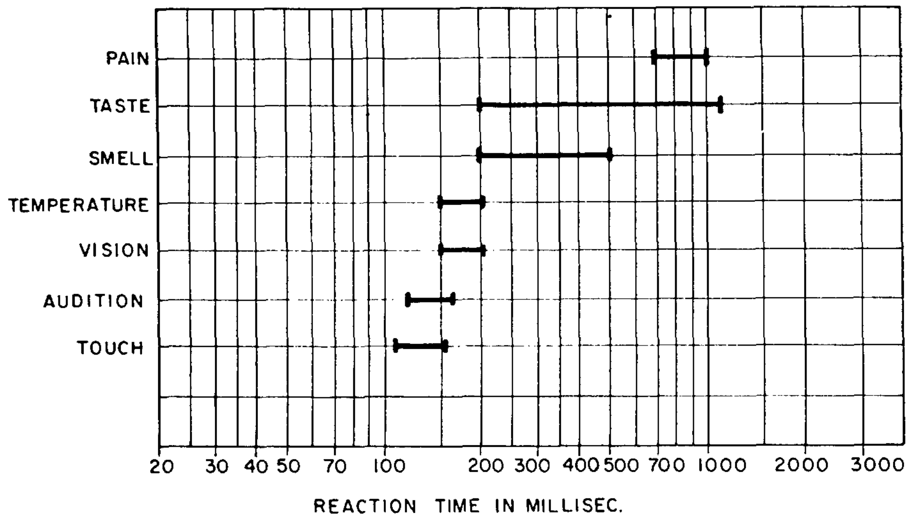


FIGURE 6–60. Range of simple and complex reaction times for the senses under various conditions (after Postman and Egan, 1949).

sound have the same intensity. Thus, Teichner (1954) stated "*. . . there is no evidence available that indicates whether or not the RT varies according to the receptor system stimulated.*" Since eighty years of research has failed to provide definitive evidence of difference, it is reasonable to conclude that if such differences do exist, they are not of practical significance to the system designer. There are, however, situational variables which will dictate a superiority of one sense over another. As shown in the studies by Bate and Bates (1967) and Bate (1969) a visual signal that is not directly in the operator's line of view may not be noticed, and an auditory signal may be required. Similarly, in certain mission phases both visual and auditory senses may be engaged. Thus, a "stick shaker" may be needed to warn a pilot of an impending aircraft stall during take-off and landing if no aerodynamic warning exists. Since no practical differences in reaction time exist between sense modalities, the only advice for determining which modality to use for discrete signals is to examine with great care what the operator will be doing when the signal could occur and what the severity of failure to detect the signal could be if undetected for a long time.

Signal Characteristics

There is a large body of literature on how various signal characteristics affect reaction times *in the laboratory*. There are very few data on the effects of various signal characteristics on response times in real-world situations. Thus, laboratory data all show reaction time as a decreasing function of intensity and of relative change in intensity. (See Figure 6–61.) In the graph, reaction time is plotted against the relative magnitude of the stimulus change ($\Delta I/I$) at four different levels of intensity (log I). The greater the relative change, the faster is the reaction time. Within the limits shown in this figure, reaction time decreases with increases in the general level of intensity.

This general relationship has been found to hold across a wide range of senses, and as Teichner (1954) noted, "Attempts have been made to fit the intensity data into mathematical, theoretical frameworks, with exponential, hyperbolic, and parabolic functions all being used more or less successfully on the same sets of data." This generalization does not, however, consider "startle" responses which can occur when extremely high-intensity signals are used. Thackray (1965) showed extremely long and variable response times in response to very loud (120–*dB*) auditory signals. The responses ranged from 356 ms to 1.80 *seconds* with an average response time of 893 msec.

Other aspects of signals such as the color of a warning light or the pitch of a warning tone have very little effect on reaction time. There

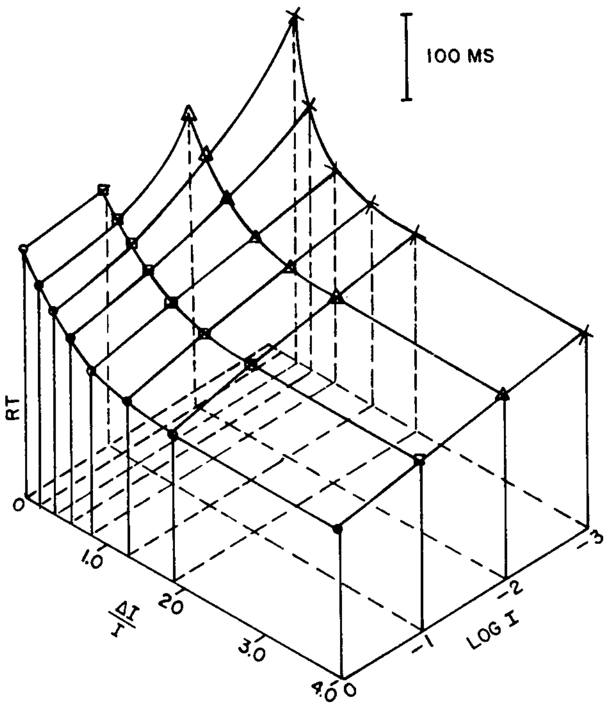


FIGURE 6–61. The relation of reaction time to magnitude of stimulus change (Postman and Egan, 1949, after A. R. Steinman, 1944).

are some data that indicate that the larger the area of a visual or thermal signal, the faster the reaction time (Teichner, 1954). To the author's knowledge there are no data showing the trade off between intensity and area. There are some data showing that flashing lights and interrupted tones have a greater attention-getting value than steady lights or tones, although simple reaction times do not differentiate between steady and intermittent signals. Geratherwohl (1951) showed that the superiority of flashing lights to steady lights increased as the brightness was decreased. That is, response to very bright lights was about the same whether they were steady or flashing, but as the brightness decreased, *response time* to the flashing lights *increased* at a much lower rate than response time to steady lights.

Signal Complexity

When the operator must do more than simply detect the existence of a signal, i.e. when he must decide which of several signals occurred and must respond appropriately, his response time is a function of the complexity of the signal. The response time in this situation depends on the number of choices, the relative probability of each of them, their discriminability one from another, and the compatability of the required response with the presented signal. Figure 6–62 shows typical choice reaction time data (Merkel, 1855) cited in Woodworth (1938). Pew (1965) points out that these data can be treated in an information-processing context by considering the reaction time as "processing time" and the number of alternatives as information to be processed. Thus, two equally likely alternatives equal one bit, four equal two bits, eight equal three bits, etc. Then the processing time, P_t, equals $a+bH$ where H is the number of bits of information. This general relationship describes a great deal of the existing data, but the functional relationships between the constants a and b and the physical variables available to the designer are not yet well defined. The constants must still be determined empirically.

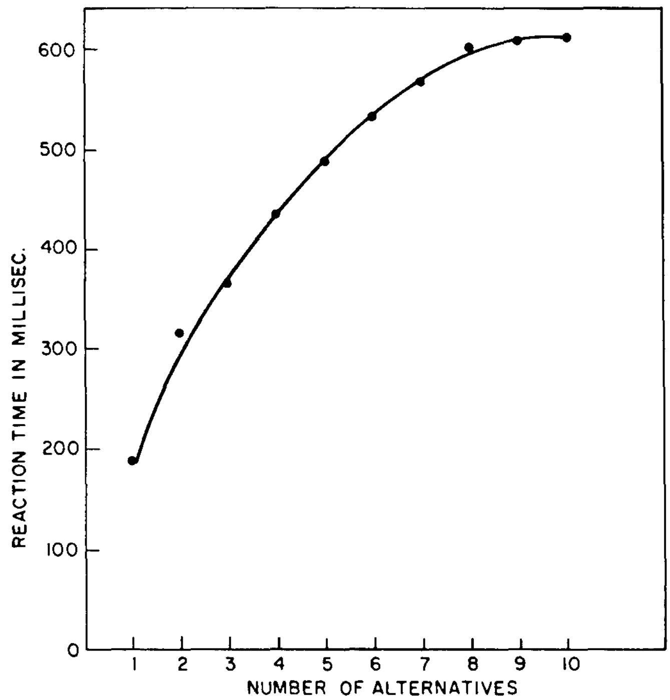


FIGURE 6–62. Reaction time as function of number of equiprobable signals (Merkel, 1855, cited in Woodworth, 1938).

Signal Location

In most practical situations an operator is not sitting, with nothing to do, waiting for a signal. Normally the operator is attending to an important task and other discrete tasks intrude upon his attention. In such a context, the location of a signal light may have a major effect upon the operator's time to detect and respond to the signal. Sharp (1967, 1968) has attempted to determine the interaction between signal light location and task load. Figure 6–63a shows reaction time to a discrete signal versus signal location, obtained while the operator was controlling a moderately difficult divergently unstable tracking task. Two sets of data from two different studies are shown. The data spanning from 0° to 75° (Sharp, 1967) represent reaction times to a combined visual and auditory signal, while the data spanning 57½° to 96½° (Sharp, 1968) represent reaction times to the visual signal alone. Figure 6–63b shows total response time, from signal onset until the operator activated a toggle switch directly below the light, for the same conditions. The most important features of these data are: the great increase in variability without the warning tone, the approximate doubling in *both* median reaction time and median response time response time without the warning tone, and the fact that over 25% of the signals at the 96½° location were entirely missed.

These location effects are particularly important in situations where an operator is not always looking in one direction. In many visual flight rules (VFR) flying situations, an operator may be looking out of the side of his aircraft. At that time, more than one-third of the cockpit may lie outside his −80° field of view. Warning lights in those areas will probably not be seen!

Response Time and Workload

The preceding section discussed the effect of warning light location on response time. It implied that response time was affected by workload. Several recent studies have shown that workload and the "unexpectedness" of a signal can greatly increase response times. Warrick et al. (1965) showed that simple reaction times to unexpected stimuli were approxi-

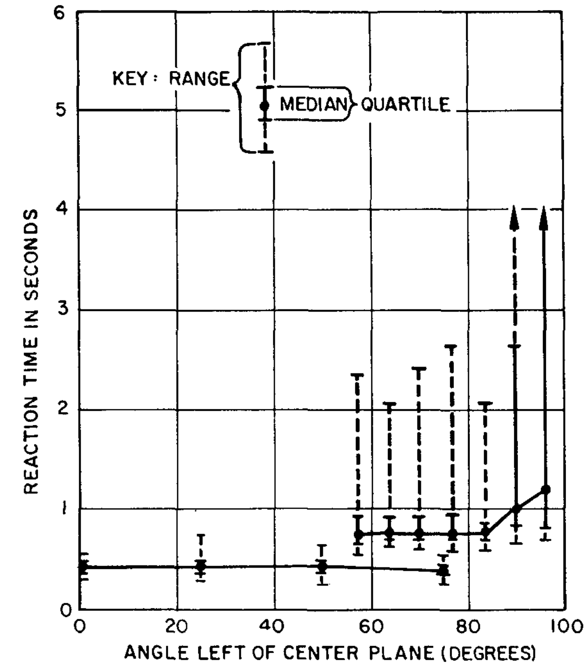


FIGURE 6–63(a). Reaction time to warning lights while tracking (Sharp, 1967 and 1968).

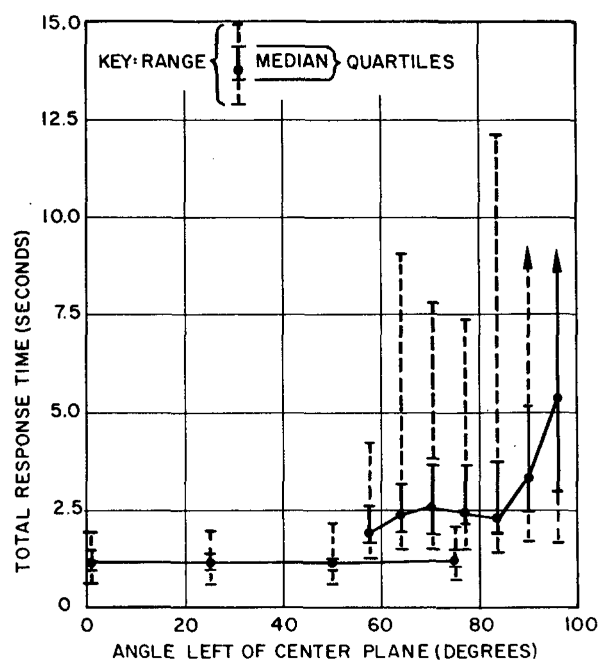


FIGURE 6–63(b). Total response time to warning lights while tracking (Sharp, 1967 and 1968).

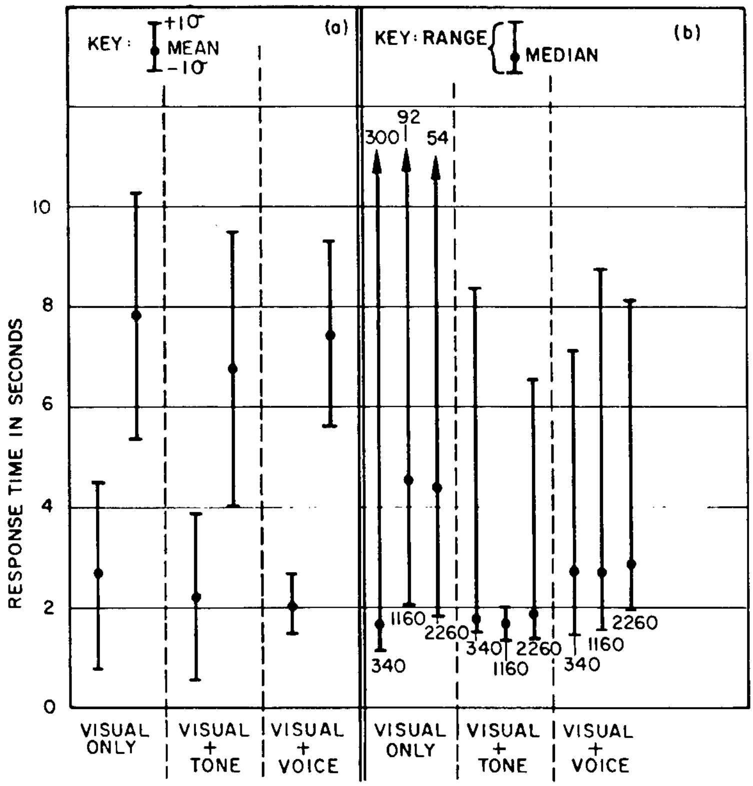


FIGURE 6–64. Response time to simulated warning signals (Bate and Bates, 1967, and Bate, 1969).

mately 0.2 sec. longer than for the identical situation when the subject was alerted prior to the signal. Variability in response times was about the same in both conditions. Bate and Bates (1967) and Bate (1969) compared several warning systems using subjects whose primary task was a very difficult simulated point-to-point navigation problem.

Bate and Bates (1967) compared a standard master caution light and annunciator panel system (per MIL-STD-411B) with the same system augmented by an auditory "sweeping pulsing tone with an 'on' time of 0.85 seconds and an 'off' time of 0.15 seconds" and with the first system augmented by a voice warning system similar to that used in the Air Force B-58 aircraft. The results are shown in Figure 6–64a. The lower set of data represents time to extinguish the master caution light, while the upper set of data represents times to depress the appropriate response switch after extinguishing the master caution light.

Bate (1969) compared a master caution and annunciator panel warning system with an audio warning and annunciator panel (no master caution light) system, and with a voice warning and annunciator panel (no master caution light) system. Only one malfunction per 1-hour trial was presented. Operators were again engaged in a point-to-point navigation task. Workload was varied by simulating three airspeeds, 340 knots, 1160 knots, and 2260 knots. Response times, shown in Figure 6–64b, were measured from onset of the malfunction until the operator depressed the lighted annunciator segment.

In both studies the effect of the primary (navigation) workload was to increase the average response times and to greatly increase the variability over what is seen in the normal laboratory reaction time experiment. The various levels of primary workload had relatively little differential effect on response time. The addition of an auditory signal, either voice or tone, served to reduce the maximum response times although it had very little effect on the minimum response times.

References

Askenas, I. L., & McRuer, D. T. The vocal adaptive controller—human pilot dynamics and opinion. *WESCON Convention Record,* Institute of Radio Engineers, 1960.

Askenas, I. L., & McRuer, D. T. A theory of handling qualities derived from pilot-vehicle system considerations. *Aerospace Engineering,* 1962, **21** (2).

Bate, A. J. Cockpit warning systems comparative study. Report No. AMRL-TR-68-193, 1969, Aerospace Medical Research Labs., WPAFB, Ohio.

Bate, A. J., & Bates, C., Jr. A comparison of cockpit warning systems. Technical Report No. AMRL 66-180, April 1967, Aerospace Medical Research Labs., WPAFB, Ohio.

Bauerschmidt, D. K., & Bescoe, R. O. Human engineering criteria for manned space flight: minimum manual systems. Technical Documentary Report No. AMRL-62-87, August 1962, Aerospace Medical Research Labs., WPAFB, Ohio.

Blackman, R. B., & Tukey, J. W. *The measurement of power spectra.* New York: Dover Publications, Inc., 1959.

Birmingham, H. P., & Taylor, F. V. A human engineering approach to the design of man-operated continuous control systems. Report No. NRL-4333, April 1954, Naval Research Lab, Washington, D. C.

Birmingham, H. P., Kahn, A., Taylor, F. V. A demonstration of the effects of quickening in a multiple-coordinated control task, NRL Report No. 4380, Naval Research Lab, Washington, D. C. June, 1954.

Bradley, J. V. Distribution-free statistical tests. Technical Report No. WADD-60-661, August 1960, Wright Air Development Division, WPAFB, Ohio.

Briggs, G. E., Fitts, P. M., & Barrick, H. P. Effects of force and amplitude cues on learning and performance in a complex tracking task. *Journal of Experimental Psychology,* 1957, **54,** 262-268.

Brown, P. B. Ground simulator studies of the effects of valve friction, stick friction, flexibility, and backlash on power control system quality. NACA Technical Note No. 3998, April 1957.

Burke, D., & Gibbs, C. B. A comparison of free-moving and pressure levers in a positional control system. *Ergonomics,* 1965, **8,** 23-29.

Chalk, C. R. Additional flight evaluations of various longitudinal handling qualities in a variable-stability jet fighter. Technical Report No. 57-719, Part I, January 1958; Part II, July 1958. Wright Air Development Center, WPAFB, Ohio.

Clement, W. F. Cardinal reconstruction theory—A tool for estimating effects of display scanning. Report No. TM 163-B, 1967, Systems Technology Inc. (Also available as ASTIA Report No. AD-661854.)

Cole, E. L., Milton, J. L., & McIntosh, B. B. Eye fixations of aircraft pilots: IX. Routine maneuvers under day and night conditions, using an experimental panel arrangement. Report No. WADC-TR-53-220, March 1954, Wright Air Development Center, WPAFB, Ohio.

Cooper, G. E. Understanding and interpreting pilot opinion. *Aeronautical Engineering Review,* 1957, **16,** 47-52.

Craig, D. R. Effect of amplitude range on duration of responses to step function displacements. Report No. AF-TR-5913, 1949, Wright Air Development Center, WPAFB, Ohio.

DiFranco, D. A. In-flight investigation of the effects of higher order control system dynamics on longitudinal handling qualities. Report No. AFFDL-TR-68-90, August 1968, Air Force Flight Dynamics Lab., AFSC, WPAFB, Ohio.

Elkind, J. I. A survey of the development of models for the human controller. In R. C. Langford & J. C. Mundo (Eds.), *Guidance and control-II.* Vol. 13 of *Progress in astronautics and aeronautics.* New York: Academic Press, 1964.

Elkind, J. I. & Darby, D. L. The statistical properties of signals and measurements of simple manual control systems. Report No. ASD-TDR-63-85, July 1963, Wright-Patterson AFB, Ohio.

Elkind, J. I. & Levison, W. H. Studies of manual control systems: Progress Report No. 4, September 1966. NASA Technical Information Division, Ames Research Center, Moffett Field, Calif.

Ellson, D. G. & Coppock, H. Further analysis of the range effect. A. F. Technical Report No. 6012, August 1951, U.S. Air Force, WPAFB, Ohio.

Ellson, D. G. & Wheeler, L. The range effect. Report No. AF-TR-5813, 1949, U. S. Air Force, WPAFB, Ohio.

Fitts, P. M., Jones, R. E. & Milton, J. L. Eye fixations of aircraft pilots: III. Frequency, duration, and sequence of fixations when flying Air Force Ground Control Approach System (GCA). Report No. USAF AMC TR 5967, February 1950.

Fitts, P. M., Jones, R. E. & Milton, J. L. Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 1950, **9** (2), 24–29.

Frost, G. G. An application of a dynamic pilot-model to system design Report No. ASD-TN-61-57, April 1961, Wright-Patterson AFB, Ohio.

Frost, G. G. & McCoy, W. K. A "predictor" display for on-board rendezvous optimization. Technical Report No. AMRL-65-81, 1965, Aerospace Medical Research Labs., WPAFB, Ohio. (Also in *Proceedings of the National Electronics Conference*, 1965, **21**

Geratherwohl, S. J. Conspicuity of flashing and steady light signals: I. Variations of contrast. Special Report, April 1951. U.S. Air Force School of Aviation Medicine, Randolph Field, Texas.

Gibbs, C. B. Controller design: interaction of controlling limbs, time lags and gains in positional and velocity systems. *Ergonomics*, 1962, **5**, 383–402.

Gibbs, C. B. & Baker, J. C. Free moving vs. fixed control levers in a manual tracking task. In A. Austin, (Ed.), *Automatic and manual control*. London: Butterworths, 1952.

Graham, D. Research on the effects of nonlinearities on tracking performance. Technical Report No. AMRL-67-9, July 1967, Aerospace Medical Research Labs., WPAFB, Ohio.

Graham, D. & McRuer, D. T. *Analysis of nonlinear control systems*. New York: John Wiley, 1961.

Harper, R. P., Jr., & Cooper, G. E. A revised pilot rating scale for the evaluation of handling qualities. Report No. 153, September 1966, Cornell Aeronautical Labs., Ithaca, N. Y.

Humphrey, C. E., Thompson, J. E. & Versace, J. Time sharing and the tracking task. Report No. APL/JHU TG-201, 1953, Applied Physics Lab., The Johns Hopkins University, Silver Spring, Md.

Hunt, D. P. Tracking performance as a function of feedback specificity. Technical Report No. WADC-58-584, March 1959, Wright Air Development Center, WPAFB, Ohio.

Hunt, D. P. The effects of curvilinear and discrete transformations of error information on human tracking performance. Technical Documentary Report No. AMRL-63-137, December 1963, Aerospace Medical Research Labs., WPAFB, Ohio.

Jenkins, W. L., & Olson, M. W. The use of levers in making settings on linear scale. Technical Report No. 6563, August 1951, Aeromedical Lab., WPAFB, Ohio.

Jones, R. E., Milton, J. L. & Fitts, P. M. Eye fixations of aircraft pilots: I. A review of prior eye movement studies and a description of a technique for recording the frequency, duration, and sequences of eye fixations during instrument flight. Report No. USAF AMC TR 5837, September, 1949.

Jones, R. E., Milton, J. L. & Fitts, P. M. Eye fixations of aircraft pilots: IV. Frequency, duration, and sequence of fixations during routine instrument flight. Report No. USAF AMC TR 5975, December 1949.

Kelley, C. R. Developing and testing the effectiveness of the predictor instrument. Technical Report No. 252–60–1, March 1960, U.S. Navy, Office of Naval Research, Washington, D.C.

Kelley, C. R. Further research with adaptive tasks. Final Report on Contract No. NONR–4986, August 1967, Dunlap and Associates, Inc., Santa Monica, Calif.

Krendel, E. S. & McRuer, D. T. A servomechanisms approach to skill development. *Journal of the Franklin Institute* 1969, **269**.

Magdaleno, R. & Wolkovitch, J. Performance criteria for linear constant-coefficient systems with random inputs. Report No. ASD–TDR–62–470, January 1963.

McDonnell, J. D. Pilot rating techniques for the estimation and evaluation of handling qualities. Report No. AFFDL–TR–68–76, December 1968, Air Force Flight Dynamics Lab., WPAFB, Ohio.

McRuer, D. T. & Graham, D. Pilot-vehicle control system analysis. AIAA paper No. 63–310, August 1963. American Institute of Aeronautics and Astronautics, 500 Fifth Ave., New York, N.Y.

McRuer, D. T. & Jex, H. R. Effects of task variables on pilot models for manually controlled vehicles. Paper presented at Advisory Group for Aerospace Research and Development Specialists' Meeting on Stability and Control, 24 September 1966, Cambridge, Eng.

McRuer, D. T. & Krendel, E. S. The man-machine systems concept. *Proceedings of the IRE*, 1962, **50**.

McRuer, D. T. & Magdaleno, R. E. Human pilot dynamics with various manipulators. AFFDL Technical Report No. 66–138, December 1966, Air Force Flight Dynamics Lab., WPAFB, Ohio.

McRuer, D. T., Graham, D., Krendel, E. S., & Reisener, W., Human pilot dynamics in compensatory systems. AFFDL Technical Report No. 65–15, July 1965, U.S. Air Force Flight Dynamics Lab., WPAFB, Ohio.

McRuer, D. T., Jex, H. R., Clement, W. F., & Graham, D., Development of a systems analysis theory of manual control displays. Report No. TR–163–1, October 1967, Systems Technology Inc.

Miller, D. C. Time-optimal control of a second order system by a human operator. In T. B. Sheridan & W. R. Ferrel (Eds.) Measurement and display of control information: Progress report, 1 October–31 March 1967. NASA Grant NsG 107–61, Engineering Projects Lab., Department of Mechanical Engineering, Massachusetts Institute of Technology.

Milsum, J. H. *Biological control systems.* New York: McGraw-Hill, 1966.

Milton, J. L., Jones, R. E., & Fitts, P. M. Eye fixations of aircraft pilots: II. Frequency, duration and sequence of fixations when flying the USAF Instrument Low Approach System (ILAS). Report No. USAF AMC 5839, October 1949.

Milton, J. L., Jones, R. E., & Fitts, P. M. Eye fixations of aircraft pilots: V. Frequency, duration, and sequence of fixations when flying selected maneuvers during instrument and visual conditions. Report No. USAF TR 6018, August 1950.

Milton, J. L., McIntosh, B. B., & Cole, E. L. Eye fixations of aircraft pilots: VI. Fixations during day and night ILAS approaches using an experimental instrument panel arrangement. Report No. USAF–TR 6570, October 1951.

Milton, J. L., McIntosh, B. B., & Cole, E. L. Eye fixations of aircraft pilots: VII. Fixations during day and night, GCA approaches using an experimenral panel arrangement. Report No. USAF TR 6709, February 1952.

Milton, J. L., & Wolfe, F. J. Eye fixations of aircraft pilots: VIII. Fixations during zero-reader approaches in a jet aircraft. Report No. WADC–TR–52–17, February 1952, Wright Air Development Center, WPAFB, Ohio.

North, J. D., & Lomnicki, Z. A. Further experiments on human operators in compensatory tracking tasks. *Ergonomics*, 1961, **4**, 339–353.

Notterman, J. M., & Page, D. E. Evaluation of mathematically equivalent tracking systems. *Perceptual and Motor Skills*, 1962, **15**, 683–716. (Also available as Report No. CU-5–60–NONR–266 942–ERL, AD 634 106, Columbia University Electronics Research Labs., New York.)

Nyquist, H. Certain topics in telegraph transmission theory. *AIEE Transactions*, 1928, **47**, 617.

Obermeyer, R. W., Swartz, W. F., & Muckler, F. A. Human operator control systems: II. The interaction of display mode, control system dynamics and course frequency. Report No. 11494–2, AD 630 881, 1960, The Martin Company, Baltimore.

Pew, R. W. Performance of human operators in a three-state relay control system with and without explicit velocity information. Paper presented at the sixth annual symposium of the Professional Group on Human Factors in Electronics of the IEEE, May 1965.

Postman, L., & Egan, J. P. *Experimental psychology: An introduction.* New York: Harper and Bros., 1949.

Rockway, M. R. The effects of variations in control-display ratio and exponential time delay on tracking performance. Technical Report No. WADC–54–618, December 1954, Wright Air Development Center, WPAFB, Ohio.

Rockway, M. R. & Franks, P. E. Effects of variations in control backlash and gain on tracking performance. Technical Report No. WADC–58–553, January 1959, Wright Air Development Center, WPAFB, Ohio.

Savant, C. J. *Basic feedback control system design.* New York: McGraw-Hill, Inc., 1958.

Seidenstein, S., & Berbert, A. G., Jr. Manual control of remote manipulators: Experiments using analog simulation. Technical Report No. AMRL–66–21, February 1966, Aerospace Medical Research., WPAFB, Ohio.

Senders, J. W. Tracking with intermittently illuminated displays. Technical Report No. WADC–55–378, October 1955 Wright Air Development Center, WPAFB, Ohio.

Senders, J. W. Tracking with intermittently illuminated stimuli (forced sampling). Report No. ARDC TR 56–8 (1956) Wright Air Development Center, WPAFB, Ohio.

Senders, J. W. The human operator as a monitor and controller of multi-degree of freedom systems. *Transactions of the Professional Group on Human Factors in Electronics, IEEE*, HFE-5, **1**, September 1964.

Senders, J. W. A re-analysis of the pilot eye-movement data. *IEEE Transactions on Human Factors in Electronics*, 1966, **7** (2).

Senders, J. W. Personal communication, June 1968.

Senders, J. W., & Ward, J. L. Multiple criteria for assessing tracking performance: The assessment of step input tracking. Report No. 1287, May 1965, Bolt, Beranek and Newman, Inc.

Senders, J. W., Elkind, J. I., Grignetti, M. C., & Smallwood, R. An investigation of the visual sampling behavior of human observers. Report No. 1246, May 1965, Bolt, Beranek and Newman, Inc. (Also available as NASA Report No. CR–434)

Sharp, E. D. Effects of primary task performance on response time to toggle switches in a workspace configuration. Report No. AMRL–TR–66–190, April 1967, Aerospace Medical Research Labs., WPAFB, Ohio.

Sharp, E. D. Unpublished original data, 1968.

Steinman, A. R., Reaction time to change. *Arch Psychol.*, 1944, **292**. p. 15.

Teichner, W. H. Recent studies of simple reaction time. *Psychological Bulletin*, 1954, **51** (2), 128–149.

Thaler, G. J. & Brown, R. G. *Servomechanism Analysis.* New York: McGraw-Hill, Inc., 1953.

Thackray, R. I. Correlates of reaction time to startle. *Human Factors*, 1965, **7** (1).

Trumbo, D., Nobel, M., Cross, K., & Ulrich, L. Task predictability in the organization, acquisition, and retention of tracking skill. *Journal of Experimental Psychology,* 1965, **70,** 252–263.

Verdi, A. P., Ornstein, G. N., Heydorn, R. P., & Frost, G. G. Effects of display quickening on human transfer functions during a dual-axis compensatory tracking task. Technical Report No. AMRL–65–174, November 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Vreuls, D., & Rheinlander, T. W. Allocation of control tasks, IB: Inflight determination of pilot performance with specific force wheel control dynamics in the T-29 Aircraft. Report No. SA 65–2, 1965, Bunker-Ramo Corp., San Antonio, Tex.

Warrick, M. J., Kibler, A. W., & Topmiller, D. A. Response time to unexpected stimuli. Technical Report No. AMRL–65–46, February 1965, Aerospaec Medical Research Labs., WPAFB, Ohio.

Wasicko, R. J., & Magdaleno, R. E. Effects of nonlinearities on human operator tracking performance: A review of the literature. Technical Report No. AMRL–65–158, October 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Webb, P. (Ed.) *Bioastronautics Data Book.* Report No. NASA SP–3006, 1964. National Aeronautic and Space Administration, Washington, D. C.

Weiss, B. The role of proprioceptive feedback in positioning responses. *Journal of Experimental Psychology,* 1954, **47,** 215–224.

Westbrook, C. B. The pilot's role in space flight. Report No. WADC–TN–59–31, February 1959, Wright Air Development Center, WPAFB, Ohio.

Wierwille, W. Improvement of the human operator's tracking performance by means of optimum filtering and prediction. *IEEE Transactions on Human Factors in Electronics,* 1964, **5**(1).

Williams, W. L. & Berbert, A. G., Jr. Control of remote manipulator rotational motion. Phase I Report, Contract AF–33–(615)–2086, 1968, Aerospace Medical Research Labs., WPAFB, Ohio.

Woodworth, R. S. *Experimental psychology.* New York: Henry Holt & Co., 1938.

Yasui, S., & Young, L. R. Manual time-optimal control for high order plants. In Third annual NASA-University conference on manual control. NASA–SP–144, 1967, National Aeronautic and Space Administration, Washington, D. C.

Chapter 7

Data Entry Devices and Procedures

Robert Seibel

The Pennsylvania State University
University Park, Pa.

A wide variety of alternative means for entering data have accompanied the automation of data handling. These range from keyboards, levers, switches, and dials to light pencils, and handwritten or voice inputs. The speed and accuracy with which data entry is accomplished using these devices depend on (a) the characteristics of the source data, (b) the design of the data entry device, and (c) the characteristics of the operator. This chapter presents recommendations and cross comparisons for the human engineering design or selection of a variety of data entry devices, procedures, and source document formats. Steering wheels and joysticks, which are continuous control data entry devices, are discussed in Chapter 6.

CONTENTS

Reviewed by George E. Briggs.

7. Data Entry Devices and Procedures

7.1 Definition of Data Entry

Data entry refers to man to machine communication. This communication is typically accomplished by using keyboards, switches, levers, or dials, mark-sense answer sheets, graphic input, constrained handwriting, or the voice. Only data which are symbolic, and in digital rather than in analog form, are considered here.

The task of entering data may consist of simple operations, such as keying from clear unambiguous hard copy, or of more complex ones, such as copying while resolving ambiguity in the original copy, or of very complex ones which may entail man-machine interaction involving problem solving or information retrieval. The emphasis here will be on tasks at the simpler end of the continuum, not because they are more important but because there are little or no data to provide design guidelines for data entry tasks at the "interacting" end.

7.2 Characteristics of the Input Source to the Operator

7.2.1 Clarity and Orderliness

The clarity and orderliness with which data are presented to a data entry operator critically affect his speed and accuracy. In the typical high-volume data entry task, an operator receives information from a source document (a printed page, written report, or notations in handwritten script) and transduces the data into some machine readable form. If the operator has to skip around or decipher partially illegible letters, numbers, or character combinations, he will tend to make errors and enter the data slowly. The occasional entry task usually gives the operator time to locate and decipher the data to be entered, but if the entry must be made under pressure, orderliness and clarity are again needed for speed and accuracy. This principle is an inference from search time and discriminability studies, and the studies cited below. There are no directly supporting data.

7.2.2 Ordering of Source and Object Documents

A different ordering of entries on the source document relative to the punched card or object document degrades performance. Klemmer and Lockhead (1960), for example, report that there is general agreement that a job which may require skipping around the source document will yield lower performance figures than one in which the source document information is in the same format as the card to be punched.

7.2.3 Source Document Legibility

Klemmer and Lockhead (1960) report general agreement among supervisors that legibility is important for speed and accuracy in key punching. Handwritten documents do not necessarily indicate poor legibility, and they find no consistent difference between printed and handwritten source documents in terms of operator performance.

7.2.4 Format of Source Document

Seemingly small changes in job format may be a hindrance to production. New operators can learn a new format about as easily as new operators formerly learned an old format. However, operators who are experienced on the old format may not regain their former speed after working on a new format for as long as six months (Klemmer and Lockhead, 1960).

7.2.5 Readability of Source Document

Poulton and Brown (1968) studied reading rates of all upper case printing (the form for most data processing systems), versus a mixture of upper and lower case printing. The printing in conventional upper and lower case characters was read 13% faster and comprehended better. While speed of reading and amount of comprehension are not directly related to data entry, both can be expected to affect speed of entry.

7.2.6 Grouping of Data Characters and Length of Message

Conrad and Hull (1967) report on the copying of "random" strings of characters by hand. Strings were either 3, 6, 9, or 12 characters long. Increased length led to more errors. Each string of digits was either grouped in sets of three characters or printed in one long string, and for the 12-digit strings, grouping led to greater accuracy. Additional supporting data are supplied by Smith (1967). Subjects entered 3-, 6-, or 10-digit messages, and errors increased with length of message. Conrad and Hille (1957, cited in Conrad, 1960b) suggest groups of three or four digits are optimum in size.

7.2.7 Isolation of Data to be Entered

Deininger (1960b) reports that the manner in which unfamiliar numbers are displayed is important. Keying unfamiliar numbers from the pages of a telephone directory can increase keying times by 75% and keying errors by 100% in in comparison with keying from a 3- x 5-in. file card on which only one number is typed.

7.2.8 Redundancy of Input

The higher the redundancy of the data being entered, at least up to the limits of normal English text, the faster the data entry and the fewer the errors. Normal English text contains a great deal of higher order sequential constraint. That is, redundancy is much higher than that indicated by the simple frequencies of occurrence of the letters of the alphabet e. g., Miller et al. (1958) and Seibel (1963b).

The effects of redundancy on a data entry task are most directly demonstrated in the study by Hershman and Hillix (1965). Normal text was typed somewhat faster than random words, and random words were typed much faster than random characters. Conrad (1960a) reports similar results for varying orders of approximation to the names of English towns.

Number of Characters Exposed

In the Hershman and Hillix (1965) study the typist was able to see 1, 2, 3, 6, or an unlimited number of characters at a time. The more characters that were exposed, the more accurate the typing for text and random word material. For random characters, however, there was little improvement beyond that obtained by exposing three characters at a time.

Age with respect to redundancy and data entry rate. Rabbitt and Birren (1967) report that people over the age of sixty appear not to take advantage of high degrees of redundancy to the same extent as people between the ages of 17 and 28 years. Differences in "taking advantage of redundancy" may be important in determining data entry proficiency, regardless of age.

Number of Different Characters

It is important to note that the principle concerning redundancy of material refers to the rate of *character* entry, not the rate of information (bits) entry. If a particular data entry job involves random (or near random) strings of alpha (26) or alphanumeric (36) characters, that job has more information per character than a job with correspondingly random numeric (10) characters only. The job with more information will tend to be entered more slowly (in terms of character entry rate), and with more error.

Alphanumeric vs. Numeric Data

Alphabetic messages are usually in a form approaching normal words or text, and as a result are highly redundant. This tends not to be true of numeric data, i. e., they have very little redundancy. Klemmer and Lockhead (1960) report no consistent difference between alphanumeric and straight numeric keypunching.

Subsets of Alphabetic Characters

Messages composed of random arrangements of just ten of the possible alphabetic characters will not be entered as well as straight numeric messages. The operator has learned to expect the occurrence of all 26 characters whenever the alphabet is used. Further, this kind of experience continues during off-work hours. The operator cannot easily take advantage of a restriction to an artificial subset of 10 of the 26 alphabetic characters. (See Conrad and Hull, 1967, and Fitts and Switzer, 1962.)

Information per Keystroke

Data entry messages should, wherever feasible, be designed to be short, by making each character entered convey a large amount of information. See Deininger et al. (1966). Such design will lead to more *messages* (not characters, however) entered per minute. The design may be accomplished by manipulating alphabet size, i.e., number of different characters used, or by manipulating the effective number of keys on the data entry keyboard (e.g., Minor and Pitman, 1965).

Source Document Redundancy and Short-Term Memory

Many data entry tasks impose a short-term memory load on the operator, or the operator may choose to perform his task utilizing his short-term memory for "chunks" (Miller, 1956) or "fields" of the message. This is most likely in situations calling for occasional data entries, in situations where the operator must search for "chunks" of a message in various parts of a source document, or in table look-up tasks such as the finding and entering of a telephone number. When an operator holds a message, or part of a message, in memory, the activity of keying the initial portion of that message is disruptive to the retention of the rest of the message. This is true even if the initial portion of the message is completely redundant (i.e., it is always the same character, or set of characters, for all messages to be keyed).

Thus, another principle is that highly redundant portions of a message handled via short-term memory should be entered last. Under these conditions the retention of the high information part of the message will not be disrupted, and overall accuracy and speed of entry will be improved. (See Conrad, 1958, and Shepard and Sheenan, 1965.)

Further evidence of the effect of short-term memory on data entry speed and accuracy indicates that the insertion of the easiest to remember (i.e., redundant) components of a message in the position most prone to error, according to short-term memory studies, is more efficient than the insertion of those easy components in any other position. The most crucial positions are those just following the middle of the message (Conrad, 1962b). (See Section 7.2.11.)

7.2.9 Chunking and Short-Term Memory

Direct evidence that people use short-term memory in the entry of telephonelike numbers is presented by Deininger (1960b and 1967) and Deininger et al. (1966). The timing of actual character-by-character entries indicates that all people divide longer messages into chunks which they then enter one chunk at a time from short-term memory. However, not all people choose the same chunks unless the messages are so formatted as to make a particular division highly persuasive. Furthermore, if the message length is near the limits of immediate memory span, e.g., about seven digits, then there are wide differences between people in terms of chunk size. Some people will store the entire message in short-term memory and enter it without referring back to the source, while others will chunk the message into at least two parts and refer back to the source between the entry of chunks. This difference in mode of entry makes for very large differences in entry rate.

7.2.10 Graphic and Pictorial Data Sources

Data entry will be faster and/or more accurate if the data to be entered contain information about the location or position of the response that is used to enter that data. For example, Leonard (1964) and Leonard and Newman (1965) report faster reaction times when there is a direct spatial or "geographic" correspondence between a stimulus light and the button which is to be touched in response to that light, than in the situation where the button to be touched is designated by a color, digit, or tone in a single centrally located stimulus cell. Other things equal, a "point-at" mode of entry is more accurate than a "type-in" mode. (See Earl and Goff, 1965.)

A series of papers, starting with Licklider (1960), call for graphic and, in some cases, speech input to data processing systems. These papers also call for outputs from the computer so that the operator and computer may actively interact in terms of all of the modes of communication which the operator would normally use. This active interaction would greatly facilitate human programming, problem solving, and designing capacities. (See Samuel, 1965; Siders, 1967; and Suppes, 1966.) Guiding principles for these modes of entry will evolve as these systems become more commonplace.

7.2.11 Recommendations for Input to the Operator

A faster entry rate with fewer errors depends on:

1. Order and clarity of source data;
2. Operator familiarity with a source's format;
3. For English text source data, a mixture of upper and lower case characters;
4. Division of long messages into chunks;
5. Clearly distinguishable source material;
6. High redundancy of source data for fast *character* entry rate, but minimum redundancy (high information per character) for fast *message* entry rate.

There is no consistent difference in data entry rate between the usual alphanumeric and straight numeric keypunching.

Messages composed of random sequences of just 10 of the possible alphabetic characters are not entered more rapidly or accurately than random sequences of all 26 characters.

The redundant portion of a message should be placed near the middle of the message in the source document. But the operator should enter that redundant portion at the end of the message. This apparently contradicts the recommendation that the sequence of source data corresponds to the sequence of data entered. The consistent sequence recommendation applies to sequences of "chunks" of a message, while this recommendation applies to sequences of characters within a "chunk," with the entire "chunk" being held in short-term memory.

Reducing the memory load requirements of the message should lead to more uniform modes of chunking and entering the data and more accurate and rapid entry.

Entry of graphically displayed data will be faster and more accurate if the data contain information concerning the spatial location or position of the corresponding act of entering that data. "Point at" entry is fast and accurate.

7.3 From Data Source to Data Entry

7.3.1 Translation, Encoding, and Chunking

Operators almost always encode source data into subjective units or "chunks" (Miller, 1956). These data are stored in short-term memory and entered by the human a chunk at a time rather than character-by-character or letter-by-letter. Whenever short-term memory is used, some form of encoding is involved.

Errors Reflect Chunking and/or Coding

In a study by Shepard and Sheenan (1965), all messages contained one of two redundant strings of four digits, along with four random digits. When errors were made in the redundant strings of digits they were frequently substitutions of one string of four for the other. When a redundant string followed the random digits, the interchange of redundant strings accounted for "60% of all incorrectly reproduced digits." Subjects handled these four-digit sequences as coherent psychological units or "chunks," and errors were not often made on characters within a chunk. Error analyses also suggest an *acoustic* encoding of *visually* or *tactually* presented messages (e.g., Conrad, 1963, 1964, 1966, and 1967). Thus, for data entry tasks with large short-term memory components, errors will tend to reflect chunk and/or acoustic confusions, and not, for example, aiming errors on the keyboard. (See Section 7.5.2.)

Encoding and Learning

The average size of subjectively encoded chunks grows with the learning or training of the operator in the particular data entry task. For example, the task of receiving Morse code and translating it into typewriter output usually starts in a one letter at a time mode of reception, gradually progresses to the point where the operator recognizes common strings of letters and short words, and eventually, for the highly skilled operator, progresses to the point where entire phrases and/or short sentences are recognized as units (Fleishman et al., 1958; Fleishman and Fruchter, 1960; West, 1955 and 1962; Woodworth, 1938). This encoding proceeds almost without attention and direction if the operator is actively engaged in the task of data entry because the encoding makes the task easier. A similar effect has been noted for typing (West, 1957 and 1962; Woodworth, 1938). A clear laboratory demonstration of this effect is reported by Leonard and Newman, 1964.

7.3.2 Information, Encoding, and Motor Activity

Character entry rate is fast for highly redundant material (e. g., normal English text relative to random letters). Conversely, the higher the information per entry, the slower the rate of entries (where information is manipulated in terms of redundancy and/or the number of alternative characters). However, when measured in *information units* rather than number of entries data entry rate decreases with redundancy and increases with the amount of information per response.

The motor activity, such as keying, of a data entry operator appears to be slower than the encoding activity. Providing the operator with a system which requires less motor activity (e.g., keystrokes) per unit of information permits the the operator to increase the overall rate of information entry. The price paid for the increased information entry rate is usually in terms of the time required to train the operator on the coding scheme, and on the special motor responses that usually go along with entering it.

The extent of on-the-job training depends upon the form of the data entry task and the material being entered. For example, if the operator's output is in a highly familiar form (e.g., reading aloud), he is almost immediately able to take advantage of the well-known redundancies in English text.

Information, Rate of Entry, Practice, and Compatibility

While entries per unit time generally decrease as the amount of information per entry increases, practice in the task generally reduces the slope of this functional relationship toward zero. That is to say, with sufficient practice (or familiarity) there is essentially no slowing down of rate of entries with increases in information per entry (Conrad, 1962a). In a similar manner, highly compatible stimulus-response relationships also lead to a slope close to zero.

7.3.3 Distribution of Errors, Discrimination, Confusions, and Information Transmitted

Difficulties in discrimination of the stimuli to be entered or the confusions among responses to be made lead to unequal distributions of operator errors. Seibel (1963a), for example, gives a formula for calculating the minimum possible transmission of information for given error rates and numbers of equally likely stimuli. In a particular example, for 1023 stimulus alternatives and a probability of error equal to 0.1, the minimum information transmission by the operator is equal to 8.53 bits (all errors equally likely, and randomly spread among all stimuli). The maximum transmission figure (only one kind of error, and always associated with one stimulus) for the same conditions is just slightly below 10.00 bits. Thus, for these conditions, the difference in the distribution of errors can lead to a difference of as much as 1.47 bits of information transmitted with each response. Clearly the distribution of errors resulting from discriminability of stimuli and/or confusions among responses must be taken into account in establishing any relationship between response time and information transmitted per response.

7.3.4 Other Factors

Motor Difficulty of Entries

Motor difficulty leads to slower responding, more variable response times, and more extensive changes in both time and variability as practice progresses (Seibel 1962b, 1962c, and 1963a). Furthermore, when mixed with easier responses, difficult motor responses produce more variable response times than would be obtained for the difficult responses in isolation. The converse is also true. That is, easier responses mixed with more difficult ones will tend to yield less variable response times than would those easy responses in isolation. Average response times for easier responses also tend to be faster in isolation than if the easy responses are mixed with difficult ones. Error rates also tend to be higher for an isolated set of easy responses, and lower for an isolated set of difficult responses, than would be obtained when the easy and difficult responses are intermixed. Clearly the motor difficulty of the particular responses cannot be ignored in designing data entry devices. See Conrad and Longman (1965) for further evidence in a typing-like task.

Speeded Production vs. Daily Production

For short speed tests of one-half hour, speed on the card punch averages more than 5 strokes per sec. on jobs with no complication (Klemmer and Lockhead, 1960 and 1962a). Over a working day an average of 2.8 strokes per sec. will be achieved during time actually spent on the machine. Thus, there is almost a 2:1 difference in data entry rate when operators are primed for taking a "speed test."

A similar effect 's reported for typists. The average typist in a 5-min. speed test will gross about 70 w.p.m. For a full day's typing the standard manuals ask for an average of approximately 35 w.p.m. (a word is equivalent to five strokes, four characters plus space).

Typing is somewhat faster than key punching according to these figures, but there is still a marked difference between daily production rate and speed test rate. Since typing tasks usually involve the entry of highly redundant English text, whereas keypunching tasks quite frequently involve material with much less redundancy, and since the typing test is only 5 min. long, the apparent difference in speed between typing and keypunching may be misleading. (See Section 7.5.1.)

Speed vs. Accuracy

Operators can be induced to exchange speed for accuracy by means of instructions, punishments, and differential pay-offs (e.g., Fitts, 1966). However, if speed stress is pushed beyond the point of achieving reasonable accuracy, performance in terms of rate of information transmission deteriorates rapidly. A similar deterioration occurs with excessive stress on accuracy (Hillix and Coburn, 1961). Experienced data entry operators work at a near optimum (for them) compromise between speed and accuracy.

Machine Lag, Delay, and Pacing

The effects of mechanically slowing down the operator, pacing the entries, or limiting the speed of entry, are also factors in data entry device design.

Delay reduces advantage of redundancy. Leonard (1958) reports that a 0.35-sec. delay (between the presentation of a stimulus and the time when the equipment would accept a response) had only a slight effect on a five-choice equiprobable discrimination reaction time task. However, when there were four equiprobable alternatives and a fifth alternative occurring 8.5 times more frequently than others, this redundant or biased input permitted markedly faster performance with no delay, but only slightly faster performance with enforced delay.

Delay reduces speed of letter sorting. Conrad (1960c) studied the distributions of operator response times on a Post Office letter sorting system. Performance continued to show improvement for a full year on the job following initial training ("Effects of Training and Time on the Job" in Section 7.5.2), with sorting rate showing a growth from approximately 36 sorts per min. to almost 60 sorts per min. The sorting system enforced a 0.55-sec. delay after each sorting response. The sorting rates are production figures, to be contrasted with sorting speeds obtained during 20-min. speed tests for which sorting rates were usually about 10 sorts per min. higher.

Examination of the distribution of actual sorting times was made for performance at the end of the year of practice. Distributions were sharply "L" shaped, with from 65% to 85% of all response times falling between the mean time and the lower limit of 0.55 sec. The mean response time for all operators was 0.88 sec. Thus, the operators had adapted to the enforced lag and were making the vast majority of their responses fall within 0.3 sec. of the minimum possible time intervals.

Utilizing the distribution observed, significantly faster sorting times are *estimated* for the unpaced condition with no artificial lag. Estimates of sorting rate show a steady decline with increases in duration of lag.

Machine pacing reduces speed of letter sorting. Estimates of sorting rate under machine-paced conditions in the Conrad (1960c) study show a maximum for a pacing rate of approximately 75 sorts per min., but this rate is still considerably below some of those actually obtained and far below those estimated for the unpaced situation. It is important to note that the sorting rates are estimates, and Conrad cautions that the assumptions under which they are made may not be entirely correct.

Conrad concludes that

> In general it seems clear that unpaced letter sorting machines ought to give a higher output than machines which either have a minimum time between one response and the next, or which are paced; the difference being fairly considerable. Where the choice is between a lag and a fixed pace, the acceptable values of each must be taken into account. But whereas reducing the lag will always increase output, increasing the rate of pacing may lead to a drop in output.

Delay reduces speed of typewriting. Minor (1964) reports on an enforced lag (an interlock system) for a typewriting task. Operators typed lists of names and addresses with each having a five-digit random number with it. The longer interlock system required 0.125 sec. between keystrokes, thus setting a typing rate ceiling of eight characters per second; the other 0.077 sec. and 13 characters per sec. Operators were still learning to adapt to the interlock systems at the end of the study, but at that time the shorter interlock condition led to approximately 42.1 w.p.m., while the longer led to only 40.3 w.p.m. (difference significant at .05 level). Total errors (detected and corrected, plus undetected) were not significantly different from each other for the two conditions, but there was a slightly greater number of total errors for the longer interlock condition. The proportion of keystrokes which were undetected errors was significantly (0.05-level) smaller for the longer interlock condition; the magnitude of the effect was approximately 15 versus 18 undetected errors per 10,000 strokes. Thus, the shorter interlock condition led to approximately 4% greater output per unit of time, a slightly lower overall error rate, but approximately three strokes in 10,000 greater undetected errors.

Operators reported that only infrequently, if at all, did they "run" into the interlock during the test period. This suggests that

> Differences in typing speed and error rates were the results of self-imposed pacing rather than a function of 'running' into interlocks. The study does not justify generalizing that longer interdigital interlock time always generates greater accuracy. In all probability the effects of different interlocks will vary as a function of key force displacement characteristics, the nature of the input material, and the skill level of the typists.

The typing of more highly redundant material would be expected to lead to a greater frequency of very short inter-keystroke times, and thus a greater debilitating effect for the long interlock condition. (See "Keystroke Timing and Key Interlocks," in Section 7.5.2.)

Continuous vs. Discrete Tasks

Another important general variable affecting data entry has to do with the fact that each entry is one of a continuous flow of entries. The operator continuously takes in and processes source data while emitting a sequence of entry responses.

In general, if the operator has the opportunity to acquire information from the source, process it, and form it into familiar sequences of entry motions, he can overlap these activities in time and take maximum advantage of the redundancy in the source (see Section 7.2.8, "Redundancy of Input"); *and* in the *sequence* of entry *movements*. A continuous data entry task apparently involves a sort of running short-term memory component as well as chunking. For non-redundant material, e.g., random strings of letters, the running memory (when coupled with the chunking and responding) is approximately three letters long. For normal English text, however, the length of the running memory component appears to have no clear-cut and obvious upper bound. (See Hershman and Hillix, 1965; Leonard and Newman, 1965; Poulton, 1958.)

The effects of the difference between discrete and continuous tasks on various functional relationships (e.g., rate of information entry as a function of information per entry, entry rate as a function of practice, information rate as a func- of practice, entry rate as a function of compatibility, etc.) have yet to be evaluated.

Error Detection Overlaps Other Data Entry Activity

Further indication that overlapping activities take place during a data entry task is provided by Rabbitt (1967). Digits were projected, one at a time, with a new digit appearing within 20 msec. of a correct key press response. Whenever a subject detected that he had made an error, he was to immediately depress a pair of error keys. Intervals between stimuli and correct responses (CRT) were recorded. Reaction times between an error response and the response signaling a detected error (EDRT) were also measured. At the end of practice a series of experimental conditions (degree of compatibility, number of alternative stimuli, number of alternative responses) produced differences in CRT's, but not in EDRT's. Error rates were generally between 3% and 5% at the end of practice, and from 86% to 90% of all errors were detected. Error detection is clearly not the same process as is the making of a correct response, but obviously some form of response checking must take place concurrently with the usual data entry activity.

Encoding to Increase Information Entry Rate

The act of encoding during a data entry task need not slow down the rate of data entry in keystrokes and/or bits. In the usual data entry task, encoding and entry overlap in time. Thus, there is an advantage in the data entry situation in which the operator can encode a redundant source message to reduce the redundancy in the sequence of data entry motions, i.e., reduce the actual number of data entry motions. Several chord keystroke systems employ this principle (Section 7.4.3. contains additional details).

Letter-sorting. Perhaps the best known are the coding systems developed and evaluated for the sorting of mail (Cornog and Craig, 1965). While no careful experimental comparisons have appeared in the open literature, various codes have been described and representative performance data have been discussed at professionsal meetings (Cornog et al., 1963).

In general, mail can be sorted, using a variety of different extraction codes and a straight memory code, with a variety of keyboards. Comparisons are difficult to make on the basis of the reported information, but it is clear that a variety of different encoding and keyboard combinations *do work.*

"Rapid-type" system. Similar demonstration-like data are reported by Seibel (1962a) for a "Rapid-type" data entry system. Operators learned to encode by substituting special abbreviations for frequently occurring strings of letters in normal English text. The abbreviations were introduced gradually and performance showed little decrement resulting from the added encoding burden. With additional practice, performance showed continuous improvement. The letter sequences to which abbreviations were assigned were derived from studies of letter-sequence frequencies (e.g., Seibel, 1963b).

Eight-key chord keyboard. In still another demonstration (Lockhead and Klemmer, 1959), operators were taught to use an eight-key chord keyboard to enter 100 common words (each with a single chord-keypress), the 35 alphanumeric characters, and two punctuation marks. After training, subjects utilized word encodings almost without exception when they were given the option of doing so, or of entering words character-by-character. The 137 chord patterns used were learned in less than 30 hours. Word patterns were entered at between 36 and 55 w.p.m.

Stenotyping system. While the stenotyping or court steno-writing systems require extensive practice to achieve acceptable steno-writing entry rates, the *minimum* expected rate for proficient operators is equal to or better than the world championship typing rate (Seibel, 1964a). The steno-typist, at 200 w.p.m. is estimated to be entering approximately 3 chord-keystrokes per sec., with between 11 and 16.7 bits per stroke, while a "very good" typist at 100 w.p.m. is entering approximately 8.3 single keystrokes per sec., with 2 to 3 bits per stroke. Thus, the typist is keying almost three times as many entries per sec., but entering information at approximately one-half the rate (for the example figures used) of the steno-typist.

Thus, anecdotal and demonstration data suggest that the encoding aspect of the task may call for additional training time, but once high levels of proficiency are attained, the data entry rate in information units or in number of messages entered, is higher than would be attained with character-by-character data entry by a conventional typewriter keyboard.

Experiments with Redundancy-Reducing Codes

Two experiments (Tirrell and Klemmer, 1962; and Deininger, et al., 1968) have examined the effects of redundancy-reducing codes utilizing single-character-per-stroke keyboards (i.e., standard typewriter) or single-character-at-a-time handwritten entry. Entry rates achieved with the coding were better than those without the coding, but the advantage was not as impressive as some of the demonstrations, or stenotyping, would lead one to expect. The impressive gains in data entry rate appear to result from a combination of message encoding and keystroke encoding, such that each single keystroke (usually a chord keystroke) carries a great deal of information. Some of the factors and trade offs involved in developing a specific system to incorporate these principles are discussed by Seibel 1964a). Experimental data are sorely needed.

Corrections, Editing, and On-Line Problem Solving

The more directly the operator can attend to the task at hand, the more efficient will be his operation. Distracting machine demands, such as specific formats, margins, unique spatial locations, special and unusual symbols, cumbersome correction procedures, etc., should be avoided if at all possible. If the important job of the operator is editing, or writing, or designing, or programming, etc., and a large and expensive system is being committed to facilitate his performance of that job, then it makes no sense to distract him from doing his important work by imposing machine-idiosyncratic data entry demands in order to save relatively trivial dollars in engineering or software costs. Unfortunately there are no experimental data for guidance in this area. What support there is for the general principle must be derived from an examination of "what sells."

7.3.5 Recommendations

Translations Requirements

In designing any data entry task, consider that the operator encodes the source data into subjective units or "chunks." Source data formats should be designed with this characteristic in mind. When short-term memory is utilized by the operator, the errors made will tend to reflect confusion among the *encoded* forms of the source data.

The size of subjectively encoded chunks grows with the experience of the operator in the particlar data entry task. This principle should be considered in source data, task, and data entry device design.

New forms of encoding may be learned rapidly if the code takes full advantage of prior learning, e.g., the redundancies of normal English text.

The motor activity of the data entry operator is slower than the encoding activity. Giving the operator a system which requires less motor activity (e.g., keystrokes) per unit of information allows the skilled operator to increase overall rate of information entry. There appears to be little or no overall reduction in entry rate with increases in information per entry for the skilled operator.

Other Factors

Discriminability among the elements in the source data, and in their encoded equivalents, has an important effect on the distribution of operator errors. Discriminability amongst the data entry motions has similar effects. More difficult discriminations lead to higher error rates. Particular confusions lead to particular errors, and these factors should be taken into account in design considerations.

The motor difficulty experienced by the operator in effecting particular entries influences overall rate of entry, error rate, and variability in the time intervals between successive entries. Awkward reaches on the standard typewriter keyboard, and particular chords on a chord keyboard, are more difficult than other strokes on the same devices.

The stage of learning for a given set of operators on given data entry tasks has a very large effect on the performance on that task. Rate of improvement clearly depends upon the motor difficulty of the task; upon discriminability of the message units, encoded message units, and responses; upon the redundancy of the source material, etc. This creates difficult practical and theoretical problems for comparing skilled entry rates for different data entry tasks and devices. Effective elimination of stage of learning as a confounding variable requires very extended periods of practice (at least six months, often a year or more) such that performance shows little or no further improvement with still more practice. (See Section 7.5.2, "Effects of Training and Time on the Job", for further consideration.)

Information entry rate is approximately constant, and maximum, only over a very limited range of speed and error trade off, with a sharp drop outside of the range. Pushing response speed above the optimum range leads to marked increases in error rate (so called "information overload") and the rate of information entered drops off. Excessive attention to accuracy leads to marked reductions in entry rate, but only slight reductions in error rate, with a consequent sharp reduction in rate of information entered.

There are optimum pacing rates for many tasks, but optimum paced rates yield performance considerably below that which the operator can achieve under unpaced conditions. Machine pacing should *not* be used to manipulate incentive.

Interlocks on keyboards introduce a lag between successive entries. Generally, interlocks degrade performance, but they *may* lead to a slightly better rate of self-detection of errors. Operators "adjust" to enforced lags by slowing down their most rapid response sequences.

Encoding to Increase Information Entry Rate

The act of encoding during a data entry task need not slow down the rate of data entry (in keystrokes and/or bits), if the operator is permitted to preview the source material and overlap in time his encoding and output motor activities. Simple encoding schemes, familiarity with material and encoding, and practice, all contribute to faster encoding activity. If encoding and motor activities can fully overlap in time, motor activities will not be slowed down by the encoding.

Information entry rate is higher for tasks with the greater amount of information per entry. This is true even if the operator is not able to preview the source data.

Given the opportunity to preview and encode redundant source data so as to reduce the redundancy in the sequence of data entry motions, (i.e., further increase information per entry), still higher information entry rates should be possible. This inference has led to the development of several data entry systems, e.g., codes and keyboards for sorting mail, "Rapid-type," an eight-key chord keyboard, etc.

Data Entry by On-line Problem Solvers

The editor, designer, programmer, etc. should not be distracted by machine specific details such as special formats, margins, unusual symbols, cumbersome correction and insertion procedures, etc.

7.4 Data Entry Devices

7.4.1 Alphanumeric Keyboards

The arrangement for the alphabetic keys, and the digits 2 through 9 and 0, on typewriters was set by tradition nearly a century ago. There are probably more than 10,000,000 typewriters in the United States with keys arranged in this traditional way. The arrangement is known as the "Sholes" or "QWERTY" arrangement. Despite demonstrated advantages for other arrangements the overall economics and retraining aspects of the situation strongly suggest that the QWERTY arrangement be considered *the standard.*

More efficient keyboard arrangements (e.g., Dvorak, 1943; Griffith, 1949) can be recommended *only* in those cases where there is rapid and complete interchangeability possible at relatively low cost. Complete interchangeability must include changes in key top designations, the characters controlled by each key, and in the case of keyboards producing encoded output for data storage or entry, the interchangeability must also include the encoding system. This complete interchangeability is not readily available in the current market place. No data are available to give reliable guidance as to the expected magnitude of advantage in daily production (as distinguished from speed-test performance) for more efficient keyboard arrangements. Obtaining these data would require lengthy and expensive experimentation. Without the information, it is impossible for the system designer to trade off equipment cost (in order to achieve interchangeability) with system throughput. A *guess* would set the *upper limit* of the daily production advantage for the modified keyboard arrangements at about 10%.

Standard Code

For electronic data processing systems, each character and function key on the alphanumeric keyboard must be assigned a unique bit pattern or code. A standard code, the American Standard Code for information Interchange (ASCII), has been adopted as a standard code for all U.S. Government users. It specifies 95 graphic characters and 33 different function messages. It is a seven-bit code and is fully described in Figure 7-1. Obviously, it should be considered *the standard.*

Special Characters and Bit-Pairing

Since bit configurations are specified for all of the characters in the ASCII code, these configurations were used for determining the arrangement of the keys on a proposed standard keyboard. Upper and lower case equivalent characters were arranged on the keys so that any upper case character differed from its lower case version only in terms of the value for a single bit ("one" for lower case and "zero" for shift). Thus the action of the shift key was supposedly simplified for the engineering of the keyboard. The characters so arranged are said to be "bit-paired." The resultant keyboard arrangement is depicted in Figure 7-2. The arrangement includes seven keys which are exceptions to the simple "bit-paired" shift action (reversals of normal action for bit five for four keys, for bit six for one key, and no effect for two keys; see Standards, 1968a, for details). Despite this "simplicity," engineering convenience is still cited as the reason for maintaining the bit-paired character arrangement on the keyboard.

The most common arrangement for the characters of the alphanumeric keyboard is illustrated in Figure 7-3a, the 1966 suggested "preferred arrangement" for the electric typewriter keyboard. Discrepancies between this arrangement and the one proposed in Figure 7-2 are highlighted in Figure 7-2 by shading. The locations for 14 special characters are involved. Highly skilled touch typists will be seriously inconvenienced by the rearrangement of these special characters.

Other arrangements for "typical" teletypewriter and manual typewriter keyboards are shown in Figures 7-3b and 7-3c.

Training Consideration and a Recommendation

If data entry terminals incorporate a character arrangement different from the electric typewriter, the data-processing industry is imposing a penalty on itself in terms of retraining, interchangeability of equipment, and reduced throughput if the same operator is to switch back and forth from a typewriter to a data entry terminal. The equipment cost for offsetting this penalty calls for special bit-code generation for approximately one dozen keys on the keyboard, particularly with respect to the upper case version of the characters on those keys. Approximately half of those keys are among the keys calling for exceptions to the simple bit-paired shift action. All things considered, it is strongly recommended that the arrangement of the special characters conform to that suggested for the electric typewriter (e.g., Figure 7-3a), and the burden of character-to-code translation be placed on the electro mechanical design of the keyboard rather than on the data entry operator.

Card Punches

The punched card and card-punch machine (i.e., the "keypunch") are the current major forms for data entry. Particular system requirements must be taken into account in deciding on the relative advantages or disadvantages of the "unit record" feature of the punched card. For typical high volume entry, however, the punched card appears to have little if any advantage.

Keyboard to Magnet Tape (or Disc)

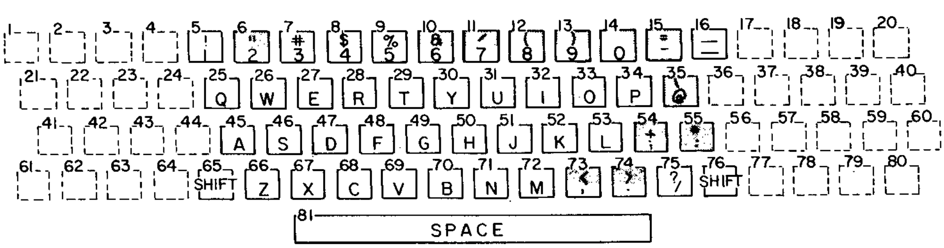
One current innovation, rapidly growing in popularity, permits data entry via a typewriter-

b4	b3	b2	b1	COLUMN → / ROW ↓	0 (b7 b6 b5 = 0 0 0)	1 (0 0 1)	2 (0 1 0)	3 (0 1 1)	4 (1 0 0)	5 (1 0 1)	6 (1 1 0)	7 (1 1 1)
0	0	0	0	0	NUL	DLE	SP	0	@	\	P	p
0	0	0	1	1	SOH	DC1	!	1	A	a	Q	q
0	0	1	0	2	STX	DC2	"	2	B	b	R	r
0	0	1	1	3	EXT	DC3	#	3	C	c	S	s
0	1	0	0	4	EOT	DC4	$	4	D	d	T	t
0	1	0	1	5	ENQ	NAK	%	5	E	e	U	u
0	1	1	0	6	ACK	SYN	&	6	F	f	V	v
0	1	1	1	7	BEL	ETB	'	7	G	g	W	w
1	0	0	0	8	BS	CAN	(	8	H	h	X	x
1	0	0	1	9	HT	EM	)	9	I	i	Y	y
1	0	1	0	10	LF	SUB	*	:	J	j	Z	z
1	0	1	1	11	VT	ESC	+	;	K	k	[	{
1	1	0	0	12	FF	FS	,	<	L	l	\	¦
1	1	0	1	13	CR	GS	—	=	M	m	]	}
1	1	1	0	14	SO	RS	.	>	N	n	^	~
1	1	1	1	15	SI	US	/	?	O	o	—	DEL

■ CHARACTERS FOR WHICH SPECIFIC KEY LOCATIONS ARE NOT PRESCRIBED

▒ EXCEPTION TO SIMPLE SHIFT ON ASCII KEYBOARD

FIGURE 7–1. USA Standard Code for Information Interchange (ASCII) per USAX 3.4–1967 as published in Standards (1968a).



NOTES:

1. The key position numbers are intended for reference purposes only.
2. The alphabetic symbols represent the lowercase letters as well as the uppercase letters.
3. The area in which the graphic keys are placed corresponds to the so-called "44-key touch-typing area."
4. Positions shown by broken lines appear for reference purposes only. Characters are not assigned to them in this standard but may be in subsequent standards.
5. Shaded symbols disagree with those proposed as a preferred arrangement for electric typewriters in 1966 (see Figure 7-3a).

FIGURE 7–2. Proposed USA Standard general purpose alphanumeric keyboard arrangement for information interchange (USASI Document X4/35. X4–A9/54, X4–A9.1/160, July 10, 1967), as published in Standards (1968a).

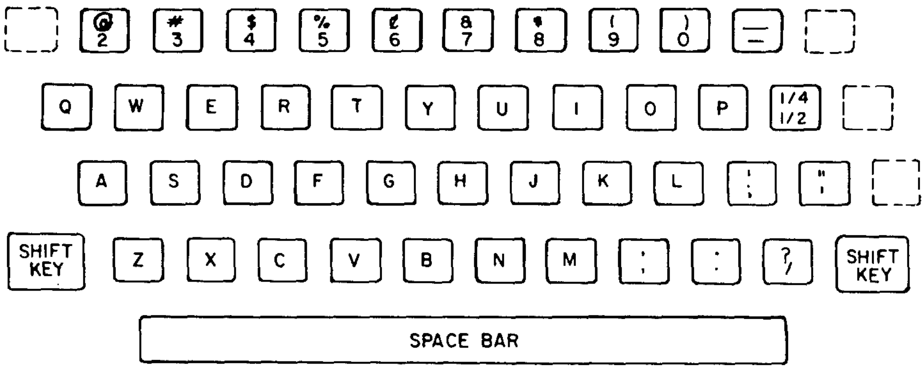


FIGURE 7–3a. Electric typewriter arrangement (preferred arrangement) as published in Standards, (1968a).

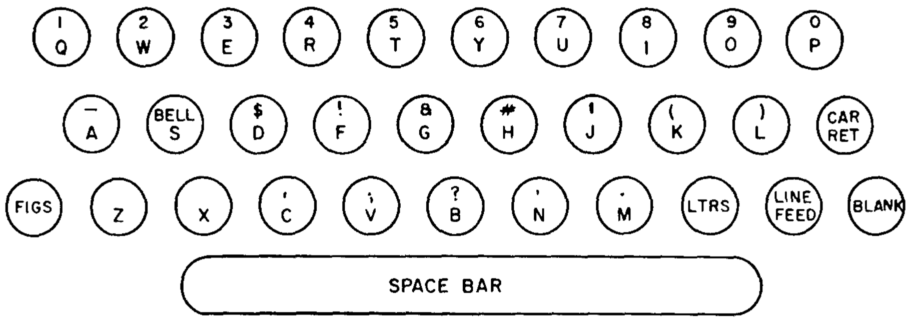


FIGURE 7–3b. Typical three-row teleprinter keyboard as published in Standards (1968a).

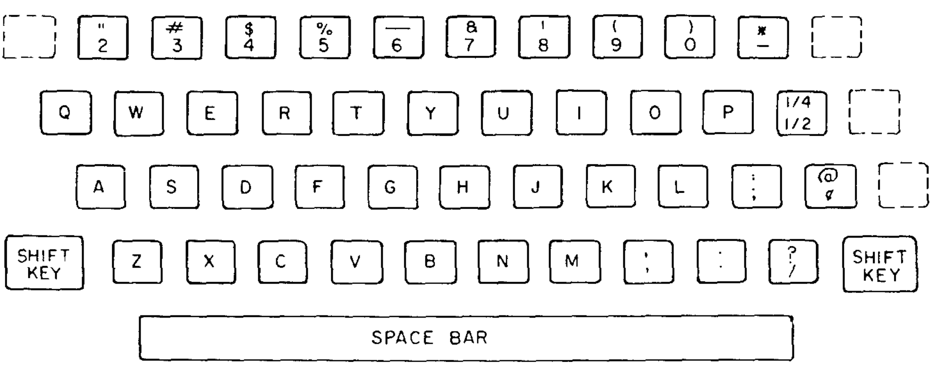


FIGURE 7–3c. Manual typewriter keyboard alternate arrangement, as published in Standards (1968a).

like keyboard and records the information directly on a magnetic tape (or disc). These systems permit the effective utilization of typists, rather than specially trained keypunch operators. They may also permit easier self-detection of errors, easier correction of error, simpler verification of data, and record lengths greater than 80 characters.

Typewriter and Character Reader

Another alternative for data entry incorporates the ordinary typewriter (perhaps with special type font) and an optical character reader. (See Section 7.4.11, "Character Readers," for details.)

Typewriter to Punched Paper Tape

Another important form of data entry utilizes a typewriter like device to simultaneously generate typed output and punched paper tape. These systems are usually less expensive, and permit endless record length. They do not offer easy insertion and removal of units of data, error correction is awkward, and speed of reading into data processing systems is slower than magnetic tape (or disc).

Choosing an Alphanumeric Data Entry Subsystem

Choices amongst these data entry subsystems must take into account the economic aspects of equipment costs themselves as well as human factors considerations, e.g.: (a) ease of self-detected error correction, (b) verification procedures, (c) ability to insert or delete parts of a message, and (d) utilization of pre-job typing training, etc. There are no data for guidance in estimating the many trade off functions involved. (See Section 7.6, "Factors in Selecting Data Entry Systems," for further considerations.)

7.4.2 Numeric Keyboard

Full Matrix vs Ten-Key Keyboard

There are two major ways of arranging keys for numeric data entry. The first is the full matrix keyboard with ten or more columns of keys and each column containing a key for each of the digits 0 through 9. The second is the ten-key keyboard containing just one key for each of the ten digits arranged in a three-by-three matrix, with the zero key either above or below the others.

Since the ten-key arrangement is the only one of the two which permits "touch" operation, there is no question at all concerning a choice between the two for high-volume, production data entry. Minor and Revesman (1962) compared the two for a non-touch data entry task where entries were made relatively infrequently by unskilled keyboard operators. The ten-key arrangement was provided with complete visual indication to the operator of the digits entered (some versions of the ten-key arrangement only provide indication of the number of digits entered). The ten-key arrangement was clearly superior to the full matrix, even for occasional entry by unskilled operators. However, if the ten-key arrangement does not provide complete indication of the digits entered, it *may* not be superior for this kind of task--there are no published data.

Ten-Key Digit Arrangements

There are many digit arrangements possible on a ten-key keyboard. Though many have been explored (Deininger, 1960a, 1960b; Lutz and Chapanis, 1955), only two have been in serious competition with one another. A third, frequently used, arrangement is the two-handed touch operation of the digit keys along the top row of the typewriter keyboard, though no comparison data are available for it in a production-like situation. The two-handed touch operation suggests that it *may* be superior for certain kinds of production tasks. Of the two arrangements currently "in competition," one has the digits 1, 2, and 3 in the top row, with 7, 8, and 9 in the bottom row. This arrangement will be referred to as the "123" arrangement and is found on the common keypunch machine. The other arrangement has the digits 7, 8, and 9 in the top row, with the digits 1, 2, and 3 in the bottom row. This latter arrangement is to be found on common adding machines and will be referred to as the "789" arrangement.

There is little question that the highly practiced data entry operator will perform about equally well with either the 123 or the 789 arrangement. The difference in arrangement becomes critical for the operator who only makes occasional entries, and for the operator who must alternate between different arrangements. Lutz and Chapanis (1955) report that people "expect to find" the 123 arrangement, with the frequency of expectation about five times higher than it is for the 789 arrangement.

Conrad and Hull (1968) compared performance on the two arrangements utilizing housewives as subjects. One group performed with the 123 arrangement, a second group with the 789 arrangement, and a third group frequently alternated between the two arrangements. The 123 group performed slightly faster and significantly more accurately than the 789 group, and both groups performed markedly better than the group that alternated. Entry rates were approximately 1 sec. per digit with error rates approximately 1% or less. Minor and Revesman (1962) report similar entry rates for production workers.

In a study utilizing Post Office clerks, Conrad (1967) reports 0.67 sec. per stroke for the 123 arrangment, with an error rate of 0.55%; the rate of entry for the 789 arrangement was 0.73 sec. per stroke with 1.16% wrong keystrokes. Paul et al., (1965) also report advantage for the 123 arrangement, using air traffic controllers as subjects.

Thus, three different populations of subjects all yield the same conclusion: namely, the 123 arrangement is both faster and more accurate than the 789 arrangement for the relatively unskilled operator.

Since push-button telephones are utilizing the 123 arrangement, all operators may be expected to be familiar with it in the not too distant future. The alternation group of housewives in the Conrad and Hull study points up the disadvantage of expecting operators to alternate between the two arrangements. Thus, the recommendation is made that the adding machine arrangement, i.e., the 789 arrangement, be changed to conform to the telephone and keypunch arrangement: the 123 arrangement. The 123 arrangement is recommended as a standard.

Zero-Key Location on Ten-Key Keyboards

Data relative to the location of the zero key on ten-key keyboards are not easily interpreted. Lutz and Chapanis (1955) report a slightly greater expectation for the zero following the nine rather than preceding the one, but performance data in actual data entry tasks are not available. The keypunch places the zero key above the three key, while the ten-key telephone arrangement places the zero key below the eight key. Performance data in controlled experiments are needed before a firm recommendation can be made.

7.4.3 Chord Keystrokes

Early Work

In 1958 Klemmer trained two subjects to type alphabetic characters on a special ten-key keyboard. There was a key for each of the ten fingers, and each alphabetic character was represented by the simultaneous depression of 2 of the 10 keys. Klemmer concluded that entry speeds were "not out of line with performance in learning to type on a conventional machine." He noted that people were capable of pressing multiple keys almost simultaneously (within 0.03 sec.). Klemmer's work represents the beginning of the investigation of chord keystrokes for for producing correctly spelled English output.

Lockhead and Klemmer (1959) examined an eight-key chord keyboard system for the entry of data. One or two keys per chord were utilized for entering the alphanumeric characters, while chords of from three through seven keys were used to represent 100 common English words. Approximately 23 hours were required to learn chords corresponding to the 100 words and the alphanumeric charactrs.

The 31 Chords of One Hand

Ratz and Richie (1961) and Seibel (1962b) studied the 31 chords possible with the five fingers of one hand, and report very similar rank orderings of the difficulties of those chords. Single key responses are the fastest, but for chords involving two or more keys it is necessray to consider the specific pattern of keys in order to assess the difficulty of the chord. Table 7-1 presents the chords, associated discrimination reaction times (DRT), and error percentages for the highly practiced subjects in the Seibel study.

TABLE 7–1. AVERAGE DRT'S AND PERCENTAGES OF ERROR FOR EACH OF THE 31 PATTERNS

Pattern* (1 2 3 4 5)	DRT (msec)	Error (%)
4	281	5.9
3	285	2.4
1	289	1.8
2	292	5.0
5	294	5.6
1 4	306	3.8
2 3	306	8.8
3 4	306	10.3
1 2	310	6.2
2 3 4	311	9.1
1 3	312	5.0
1 2 3 4	314	4.1
1 2 3	315	5.3
1 5	315	5.6
4 5	316	11.5
2 4	316	12.1
2 3 4 5	317	4.4
1 3 4	320	10.6
3 4 5	321	7.6
1 2 3 4 5	325	7.4
2 5	326	12.4
1 4 5	328	8.2
1 2 4	328	13.2
1 3 4 5	330	12.4
1 2 5	335	11.8
3 5	343	13.2
1 2 3 5	345	18.8
1 3 5	349	15.0
2 4 5	349	20.9
1 2 4 5	351	25.9
2 3 5	352	22.1

Seibel (1962b).
*1 = thumb; 5 = small finger.

It is difficult for the subject to manipulate his fingers to strike some of the 31 chords.

Reaction Time and Number of Different Chords

If the effect of motor difficulty is balanced out, the number of alternative chords involved in a given reaction time task makes little or no difference in the reaction time for numbers of alternatives from 5 through 31 (Seibel, 1962c). The overall average motor difficulty of a set of responses, however, does influence the reaction times for the specific chord responses in the set.

The lack of dependence of reaction time on the number of alternative responses is further emphasized in a report by Seibel (1963a) in which reaction times for 1,023 alternatives are only approximately 25 msec. slower than the reaction times for 31 alternatives, and part of this small difference is attributable to the fact that the larger set contained more difficult chord patterns.

Thus, laboratory data support what is obvious from listening to a piano concert: humans can simultaneously strike several keys in a keyboard and produce a chord. The laboratory data also indicate:

1. That practiced subjects can strike chords within 0.3 to 0.4 sec. after being shown which chord to strike;
2. The relative difficulties of the chords of one hand;
3. Simultaneity can mean less than 30 msec.;
4. The number of different chords (beyond five or so) has little or no effect on speed of response for practiced subjects.

Chord Keystrokes on a Modified Typewriter

Seibel (1962a) reports on a practical system for utilizing chord keystrokes on a modified standard typewriter with two extra shift keys. The system was designed to utilize trained typists who could continue productive work while learning to utilize the additional advantage of the special chord keystroking. The system utilized chord strokes to stand for commonly occurring words, phrases, prefixes, and/or suffixes. Lists of these abbreviations could be entered at from 100 to 120 chord entries per min. This is almost twice the rate reported by Lockhead and Klemmer (1959) for their eight-key word-writing system. The *demonstration* data suggest data entry rates one and one-half times standard typing rates with only 150 special chords.

Mailsorting by Chord Keystroke

The major use of chord keystroke data entry has been in systems designed to facilitate letter sorting in Post Offices in Canada, England, and the United States.

Conrad (1960b and 1960c) reports on a letter-sorting machine involving simultaneous depression of two keys (one for each hand) in order to sort letters into 1 out of 144 possible destinations. Each hand had one of 12 keys to select (arranged in two rows of six keys each). The machine was limited to approximately 110 letter sorts per min., i.e., approximately 0.55 sec. between sorts. After nine months of practice, letter-sorting performance was examined in terms of times between successive chord entries. Most times were close to the 0.55 sec. machine limitation. Sorting rate proceeded from approximately 35 sorts per min. to about 60 sorts per min. over the practice period.

Cornog et al. (1963) and Cornog and Craig (1965) report on various chord keyboards and coding schemes utilized by the U.S. Post Office. Some representative speeds and error rates are presented for one of the keyboards and codes, but no experimental data are reported.

Conrad and Longman (1965) report on a direct experimental comparison of a chord keyboard versus a standard typewriter keyboard for the sorting of mail by postmen who were naive with respect to the use of a typewriter. The keyboard called for the simultaneous (within 0.05 sec.) depression of two keys for each chord, one key affected by each hand. There was a logical sequential pairing of the alphabet to the chords. Postman were trained for 3.5 hours per day, five days per week, for approximately seven weeks. One group was taught to encode mail for sorting while using the chord keyboard while another was taught to do so using the standard typewriter keyboard.

At the end of 33 days of total training, the chord keyboard class was entering an average of approximately 98 strokes per min., about 6 key strokes per min. faster than the average of the typewriter class. The superiority of the chord keyboard was present despite the lack of immediate feedback for operators utilizing the chord system, while operators of the typewriter system did get immediate feedback with respect to what they had keyed. However, the typists were more accurate at the end of practice (10% verus 20% error rate). Only about 1% of the chord keystrokes were in error due to lack of simultaneity.

The chord keyboard system appears to have decided advantages in cases where the operator population has no typewriter training prior to the job though the opposite might be true if the operator population were trained to type prior to starting their job. In contrast, the system described by Seibel (1962a) is specifically designed to take advantage of prior typewriter training.

Stenotyping

An attempt to utilize the phonetically encoded stenotyping system for data entry is described by Galli (1960). A modified Stenotype keyboard was used to feed modified Stenocode into a computer for translation. No experimental data are reported. Drawbacks to the system are related to the extensive training period to learn the modified Stenotype code, and the extensive and sophisticated computer and computer program necessary to translate the somewhat ambiguous phonetic code into correctly spelled English text.

Application and Potential of Chord Keyboards

A variety of keyboards and coding systems have been developed for letter sorting. None stands out as significantly better than others, so far as the published experimental literature reports. There are no other commercially available chord keyboards for data entry.

The potential of chord keyboard data entry is very high. Indications are that entry rates of 150% of standard typing are relatively easy to achieve. It is recommended that chord keyboard systems be explored more fully for data entry purposes, especially in systems where small special-purpose digital computers can be linked with the keyborads to effect immediate translation of the chord keystrokes and produce "on-line" visual feedback of the data being entered.

7.4.4 Levers, Dials, and Special Pushbuttons and Switches

Comparison with Keyboards

The major conclusion to be reached with respect to levers, dials, pushbuttons, and switches for data entry is that keyboards are better. This is true even for unskilled operators making occasional entries. The major comparison study is by Minor and Revesman (1962) in which a ten-key keyboard, a matrix keyboard, levers, and rotary knobs were compared. The ten-key keyboard was best in terms of accuracy and preference, and it was faster than all others except for the full matrix keyboard which was approximately equal in speed of entry.

If the frequency of data entry and/or the skill of the data entry operator calls for "touch" operation, the ten-key keyboard is to be preferred for the entry of numeric data, and the typewriterlike keyboard for the entry of alphabetic information. Only in those cases where design or economy considerations rule out a keyboard should consideration be given to other forms of pushbuttons, switches, dials, and levers for the entry of digital or numerical data.

Keyboard Overlays and Modular Keys

A great variety of switches, dials, etc. are available on the commercial market. (See Hillenbrand 1967.) A recently announced system incorporates a book of keyboard overlays with coded contacts actuated as the pages in the book are placed over the keys. With this overlay system, relatively few keys may control many different functions as the different page overlays are used. Other recent announcements involve modular keys to be arranged in patterns for panels and keyboards according to design requirements. By means of interrupted light beams, magnetically actuated reed switches, and/or magnetically actuated solid-state circuitry, each modular key can create its own coded output, as required by the overall system design.

Dials and Thumb wheel Switches

Several studies have explored data entry by means of devices other than keyboards. Conrad (1958) found the rotary telephone dial less accurate than a ten-key pushbutton keyset if the data message had to be held in memory until entered. This effect is attributable to the delay introduced by the return of the dial before the next digit can be entered. Deininger (1967) compared the rotary telephone dial with ten thumb wheel switches for entering phone numbers and found the dial to be faster, and preferred, for random sequences of phone numbers. Plath and Kolesnik (1967) examined the use of thumb wheel switches for entering navigational coordinates in an aircraft. Entries took approximately 2.75 sec. per wheel setting, with approximately a 2% error rate. The wearing of flight gloves made no significant difference.

Pushbutton Keyset

Deininger (1960a and b) explored the entry of phone numbers on a telephone pushbutton keyset. Little or no difference in performance was found as a function of the size, force-displacement, presence of snap action, and presence of auditory feedback. (See "Keystroke Touch and Stiffness," in Section 7.5.3 for further discussion.) There are no design-relevant data.

Applications of Levers, Dials, Pushbuttons and Switches

Use keyboards wherever possible in preference to levers, dials, pushbuttons, and switches for the entry of numeric and/or alphabetic data. For unskilled operators and/or occasional entry of information it is desirable to provide a visual indication of the characters entered so that the operator may sight check the data prior to its actual entry into the data processing system. In situations such as the telephone where the entire message is not stored in the terminal prior to "entry," visual feedback is apparently unnecessary. Recommendations concerning visual feedback are based on "educated guesses"; there are no data.

7.4.5 Consoles for Data Entry

General Considerations

Consoles for data entry are relatively new. Each is typically designed for a specific job or class of jobs. There are too many varieties and variations possible both in experimental and commercially available units to permit a meaningful classification and recommendation. Almost without exception, the results of any study comparing two or more console configurations must be interpreted within the limits set by the task for which the system was designed, and by the task used to evaluate human performance at the console. What is needed are several "standard" tasks which are designed to represent the spectrum of tasks for which consoles are expected to be used. To the extent that investigators incorporate these standard tasks into their experimental evaluations, would comparisons of performance data be meaningful for making evaluations amongst the console configurations. The design of such "standard" tasks is *not* an easy job.

Several reports (Barmack and Sinaiko, 1966; Devoe, 1967; Devoe et al., 1966; and Technology Profile, 1968) describe many commercially available, and experimental, data entry consoles. Typical applications are also described in Technology Profile (1968). Dolotta and Selfridge (1968) provide a listing of features for a typewriterlike time-sharing terminal, with accent on the user's (programmer) point of view. Though they provide no performance data, their arguments and opinions are convincing. Suppes (1966) describes some of the features he considers desirable for consoles, in systems designed for computer-assisted instruction. While Dolotta and Selfridge concentrated on the features for a typewriterlike terminal, Suppes indicates that a console terminal limited to a typewriterlike mechanism would be just a minimal console for his purposes. He favors analog, graphic, and voice inputs to the system and outputs to the operator. Siders (1967) provides similar descriptive information for console features for active interaction between computer and designer in carrying out engineering design tasks.

Semling (1968) describes the data processing system at the Internal Revenue Service (IRS), where the system currently punches and transcribes to magnetic tape more than 600 million punched cards each year. He reports that there are about 500 keypunch machines at each of seven regional IRS centers. An experimental system is currently under evaluation which includes a cathode ray tube and keyboard console along with additional hardware to enter and verify the volume of data directly to magnetic tape. Though this appears to be an ideal situation for a careful performance evaluation of competitive systems of data entry, once again, no performance data are available.

Hand Printing vs. Keyboard Entry for Console Tasks

Performance data for a simulated console task are reported by Devoe (1967). A task involving the making of measurements on a drawing, table look up, computations, and data entry was studied with respect to the relative desirability of hand printing the data entries on printed forms versus making the entries via a typewriterlike keyboard. Keying was only slightly slower than hand printing for the non-typist subjects. Practice would presumably eliminate that difference in less than a week of full-time work. Errors were relatively few, not systematically distributed, and not reported. The need to keypunch the hand-printed material adds a further time delay for that mode of data entry, and in somes cases this is a critical delay before the system can interact with the console operator. It is important to note that in this study data entry was a small though crucial part of the total task performed. In this respect it is the only study of its kind in the literature.

Devoe and Graham (1968) further explored the possibility of utilizing hand-printed entry of data at a console. Operators made printed entries on paper placed over a tabletlike entry device. This gave information to the computer analysis system with respect to the sequence of printed strokes as well as to their form, making machine recognition of the capital alphabetic characters and numbers considerably simpler. The character recognition task was made still simpler by imposing certain constraints on the shapes and sequences of strokes for each of the characters which were acceptable. The authors conclude, "With relatively little practice, most subjects were able to learn all constraint sets and to copy difficult messages with reasonably good speed and accuracy." The data indicate printing rates just under 30 characters per min. with about a 3% error rate for the on-line hand printing. The incorporation of a correction cycle reduces the error rate close to zero, with an entry rate of about 20 characters per min. overall. The authors recommend on-line hand printing as a mode of data entry for tasks calling for occasional data entry on the part of the operator. An important feature of hand-printed data entry is the fact that the entry may be made with just one hand.

Barmack and Sinaiko (1966) mention a five-button keyboard developed by Englebart for one-hand entry of alphabetic information. They report data for one subject (Englebart) who eventually reached 35 w.p.m. with his right hand. Dvorak (1950) also describes a one-hand keyboard, but reports no performance data.

Whether or not a typewriterlike keyboard is an appropriate data entry device for a console depends in part at least on the proportion of people in the user population who know how to type. A small informal survey by Barmack and Sinaiko (1966) indicates that about 75% of professional scientists and engineers have a "working familiarity with a standard typewriter."

It is recommended that in most cases a typewriterlike keyboard be utilized if the volume of data entry is at all heavy. Further, the keyboard should conform to the recommended typewriter standard shown in Figure 7-3a. However, in those situations where entries are relatively infrequent, or are a small part of the total task being performed, or where it is important to keep one hand free, it is strongly suggested that consideration be given to on-line hand-printed data entry. This set of conditions however, also suggests voice entry as an ideal mode of data entry. When practical systems are developed for voice entry, this mode should also be evaluated. (See Section 7.4.7, "Handwritten Inputs," for further information.)

7.4.6 Graphic and Analog Inputs

There are many hardware and programming systems for achieving graphic and analog inputs to data processing systems (e.g., Barmack and Sinaiko, 1966; English et al., 1967; Licklider, 1960; Ridinger, 1967; Siders, 1967; Suppes, 1966; Sutherland, 1966a, 1966b). The March 1967 issue of *IEEE Transactions on Human Factors in Electronics* (HFE 8) is devoted to man-computer, input-output techniques, and many of the articles contain descriptions of on-going work and developments.

Some of the devices utilized for entering analog information are described by English, et al., (1967) and include a joystick, the Grafacon, a "mouse," a knee control, and a light pen. The "mouse" is a small box which is moved by the operator on a tablelike surface in order to transmit the X-Y coordinates for the cursor. The knee control uses a left-right and an up-down movement for the X and Y coordinates. These devices were compared in a task calling for the manipulation of a cursor in a text manipulation system. The results, however, are specific to the particular configurations used.

While devices for graphic and analog entry are many, comparative data are few. Compatibility and spatial correspondence between display and control movements are the only general principles which appear to be supported with data (e.g., Earl and Goff, 1965; Leonard, 1964).

7.4.7 Handwritten Inputs

Comparison Among Modes

Several studies have examined characteristics of handwritten and printed characters for data entry. For example, Devoe (1967) reports free cursive handwriting at a rate of approximately 80 characters per min., typing by non-typists at about 50 characters per min., unconstrained printing at about 60 characters per min., and constrained printing at about 35 characters per min. All rates but cursive handwriting were still showing improvement at the end of the experiment. Errors were at the 1% level or less at the end of the experiment except for the constrained printing where it was about 3% (almost entirely due to lack of conformity with constraints). (See Section 7.4.5, "Handprinting vs. Keyboard Entry for Console Tasks" for additional information and comparisons.)

In another experiment (Devoe, 1967), subjects printed random numbers at about 100 per min. and random letters at about 75 per min., where each character was printed in a ½-in. block on a specially provided form. These entry rates may be compared with those reported by Hirsch et al. (1960) for constrained numeric printing. The digits were printed at about 20 per min., with an error rate of 0.32% for automatic recognition, but an error rate of only 0.11% attributable to incorrect or transposed digits. The comparison between the two reports shows very large differences in rate of printing and in rate of error production. The Hirsch et al. experimental conditions obviously placed greater emphasis on careful and correct printing. Conrad and Hull (1967) report the copying of digits at approximately 90 to 100 digits per min. with no constraints on the printing, and between 0.1% and 0.3% of characters in error. Thus, error rates are comparable in the Conrad and Hull study and in the Hirsch et al. study. The constraints for the Hirsch et al. study very markedly reduce the rate of printing. More powerful character recognition systems, permitting less constraint in printing, should permit higher rates of hand printed data entry.

Errors in Handwriting

While overall production and error rates are of interest for system design, more detail is needed with respect to the errors that are made in order to guide the design of character recognition systems, and error detecting and correcting codes. The first feature of the error data worth noting is that a few operators produce a large proportion of the total number of errors observed. Hirsch et al. (1960), for example, report that three out of their 100 subjects accounted for 22% of the errors made, and these subjects tended to be consistent in the kind of error they did make. McArthur (1965) and Crook and Kellogg (1963) report similar results. Identifying error-prone operators can reduce overall error rates.

The frequency with which a given class of error is found is heavily dependent upon the details of the copying task. Conrad and Hull (1967) report, e.g., where short-term memory is not involved, transposition errors account for about 5% of the total errors, while they account for more than 20% of the errors if short-term memory is involved. If the copying task is such that the operator may easily lose his place in the original message, then as many of 50% of the errors will be the omission of one of the digits, while in cases more typical, this percentage is usually between 5 and 10%. The most frequent error, by far, is the substitution of one character for another. This may run as high as 80% of the total digit errors when strings of digits are copied directly below those presented.

When capital alphabetic characters are included in the material to be hand printed, substitution errors are by far the most frequent, accounting for more than 80% of the errors in the Conrad and Hull study. Illegible and omitted characters are next in frequency. Chapdelaine (1963) and McArthur (1965) report data for similar tasks.

Substitution errors may be further analyzed in terms of which characters are substituted. McArthur (1965) reports 0 (zero), 8, B, D, I, O (Oh), and Z as the characters which most frequently had other characters substituted for them. These 7 characters accounted for 72% of the substitution errors, with 8, I, and O (Oh) as the most frequent three. Characters which were substituted for these eight characters were almost always "iconic" (i.e., "look-alike"): e.g., O (Oh), 6, D, and U for 0 (zero); B for 8; etc. These data are important but are difficult to place in context because there are no reports of the rates at which the letters and numbers were transcribed, i.e., only error data are reported.

Off-Line Handprinted Input

At least two different automatic off-line handprinted character recognition devices have been described in the literature (Greanias et al., 1963; Simek and Tunis, 1967). Recognition rates greater than 99% of the digits written are reported. Commercial units for reading hand printed digits (and several alphabetic characters as well) were announced as early as 1966.

7.4.8 Mark-Sensed Answer Sheets

The only performance data comparing mark-sensed answer sheet marking with other forms of data entry are reported by Devoe (1967). While hand printing led to entry rates of 100 digits per min. or 75 alphabetic characters per min., equivalent entries on mark-sensed answer sheets were 65 per min. and 15 per min.

Mark-sensed answer sheets should not be used as a routine form of data entry for either numeric or alphabetic information. However, in situations where each of a great many operators are to enter a relatively small amount of data, the mark-sensed answer sheet may be a desirable mode for data entry. When the automatic recognition of hand-printed characters becomes more economical and readily available, it is anticipated that the need for the mark-sensed form of data entry will be completely eliminated.

7.4.9 Preperforated Punched Cards and Mark-Sensed Cards

The only performance data available are in a study by Kolesnik and Teel (1965) in which navigational data were entered onto cards by means of three simple manual methods. The results suggest that preperforated punched cards or mark-sense cards are preferable to thumb wheels, push buttons, or hand-punched cards. Improved availability and cost of automatic recognition devices for hand printed characters should lead to a preference for that mode of entry in most situations which utilize the prepunched card or mark-sensed card.

7.4.10 Acoustic Input

Experimental acoustic recognition systems for the ten digits (and several additional words) have been demonstrated with at least 98% recognition capability (Licklider, 1960; King and Tunis, 1966.) Commercial systems are not yet available. Assuming that acoustic recognition systems are feasible and practical, under what conditions would they be desirable as means for data input?

Braunstein and Anderson (1959) studied the relative speed and accuracy of reading digits aloud and keypunching digits for five experimental subjects with no prior training. Subjects read digits at about twice the speed at which they could keypunch, even after several hours of practice; but the keypunching was an easier task. When given the option, four of the five subjects chose to punch rather than read the 1,050 digits which had to be entered. Subjects averaged a reading rate between 2.5 and 3 digits per sec., and a keypunching rate between 1 and 1.5 digits per sec. Reading errors were not very different in number from keypunching errors. The reading was a more tiring task for all of the subjects, but it is uncertain whether additional training (e.g., in breath control) would change this. A trained keypuncher performing the same task entered digits at 2.8 per sec., and this was probably slightly below her usual rate due to a difference in machines. The authors concluded that voice input does not offer speed or accuracy advantages over conventional keypunching by experienced operators. Thus, vocal entry of digital data appears to be highly desirable only where the amount of data to be entered is relatively small and the frequency of entry is low. If the volume of data entry is large, some form of keyboard entry is preferable. Performance data for data entry of larger vocabularies are not available. The reading of words might be both a preferred and more accurate mode for alphabetic data entry, but there are no performance data nor are there devices readily available for accurately recognizing a large acoustic vocabulary.

7.4.11 Character Readers

Character recognition machines are ideally suited for reading straight typed copy, where each page of copy is a loose unit, i.e., a page or card. The recognition system becomes less complex and less expensive if all the data appear in a *single stylized* type font. A special type font has been suggested as a standard (Standards, 1968b). Newer and more powerful units, however, are able to read as many as 200 different type fonts, for example, and a standard *stylized* font appears superfluous. Character readers still have problems, however, in dealing with bound volumes, illustrations, citations, and footnotes. Thus, completely automatic read-in of periodicals and books is not yet feasible with commercially available units.

The character reader coupled with a team of typewriter operators is competitive with other high-volume data entry systems. Helweg (1962) estimated a break-even point at approximately 10,000 pages per day in comparing key-punching with character reading at that time. Even at that point, he indicates, the character recognition equipment would provide higher speed and greater accuracy.

Fein (1967) reports that the optical reader used at the Social Security Administration has replaced 150 keypunch operators, and is reported to have an error rate of better than 1 error in 1,728 characters.

If the volume of raw data to be entered into the computer system is very large, and is available in typed or printed form on separate sheets of paper or card, an optical character recognition system may appear warranted to achieve higher entry rates and greater accuracy. Character recognition devices for hand-printed, alphanumeric characters are currently in experimental development. For digital data they are already available commercially. Thus, even hand-printed data may be read directly into computer systems. However, the current character recognition systems are quite expensive, and alternative means of data entry may be preferable for those situations involving relatively small volumes of data. (See Section 7.6.4 for additional comments.)

7.5 Data Entry Rates

7.5.1 Production

High-Volume Entry

For high-volume data entry of redundant data such as English text, a summary by Devoe (1967) reports that a typewriterlike keyboard provides the highest *character* entry rate, with speed test rates of 60 words (300 characters) per min. quite common, 100 w.p.m. more or less the upper limit likely to be found in production situations, and championship speed approaching 150 w.p.m. (See Figure 7-4.) These figures should be divided by approximately two to estimate daily production (as contrasted with speed-test performance). Experimental "abbreviation" typewriterlike systems (e.g., Klemmer, 1958; Lockhead and Klemmer, 1959; Seibel, 1962a) promise somewhat higher *information* entry rates, but commercial units are not available. The steno-type system provides the highest known *information* entry rate for English text, but it produces a phonetically encoded output which must then be translated to correctly spelled English (requiring sophisticated hardware and programming). The entry of straight numeric data is somewhat slower than that for highly redundant English text, primarily because numbers have close to a random distribution.

Keypunching

Considering many different data entry jobs (and several different keypunch installations), a rate of 170 characters per min. (2.8 per sec.) is a good estimate (Klemmer and Lockhead, 1960, 1962a, 1962b) of the mean rate of *daily* entry for time "spent at the machine," with an estimated range from 127 to 206 characters per min. including the middle 95% of the keypunch operators. Better operators will produce a daily average of more than 250 characters per min. for some (easier) data entry jobs. Klemmer and Lockhead (1960) report that there are no consistent differences between alphanumeric and straight numeric keypunching; hence, these production figures may be considered to apply to either. Straight English text "should" yield somewhat higher rates, but there are no data.

Typing

Droege and Hill (1961) report that typists, doing straight-copy 10-min. tests, average about 326 characters per min. (about 5.4 per sec., or 65.3 w.p.m.), with a range of 215 to 435 characters per min. (43 to 87 w.p.m.) including the middle 95% of the typists (and 36 to 94 w.p.m. for the middle 99%). Adjusting for *daily* production by dividing by "approximately two" indicates typing and keypunching have very similar daily production rates. However, daily production typing figures are not available, and it is not possible to determine whether "two" is the correct number by which to divide for estimating daily typing production. "Approximately two" (1.8) does work for keypunching and 20-min. to 30-min. speed tests.

Production Rate and Error Correction

The ease with which self-detected errors may be corrected, can have a marked influence on overall production rate. The common keypunch is notoriously awkward in this respect, as is the typewriter. The newer keyboard to-tape systems (and other similar systems) are decidedly better in this respect, and in terms of correcting errors found by the verification process. Comparative data are not available.

Unskilled Operator Production

The entry of data by relatively unskilled operators making occasional entries varies from approximately 100 digits per min. for unconstrained hand-printed numbers down to approximately 20 alphanumeric characters per min. for constrained hand printing with an error-correction cycle to reduce errors close to zero. Typewriter and 10-key numeric keyboards (used by unskilled operators) provide entry rates which are approximately equivalent to hand printing, and perhaps slightly better if one takes into account the

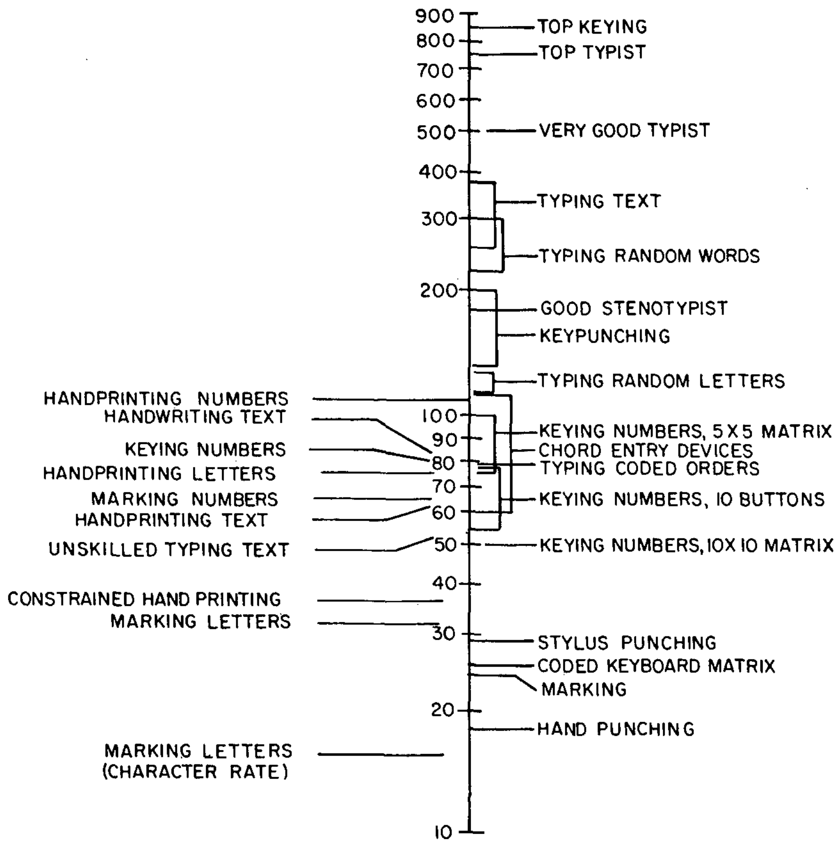


FIGURE 7-4. Representative manual entry rates (adapted from Devoe, 1967).

necessary constraints to handwritten material in order for it to be machine readable. The hand printing however, requires only a pencil (or its equivalent) and the use of just one hand.

Figure 7-4 shows representative manual entry rates for various types of devices in terms of characters entered per minute.

7.5.2 Errors

Keypunching

Production data (Klemmer and Lockhead, 1960, 1962a, 1962b) indicate average error rates for different keypunch installations ranging between 1600 and 4300 keystrokes per undetected error (0.2% and 0.06% keystrokes in error). Self-detected, and self-corrected, errors ("restarts") are four or more times as frequent. Whether most operators make many errors and a few are very accurate, or most operators are quite accurate and a few make many errors, depends on the error measure used. Adjusting for differences in measures, the logarithm of percent keystrokes in error suggests an approximately symmetrical distribution about the average error rate for a given installation, with a few very accurate operators.

Distributions of operator error rates for different installations, and groups of operators at grossly different skill levels, are essentially equivalent except for differences in median error rates. The distributions indicate that the operators at the high end of the distribution (90th percentile) make six to ten times as many errors as the operators at the other end of the distribution (10th percentile). The higher an operator's average error rate, the more variable will be her error rate from week-to-week (essentially a linear relationship). The fastest operators tend to be the most accurate, though the relationship is not very strong (correlation of 0.4 to 0.5). Operators continue to show improvement in error rates for at least a year or two on the job.

Approximately 70% of keypunching errors (Klemmer and Lockhead, 1962b) involve a single character. Two or more characters were transposed for about 15% of the errors, a character was omitted for 4% of the errors, an extra character was inserted for 1% of the errors, and procedural errors accounted for the remaining 10% of the errors. Approximately 80% of the single-character numeric errors were "aiming errors" in that adjacent keys (horizontally or vertically) were struck instead of the intended one. Westhoff (1962, as described in Conrad, 1967) reports a similar high frequency for aiming errors.

Typewriting

Typewriting data analyses (e.g., Flynn, 1959) suggest similar aiming errors as most frequent, some "opposite hand" errors, omit errors, and some other particular substitutions of one letter for another. These studies also indicate that the frequency with which a letter or word is involved in an error is very closely related to the frequency of use of that letter or word in the English language.

Effects of Memory Load

If data are held in short-term memory, and entered from memory, then the aiming errors tend to disappear and memory errors become more frequent. Memory errors involve "chunk" confusions, "acoustic" confusions, and forgetting of central and latter parts of messages.

Verification, Proofreading, and Error Correction Procedures

Verification and error correction procedures are much more convenient if the correction procedure does not involve duplicating a portion of the data and the removal of a faulty data card. A simple back up and write over is the preferred mode of error correction.

Some data entry systems call for proofreading rather than verifying. In proofreading, certain omitted letters are more likely to be overlooked than others; missing letters which *are not* pronounced are much more likely not to be detected than are missing letters which *are* pronounced; the probability of failing to detect a missing silent letter is greater when the letter occupies positions early in a word than it is when a letter occupies positions toward or at the end of the word. The final pronounced "e" in the word "the" is a special case with a high probability of of failure to detect the missing letter. (See Corcoran, 1967, and Corcoran and Weening, 1968.)

There are no data giving comparative error detection rates for verification versus proofreading procedures.

Damerau (1962) reports on a computer program to detect and correct common data entry spelling errors in English text. For one test sample, better than 90% of the misspelled words were identified. Thus, a computer could be used to compensate for a very large percentage of data entry errors. Whether or not to use it is a question of cost. Commercial packages are not available on the market.

The commonly used "*net* words per min." score for combining speed and error rates is based on very untenable assumptions (Hirsch, 1958), and should not be used. Better combination scores have been developed, but gross speed and error data should *always* be reported in addition to any combination score that may be used.

Keystroke Timing and Key Interlocks

Inter-keystroke intervals of 50 msec., and less, will be found for skilled operators in high-volume data entry tasks (Fox and Stansfield, 1964). The present author *estimates* that enforced inter-keystroke intervals of 10 msec. or less will cause essentially no detrimental effects, while enforced inter-keystroke intervals of 100 msec. or greater will most certainly cause detrimental effects. There are no data adequate to the task of providing the design engineer with the necessary trade-off functions for enforced inter-keystroke intervals between 10 and 100 msec. "Armchair estimates" do not appear to be advisable. (See Section 7.3.4., "Delay reduces speed of typewriting.")

Effects of Training and Time on the Job

Performance on data entry tasks shows improvement with practice over periods at least two years long (Klemmer and Lockhead, 1960, 1962a, 1962b), and estimates of high-volume data entry production should not be based on operator performance unless at least 6 months of practice on the particular task has been taken into account. There are no *reliable* "short-cuts" or "cheap ways out" (Seibel, 1964b).

The actual mode of operation and abilities involved in performing data entry tasks, shifts with practice and improvement in performance (Fleishman, 1960 and 1965; Leonard and Newman, 1964). At higher levels of skill, data entry operators are utilizing the redundancies in the input data to encode that data into "higher order units" which are then processed as units. This is to be contrasted with the character-by-character mode of entry which is typical of the operator at a low level of skill, early in practice. Thus, the same prediction and screening tests which worked for low levels of performance (early in practice) will probably not work for higher levels of performance (later in practice), and the effects of variables which influence data entry performance at low levels of skill will most likely not act in the same way in influencing performance at higher levels of skill (e.g., response speed as a function of information per response). *Reliable* and *valid* data for highly skilled high-volume data entry performance are expensive to obtain.

7.5.3 Effects of Special Modifications in Entry Devices

Keystroke Storage

Though single keystroke storage is now available in many commercial keyboards, there are no published data known to the present author which compare the single stroke storage units with other units in terms of data entry performance.

Erase and Delete Actions

Erase and delete features for data entry tasks are reported to be highly desirable (e.g., Semling, 1968), but no experimental data are available in the open literature. Since operators at higher levels of skill tend to work with higher order units of the data, e.g., words rather than single characters, it is reasonable to expect a "word erase feature" would be highly desirable. At lower levels of skill, or for the entry of data with very little redundancy (e.g., the usual numerical data) a single character erase feature *may* be more desirable. Data are not available in the open literature.

Visual and Auditory Feedback

Visual and auditory feedback appear to be unnecessary for the highly skilled high-volume data entry operator performing under speed-test conditions (e.g., Diehl and Seibel, 1962). However, other studies have shown that the removal of visual feedback does have a detrimental effect on the number of errors which the data entry operator makes, or fails to detect (e.g., Devoe et al., 1966; Devoe, 1967; and West, 1967). These and two studies by Chase, Harvey et al. (1961) and Chase, Rapin et al., (1961), lend support to the conclusion that the major source of feedback for the highly skilled operator is the kinesthetic-proprioceptive-tactual feedback which the operator gets from actually making the movement and striking the key. The West data suggest that reliance on these proprioceptive cues increases with skill level, but even at high levels of skill the operators utilize visual feedback for further reductions in errors. Thus, the removal of auditory feedback may be expected to have relatively little effect on performance, but the removal of visual feedback can be expected to lead to higher error rates.

Keystroke Touch and Stiffness

There are no data in the open literature with respect to keystroke touch, stiffness, and "feel" on high-volume data entry tasks.

Descriptions of the force-displacement characteristics of some commercially available switches and push buttons are found in Pollock and Gildner (1963), but the effects of variations in force-displacement characteristics on data entry performance are not available.

Two studies by Deininger (1960a and 1960b) describe force-displacement characteristics for keys to be used in the entering of telephone numbers, and occasional-entry performance data indicate no difference as a function of the force-displacement variations.

In general, most keyboards exhibit an increasing amount of force required as the key is depressed, with a marked drop-off in necessary force as the entry (typing) mechanism is tripped, and then a relatively sharp further increase in force requirement as the key is "bottomed." Most electrically operated keyboards call for a relatively light touch in order to trip the entry mechanism, frequently in the range of 2 to 3 oz. There are no performance data to indicate the effects of increasing, or decreasing, this force requirement; nor are there data to indicate the effects of changing the characteristics of the force-displacement function, or its variations under dynamic conditions.

7.6 Factors in Selecting Data Entry Systems

7.6.1 Training and Type of Keyboard

The highest levels of high-volume data entry performance are achieved only by extensively trained and highly skilled data entry operators. If this training is done by the data entry employer, it often involves very extensive investments in time before high rates of production are achieved. To the extent that typing skill can be employed in the data entry task, the data entry employer saves a great deal of time and money, as most applicants have already been taught to type.

Thus, data entry devices should be as much like typewriters as possible. Even for the occasional entry of data, the majority of high school graduates in the United States has some familiarity with the standard typewriter keyboard, and thus would be better able to use it than to use some other arrangement.

If special non-typing populations of operators are to be considered (e.g., postal mail-sorting clerks), then other forms of keyboard arrangements may be defensible (e.g., Conrad, 1960b and 1960c).

Entry rates higher than those which may be achieved with a standard typewriterlike device can be achieved with some chord keystroke devices. Typically, such chord keystroke devices require more extensive training periods in order to achieve this higher data entry rate. Thus, if personnel turnover is expected to be high, the investment in additional training may not be warranted. A compromise system, which is a variation of a standard typewriter keyboard, and which permits useful data entry production while the chord strokes are being learned, appears to be a reasonable way to overcome this "bind" (Seibel, 1962a and 1964a).

7.6.2 Instructions and Incentives

Instructions and incentives appear to have a relatively potent effect on the production of data entry operators. Pay, or bonuses above basic pay, which are related to level of production, should lead to higher overall production. For a variety of reasons this does not always work out that way. A major shift from this "operator-is-a-machine-component" attitude has recently been reported (The Christian Science Monitor, 1968). The report indicates that the output of key punch operators can be radically improved by "enriching" the job. Until recently "the key punch girl punched information onto a data card. Then usually her work was 100% verified. . ." In an experimental installation, operators are "given responsibility for a certain task (say, the payroll of one department) and are recognized as experts in this job. Any mistakes are fed back directly to the girl instead of to a supervisor. The key punch operator may decide, only if she likes, to have her work sample verified. She schedules her down day" (The Christian Science Monitor, 1968). With this arrangement, output was up, there was no loss in accuracy, and absenteeism and turnover dropped. The results may be reflecting only the classic "Hawthorne effect" but, if the phenomena are "real", the potential pay-off is very great. More careful evaluation is needed.

7.6.3 Machine Pacing

On occasion, incentive or machine-design considerations have led to the utilization of a machine-paced data entry station (e.g., mail-sorting machines). Machine pacing should never be used as an incentive device. And, permitting self-pacing by the operator, without artificial delays or speed-up attempts, is worth a considerable amount of additional engineering effort and cost, as it will lead to sizeable increments in throughput. (See Section 7.3.4, "Machine pacing reduces speed of letter sorting.)

7.6.4 Keypunch, Paper Tape, Magnetic Tape Typewriter, and Character Recognition

Four major equipment systems are competing in the current market for high-volume data entry. The popular key punch, and somewhat less popular punched paper tape, appear to be losing ground to newer systems in which data are transferred from keyboard to some form of magnetic tape or disc storage.

A fourth system is the production of typewritten material followed by automatic character recognition equipment. One characteristic of the keyboard-to-magnetic-storage systems which is an important advantage, is the feature which allows for rapid and easy correction of self-detected and of verification-detected errors. If the *original* data source is in the form of typewritten material, with relatively clean and error-free copy, then character recognition appears to be highly promising. If the original data are handwritten, scattered here and there, and/or otherwise irregular and calling for organization and interpretation by the data entry operator, then the keyboard-to-magnetic-storage techniques appear to represent the best of the systems currently available.

The present author suggests that it may be highly advantageous to use small special purpose computers for on-line data entry so that abbreviations, standard formating, phrase writing, and certain levels of editing, and error correction, can be accomplished at the point of data entry.

7.6.5 Special Abbreviation Codes

The utilization of special "abbreviation" codes, with either standard or special chord keyboards, involves an additional set of trade offs and compromises. Several of the coding schemes which have been developed for the sorting of the U. S. mail are described by Cornog and Craig (1965) and Cornog et al. (1963). Seibel (1962a and 1964a) discusses the variables and trade offs involved. They include frequency of usage of each "abbreviation," number of keyboard strokes involved and/or the motor difficulty of the chord entry, and the size of the abbreviation vocabulary. Decisions with respect to these variables are certainly expected to affect throughput and training time, but the trade off functions are largely unknown. One reasonably well established principle, however, is that the more information transmitted with each keyboard entry, the faster will information be entered.

7.6.6 Occasional Data Entry

For situations involving occasional data entry, there are two major competing systems to be considered. The first is the utilization of the standard keyboard arrangements which are used in the high-volume data entry situation. The second is the utilization of handwritten characters and on-line character recognition devices, or off-line character recognition devices. In situations where the keyboard is a major distraction and inconvenience (e.g., man–computer interacting systems for problem solving, design, editing, etc.), the use of handwritten characters for data entry appears highly desirable. The current expense of the character recognition devices and systems limits their current usefulness. Where the keyboardlike device is not a major distractor, it is preferred, as it is currently available, relatively cheap, and can produce highly accurate and relatively rapid data entry. A third approach for the occasional entry of data is via voice. Experimental systems have been built for recognizing the spoken digits and some additional vocabulary, but commercial systems are not yet available. This method of data entry should prove least distracting of all, and leaves both hands free. It is currently limited to small vocabularies and experimental devices, and is definitely not preferred by the operator if there is any appreciable volume of data to be entered, but it should prove highly desirable for man-computer interacting systems if and when it becomes commercially available.

References

Barmack, J. E., & Sinaiko, H. W. Human factors problems in computer-generated graphic displays. Study S-234, Report No. IDA/HQ 66-4820, April 1966, Institute for Defense Analysis, Research and Engineering Support Division, Washington, D.C.

Braunstein, M., & Anderson, N. S. A comparison of reading digits aloud and keypunching. Research Memo No. RC-185, November 1959, IBM Research Center, Yorktown Heights, New York.

Chapdelaine, P. A. Accuracy control in source data collection. Report No. AF-WP-O-JUL 63 475, AFSC Applied Research Objective, 73OJ, Task 152306, Air Force Logistics Command, Data Management Division (MCCD), Research Branch (MCCDR).

Chase, R. A. Harvey, S., Standfast, S., Rapin, I., & Sutton, S. Studies on sensory feedback. I: Effect of delayed auditory feedback on speech and keytapping. *Quarterly Journal of Experimental Psychology*, 1961. **13,** 141–152.

Chase, R. A., Rapin, I., Gilden, L., Suton, S., Guilfoyle, G. Studies on sensory feedback. II: Sensory feedback influences on keytapping motor tasks. *Quarterly Journal of Experimental Psychology*, 1961, **13,** 153–167.

The Christian Science Monitor, "Focus on business, What's ahead." Eastern edition, Boston, Weekend Issue, May 11–13, 1968.

Conrad, R. Accuracy of recall using keyset and telephone dial, and the effect of a prefix digit. *Journal of Applied Psychology*, 1958, **42,** 285–288.

Conrad, R. Ergonomics in the Post Office, the design of keyboards for non-language material. Report No. APU 406/60, September 1960(a).

Conrad, R. Experimental psychology in the field of telecommunications. *Ergonomics*, 1960 (b) **3,** 289–295.

Conrad, R. Letter sorting machine-paced, 'lagged' or unpaced? *Ergonomics*, 1960 (c), **3,** 149–157.

Conrad, R. Practice, familiarity and reading rate for words and nonsense syllables. *Quarterly Journal of Experimental Psychology*, 1962 (a), **14,** 71–76.

Conrad, R. The location of figures in alphanumeric codes. *Ergonomics*, 1962 (b), **5,** 403–406.

Conrad, R. Acoustic confusions and memory span for words. *Nature*, 1963, **197,** (4871), 1029–1030.

Conrad, R. Acoustic confusions in immediate memory. *British Journal of Psychology*, 1964, **55,** 75–84.

Conrad, R. Short-term memory factor in the design of data-entry keyboards: an interface between short-term memory and S-R compatibility. *Journal of Applied Psychology*, 1966, **50** (5), 353–356.

Conrad, R. Performance with different pushbutton arrangements. *HET PTT-BEDRIJF DEEL*, 1967, **15,** 110–113.

Conrad, R., & Hull, A. J. Copying alpha and numeric codes by hand: An experimental study. *Journal of Applied Psychology*, 1967, **51** (5), 444–448.

Conrad, R., & Hull, A. J. The preferred layout for numeral data-entry keysets. *Ergonomics*, 1968, **11** (2), 165–174.

Conrad, R., & Longman, D. S. A. Standard typewriter versus chord keyboard–an experimental comparison. *Ergonomics*, 1965, **8,** 77–88.

Corcoran, D. W. J. Acoustic factor in proof reading. *Nature*, 1967, **214,** 851–852.

Corcoran, D. W. J., & Weening, D. L. Acoustic factors in visual search. *Quarterly Journal of Experimental Psychology*, 1968, **20,** 83–85.

Cornog, J. R., & Craig, J. C. Keyboards and coding systems under consideration for use in the sorting of United States mail. Paper presented at Sixth Annual Symposium of the Professional Group on Human Factors in Electronics, IEEE., May 6–8, 1965.

Cornog, J. R., Hockman, J. F. & Craig, J. C. Address encoding—A study of the double-binary keyboard as a link in the machine-sorting of mail. Paper No. 63, WA 338, October 1963, American Society of Mechanical Engineers, New York.

Crook, M. N., & Kellog, D. S. Experimental study of human factors for a hand-written numeral reader. *IBM Journal*, 1963, **7**, 76–78.

Damerau, F. A technique for computer correction of spelling errors. Research Note No. NC-65, February 1962, Thomas J. Watson Research Center, Yorktown Heights, New York.

Deininger, R. L. Desirable push-button characteristics. *IRE Transactions*, 1960 (a), **1**, 24–30.

Deininger, R. L. Human factors engineering studies of the design and use of push-button telephone sets. *Bell System Technical Journal*, 1960 (b), **39**, 995–1012.

Deininger, R. L. Rotary dial and thumbwheel devices for manually entering sequential data. *IEEE Transactions on Human Factors in Electronics*, 1967, **8** (3), 227–231.

Deininger, R. L., Billington, M. J., & Michaels, S. E. Operator speed and accuracy with literal and combination codes for processor input. *IEEE Transactions on Man-Machine Systems*, 1968, **9** (1), 10–14.

Deininger, R. L., Billington, M. J., & Riesz, R. R. The display mode and the combination of sequence length and alphabet size as factors speed and accuracy. *IEEE Transactions on Human Factors in Electronics*, 1966, **7**, 110–115.

Devoe, D. B. Alternatives to handprinting in the manual entry of data. *IEEE Transactions on human Factors in Electronics*, 1967, **8** (1), 21–32.

Devoe, D. B., & Graham, D. N. Evaluation of hand printed character recognition techniques. Final Report Contract AF30 (602)-4385, Project 4594 (Sylvania Project 171), January 1968, Sylvania Electric Products, Inc., Waltham Mass.

Devoe, D. B., Eisenstadt, B. & Brown, D. E., Jr. Manual input coding study. Report No. RADC-TR-66-476, 1966 Rome Air Development Center, Griffis Air Force Base, New York.

Diehl, M. J., & Seibel, R. The relative importance of visual and auditory feedback in speed typewriting. *Journal of Applied Psychology*, 1962, **46**, 365–369.

Dolotta, T. A., & Selfridge, O. G. Functional specifications for typewriter-like time sharing terminals, April 1968, Princeton Computer Center, Princeton, New Jersey.

Droege, R. C., & Hill, B. M. Comparison of performance on manual and electric typewriters. *Journal of Applied Psychology*, 1961, **45**, 268–270.

Dvorak, A. There is a better typewriter keyboard. *National Business Education Quarterly*, 1943, **12-2**, 51-58.

Dvorak, A. A one-handed keyboard for one-handers. *Handicap*, 1950.

Earl, W. K., & Goff, J. D. Comparison of two data entry methods. *Perceptual and Motor Skills*, 1965, **20**, 369–384.

English, W. K., Engelbart, D. C. & Berman, M. L. Display-selection techniques for text manipulation. *IEEE Transactions on Human Factors in Electronics*, 1967, **8** (1), 5–15.

Fein, L. IEEE pattern recognition workshop. *Datamation*, 1967, **13**, 84–86.

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, **71**, 849–857.

Fitts, P. M., & Switzer, G. Cognitive aspects of information processing: I. The familiarity S–R sets and subsets. *Journal of Experimental Psychology*, 1962, **63**, 321–329.

Fleishman, E. A. Abilities at different stages of practice in rotary pursuit performance. *Journal of Experimental Psychology*, 1960, **60**, 162–171.

Fleishman, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), *Training research and education.* New York: Wiley, 1965.

Fleishman, E. A., & Fruchter, B. Factor structure and predictability of successive stages of learning Morse code. *Journal of Applied Psychology*, 1960, **44**, 97–101.

Fleishman, E. A., Roberts, M. M. & Friedman, M. P. A factor analysis of aptitude and proficiency measures in radiotelegraphy. *Journal of Applied Psychology*, 1958, **42**, 129–135.

Flynn, M. A. Analysis and comparison of the errors of a one-year typewriting group with the errors of a three-year typewriting group in typing selected timed writing tests. *The National Business Education Quarterly*, 1959, **28**, 20.

Fox, J. G., & Stansfield, R. G. Diagram keying times for typists. *Ergonomics*, 1964, **7**, 317–320.

Galli, E. J. The stenowriter system (abridged). Report No. RC-308, July 5, 1960, IBM Research Lab,. Yorktown Heights, N.Y.

Greanias, E. C., Meagher, P. F., Norman, R. J., & Essinger, P. The recognition of handwritten numerals by contour analysis. *IBM Journal*, 1963, **7**, 14–21.

Griffith, R. T. The minimotion typewriter keyboard. *Franklin Institute Journal*, 1949, 399–436.

Helweg, T. W. When can you justify optical character readers? *Control Engineering*, 1962, **9**, 119–121.

Hershman, R. L., & Hillix, W. A. Data processing in typing: typing rate as a function of kind of material and amount exposed. *Human Factors*, 1965, **7**, 483–492.

Hillenbrand, W. I. A special report on switches. *Electronic Products*, 1967, **9** (8), 27–47.

Hillix, W. A., & Coburn, R. Human factors in keyset design. Research Report No. 1023, March 1961, U.S. Navy Electronics Lab., San Diego, Calif.

Hirsch, R. S. Measurement and estimation of human input performance in mechanized systems. *IBM Report*, RJ 152, October 22, 1958, San Jose, Calif.

Hirsch, R. S., Teichroew, D., & Paige, J. C. Experimental evaluation and statistical analysis of human performance in a constrained handwriting method of data recording. Technical Report No. 16.10.070.019, June 1960, IBM Advanced Systems Development Division, San Jose, Calif.

King, J. H., & Tunis, C. J. Some experiments in spoken word recognition. *IBM Journal* of Research Development, 1966, **10**, 65–79.

Klemmer, E. T. A ten-key typewriter. *IBM Research Memorandum*, November, 1958, RC-65, IBM Research Center, Yorktown Heights, N. Y.

Klemmer, E. T. & Lockhead, G. R. An analysis of productivity and errors on keypunches and bankproof machines. *IBM Research Report*, November 1960, RC-354, IBM Research Center, Yorktown Heights, N. Y.

Klemmer, E. T., & Lockhead, G. R. Productivity and errors in two keying tasks: A field study. *Journal of Applied Psychology*, 1962(a), **46**, 401–408.

Klemmer, E. T., & Lockhead, G. R. Further data on card punch operator performance. Research Note No. NC 39, 1962(b), Thomas J. Watson Research Center, Yorktown Heights, N. Y.

Kolesnik, P. E., & Teel, K. S. A comparison of three manual methods of inputting navigational data. *Human Factors*, 1965, **7**(5), 451–456.

Leonard, J. A. The effects of 'machine lag' on a serial choice task with balanced and biased input frequencies. *Ergonomics*, 1958, **2**, 44–51.

Leonard, J. A. On the relationship between coding and the ability to see what one is doing in a sensori-motor task. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 325–331.

Leonard, J. A., & Newman, R. C. Formation of higher habits. *Nature*, 1964, **203**, (4944), 550–551.

Leonard, J. A., & Newman, R. C. On the acquisition and maintenance of high speed and high accuracy in a keyboard training task. *Ergonomics*, 1965, **8**, 281–304.

Licklider, J. C. R. Man-computer symbiosis. *IRE Transactions on Human Factors in Electronics*, 1960, **1**, 4–11.

Lockhead, G. R., & Klemmer, E. T. An evaluation of an 8-key word-writing typewriter. *IBM Research Report*, November 1959, RC-180, IBM Research Center, Yorktown Heights, N.Y.

Lutz, M. C., & Chapanis, A. Expected locations of digits and letters on ten-button keysets. *Journal of Applied Psychology*, 1955, **39**, 314–317.

McArthur, B. N. Accuracy of source data human error in hand transcription. Technical Report No. ASD-TR-65-10, May 1965, Air Force Logistics Command, WPAFB, Ohio.

Miller, G. A. The magical number seven, plus or minus two; some limits on our capacity for processing information. *Psychological Review*, 1956, **63**, 81–97.

Miller, G. A., Newman, E. B., & Friedman, A. E. Length-frequency statistics for written English. *Information Control*, 1958, **1**, 370–398.

Minor, F. J. Experimental comparison of two keyboard interlock systems. *Journal of Engineering Psychology*, 1964, **3**, 9–15.

Minor, F. J., & Pitman, G. G. Evaluation of variable format entry terminals for a hospital information system. Paper presented at Sixth Annual Symposium of The Professional Group on Human Factors in Electronics, IEEE May 6–8, 1965.

Minor, F. J., & Revesman, S. L. Evaluation of input devices for a data setting task. *Journal of Applied Psychology*, 1962, **46**, 332–336.

Paul, L. E., Sarlanis, K., & Buckley, E. P. A human factors comparison of two data entry keyboards. Paper presented at Sixth Annual Symposium of the Professional Group on Human Factors in Electronics, IEEE, May 6–8, 1965.

Plath, D. W., & Kolesnik, P. E. The thumbwheel switch as a data entry device. *Human Factors*, 1967, **9** (5), 479–482.

Pollock, W. T., & Gildner, G. G. Study of computer manual input devices. Report No. ESD TDR 63 545 & BSC Report No. 40138, September 1963, U.S. Air Force Decision Sciences Lab., Hanscom AFB, Mass.

Poulton, E. C. Copying behind during dictation. *Quarterly Journal of Experimental Psychology*, 1958, **10**, 48–55.

Poulton, E. C., & Brown, C. H. Rate of comprehension of an existing teleprinter output and of possible alternatives. *Journal of Applied Psychology*, 1968, **52**, 16–21.

Rabbitt, P. M. A. Time to detect errors as a function of factors affecting choice-response time. *Acta Psychologica*, 1967, **27**, 131–142.

Rabbitt, P. M. A., & Birren, J. E. Age and responses to sequences of repetitive and interruptive signals. *Journal of Gerontology*, 1967, **22**, 143–150.

Ratz, H. C., & Ritchie, D. K. Operator performance on a chord keyboard. *Journal of Applied Psychology*, 1961, **45**, 303–308,

Ridinger, H. J. The Grafacon man-machine interface. *Datamation*, 1967, **13** (7), 44–51.

Samuel, A. L. Time-sharing on a computer. *New Scientist*, 1965, **26** (445), 583–587.

Seibel, R. A feasibility demonstration of the Rapid-Type data entry station. Research Report No. RC 845, December 1962 (a), Thomas J. Watson Research Center, Yorktown Heights, N.Y.

Seibel, R. Performance on a five-finger chord keyboard. *Journal of Applied Psychology*, 1962(b), **46**, 165–169.

Seibel, R. Discrimination reaction time as a function of the number of stimulus-response pairs and the self-pacing adjustment of the subject. *Psychological Monographs*, 1962(c), **76**, 561.

Seibel, R. Discrimination reaction time for a 1,023-alternative task. *Journal of Experimental Psychology*, 1963(a), **66**, 215–226.

Seibel, R. Letter sequences (N-grams): I. Frequencies in samples of text from the (London) Times. Research Paper No. RC 972, June 27, 1963(b), IBM, Thomas J. Watson Research Center, Yorktown Heights, N.Y.

Seibel, R. Data entry through chord, parallel entry devices. *Human Factors*, 1964(a), **6,** 189–192.

Seibel, R. Levels of practice, learning curves, and system values for human performance on complex tasks. *Human Factors*, 1964(b), **6,** 293–298.

Semling, H. V., Jr. Data processing at the Internal Revenue Service. *Modern Data Systems*, 1968, **1,** 78–81.

Shepard, R. N., & Sheenan, M. M. Immediate recall of numbers containing a familiar prefix or postfix. *Perceptual and Motor Skills*, 1965, **21,** 263–273.

Siders, R. A. Computer-aided design. *IEEE Spectrum*, 1967, **4** (11), 84–92.

Simek, J. G., & Tunis, C. J. Handprinting input device for computer systems. *IEEE Spectrum*, 1967, **4** (7), 72–81.

Smith, W. A., Jr. Accuracy of manual entries in data-collection devices. *Journal of Applied Psychology*, 1967, **51** (4), 362–368.

Standards. Proposed USA standard general purpose alphanumeric keyboard arrangement for information interchange. *Communications of the ACM*, 1968a, **11** (2), 126–129.

Standards. Proposed USA standard USASC-SOCR dual case keyboard arrangement. *Communications of the ACM*, 1968b, **11,** (2), 130–131.

Suppes, P. The uses of computers in education. *Scientific American*, 1966, **215** (3), 206–220.

Sutherland, I. E. Computer inputs and outputs. *Scientific American*, 1966 (a), **215** (3), 86–96.

Sutherland, I. E. Computer graphics: Ten unsolved problems. *Datamation*, 1966(b), **12** (5), 22–27.

Technology Profile: CRT displays, Part 2. *Modern Data Systems*, 1968, **1** (6), 46–52.

Tirrell, J. A., & Klemmer, E. T. The effects of coding on keying rate. IBM Research Report No. RC-775, August 1962, Thomas J. Watson Research Center, Yorktown Heights, N.Y.

West, L. J. Review of research on Morse code learning. Research Report No. AFPTRC-TN-55-52, December 1955, Air Force Personnel and Training Research Center, Lackland AFB, San Antonio, Tex.

West, L. J. Review of research in typewriting learning with recommendations for training. Research Report No. AFPTRC-TN-57-69, June 1957, Air Force Personnel and Training Research Center, Lackland AFB, San Antonio, Tex.

West, L. J. Implications of research for teaching typewriting. Delta Pi Epsilon Research Bulletin No. 2, 1962 (Available from: J. E. Binnion, University of Denver, Colorado).

West, L. J. Vision and kinesthesis in the acquisition of typewriting skill. *Journal of Applied Psychology*, 1967, **51,** 161–166.

Woodworth, R. S. *Experimental psychology*. New York: Henry Holt, 1938.

Selected Readings

Damon, A., Stoudt, H. W.,& McFarland, R. A. *The human body in equipment design* Cambridge, Mass.: Harvard University Press, 1966.

Fitts, P. M., & Posner, M. I. *Human performance* Belmont, Calif.: Brooks/Cole (Div. of Wadsworth). 1967.

Chapter 8

Design of Controls

Alphonse Chapanis

The Johns Hopkins University
Baltimore, Md.

Robert G. Kinkade

American Institutes for Research
Washington, D.C.

This chapter provides guidelines for various aspects of control design and selection. The principles and design recommendations presented are the result of human engineering studies and operational field experience. General principles of control design and selection, as well as specific design recommendations for commonly used controls, are presented. In the final section, unusual control methods are discussed.

CONTENTS

This chapter was based in part on Chapter 6 of the previous *Guide*. It was reviewed by Robert Thomson.

8. Design of Controls

8.1 Selecting the Best Control

Selecting the best control from the wide variety currently available is critical to designing for effective man–machine performance. Control selection should be dictated by suitability for a particular task. The best control for one task may be inadequate for another.

Before a control is selected, the following information should be obtained:

1. The function of the control. What general purpose and function will be served by the control? How important is this to the system? What specifically is to be controlled? What type of change is to be accomplished by using the control? What are the extents, levels, and directions of change?
2. The requirements of the control task. With what precision, speed, or force will the control be used over what range? What and how serious are the consequences of not meeting these requirements?
3. The informational needs of the operator. What are the requirements for locating and identifying the control, for determining its position (setting), and for sensing a change in its position?
4. The requirements imposed by the workplace. Where is the control to be located? How much space is available? How important is it to locate the control in a certain position for proper grouping and association with other equipment, controls, and displays?
5. What are the consequences of inadvertent or accidental operation of the control?

The following general rules should be followed in selecting controls:

1. Controls should be distributed so that no one limb is overburdened. Controls requiring rapid, precise setting should be assigned to the hands. Controls requiring large or continuous forward applications of force generally should be assigned to the feet. Although a considerable number and variety of controls can be assigned to the hands, each foot should not have more than two controls assigned to it, and these should require only fore-aft or ankle flexion movement.
2. Select, locate, and orient controls so that their motion is compatible with the movement of the associated display element, equipment component, or vehicle. The general situations in which linear (or nearly linear) and rotary controls should be used are shown in Figure 8–1.
3. Select multi-rotation controls (e.g., cranks) when precise settings are required over a wide range of adjustment. Because the range of movement of a linear control is limited, it does not permit high precision over a wide range of adjustment. With a multi-rotation control, any desired precision can be obtained by appropriate gearing (although this might affect operating time).
4. Select discrete-adjustment (detent) controls or pushbutton arrays rather than continuous-adjustment controls when the controlled object is to be adjusted for discrete positions or values only. Discrete-adjustment controls are preferred when a limited number of settings is required, or when precision requirements are such that a limited number of settings can represent the entire continuum (Jenkins, 1953).
5. Continuous-adjustment controls should be selected when precise adjustments along a continuum are needed, or when a large number of discrete settings (usually more than 24) is required. Discrete-adjustment controls (i.e., controls that snap into place) can be positioned with one gross movement; continuous-adjustment controls require a slewing movement and a fine adjustment movement and thus more time and attention.
6. When force and range of settings are

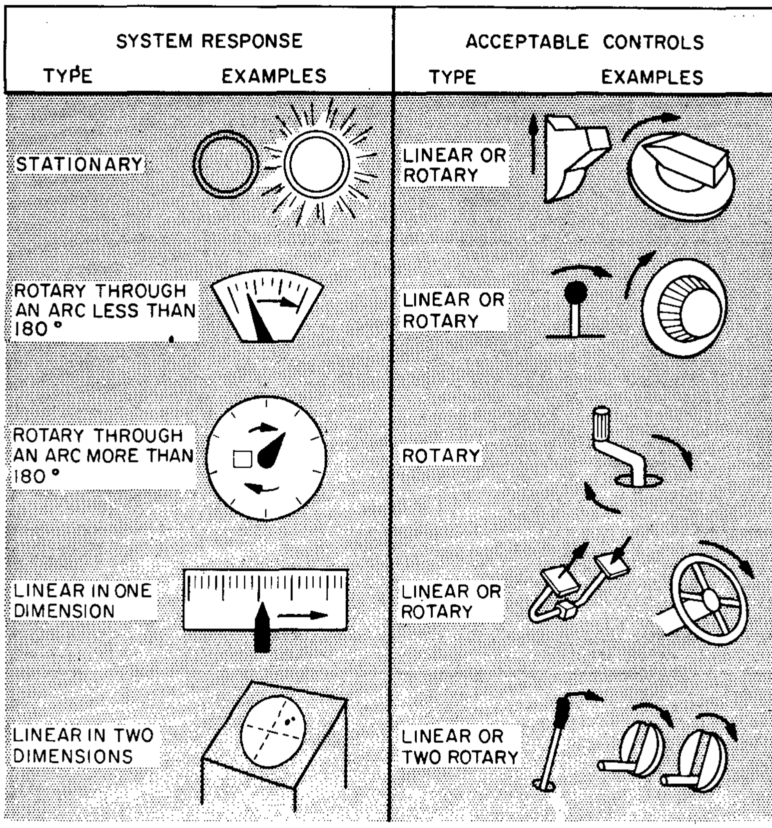


FIGURE 8–1. Examples of acceptable controls for various types of system response.

primary considerations, select the type of control recommended in Table 8–1.

7. Select controls that can be easily identified. All controls should be made identifiable, in part by standardizing their locations. All critical and emergency controls should be identifiable both visually and by touch (shape coding). Identification information should not hinder the manipulation of the control nor increase the likelihood of accidental activation.

8. Combine functionally related controls to reduce reaching movements, to aid in sequential or simultaneous operations, or to economize in panel space.

8.2 Primary Concepts Associated with Controls

This section describes general concepts that are applicable to most types of controls. These are:

1. The control-display (C/D) ratio.
2. Direction-of-movement relationships.
3. Control resistance.
4. Control coding.
5. Preventing accidental activation.

8.2.1 The Control-Display Ratio

The control-display (C/D) ratio is the ratio of the distance of movement of the control relative to that of the moving element of the display (pointer, cursor, etc.). It applies to continuous controls, not to discrete controls. For position (zero-order) controls, the C/D ratio is the reciprocal of "gain" or "sensitivity," expressed in terms of the ratio between display and control movements (See Chapter 6.) The C/D ratio is a critical design factor affecting operator performance; a good C/D ratio has been shown to save from 0.5 to 5 sec. in positioning time when compared with a poor C/D ratio (Jen-

Table 8–1. Recommended controls for the case where both force and range of settings are important

For *small* forces and—	Use—
Two discrete settings	Pushbutton or toggle switch.
Three discrete settings	Toggle switch or rotary selector switch.
Four to 24 discrete settings	Rotary selector switch.
Small range of continuous settings	Knob or lever.
Large range of continuous settings	Crank or multi-rotation knob.
For *large* forces and—	**Use—**
Two discrete settings	Detent lever, large hand pushbutton, or foot pushbutton.
Three to 24 discrete settings	Detent lever.
Small range of continuous settings	Handwheel, rotary pedal or lever.
Large range of continuous settings	Large crank.

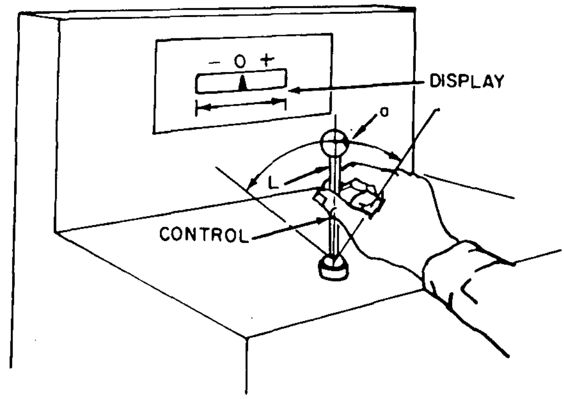


Figure 8–2. An illustration of C/D ratio relationships with a linear display and ball lever.

kins and Connor, 1949; Jenkins and Karr, 1954; Jenkins et al., 1950; and Jenkins and Olson, 1952).

For linear controls (e.g., levers) that affect linear displays, the C/D ratio is defined as the ratio of the control's linear displacement to the resulting display displacement. For controls that involve appreciable rotational movement (e.g., ball controls), that affect linear displays, the C/D ratio is defined by the formula:

$$C/D = \left(\frac{a}{360}\times 2\pi L\right) \Big/ \text{display movement} \qquad (8\text{–}1)$$

where a is the angular movement of the control in degrees and L is the length of the lever arm. (See Figure 8–2.)

For rotary knobs, the C/D ratio is the reciprocal of the display movement (in inches) for one complete revolution of the knob.

Optimizing the C/D ratio. In positioning a continuous-adjustment (non-detent) control, the skilled operator makes the following movements:

1. A slewing movement by which he rapidly moves his control close to the final desired position. An increase in the C/D ratio will slightly increase slewing time because of the longer movements required. If a 4-in. movement takes 0.8 sec., a 16-in. movement may take only 0.2 sec. more. This is because reaction and starting times are nearly constant regardless of distance moved, and longer movements permit higher rates of movement.

2. Fine-adjusting movements by which the operator places his control precisely in the final, desired position. Fine-adjusting time is reduced either by increasing the C/D ratio, or by easing the tolerance requirements (Jenkins and Connor, 1949). The optimum C/D ratio is that which minimizes the total time (slewing plus fine-adjusting) required to make the desired control movement. (See Figure 8–3.)

Factors affecting C/D ratio. For ball-levers the optimum C/D ratio usually ranges from 2.5:1 to 4:1. For knobs, on the other hand, the optimum C/D ratio is independent of knob diameter and thus independent of displacement around the circumference of the knob. Optimum C/D ratios for knobs usually range from 0.2 to 0.8. Because of the many complexities involved, optimum C/D values should be established empirically at least for those applications where time and precision are critical. This can be accomplished by indicating various target positions for the display element (pointer, cursor, etc.) and measuring the time operators take to move the display element to the target position within the required level of accuracy using different C/D ratios, or by a similar experiment designed to simulate the anticipated control

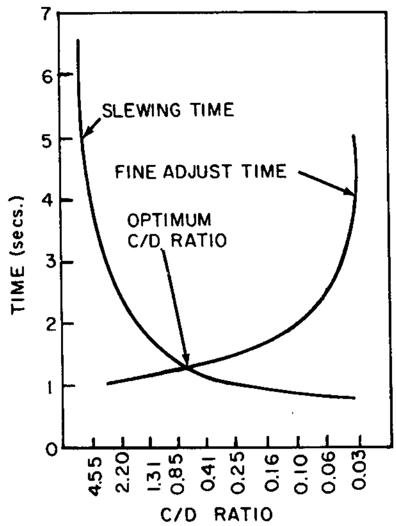


FIGURE 8–3. Slewing time and fine-adjusting time for knob settings with various C/D ratios. Ratios around 0.5 were optimum (adapted from Jenkins and Connor, 1949).

task. These ratios will be affected by the following factors:

1. Display size. With a constant level of control placement accuracy, changing the display size changes total adjustment time, i.e., increasing the display size will decrease adjustment time.

2. Tolerance. Fine-adjusting time is reduced by easing tolerance requirements. Slewing time may also be reduced because the operator tends to move his control into position more rapidly when he knows that it will not have to be positioned precisely. Thus, changing tolerance changes the optimum C/D ratio.

3. Time delay. The type and extent of any time delay (lag) in the system may affect the optimum C/D ratio. For exponential lags between the control movement and the resulting display response, the longer the lag the smaller the optimum C/D ratio, within reasonable limits (Rockway, 1954).

8.2.2 Control-Display Direction of Movement Relationships

Correct control-display movement relationships will enhance system performance by:

1. Decreasing reaction or decision time.
2. Reducing the frequency of incorrect initial control movements (so-called reversal errors).
3. Improving the speed and precision of control adjustment.
4. Decreasing learning time.

The importance of control-display relationships increases directly with:

1. The number of control actions.
2. The nonsequential nature of control actions.
3. The discontinuity, or number of interruptions, in the control sequence.
4. The operator's stress or anxiety.

The recommended direction-of-movement relationships are intended to satisfy one or more of the following: (a) natural relationships, (b) existing design practice, and (c) standardization and consistency. Natural relationships refer to control-movement habit patterns that are consistent from person to person without special training or instructions, i.e., they are responses that individuals make most often and are called "population stereotypes."

Direction of movement rules. The preferred direction of movement for most hand controls is horizontal, rather than vertical, and is fore-and-aft, rather than lateral. Horizontal movements are generally faster than vertical ones (Woodson and Conover, 1966), and for right-handed operators right-left movements are slightly faster than right-left ones (Brown and Slater-Hammel, 1949). For short precise linear movements of 2 to 3 in., vertical movements are fastest, followed by lateral and fore-aft movements (Herbert, 1957).

The following general direction of movement rules are applicable:

1. The directoin of movement of a control must be considered in relation to: (a) The location and orientation of the operator relative to the control, the controlled element, and the vehicle; (b) The position of the display relative to the control and the nature and direction of the display's response; (c) The change resulting from the control movement—either in terms of motion of moving components (landing gear, gun turret, etc.) or in terms of some dimensional quantity (volume of a radio receiver, brightness of a radar scope, etc.).

2. All equipment that the same operator uses should have the same control-display motion relationship.

3. The direction of movement of controls and displays should generally be related to the purpose or function of the control action rather than to the particular intervening mechanism used to accomplish the desired function. For example, if the operator wants to lower temperature, opening a vent that will admit cool air or closing a vent that has been admitting hot air need not itself be related to the control movement. The control movement should be related to the basic purpose: to raise temperature (upward movement, clockwise movement, etc.) or to lower temperature (downward movement, counterclockwise movement, etc.).

4. Movement relationships are particularly important when they result in vehicle movement. A movement of a control to the right should result in a movement to the right, a right turn, or right bank of the vehicle, etc. Recommended relationships between control movement and system or component response are shown in Table 8–2.

5. Controls that are related to the direction of movement of the vehicle in which they are mounted should not be located so that the operator faces toward the rear. An operator inside a vehicle controlling its direction of movement should face in the direction in which the vehicle normally moves. Otherwise the movement relationships will be ambiguous with respect to left-and-right, fore-and-aft, etc.

TABLE 8–2. RECOMMENDED CONTROL MOVEMENTS

Function	Control action
On	Up, right, forward, pull (switch knobs).
Off	Down, left, rearward, push (switch knobs)
Right	Clockwise, right.
Left	Counterclockwise, left.
Up	Up, rearward.
Down	Down, forward.
Retract	Rearward, pull, counterclockwise, up.
Extend	Forward, push, clockwise, down.
Increase	Right, up, forward.
Decrease	Left, down, rearward.

U.S. Army Materiel Command (706–134).

8.2.3 Control Resistance

Some force must always be applied to move a control. The kinds of resistance offered by a control (and the device to which it is physically coupled) are:

1. Elastic (spring loading).
2. Static and sliding friction.
3. Viscous damping.
4. Inertia.

Virtually all controls have mass and inertia. Most controls move on a slide, or around a shaft or pivot, and, hence, have static and sliding friction. In addition, there are interactions between the various kinds of resistance and operator performance. Friction or viscous damping, for example, can sometimes be helpful in counteracting the adverse effects of inertia or vibration.

Depending on the kind and amount, resistance can affect the following:

1. The precision and speed of control operation.
2. The "feel" of the control.
3. The smoothness of control movement.
4. The susceptibility of the control to accidental activation, to the effects of vibration and G forces, etc.

The designer should build into the control the kind or kinds of resistance that best satisfy performance requirements. The following paragraphs discuss the most important characteristics of the four major kinds of resistance.

Elastic resistance. Elastic, spring-loading, resistance has the following characteristics:

1. It varies with control displacement and thus provides a cue to the amount of displacement independent of the velocity and acceleration of the control.
2. It applies force toward the null position of the control when it is displaced; hence, it aids in identifying the null position and in making adjustments around it.
3. It returns the control automatically to the same (null) position when the operator's hand or foot is removed; hence, it is ideal for momentary-contact or "dead-man" switches.
4. It permits quick changes to be made in the direction of the control movement.
5. It allows the operator's hand or foot to rest on the control without activating it, if

there is sufficient resistance (preloading) at the null position.

6. It reduces the likelihood of undesired activation caused by accidental contact with the control or by G forces, vibration, etc.

7. By modifying the force gradient, special cues about critical positions of the control can be provided (e.g., resistance suddenly increases as a limit is approached).

Frictional resistance. Static- and sliding-friction resistance have the following characteristics:

1. Static friction decreases sharply when the control starts to move, hence sliding friction is independent of displacement velocity and acceleration. By increasing or decreasing friction as a function of control position, gross control positions can be identified by feel.
2. Static friction tends to hold the control in position and thus reduces the likelihood of undesired activation caused by accidental contact with the control, or by G forces, vibration, etc.
3. Sufficient static friction allows the operator's hand or foot to rest on the control without activating it.
4. It increases the difficulty of making precise settings and small changes in position of the control.
5. It is difficult to design so that it has a constant amount of friction. However, a "locking" device can be provided which allows friction to be varied.

Viscous-damping resistance. Viscous-damping resistance has the following characteristics:

1. It varies directly with control velocity but is independent of displacement and acceleration.
2. It resists quick movements.
3. It reduces the likelihood of undesired activation caused by accidental contact with the control or by shock, G forces, vibration, etc.
4. It assists the operator in making smooth control movements.
5. It permits rapid changes in direction and aids the operator in making precise settings and small changes in control position.
6. It provides the operator with feedback information ("feel") about the velocity of control movement. (It is questionable, however, whether the operator can use this information precisely.)

Inertial resistance. Inertial resistance has the following characteristics:

1. It varies directly with control acceleration but is independent of displacement and velocity.
2. It resists sudden changes in control velocity; hence, it aids in making smooth movements or gradual changes in velocity.
3. It reduces the likelihood of undesired activation caused by vibration, etc.
4. It requires that large forces be applied to stop or start control movements quickly; hence, it hinders any changes in direction and speed.
5. It provides the operator with feedback information ("feel") about the acceleration of control movements. (It is questionable, however, whether the operator can use this information precisely.)
6. It increases the difficulty of making small or precise adjustments quickly because of the danger of overshooting.
7. It can be used to maintain control movement without requiring the continual application of force, e.g., spinning a handwheel.

Speed of movement. For maximum speed (not accuracy) of control movement, control resistance should be minimal, because control speed decreases as the load increases. Pilots can move "joysticks" at rates up to 75 in./sec. when the control resistance is less than 35 lb.; 50 in./sec with a resistance of 35 lb.; and 10 in./sec with a resistance of 100 lb. (Orlansky, 1948).

Resistance for hand controls. In general, resistance for hand controls (except finger-operated controls) should not be less than 2 lb. Below 5 lb. the pressure sensitivity of the hands is poor thereby reducing cutaneous (touch) feedback associated with control actuation. If the arm and hand are used on a control, the minimum resistance should be 10 to 12 lb.; if only the forearm and hand, 5 lb.; if only the hand, 2 lb. (Dempster, 1955).

Limits of the resistance of hand controls are difficult to determine because of (a) wide variations in the operator population; (b) type and location of controls; and (c) frequency, duration, direction, and amount of control movement. For example, there is more than a fourfold difference in the push exertable on a control stick depending on whether it is located on the midline of the body away from the operator (maximum), or to the left of a right-handed operator (minimum). In spite of this, some recommendations for maximum control resistance can be given. These will be found among the recommendations for specific controls.

Operators may be affected by the esthetic properties of controls, and comment about a "smooth," "flimsy," "stiff," or "sloppy" feel. In an investigation of control "feel," the subjective response to the mechanical characteristics of a control, Knowles and Sheridan (1966) found:

1. In continuous rotary controls, a change of about 10% in both friction and inertia is required for threshold, i.e., 75% detection. These results suggest that design changes of this amount or less are not usually worth worrying about. Before the human operator can really detect a difference, a 15% to 20% change is required.
2. Operators prefer: (a) low friction levels, (b) viscous to sliding friction, and (c) at least some inertia in continuous rotary controls.

8.2.4 Control Coding

The purpose in coding controls is one of identification. Making controls easy to identify decreases the number of times a wrong control is used, and reduces the time required to find the correct control. Control coding not only improves operator performance, but also reduces training time.

There are a number of methods of coding controls. The five most common methods are location, labeling, color, shape, and size. Several methods of coding should be combined to achieve maximum differentiation and identification. For example, an emergency control labeled EMERGENCY that is larger than surrounding controls and is colored red, would be more readily identified than if only one of these coding methods were applied. In selecting which method or combination of methods to use, standard codes are preferred (e.g., do not use the color green to indicate an emergency control).

To some extent, coding controls by location is the natural result of good workplace design (e.g., standardized control positions aid in identification). Which specific coding method to use depends on the following:

1. The total demands on the operator during the time when the control must be identified,
2. The extent and methods of coding already in use,
3. The illumination of the operator's workplace,
4. The speed and accuracy with which controls must be identified,
5. The space available for the location of controls, and
6. The number of controls to be coded.

Labeling. Labeling is a simple and effective way of coding controls and usually requires little or no special training of the literate operator. Adequate space, visibility, and lighting are prerequisites for using this method (Chapanis et al., 1949). The following general recommendations should be observed:

1. Labels should be located systematically in relation to controls (usually above the controls).
2. Labels should be brief, but only common abbreviations should be used.
3. Labels should tell what is being controlled (e.g., landing gear position, brightness level, etc).
4. Unusual technical terms should be employed only when absolutely necessary and only when they are familiar to all operators.
5. Abstract symbols (squares, stars, etc.) should not be used when they require special training. Common symbols, used in a conventional manner, are acceptable (e.g., a red cross, a poison symbol, an arrow).
6. Letter and numeral style should be standard and easily readable under all conditions of possible use. (See MIL–C–10812.)
7. Labels should be located so they can be observed while the operator is adjusting controls.

Color coding. Color coding is most effective when a specific meaning can be attached to the color (e.g., red for emergency controls). The use of color coding depends on ambient or internal illumination. Color should not normally be used as the sole or primary method for coding controls; it is effective, however, when combined with other methods.

In general, only five colors should be used for single-color coding; red, orange, yellow, green, blue. Even under ideal conditions, an operator has difficulty attaching a name, or function, to more than 10 or 12 colors. However, by patterning a few colors, such as alternating red and black stripes, many distinctive combinations are possible.

Shape coding. Shape coding provides for tactile (touch) identification of controls and aids in identifying them visually. When feasible, it is desirable to select functional shapes that suggest the purpose of the control. The following rules should be observed in shape-coding controls.

1. Use shapes that are easily associated with or that resemble the control function.
2. Use shapes that can be easily identified visually.
3. Avoid sharp edges on the parts of the control that must be grasped (Brennan and Morant, 1950).
4. Use shapes that are easily distinguished from one another.

Size coding. Controls can be coded by size, but if the operator must rely on touch alone, usable sizes are quite limited. The ability to discriminate size by touch is relatively independent of shape discrimination; hence, size coding can be superimposed on shape coding (Eriksen, 1954). Tactile coding by size and shape is less effective if the operator must wear thick gloves.

Information is available for "relative" (as against "absolute") size discrimination. Figure 8–4 shows the diameters required to discriminate between two round knobs when the operator can feel each before deciding which is larger. For knobs ranging from 0.5 to 6 in. in diameter, the larger knob should be at least 20% larger than the smaller.

When the operator cannot feel all the controls before selecting the proper one, only two or at most three different sizes of controls should be used (viz., small, medium, and large). When the operator always feels various controls before selecting one, as many as five different sizes can be used (Hunt, 1953).

8.2.5 Preventing Accidental Activation

It is always desirable to reduce the possibility of accidentally activating controls. However,

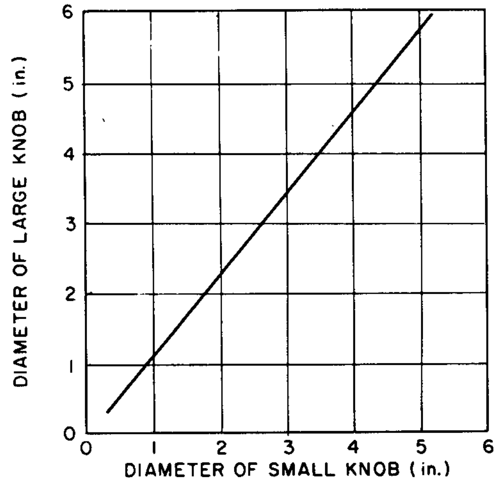


FIGURE 8–4. Desirable relationship between diameters of large knobs and small knobs when the operator must compare each by touch alone to decide which is larger.

in the application of any of the various methods of physically protecting controls against inadvertent activation, it is necessary to consider the extent to which other human engineering design criteria are compromised. For example, protection against accidental activation could increase the time required to operate the control. In some situations, the consequences of accidental activation are so serious that it is mandatory to design the control assembly to eliminate the possibility.

Methods of physically protecting controls against inadvertent activation include:

1. Recessing. Controls can be recessed so that they do not protrude above the control panel surface. A related technique is to place raised barriers around the control. A disadvantage of this method is the amount of panel space that must be used.
2. Location. Controls can be located so that they are unlikely to be hit accidentally. This can be accomplished by isolating one control from others, and by arranging controls so that the sequence of operations is not conducive to accidental activation.
3. Orientation. The direction of movement of the control can be oriented along an axis in which accidental forces are least likely to occur, but care should be taken to insure that recommended direction-of-motion relationships are not violated.

4. Covering. Protective covers or guards can be placed over the control. If the control is to be operated frequently, however, this method probably cannot be used or will be disabled.

5. Locking. Controls can be locked in position. (See Figure 8–5.) This method generally requires the sequential application of force in at least two directions to release and operate the control. This method is undesirable, however, if the control is used frequently.

6. Operation Sequencing. A series of interlocks can prevent Step 2 from being performed before Step 1, Step 3 before Step 2, etc. There are two variations of this method. In the first, all the steps prior to the last one have a direct effect on system output (e.g., a bomb-release control that cannot be operated unless the control that arms the bomb has already been activated). This variation cannot, of course, be used when the sequence of operation differs from situation to situation. In the second variation, the steps prior to the last one have no direct effect on system output other than to permit the next control operation to be performed. In the simplest situation, a preliminary operation (e.g., pushing a button, squeezing a trigger) releases the control for its normal operation.

7. Resistance. Use of the proper kind or kinds and amount of resistance prevents accidental forces (e.g. below the breakout force) from activating the control.

FIGURE 8–5. One method of preventing accidental activation is by requiring sequential application of force in at least two directions.

8.3 General Principles of Control Design

8.3.1 For All Control Types

The following general principles apply to the design of all types of controls, whatever their purpose or mode of operation.

1. The maximum force, speed, accuracy, or range of body movement required to operate a control should not exceed the limits of the least capable operator, and normal requirements for control operation should be considerably less than the maximum capabilities of most operators. (See Chapter 11.)

2. The number of controls should be kept to a minimum, and the control movements should be as simple and as easy to perform as possible.

3. Control movements that seem "natural" for the operator are more efficient and less fatiguing than those that seem awkward or difficult.

4. Control movements should be as short as possible, consistent with the requirements of accuracy and "feel."

5. Controls should have sufficient resistance so as to reduce the possibility of inadvertent activation by the weight of a hand or foot. For controls requiring single applications of force, or short periods of continuous force, a reasonable maximum resistance is half of the operator's greatest strength. For controls operated continuously, or for long periods, resistances should be much lower.

6. When an operator cannot apply enough unaided force to operate the controls and power-boosted or fully powered control systems are necessary, artificial resistance cues should be provided.

7. Controls should be designed to stand abuse; for example, emergency or panic responses frequently impose large forces on controls.

8. Controls should provide a positive indication of activation so that malfunction will be obvious to the operator.

9. Control actions should result in a positive indication to the operator that there has been a system response.

10. Control surfaces should be designed to prevent the activating hand, finger, or foot, from slipping.

8.3.2 For Specific Control Types

The following paragraphs give general design principles for vehicle, valve, and handgrip controls. Push-pull controls, which project from the panel surface on which they are mounted (e.g., a hand choke on an automobile, a throttle on some types of small aircraft), are not recommended because the pushed-in position on these controls is generally associated with "off" or "decrease," and this control movement, particularly when the control is mounted on the front plane, conflicts with the usual forward-to-increase control movement. Even more important in some applications is the fact that it is usually difficult to tell how a push-pull control is set just by looking at it or feeling it. These objections, of course, do not apply to push-button controls, or to push-pull levers.

Vehicle controls. When using a rotary control such as a steering wheel, the operator orients himself with respect to a certain point on the control; he perceives the control as moving in the direction in which this point is moving. When rotary control movements are in a horizontal plane, the operator orients himself to the forward point of the control. (See Figure 8–6.) With rotary control movements in a vertical plane, the operator orients himself to the top of the control. (See Figure 8–7.) With these points in mind, the following recommendations should be observed:

1. When the control affects the direction of movement of a vehicle, the point of control to which the operator is oriented should move in the same direction as the desired direction of the vehicle.
2. The axis of rotation of the control should parallel the corresponding axis of rotation of the vehicle (provided that the resulting control movements do not cause undue operator discomfort).
3. With a rotary display that has a moving pointer and a stationary dial, a clockwise rotation of the rotary control should result in a clockwise rotation of the display pointer.
4. A rotary display that has a moving dial and a fixed pointer usually will cause direction-of-movement ambiguities and should be replaced, when feasible, by a fixed dial and moving pointer. If it cannot be replaced, the following are recommended for the case where the indicator (pointer) is fixed at the 12 o'clock position (Bradley, 1954):

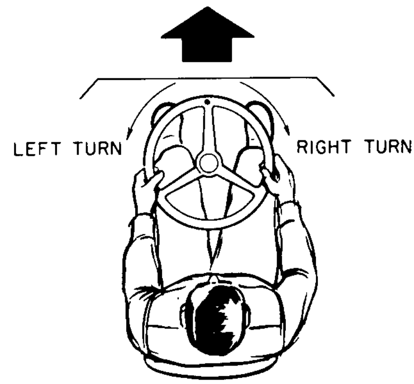


FIGURE 8–6. The operator normally orients himself towards a forward point of the control when using a rotary control in a horizontal plane.

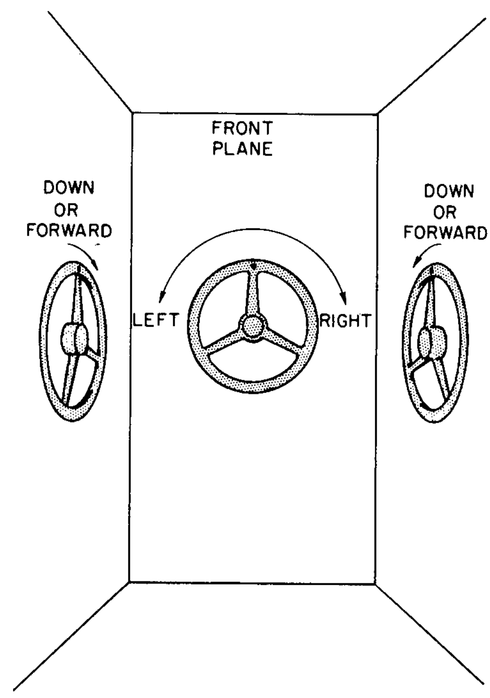


FIGURE 8–7. Recommended movement relationships for rotary controls mounted in a vertical plane.

a. If the control-display combination is not related directly to the movement of the vehicle and if the operator always views the display when operating the control, scale numbers on the display should progress in a clockwise direction, and a clockwise rotation of the control should result in a clockwise rotation of the display.

b. If the control directly affects the movement of the vehicle and if the operator does not have to view the display when operating his control, scale numbers on the display should progress in a clockwise direction, and a clockwise movement of the control should result in a counterclockwise movement of the display.

Valve controls. Most valves turn clockwise rather than counter-clockwise for "close" or "off." This can be confusing to an operator for two reasons: (a) a clockwise rotation to decrease something conflicts with recommendations for all other controls that a clockwise rotation increase whatever is being controlled; (b) opening a valve may increase pressure in one case and decrease pressure in another. For these reasons, valve controls should operate and be labeled according to the end purpose they serve (for example, "raise" or "drain") and direction of movement required to perform the function. In this way, direction-of-movement relationships for controls and displays can be consistent with each other and with other control-display movements in the system.

Handgrips. A handgrip where thumb and forefinger overlap is much better than a wider grip separating the two. Handgrips may be either round or oval in cross section or contour molded to the shape of the hand. An acceptable diameter for a handgrip is between 0.75 and 1.5 in. (Muller, 1935). Minimum diameters depend on the forces exerted: for 10 to 15 lb. the diameter should be no less than 0.25 in. (preferably larger); for 15 to 25 lb., a minimum of 0.5 in.; for 25 lb. or more, a minimum of 0.75 in. is recommended. Maximum diameters should not greatly exceed 1.5 in., and the length of a handgrip should be at least 3.75 in. to accommodate the full breadth of the hand.

For whole-hand grasping or "trigger" grips in which two elements are squeezed together, the fingers should be placed around the main shaft and the heel of the hand used to close the movable part (Dupuis et al., 1955). Maximum force can be applied when 2.5 in. separate the "trigger" and the heel in the open position (Hertzberg, 1955). In the closed position, this distance should be 1.5 to 2 in. (Dupuis et al., 1955).

8.4 When to Use Certain Controls

Hand controls should be used in preference to foot controls in the following situations:

1. If accuracy of control positioning is important.
2. If speed of control positioning is required.
3. If continuous or prolonged application of moderate-to-large forces (20 lb. or more) is not necessary.

For precision and speed, one-hand controls are preferable to those operated with both hands. For start-stop or on-off controls, use either push buttons or toggle switches. When speed of operation is important, use push buttons, but if the controls must be spaced closely together (less than 1 in. between centers), use toggle switches with small dimensions and relatively large resistance to prevent inadvertent operation (Bradley and Wallis, 1959).

Foot controls should be considered in the following situtations:

1. When continuous control is required (if precision of control positioning is not of primary importance).
2. When the application of moderate-to-large forces (greater than about 20 to 30 lb.), whether intermittent or continuous, is necessary.
3. When the hands are in danger of becoming overburdened with control or other tasks.

For fine-adjustment tasks requiring small forces, such as radio tuning, use either finger- or hand-operated rotary knobs. A single control moving in two or three dimensions is better than separate controls, each moving in one dimension, if the control order is the same for each axis of control and if cross coupling is not a problem (Orlansky, 1948).

For discrete control settings operated by applying moderate-to-large forces, such as gear shifting or hand braking, use control sticks or levers.

Where great force must be exerted on a control, push-pull movements should be used. Push-pull controls should be located above the waist, near the chest (Konz and Day, 1966). Rotation produces the next highest force, followed in descending order by up-down and right-left movements. In the weakest direction (right-left), control movements are only about one-third as strong as those exerted in the strongest direction (push-pull).

Large-diameter handwheels with reciprocal rotary motion (such as vehicle steering wheels) are best operated with two hands. These controls are used when large forces are required. The forces that can be applied to two-hand controls compared to a control operated by one hand, depend on control type and location and on the kind and direction of movement, as follows (Provins, 1955):

1. On wheel controls, rotary forces are effectively doubled in most cases.
2. On stick or lever controls located along the body midline, pull is generally almost doubled, push is doubled near the body midline, but is only slightly stronger at distances away from the body, and push right or left is increased about 50%.
3. On stick or lever controls located on either side of the body midline, at or beyond the shoulder, pull is approximately doubled, push is not greatly increased except at close distances, pull right on controls located to the left is slightly better with two hands than with only the right hand, and push right on controls located to the right is slightly better with two hands than with only the left hand.

Pedals on which pressure is applied by the whole leg, such as an automobile brake pedal, should be used when forces above 10 to 20 lb. are required. Pedals on which pressure is applied mainly from the ankle, such as an automobile accelerator pedal, should be used when small forces (about 10 lb. or less) and continuous operation are required. Foot push buttons, such as headlight dimmer switches, are advantageous where small forces and intermittent operation are required. For all foot controls, the direction of movement should be down (or away from the body) and in line with the long axis of the lower legs or roughly parallel to the midplane of the body.

8.5 Design Recommendations for Commonly Used Controls

These design recommendations apply to commonly used controls, such as pushbuttons, toggle switches, rotary selector switches, knobs, handwheels, pedals, etc. The design of controls similar to these can be helped by extrapolating from the recommendations given here. Recommendations for combined controls (e.g., a hand push button mounted on a lever) can be established from the recommendations given for individual controls.

Along with recommended size and displacement (range of control movement), recommended amounts and kinds of resistance to be built into each control are given. For toggle switches, handwheels, cranks, and levers, resistance is described in terms of linear resistance (i.e., the resistance at the point where the operator applies force to the control) rather than torque. For these controls, operator output normally can be considered as a force relatively independent of control radius. For circular knobs, resistance is described in terms of torque. The force that can be brought to bear on these controls is a function of the "efficiency" of the operator's grasp (i.e., the amount by which the fingers must be spread, etc.), which, in turn is related to knob diameter.

Controls are generally of two types: discrete action and continuous action. The features of each type will be discussed in the following sections.

8.5.1 Discrete Action Controls

Discrete action controls may be set to any one of a limited number of exact positions. Controls of this type are typically used for turning equipment on or off, selecting modes of operation, and choosing meter scales. This class of controls includes push buttons, toggle switches, thumbwheels, rotary selector switches, and knife switches.

Most of these discrete action controls are available in modular components, which allow greater freedom in the design of an integrated control panel. These modular components are designed as some multiple of a standard unit, for example, 0.75 in., which permits a wide range of alternative placements for different controls. By selecting alternate mechanisms within the modular control components, almost the entire range of switch requirements is available: alternate-action, momentary power, multiple pole, mechanical interlocking, and encoding (Harkins, 1966).

Pushbuttons

Pushbuttons are of three major types: latching (push-on, lock-on), momentary (push-on, release-off), and alternate action (push-on, push-off). Latching pushbuttons may be mechanical or electromechanical. When the button is depressed, activation should be indicated by a sudden drop in resistance and, if possible, an audible click. The button should remain depressed until the circuit is deactivated. The position of a latching pushbutton cannot be easily identified visually. For this reason hand pushbuttons that are used for critical functions or that are located in dimly lighted areas should be either back-lighted or used with displays that indicate the position of the control.

Momentary contact pushbuttons are operated by depressing the button to its limit and holding it at this position. Activation can be indicated only by an associated display. The control is deactivated and returns to normal position when the operator's finger is removed. Generally, lamp-test controls are of this type.

Alternate action pushbuttons are operated by depressing the button to its limit and releasing it, thereby activating a circuit. Activation is indicated by feel, an auditory click, and by an associated display action. Reactivation is accomplished by depressing the button again. A common example of this type of control is the headlight beam control on an automobile.

All three of these types of pushbutton controls are generally used for hand operation, while the foot-operated buttons are usually restricted to momentary and alternate action controls. Hand pushbuttons generally require only a small amount of panel space. A hand pushbutton can be coded by color, size, shape, location, or labeling, while foot pushbuttons are usually coded by location only. Hand pushbuttons can be operated quickly and simultaneously with other pushbuttons in an array and are identified easily by their position within an array or by their associated display signal.

Foot pushbuttons require a large amount of space for their operation because of the swept volume of the foot. (See Table 8-3.) They do, however, free the hands for other operations.

The "on" or "off" hand pushbutton control setting is not easily identified, visually or tactually (Hunt, 1953, and Orlansky, 1949), and neither the foot pushbutton control nor its setting is easily identified.

The foot pushbutton is quickly activated when the foot is resting on it. Toe-operated controls are activated slightly faster than heel-operated ones (Barnes et al., 1942). The operating time for both hand- and foot-operated pushbuttons increases with an increase in required displacement and/or resistance.

Hand pushbuttons are designed to be operated by the fingers, thumb, or the palm or heel of the hand. In most applications pushbuttons are operated by the finger or thumb. Various sequences of fingertip control operation are possible: one finger can be used to operate a series of pushbuttons in a random manner, such as in the touch-tone entry on a telephone, or sequentially where the pushbuttons are arranged in order of operation; or different fingers can be used randomly or sequentially in some matrix configuration, such as in computer entry devices.

Some design recommendations are common to both hand and foot pushbuttons. These are:

1. Use elastic resistance. Resistance should start low, build up rapidly, then drop suddenly to indicate that the control has been activated. Minimize viscous damping and inertial resistance. The elastic resistance is aided by (a) a slight amount of sliding friction in the case of the hand pushbutton and (b) by static friction to support the foot in the case of the foot pushbutton.

2. The surface of the pushbutton should have a high degree of frictional resistance to prevent slipping. For finger-operated buttons a concave surface is even more practical. (See Figure 8-8.)

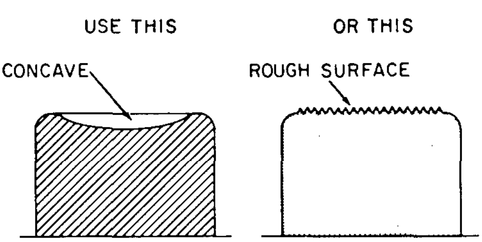


FIGURE 8-8. Two recommended methods for preventing accidental slippage on pushbuttons.

TABLE 8–3. COMPARISON OF THE CHARACTERISTICS OF COMMON CONTROLS

	Type of Control										
	Discrete adjustment					Continuous adjustment					
Characteristic	Rotary selector switch	Thumb-wheel	Hand push-button	Foot push-button	Toggle switch	Knob	Thumb-wheel	Hand-wheel	Crank	Pedal	Lever
Large forces can be developed	------	------	------	------	------	No	No	Yes	Yes	Yes	Yes
Time required to make control setting	Medium to quick.	------	Very quick.	Quick	Very quick.	------	------	------	------	------	
Recommended number of control positions (settings).	3 to 24	3 to 24	2	2	2 to 3	------	------	------	------	------	
Space requirements for location and operation of control.	Medium	Small	Small	Large	Small	Small to medium.	Small	Large	Medium to large.	Large	Medium to large.
Likelihood of accidental activation	Low	Low	Medium	High	Medium	Medium	High	High	Medium	Medium	High.
Desirable limits to control movement	270°	------	⅛″ × 1½″.	½ × 4″	120°	Un-limited	180°	±60°	Un-limited	Small*	±45°.
Effectiveness of coding	Good	Poor	Fair to good.	Poor	Fair	Good	Poor	Fair	Fair	Poor	Good.
Effectiveness of visually identifying control position.	Fair to good.	Good	Poor†	Poor	Fair to good.	Fair‡ to good.	Poor	Poor to fair.	Poor§	Poor	Fair to good.
Effectiveness of non-visually identifying control position.	Fair to good.	Poor	Fair	Poor	Good	Poor to good.	Poor	Poor to fair.	Poor§	Poor to fair.	Poor to fair.
Effectiveness of check-reading to determine control position when part of a group of like controls.	Good	Good	Poor†	Poor	Good	Good‡	Poor	Poor	Poor§	Poor	Good.
Effectiveness of operating control simultaneously with like controls in an array.	Poor	Good	Good	Poor	Good	Poor	Good	Poor	Poor	Poor	Good.
Effectiveness as part of a combined control.	Fair	Fair	Good	Poor	Good	Good¶	Good	Good	Poor	Poor	Good.

Adapted from AFSCM 80–3.

*Except for rotary pedals which have unlimited range.

†Exception: when control is back-lighted and light comes on when control is activated.

‡Applicable only when control makes less than one rotation. Round knobs must also have a pointer attached.

§Assumes control makes more than one rotation.

¶Effective primarily when mounted concentrically on one axis with other knobs.

TABLE 8–4. DESIGN RECOMMENDATIONS FOR PUSHBUTTON CONTROLS

	Diameter (in.)		Displacement (in.)		Resistance		Control separation* (in.)	
	Min.	Max.	Min.	Max.	Min.	Max.	Min.	Preferred
Fingertip:								
One finger—randomly	0.5	†	0.125	1.5	10 oz.	40 oz.	0.50	2.0
One finger—sequentially	----	----	----	----	----	----	0.25	1.0
Different fingers—randomly or sequentially	----	----	----	----	5 oz.	20 oz.	0.50	0.5
Thumb (or palm)	0.75§	†	0.125	1.5	‡	‡	‡	‡
Foot:								
Normal	0.50	†	0.50	----	4–10 lb¶	----	‡	‡
Heavy boot	----	----	1.0	----	----	----	----	----
Ankle flexion	----	----	----	2.5	----	20 lb.	----	----
Leg movement	----	----	----	4.0	----	----	----	----

Adapted from MIL STD-803A-3 (USAF) and AFSCM 80-3.
* Edge-to-edge measurement.
† No limitation set by operator performance.
‡ Not available.
§ Usually for emergency use (NAV).
¶ For foot not resting on control—foot resting on control.

3. To indicate that the control has been activated, an audible click should be provided if the working environment is not too noisy to mask it.

Design recommendations concerning the diameter, separation, displacement, and resistance characteristics for both hand and foot-activated pushbuttons are given in Table 8–4.

Toggle Switches

Toggle switches can be used as either two- or three-position controls. They may have either a lever or a rocker action. Lever-action toggle switches are two- or three-position controls. They may be positive-action or spring-loaded controls. Positive-action controls are operated by applying force to the lever of the toggle switch until the control snaps into the next position. Activation is indicated by an audible click and a release of pressure. Control setting for two-position toggle switches is usually easier to identify than for three-position toggle switches. However, three-position toggle switches are usually acceptable if the control arm is ½ in. or longer, and positions are separated by at least 30°. Detent-position toggle switches should be used, for example, for interlock and two-mode power controls.

In spring-loaded controls, the lever remains at the next position only as long as force is applied, and it returns to its original position when the operator's hand is removed. Activation is indicated by feel and by an associated display.

A rocker-action switch is a special kind of toggle switch with two control faces that meet at an obtuse angle. The face that is depressed indicates the present position of the control. Switch position is changed by applying force to the opposite face until it is depressed and locked into position. Indication of activation is not as visible as it is for the positive-action lever toggle switch. For this reason rocker-action switches should usually be backlighted to help the operator identify the position of the control.

Toggle switches require only a small amount of space of their location and operation. (See Table 8–5.) They can be operated quickly and simultaneously with other toggle switches in a row and are identified easily by their proximity to the associated display or by their location within an array.

Some design recommendations are given in Table 8–5; others are as follows:

1. Use elastic resistance in such a way that it first builds up and then decreases as the desired position is approached, so that the control will snap into position and not stop between adjacent positions.
2. Minimize frictional and inertial resistances.

TABLE 8–5. DESIGN RECOMMENDATIONS FOR TOGGLE-SWITCH CONTROLS

Parameter	Minimum	Maximum
Size:		
Lever tip diameter (in.)	0.125	1.00
Lever arm diameter (in.)	0.50	2.00
Displacement (degrees):		
Between adjacent positions	30	------
Total displacement	------	120
Resistance (oz.)	10	40
Number of positions	2	3
	Minimum	Preferred
Control separation (in.):		
One finger—random order	0.75	2.00
One finger—sequential order	0.50	1.00
Different fingers—random or sequential	0.625	0.75

Adapted from MIL STD 803 (USAF).

3. Install toggle switches for vertical orientation: up for "on," "go," or "increase"; down for "off," "stop," or "decrease." Install toggle switches for horizontal orientation *only* if necessary to be consistent with the orientation of a control function, equipment location or display, or to prevent accidental activation.

4. Provide an audible click to indicate that the control has been activated in areas where the ambient noise level is low enough for a click to be heard.

5. A series of toggle switches should be mounted in a horizontal (rather than vertical) array for speed and ease of operation. For vertical arrays, larger spaces between switches are desirable to minimize the possibility of accidental activation.

Rotary Selector Switches

Rotary selector switches may have from 3 to 24 control positions. Speed and accuracy of setting controls and checking their position are sacrificed if there are more than 24 settings (Chapanis, 1951b and 1951c). The control is activated by applying force to the knob of the switch until the switch snaps into the next position. Activation is indicated by an audible click and tactually by detent action. The position of this control can in some cases be identified by feel and, if a scale is provided, visually. However, critical controls or controls in dimly lit areas may require indicators to identify the position of the switch. Detent locations on a rotary switch should be separated by a *minimum* of 15°. Separations of 30°–35° are to be preferred.

Rotary switches require a medium amount of space for operation because of the swept volume of the hand. (See Figure 8–9.) When a large number of discrete settings is needed, one rotary selector switch requires less space than an array of pushbuttons or toggle switches.

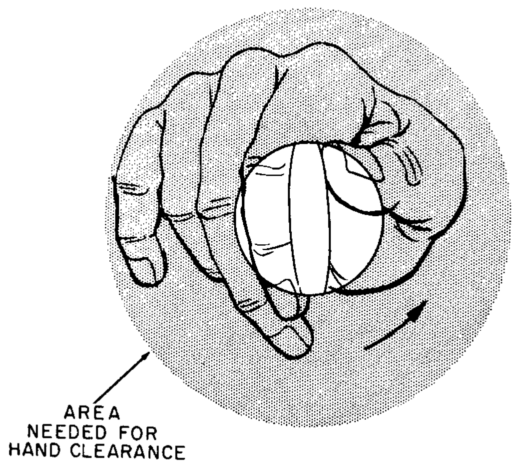


FIGURE 8–9. The diameter of the area needed for hand clearance when using a rotary switch should be 4.6 in.

Rotary switches can have either a moving pointer and fixed scale or a moving scale and fixed index, but a moving pointer with a fixed scale is preferred for most tasks. The moving-pointer type conforms with direction-of- motion relationships without violating other principles, and it facilitates the check reading of control positions for individual controls and for arrays of controls. (See Figure 8–10.)

With the moving-scale type, the index remains in the same position and the scale moves when the knob is rotated. (See Figure 8–11.) The index can be located at any one of the four cardinal points, depending on which is most desirable for the specific situation. In addition, a small segment of the entire scale is all that need be shown. With the open-window arrangement shown in Figure 8–12, the clutter of numbers on the panel is reduced to a minimum.

The knob of a moving-pointer rotary switch usually is a bar-type knob with a tapered tip

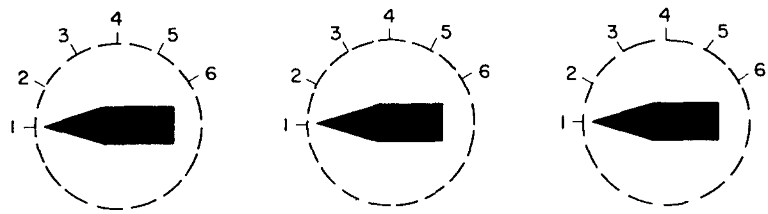


FIGURE 8-10. Moving-pointer rotary switches are easy to check-read.

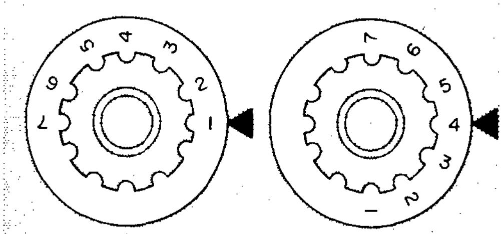


FIGURE 8-11. Moving scale rotary selector switches.

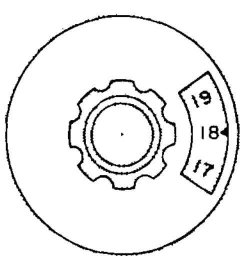


FIGURE 8-12. An open window on a moving scale of rotary selector switch helps to reduce clutter on the panel.

TABLE 8-6. DESIGN RECOMMENDATIONS FOR ROTARY SELECTOR-SWITCH CONTROLS

Type	Pointer*	Knob†
Width or diameter:		
Minimum	0.5 in.	1 in.
Maximum	1 in.	4 in.
Depth:		
Minimum	½ in.	½ in.
Maximum	3 in.	3 in.
Length:		
Minimum	1 in.	n.a.
Maximum	3 in.	n.a.
Displacement:‡		
Minimum	15–30°§	15–30°§
Maximum	40°¶	40°¶
Resistance:		
Minimum	12 oz.	12 oz.
Maximum	48 oz.	48 oz.
Number of positions:		
Minimum	3	3
Maximum	24	24
Control separation:**		
One hand random order:		
Minimum	1 in.	
Preferred	2 in.	
Two hands simultaneous:		
Minimum	3 in.	
Preferred	5 in.	

Adapted from MIL STD 803A (USAF)
*Moving pointer, fixed scale.
†Moving scale, fixed index.
‡Between adjacent detents.
§First number is for visual positioning; second number is for tactile positioning.
¶When special requirements demand large separations, max. should be 90 deg.
**Edge to edge separation of controls.

serving as a pointer, or a round knob with a pointer attached. The moving-scale type is usually a round knob with a scale attached. Shape coding for both types of rotary switches is effective.

Some design recommendations are given in Table 8-6; others are as follows:

1. For most applications selector switches should have fixed scales and moving pointers.

2. Detents should be provided at each control position (setting).

3. Elastic resistance that builds up and then decreases as each position (detent) is approached should be used so that the control will fall into each detent and will not stop between detents. Minimize frictional and inertial resistances.

4. Round knobs should not be used for high-resistance detent controls because they are not effective for the application of large torques.

5. For speed and ease of operation, there should be no less than 15° between settings.

When more than 24 positions must be made available, the minimum separation between adjacent positions should be ¼ in. at the index marks.

6. When few control settings are required, the settings should be separated by about 30° from each other. (See Figure 8-13.) If followed, this rule will help to reduce errors in setting and reading the wrong end of the moving pointer.

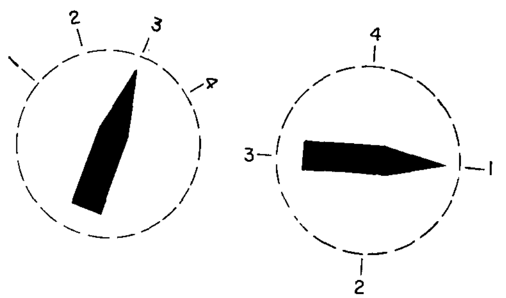


FIGURE 8-13. When only a few control settings are required, use the spacing on the left rather than on the right.

7. When less than 24 settings are required, the beginning and end of the scale should be separated by a gap larger than the displacement between adjacent detent positions.

8. When settings are to be selected sequentially, always going from lower to higher or higher to lower, stops should be placed at the beginning and end of the range of control positions. This will facilitate blind-positioning by enabling the operator to count the number of the clicks from a starting position. The stops should be capable of withstanding at least 25 in.-lbs. of rotational torque without damage.

9. For most applications moving pointers should be bar-type knobs with tapered tips.

10. For areas where ambient noise levels are low, provide an audible click to indicate that the control has been activated.

11. Setting values should increase with a clockwise rotation.

12. Minimize parallax by designing the end of the pointer to come close to the scale index.

13. Index numbers should not be obscured when the hand is on the control.

14. For round knobs the grasp area of the knob should be either serrated or otherwise provided with a high degree of frictional resistance to prevent the fingers from slipping.

15. Use ganged selector switches only when space is extremely limited.

16. Code selector switches by shape, location, and labeling. Rotary selector switches used as emergency controls should be red.

Detent Thumbwheels

Detent thumbwheel controls may be used to provide a compact manual digital control-input device (for a series of numbers) and a readout of these inputs for verification. The positions around the circumference of a discrete-operating thumbwheel should have a slightly concave surface on which the numbers are located, or should be separated by a serrated or high-friction area which may be raised at least $\frac{1}{16}$ in. from the periphery on the thumbwheel. (See Figure 8-14.) The thumbwheel is operated by placing the thumb in the concave surface and rotating the wheel. The detent action and resultant visual setting of this control provide setting feedback.

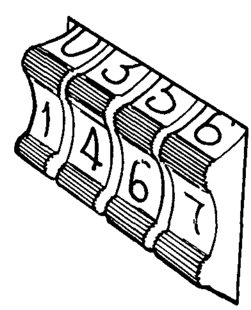


FIGURE 8-14. An array of four discrete thumbwheels.

For areas in which ambient illumination yields display-lettering brightness below 1 ft.-L., internal illumination is required and the numbers located in the concave surfaces should appear as illuminated characters on a black background with the following approximate dimensions: (a) height, at least 0.3 in., (b) height-to-width ratio, 3:2, (c) height-to-stroke-width ratio, 10:1. Where ambient illumination provides a display brightness greater than 1 ft.-L., the digits may be black numerals engraved on a light background. In this case the height-to-stroke-width ratio should be 5:1.

Where possible, detent thumbwheel controls should be placed vertically on the control panel and the next number in sequence should be visible to indicate direction of increase or decrease. Usually, a downward control movement is associated with a decrease and an upward motion is associated with an increase for detent thumbwheel controls. These design recommendations are applicable to fore-aft and right-left orientations of thumbwheels.

Other design recommendations are given in Table 8–7.

TABLE 8–7. DESIGN RECOMMENDATIONS FOR DISCRETE THUMBWHEEL CONTROLS

	Minimum (in.)	Maximum (in.)
Size:		
Diameter	1.5	2.5
Width	0.25	0.5
Protrusion from surface	0.125	0.25
Displacement between adjacent detents.	(*)	(*)
Resistance (torque)	1 in.–lb.	3 in.–lb.
Control separation (in.)	0.25 (0.375 preferred)	

Adapted from MIL STD—803A (USAF) and Chambers & Preusser (1962).
*As determined by the number of positions.

8.5.2 Continuous Action Controls

Continuous action controls may be set at any position between the limits of movement of the control. They are used for varying potentiometer settings, adjusting displays, and opening and closing valves. Controls of this kind include control knobs, thumbwheels, handcranks, handwheels, levers, pedals, and screwdriver adjustments.

Control Knobs

Control knobs are used for making small turning movements that do not require large forces. They have a virtually unlimited range and can be used for either gross or fine positioning over a wide range of adjustments.

Knobs may or may not require symbols to indicate position on a stationary skirt. If position must be indicated, the knob should have a movable indicator with symbols on a stationary skirt. Although control knobs are discussed here under continuous action controls, note that rotary selector switches also require some form of knob, preferably some form of bar knob, for their activation. For continuous action knobs, a folding crank handle may be attached to the knob to aid in rapid slewing. (See Figure 8–15.) Two or more knobs may sometimes be ganged by mounting them on concentric shafts. Mounting more than two knobs on the same shaft is likely to be wasteful of panel space; however, it might be desirable for other reasons such as facilitating a sequence of operations.

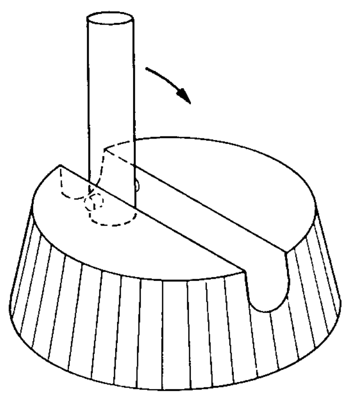


FIGURE 8–15. A crank attached to a continuous-action knob aids in rapid slewing.

Knobs require only a small-to-medium amount of space for their operation because of the swept volume of the hand. They are easily coded by color, size, or shape. Control setting (position) is visually identifiable if the control makes less than one rotation and has a pointer or marker attached.

Knobs lend themselves well to shape coding. For shape-coding purposes, knobs can be divided into the following three classes (Hunt and Craig, 1954):

1. Those for twirling or spinning (more than one full turn is required), for which knob position is not important (Class A).

2. Those for which less than one full turn is required and knob position is not important (Class B).

3. Those for which less than one full turn is required and knob position is important (Class C).

Figure 8–16 shows examples of all three classes. Each knob can be identified by touch alone with the bare hand or while wearing lightweight gloves. All of these knobs can be used together without confusing one with the other, with the following exceptions:

1. Do not use knobs of Class A-3 with those of B-4.
2. Do not use knobs of Class B-1 with those of B-5.
3. Do not use knobs of Class B-2 with those of B-3 or B-4.

Within the ranges recommended in Table 8-8, knob size is relatively unimportant provided the C/D ratio is optimum, the resistance low, and the knob easily grasped (Jenkins and Connor, 1949). When panel space is limited, the use of minimum values for knob size will not degrade performance, provided that knob resistance is very low (Stump, 1953).

Knob diameters have little effect on speed and accuracy. Diameters from 0.5 to 2 in. are generally acceptable (Jenkins and Connor, 1949, and Craik and Vince, 1945). Diameters of 2 in. provide smooth operation at any resistance, though smaller diameters can be used with moderate resistances (1.75 to 3.5 oz. or less). If resistances are above 5.25 to 7 oz., knob diameters should be at least 1.5 in. (Bradley and Arginteanu, 1956).

For concentric, ganged knobs, the best arrangement is with knobs greater than 0.5 in. in diameter and 0.75 to 1.25 in. between their edges (Bradley, 1957). Where three knobs are concentric, the best diameters are 0.5 to 1 in. for the front or top knob, 2 in. for the middle one, and 3.25 in. for the back or bottom one (Bradley and Stump, 1955). Minimum knob depth should be about 0.5 in., and the best depth is about 0.75 in.

The kind or kinds of resistance that should be provided depends, primarily, on performance requirements. When other kinds of resistance are satisfactory for precise positioning tasks, changes in inertial resistance have little practical effect on performance until an excessive level is reached (Jenkins et al., 1951), but the addition of inertial resistance can counteract some of the harmful effects of friction, and vice versa (Searle and Taylor, 1948). When knobs are used for direct positioning, dry frictional resistance should be zero and the torque required to overcome inertia should be the maximum that the physical capabilities of an operator will permit. Control resistance should be applied evenly; there should be no sticky spots or detents.

When bracketing is used for locating a visual or auditory null position (e.g., tuning a transmitter), the knob should move through an arc of 30° to 60° on either side of the null position for a misalignment to be just noticeable (Craik and Vince, 1945). Finger-operated knobs used for fine adjustment should have from 1 to 2 in. of pointer movement for one complete turn of the knob. If less pointer movement is required, lower ratios (less pointer movement per turn of knob) should be provided; for more pointer movement, higher ratios should be provided. In general, accuracy increases as the ratio decreases (Jenkins and Connor, 1949).

Accidental activation can be prevented most easily by careful choice of location. Recessing, covering, and resistance can also be used to prevent accidental activation. Other recommendations follow:

1. The scale should be visible when the operator's fingers are on the knob.
2. Setting values should increase with clockwise rotation of the knob.
3. If a pointer or index is used on the control, it should be close to the scale index mark to minimize parallax.
4. If the control is not used for multi-turn operation there should be a gap between the beginning and end of the scale larger than the separation between consecutive numbered index marks on the scale. Start and end stops should be provided.

Continuous Thumbwheels

Continuous thumbwheels are used for making small turning movements that do not require large forces. They have a virtually unlimited range and can be used for either gross or fine positioning over a wide range of adjustments. They require only a small-to-medium amount of space for their operation because of the swept volume of the hand.

The general design recommendations with respect to size and control separation are

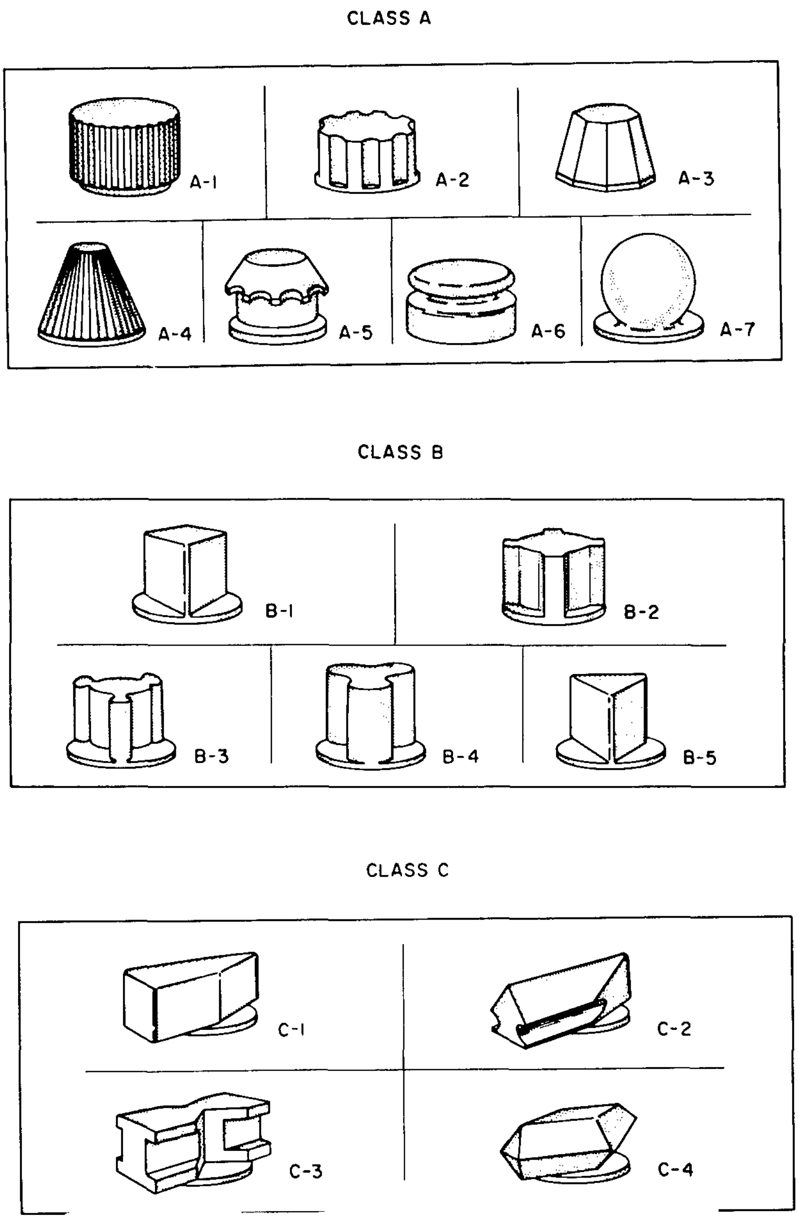


FIGURE 8–16. Examples of three classes of knobs: (A) those for twirling or spinning; (B) those to be used where less than a full turn is required and position is not so important; and (C) those where less than a full turn is required and position is important.

TABLE 8–8. DESIGN RECOMMENDATIONS FOR KNOB CONTROLS

Grasp	Fingertip	Palm of hand
Diameter:		
Minimum (in.)	0.375*	1.5
Maximum (in.)	4.0	3.0
Depth:		
Minimum (in.)	0.5	(†)
Maximum (in.)	1.0	
Displacement	Depends on C/D ratio.	
Resistance:		
Minimum	No limitation	
Maximum	4½–6 in.-oz.‡	
Control separation (in.):§		
One hand, randomly:		
Minimum	1.0	
Preferred	2.0	
Two hands, simultaneously:		
Minimum	3.0	
Preferred	5.0	

Adapted from AFSCM 80-3.
*0.25 when resistance is made very low.
†No limit set by operator performance.
‡First number is for 1 in. diameter; second number is for larger knobs.
§Edge to edge.

essentially the same as those for the detent thumbwheel. (See Table 8.7.) Other considerations are:

1. Use a fluted or knurled peripheral surface to prevent slipping when operating the control.
2. Pay attention to the C/D ratio, so that it is not too high or too low, remembering that control feedback is obtained from a display and not the control itself.
3. Since the display is not itself contained in the control, the movement of the thumbwheel *up*, *forward*, or to the *right* should produce an *increase* or a corresponding movement in the display.
4. A visible dot on the peripheral surface should be used to indicate the "ON"-"OFF" boundary when less than one revolution of the thumbwheel is planned.
5. Normally a breakout force of 3 to 4 oz. should be required, with operating torque between 3 and 6 in.—oz., depending on the thumbwheel diameter.
6. Low friction and high inertia aid precise setting and slewing.

Handcranks

Cranks are effective in making adjustments on a continuum when large distances must be covered and high rates of turning are required; for slower rates, a knob or handwheel is more effective (Baines and King, 1950). Cranks require a medium-to-large amount of space for their location and operation (turning) because of the swept volume of the hand and crank arm. Because cranks are usually multi-rotational, the position of the crank handle generally does not indicate the control setting (position). Cranks can have an unlimited range of control movement, and, with proper gearing, can be used for either gross or fine positioning over a wide range of adjustments. (See Table 8-3.)

Cranks can be attached to knobs or handwheels to increase the versatility of those controls. Under no-load conditions, small cranks can be turned more rapidly than large ones. As the load increases, however, the crank size that maximizes turning rate also increases. (See Figure 8–17.) For rotating cranks at a constant rate, larger cranks (about 4.25 in. in radius) are better than smaller ones (Foxboro, 1943).

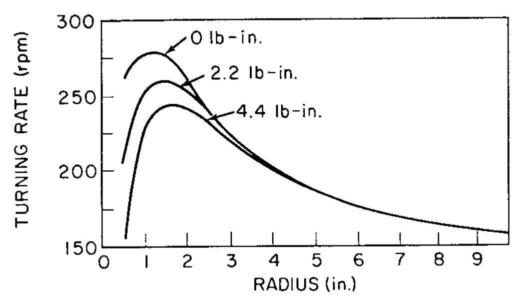


FIGURE 8–17. Turning rate as a function of crank-arm radius and load conditions.

Cranks can be coded by location, labeling, and color. Cranks used as emergency controls should be red. Accidental activation can be prevented by locking, location, and resistance.

Handcranks are turned most rapidly when the plane of rotation is vertical and parallel to the frontal plane of the operator (Lehmann, 1958). Speed of movement varies with control resistance and radius (slower speed with larger resistance and radius), but 180 r.p.m. is a good "average" speed (Helson, 1949, and Reed, 1949). Handcranks should be turned forward or clockwise for maximum speed and efficiency (Provins, 1953, and Baines and King, 1950). Two-hand, simultaneous, positioning movements, are fastest when reaching about 30° to the left and right of the midplane of the body (Barnes and Mundell, 1939).

Some design recommendations are given in Table 8–9; others are as follows:

1. Cranks, rather than knobs or handwheels, should be used for tasks involving two or more rotations of control movement; knobs and handwheels are better when the task requires less movement.

2. For tasks involving large slewing movements plus small fine adjustments, a crank handle may be mounted on a knob or handwheel for the gross adjustments, with the knob being used for fine adjustments. (An alternative practice would be to provide rate control rather than position control for such tasks.)

3. For small cranks (less than 3.5 in. radius), when rapid, steady turning is involved, the minimum resistance should be 2 lb. and the maximum 5 lb.

4. For large cranks (5- to 8-in. radius) where rapid, steady turning is involved, the minimum resistance should be 5 lb. and the maximum 10 lb.

5. For large cranks where precise settings are required (adjusting between one-half and one rotation), the minimum resistance should be 2.5 lb. and the maximum 8 lb.

6. The kind or kinds of resistance to be provided depends primarily on performance requirements. In general, however, any resistance will decrease the maximum rate of turning (about 275 r.p.m.). Friction of 2 to 5 lb. reduces the effects of shock (Hick, 1945), but friction degrades performance in rotating cranks at constant rates: primarily at low rates (3 to 10 r.p.m.), slightly at moderate rates (about 30 r.p.m.), and negligible at high rates (above 100 r.p.m.). Inertial resistance aids performance in rotating cranks at constant rates, particularly for small cranks and for low rates.

7. The crank handle should be designed so that it turns freely around its shaft.

8. The surface should be provided with a high degree of frictional resistance to prevent slipping.

TABLE 8–9. DESIGN RECOMMENDATIONS FOR CRANKS.

Parameter	Minimum	Maximum
Radius or length (in.)	0.5	4.5–20.0*
Handle diameter (in.)	1.0	3.0
Handle depth (in.)	3.0	----
Displacement	(‡)	
Resistance	(§)	----
	Minimum	**Preferred**
Control separation (in.):†		
One hand, randomly	2.0	4.0
Two hands, sequentially	3.0	5.0

Adapted from AFSCM 80–3.
*First number is for light loads; second number is for heavy loads.
†Edge to edge.
‡Determined by desired C/D ratio.
§See text.

Handwheels

Handwheels are designed for two-handed operation. Handwheels are useful for exerting greater rotary force than is possible with knobs and cranks, but they require a large amount of space for their location and operation. (See Table 8–3.) They can be coded by size, location, labeling, and color. Handwheels used as emergency controls should be red. Identification of control settings (positions) is poor or impossible if multiple rotations are permitted.

Accidental activation of handwheels is most easily prevented by locking the handwheel in place with a pin. Accidental activation may also be prevented by careful selection of location and by resistance.

Some design recommendations are given in Table 8–10; others are as follows:

1. For most effective use, handwheel displacement should not exceed ±60° from the normal (null) position because larger arcs require the hands to shift position on the control.

2. Displacement should be determined by the desired C/D ratio. When the handwheel must move through a large arc, the C/D ratio can be increased by increasing the handwheel diameter. In the latter situation, control movements are nearly linear, so that increasing the extent of control movement will increase the C/D ratio even though the arc of rotation is not increased.

3. The diameter of the handwheel rim should not exceed 0.75 to 2 in., and should ordinarily increase as handwheel diameter increases.

TABLE 8–10. DESIGN RECOMMENDATIONS FOR HANDWHEELS

Parameter	Minimum	Maximum
Size:		
Diameter (in.)*	7.0	21.0
Rim thickness (in.)	0.75	2.0
Displacement	(†)	‡90–120°
Resistance (lb.):		
Precision operation:		
Up to 3.5 in. radius	n.a.	n.a.
5.0 to 8.0 in. radius	2.5	8.0
Resistance at rim:		
One hand	5.0	30.0
Two hands	5.0	50.0
	Minimum	Preferred
Control separation (in.)§	3.0	5.0

Adapted from MIL STD–803A (USAF) and AFSCM 80–3.
*Two-hand grasp.
†Determined by desired C/D ratio.
‡Provided optimum C/D ratio is not hindered.
§Edge-to-edge separation.

4. The kind or kinds of resistance that should be provided depends, primarily, on performance requirements, but, for controls moving through small arcs, inertial resistance should be minimized, and, for aircraft handwheels, resistance should be elastic and increase nonlinearly from 5 to 30 lb.

5. Contour molding should be provided on the handwheel rim to aid in holding it. The surface of the rim of the handwheel should also be provided with a high degree of frictional resistance.

6. When the maximum displacement is less than 120°, only the two sections of the handwheel that the operator grasps need be provided. These parts are usually the chords of arcs, each approximately 6 in. long, across from one another. Eliminating the rest of the control increases the visual and pedal areas.

7. When large displacements must be made rapidly, a crank handle may be attached to the handwheel.

Levers

Levers include "joysticks," gear shifts, and controls such as aircraft throttles. Levers are usually designed to move when force is applied, but they also may be designed to remain fixed in one position. For these "rigid" (or "pressure") controls, the amount of force being applied is used as the input to the system. Both rigid and spring-loaded levers are characterized by their elastic resistance. Spring-loaded levers are generally preferred because their control positions (settings) can be determined visually and because they provide the operator with feedback information about both control position and resistance (thus giving him better "feel"). The primary advantage of rigid levers is that they permit a rapid response to be made and also require no extra space for displacement.

Spring-loaded levers require a medium-to-large amount of space for their location and operation; they are easily coded by color, size, labeling, location, or shape. Levers used as emergency controls should have red handles. Control setting (position) can be identified fairly well, both visually and nonvisually. Because they generally have a limited range of movement, however, levers are usually unsatisfactory for precise positioning over a wide range of adjustments.

The best length of a lever is a function of each specific situation and is often determined by the mechanical advantage that is needed. For making large fore-and-aft movements, a long lever is usually more desirable than a short one because the movements of a long lever are more nearly linear. Accidental activation can be prevented by locking, orientation, location, or resistance.

Some design recommendations are given in Table 8–11; others are as follows:

1. The maximum resistance for one-hand push-pull (fore-and-aft) movements with the control along the midline of the body is 30 to 50 lb., depending on how far away the control is from the body (the farther away, the greater the recommended resistance).

2. The maximum resistance for two-handed push-pull (fore-and-aft) movements is twice as much (60 to 100 lb.) as for one-handed operation.

3. The maximum resistance for one-handed right-left (lateral) movements is 20 lb.

4. The maximum resistance for two-handed right-left movements is 30 lb.

5. Levers are most effective when they move through an arc of not more than 90° but in any event, the range of movement should never exceed the convenient reach of the arm.

6. For making fine adjustments, support should be provided for the body part being used:

TABLE 8–11. DESIGN RECOMMENDATIONS FOR LEVER CONTROLS

Parameter	Minimum	Maximum
Handle diameter (in.):		
Finger grasp	0.5	3.0
Hand grasp	1.5	3.0
Length of grasp area (in.)	3.0	(*)
Length of lever	(†)	----
Displacement:		
Floor-mounted levers (in.):		
Fore–aft movements	none	14
Lateral movements.	none	38
Joysticks	none	60° from null position
Resistance:		
Finger grasp	12 oz.	32 oz.
Hand grasp	2 lb.	20–100 lb.‡
Control separation (in.):		
One hand random order	2.0	4.0
Two hands simultaneous	3.0	5.0
Maximum simultaneous operated one hand-span	----	6.0

Adapted from MIL STD–803A (USAF) and AFSCM 80–3.
*No limit set by operator performance.
†Depends on mechanical advantage; see text.
‡See text.

elbow support for large hand movements, forearm support for small hand movements, and wrist support for finger movements.

7. In making very fine adjustments with a small joystick, operators rest their wrists on the control panel and grasp the control, pencil-style, below the tip rather than on it. For such situations, the pivot point should be recessed below the surface on which the wrist rests.

8. When levers are used as discrete adjustment controls, if the lever arm is longer than 6 in., the minimum separation between control positions should be 2 in. When the control also serves as a visual indicator, control positions may be placed closer to each other, their minimum separation being largely determined by the operator's ability to see them. (See Figure 8–18.)

9. The lever should return to its null position when force is removed, that is, it should be self-centering.

10. The surface of the handle of a lever should be provided with a high degree of frictional resistance to prevent slipping.

Pedals

Pedals are of three major types: rotary, reciprocating, and translatory. (See Figure 8–19.)

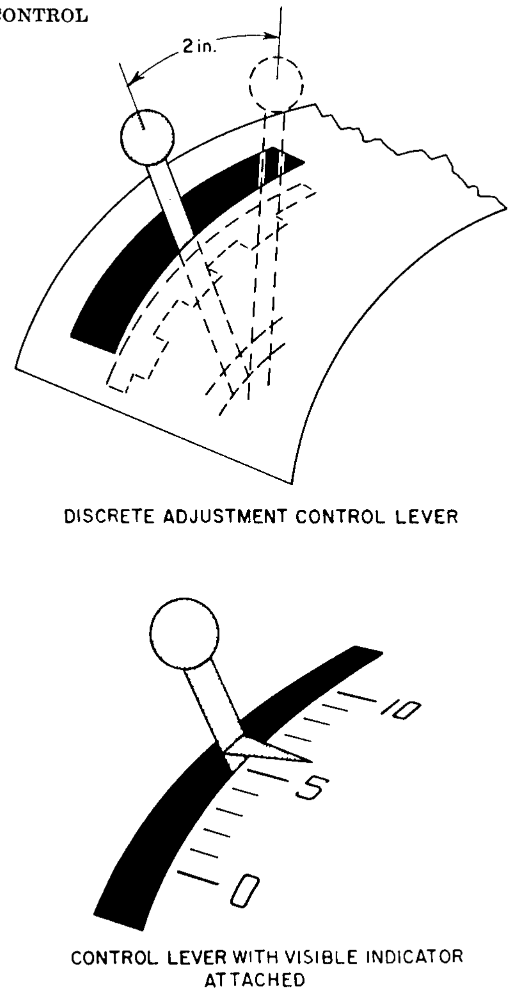


FIGURE 8–18. Examples of two ways of using a lever: for a discrete adjustment control and when a visual indicator is attached to the control.

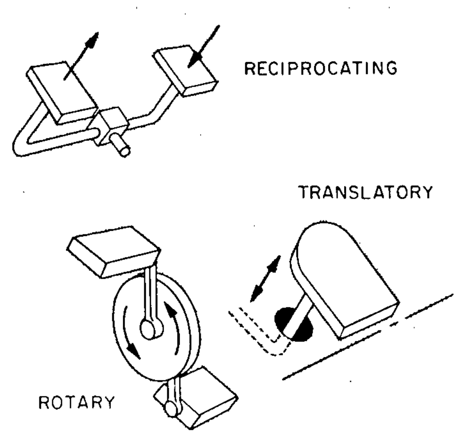


FIGURE 8–19. Three types of pedal controls.

They require a large amount of space for their location and operation, and, because they usually cannot be seen or felt (without danger of activating them), neither the controls nor their settings (positions) are easily identified. Pedals permit more force to be applied than do hand controls, but with less precision and speed. Rotary pedals have an unlimited range of control movements and, with proper gearing, can be used for either gross or precise positioning over a wide range of adjustments.

Pedals can be coded only by location. Accidental activation can be prevented by location and resistance.

Most design recommendations are given in Table 8–12; others are as follows:

1. The kind or kinds of resistance that should be provided depends, primarily, on performance requirements, but, in most situations, elastic resistance should be used and the pedal should return to its null position when force is removed.

2. Maximum pedal resistance should never exceed the maximum pressure exertable by the weakest operator. Accordingly, for male military personnel and for most male workers, resistance for leg-operated (as opposed to ankle-operated) pedals should not exceed 200 lb. for a single, brief application (Elbel, 1949, and Hugh-Jones, 1947).

3. For frequently but not continuously used leg-operated pedals, a force of about 30% of the maximum exertable is reasonable (Dupuis et al., 1955).

4. For ankle-operated pedals in continuous use, such as an automobile accelerator, maximum resistance should be about 10 lb. (Dupuis, 1958).

5. The minimum resistance for leg-operated pedals, when the leg normally rests on the pedal and the other leg does not rest on an opposing pedal, should be about 10 lb. to exceed by a good safety margin the 7 lb. average force exerted on the pedal by the weight of the leg alone (Orlansky, 1948).

6. Minimum resistance for ankle-operated pedals may be about 4 lb. less than that for leg-operated pedals.

7. The optimum range of resistance for leg-operated pedals is 8 to 60 lb. (Orlansky, 1948).

8. For ankle-operated pedals, the optimum resistance is 6.5 to 9 lb. (Lehmann, 1958).

TABLE 8–12. DESIGN RECOMMENDATIONS FOR PEDALS

Parameter	Minimum	Maximum
Size:		
Length (in.)	3.5	(*)
Width (in.)	1.0	(*)
Displacement (in.):		
Normal	0.5	----
Heavy boot	1.0	----
Ankle flexion	----	2.5
Leg movement	----	7.0
Resistance (lb.):		
Foot rested off control	4	----
Foot rested on control	10	----
Ankle flexion	----	10
Leg movement	----	180
	Minimum	**Preferred**
Control separation (in.):		
One foot randomly	4	6
One foot sequentially	2	4

Adapted from MIL STD–803A (USAF).
*Depends on space available.

9. Pedals operated by the entire leg generally should have a 2- to 4-in. displacement, except for an automobile brake type of pedal, for which an additional 2- to 3-in. of travel may be added (Dupuis et al., 1955). Caution: displacements of 3 to 4 in. or more should never be coupled with resistances of less than 10 lb.

10. Pedals operated by ankle action should have a maximum travel of about 2 in., corresponding to a motion angle of about 10°. Caution: no motion angle greater than 30° should be used because this is about half the total range of ankle movement (Dupuis et al., 1955, and Dupuis, 1958).

11. Since heavy footgear makes it difficult to gauge pedal travel, excessive movement and force usually result. Under such conditions, pedal travel should be increased (Clark and Weddell, 1944). The minimum should be increased by at least 0.5 in. (Ely et al., 1956).

12. The angulation of most pedals operated by leg action at the hip and knee should permit the foot to be placed on the pedal surface with the ankle at a 90° angle, but pedal angulation will vary considerably with vertical and fore-and-aft pedal location. For example, it has been found that the forces applied to aircraft brake pedals dropped off sharply as the pedal angle (with the vertical) decreased below 20° or increased above 40° and maximum forces were at pedal angles of about 30° (Hertzberg, 1954).

13. The angulation of most pedals operated by ankle action also varies widely with vertical and fore-and-aft location. In general, however, the angle of pedal surface should permit the foot and lower leg to form an angle of at least 90° but never more than 130°. Foot–leg angles of less than 90° should be avoided, except when greater forces are needed for brief intervals.

14. Most pedals should be as wide, or almost as wide, as the sole of the shoe, i.e, at least 3.5 in. The maximum width matters little as long as there is enough clearance between adjacent pedals.

15. Pedals used intermittently or for short periods should be at least 3 in. long. Pedals used continuously or for long periods should be 11 to 12 in. long.

16. Pedal shape can be square, rectangular, circular, or oval as long as it is flat and affords enough area of contact with the shoe.

17. For pedals with which large forces must be exerted, i.e., 200 lb. or more, a pedal bar (or recessed heel section) will prevent the foot from slipping off the pedal and will assist the operator in locating the pedal by feel. (This is particularly advantageous in cold weather or with large, heavy boots.)

8.6 Unusual Control Methods

The hands and feet are the only parts of the body that have been used routinely for operating controls. According to an old principle of psychology, however, any stimulus that an organism can sense can be linked to any response that the organism is capable of making. Theoretically, therefore, any of a very large number of body responses—movements of the knees, hips, elbows, shoulders, head, eyes, mouth, and tongue, for example—could be used to actuate control devices. In fact, knee levers have been used for years as a standard control on certain sewing machines and pipe organs. In the main, however, these unconventional control possibilities have been largely ignored in man–machine systems.

Situations occasionally arise when an operator may not be able to use his hands or feet, either because they are already fully occupied or because they are largely immobilized, as, for example, that of an astronaut who is encased in a highly restrictive pressure suit. In addition, some people lack the use of one or more limbs either because they are multiple amputees or are paralyzed. For these reasons there is interest in the possibilities of unconventional control methods. This section considers some of these control methods and summarizes what little is known about them from the human factors point of view.

It is important at the outset to distinguish between the purely *engineering* and the *human engineering* aspects of unusual control methods. A nod of the head, for example, can be used to activate an accelerometer. The coupling between the operator's head and the mechanical sensing device that responds to his head movements is clearly a human factors problem. The output of the accelerometer, in turn, can be used to do work or produce any of a large number of effects: start a motor, summon an elevator, open a refrigerator door, or turn on a television set. The latter are, almost exclusively, straightforward engineering problems. In this section we shall be concerned only with the human engineering aspects of unconventional control devices—the linkages between body responses and the corresponding mechanisms that serve as controls. We shall not look at what one can do with the outputs of the controls after they have been activated.

Since there have been practically no studies at all on unconventional control methods, this section is based largely on analytical studies. Given the present state of our knowledge, extensive engineering development, or more detailed human engineering studies, or both, would be required to make such systems operational.

8.6.1 A Survey of Some Unconventional Control Possibilities

Table 8–13 compares various unconventional control methods that were analytically studied for possible usefulness in a unit maneuvered by an astronaut. The control methods are compared in the following ways:

1. Mechanization. The kind of mechanism that would probably be used to sense the human body response.

TABLE 8-13. ANALYTICAL COMPARISON OF SOME UNCONVENTIONAL CONTROL METHODS

Concept		Characteristic							
Controller	Type	Mechanization	Output	Location	Command capability	Provides hand freedom	Accessibility	Natural direction of operation	Accuracy
Hand	On chest or side	Electro-mechanical pencil stick.	Continuously variable or on-off.	On chest, stomach or hip.	Complete	One hand free	Good for both hands.	Yes	± 1° or less.
	In auxiliary glove.	Possibly push buttons.	Continuously variable or on-off.	Back of glove	Complete	One hand free	With one controller for each hand, accessibility good for either hand.	No	± 1° if continuously variable; on-off N/A.
Mouth	Voice	"Audrey," "Sceptron," "Shoe Box," etc.	On-off or incremental.	At throat	Complete	Both hands free	Good	No	N/A.
	Tone	Resonant transducers.	On-off or incremental.	At throat	Complete	Both hands free	Good, depending upon musical ability of operator.	No	N/A.
	Breath	Sensitive diaphragms in "mouth organ" configuration.	On-off or cremental.	In front of mouth.	Complete	Both hands free	Good	No	N/A.
	Tongue	Switches, ports, slide-wires, etc.	On-off or continuously variable.	On lips or in mouth.	Unknown	Both hands free	Good	Perhaps	N/A.
Eye	Reflected beam.	Light beam reflected from cornea.	Continuously variable.	At eye	Forward translation only.	One hand may be required for supplemental control and switching.	Good	Yes	±10 min at center of field, ±1° at edge of field.
	Corneal-retinal potentials.	Electrical potential across eye.	Continuously variable.	At eye	Forward translation only.	One hand may be required for supplemental control and switching.	Good	Yes	±1 − 2°.
	Muscle action potentials.	Electrical signals in eye muscles.	Continuously variable.	At eye	Forward translation only.	One hand may be required for supplemental control and switching.	Good	Yes	Unknown.

TABLE 8–13. ANALYTICAL COMPARISON OF SOME UNCONVENTIONAL CONTROL METHODS—

Concept		Characteristic								
Controller	Type	Cross coupling between rotational axes	Motion coupling	Acquisition actuation	Inadvertent actuation	Type of feedback	Response time (sec)	Reliability	Size*	Weight*
Body	Head	Electrical or mechanical pickoffs sensing position of head relative to body.	Continuously variable.	Sight at eye; pickoffs at neck or on head.	Forward translation only or rotation only.	One hand may be required for control or switching.	Good	Yes	With sight ± 3 mils, without sight ± 1°.	
	Limb motion	Force or displacement sensors attached to limb.	Continuously variable or on-off.	At controlling limb.	Can be complete.	One hand may be required for control or switching.	Good	Perhaps	± 1 − 2°.	
	Myoelectronics	Skin electrodes sensing muscle action potentials.	On-off or incremental.	At controlling limb.	Can be complete.	Both hands free	Good	Perhaps	N/A.	
Hand	On chest or side	Not significant with small stick excursions.	Not significant at small accelerations.	Probable	Not probable except in cramped quarters.	Visual, force	1.0 − 1.5 (0.5 − 1.0 reach plus 0.5 operate.)	Good. Can be simple, rugged.	1	1
	In auxiliary glove.	Not significant	Not significant	None if properly designed.	Not probable	Visual, possibly force.	1.5 − 2.5 (1.0 − 2.0 reach plus 0.5 operate).	Good. Simple, rugged, repeatable.	1	1
Mouth	Voice	Not possible	None	None	Not probable if unique sounds reserved for commands.	Visual	0.5	Good. Simple, repeatable.	5	4
	Tone	Not possible	None	None	Not probable	Visual	0.5	Good. Simple, repeatable.	3	2
	Breath	Not possible	None	None	Not probable if sensor thresholds are high enough.	Visual	0.5	Good. Simple, repeatable.	3	2
	Tongue	Not possible	None	Possible	Possible	Visual	0.5	Probably good	3	2

TABLE 8–13. ANALYTICAL COMPARISON OF SOME UNCONVENTIONAL CONTROL METHODS—Continued

Concept		Characteristic								
Controller	Type	Cross coupling between rotational axes	Motion coupling	Acquisition actuation	Inadvertent actuation	Type of feedback	Response time (sec)	Reliability	Size*	Weight*
Eye	Reflected beam	Not possible	None	None	Not probable if lockout switch included.	Visual	0.1	Fair. Excessive equipment, tendency to misalignment.	5	3
	Corneal-retinal potentials.	Not possible	None	None	Not probable	Visual	0.1	Fair. Extraneous signals, spurious variations, frequent miscalculations.	5	3
	Muscle action potentials.	Not possible	None	None	Not probable	Visual	0.1	Fair. Variable contact resistance, excessive electronics.	5	3
Body	Head	Not possible	Not significant	None	Possible unless lockout is used.	Visual	0.5. Movement plus fixation time.	Probably good	1.5	1.5
	Limb motion	Possibly some	Possible; not probable at low thrust levels.	None	Possible	Visual	0.3 to 0.5	Probably good	1.5	1.5
	Myoelectronics	Not significant	Possible	None	Possible	Visual	0.2	Fair. Variable contact resistance, excessive electronics.	5	3

*Normalized with respect to hand controllers.
Adapted from Lovinger.

2. Output characteristics. Whether the output is continuous or discrete, and, if discrete, whether it is simply an on-off or an incremental response, that is, one with various steps possible.

3. Location. The most likely location to put the sensing mechanism on the body.

4. Command capability. Whether the control would permit complete movement over three degrees of freedom in rotation, three degrees of freedom in translation, and switching.

5. Hand freedom. Whether both hands would be free or one would be required for controlling or switching.

6. Accessibility. Whether the control can be made easily accessible to the actuating part of the body.

7. Compatibility. Whether the resulting system movement can be made compatible with the movement of the operator.

8. Accuracy. The precision with which the controlling body movements can be made.

9. Cross coupling. Whether movements of the control in one direction could be confusing to the operator and so cause him to make incorrect control movements in other directions.

10. Motion coupling. Whether movement of the system in any direction would be likely to set up inertial forces that would tend to cause inadvertent actuation of the control.

11. Acquisition actuation. Whether activation of the control would be accompanied by jostling or fumbling.

12. Inadvertent actuation. Whether the control would likely be actuated inadvertently when it is not in use.

13. Type of feedback. The kind of feedback the operator would get from the control.

14. Response time. The time required to reach and actuate the control, that is, to generate an electrical command signal to the system.

15. Reliability. An estimate of the simplicity, ruggedness, and repeatability of the body movement and control combination.

16. Size and weight. An estimate of the size of the control mechanism as compared with a comparable hand control.

8.6.2 A Head-Actuated Control System

One unconventional control mechanism that has been developed and put to use is a head-actuated control system. The system was designed to permit a quadriplegic to control a motor-driven wheelchair and several appliances, such as a tape recorder. The principal human engineering requirements were that the system must be safe, reliable, inconspicuous, detachable from its user, and usable by similarly handicapped persons.

The system makes use of four simple, unidirectional, angular accelerometers that can be actuated by nods of the user's head. The accelerometers are spaced 90° apart on a helmet. An enabling switch, a switch that turns the system on, is also mounted in the helmet. The purpose of the enabling switch is to prevent incidental head movements from operating the equipment.

In practice, the operator enables the system and then nods his head forward. This signal is applied to the motors in the wheelchair causing it to move forward. To stop, the operator may disable the system or brake it by nodding his head backward. A forward left turn is made by nodding the head midway between forward and left (315°) while the system is enabled. Different head commands produce other turns and maneuvers.

Of the eight possible commands that can be generated by nods of the head (forward, right forward, right, right rear, rear, left rear, left, left forward), six are used for controlling the chair. A nod directly to the left without tilting the head forward or backward sounds a horn. A similar nod to the right triggers an eight-channel switch, disconnects the system from the chair motors and connects it to the appliances. A second nod to the right steps the switch to the first appliance, a third nod steps the switch to the next appliance, and so on. A seventh nod restores control to the motors in the chair.

Experience shows that such a control system can easily be learned and operated. This type of control system could, therefore, be used to actuate many other kinds of mechanisms.

References

Baines, R. M., & King, E. S. A study in the relationship between maximum cranking speed and cranking radius. *Motor Skills Research Exchange*, 1950, 2, 24.

Barnes, R. M., & Mundell, M. E. A study of simultaneous symmetrical hand motions. *Studies in Engineering*, Bulletin No. 17, 1939, University of Iowa, Iowa City.

Barnes, R. M., et al. Which pedal is best? *Factory Management and Maintenance*, 1942, **100**, 98.

Bradley, J. V. Desirable control-display relationships for moving-scale instruments. Report No. WADC-TR-54-423, 1954 b, Wright Air Development Center, WPAFB, Ohio.

Bradley, J. V. Glove characteristics influencing control manipulability. Report No. WADC-TR-57-389, 1957, Wright Air Development Center, WPAFB, Ohio.

Bradley, J. V., & Arginteanu, J. Optimum knob diameter. Report No. WADC-TR-56-96, 1956, Wright Air Development Center, WPAFB, Ohio.

Bradley, J. V., & Stump, N. E. Minimum allowable dimensions for controls mounted on concentric shafts. Report No. WADC-TR-55-355, 1955, Wright Air Development Center, WPAFB, Ohio.

Bradley, J. V., & Wallis, R. A. Spacing of on-off controls: II. Toggle switches. Report No. WADC-TR-59-475, 1959, Wright Air Development Center, WPAFB, Ohio.

Brennan, T. N. N., & Morant, G. M. Selection of knob shapes for radio and other controls. Report No. FPRC-702, 1950, Flying Personnel Research Committee, RAF Institute of Aviation Medicine, Farnborough, England.

Brown, J. S., & Slater-Hammel, A. T. Discrete movements in the horizontal plane as a function of their length and direction. *Journal of Experimental Psychology*, 1949, **39**, 84.

Chambers, A. N., Preusser, P. B., Lenzycki, H. P., & Wahl, J. G., Human factors design standards for the fleet ballistic missile weapon system, Volume 1: Design of systems; Volume 2: Design of equipment. Report No. OD 18413, 1960–1962, Bureau of Naval Weapons. (Also available as a U.S. Government Printing Office document.)

Chapanis, A. Studies of manual rotary positioning movements: I. The precision of setting an indicator knob to various angular positions. *Journal of Psychology*, 1951 (b), **31**, 51.

Chapanis, A. Studies of manual rotary positioning movements: II. The accuracy of estimating the position of an indicator knob. *Journal of Psychology*, 1951 (c), **31**, 65.

Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied Experimental Psychology*. New York: John Wiley and Sons, 1949.

Clark, W. E. leG., & Weddell, G. The pressure which can be exerted by the foot of a seated operator with the leg in various positions. Report No. 44/153, 1944, Royal Naval Personnel Research Committee, Medical Research Council, London, England.

Craik, K. J. W., & Vince, M. A. A note on the design and manipulation of instrument knobs. Report No. APU-14, 1945, Applied Psychology Research Unit, Medical Research Council, Cambridge, England.

Dempster, W. T. The anthropometry of body action. *Annals of the New York Academy of Science*, 1955, **63**, 559.

Dupuis, H. Some standards for the design of the tractor driver's workplace, a paper presented at the annual meeting of the American Society of Agricultural Engineers in Chicago, 1958.

Dupuis, H., et al. Zweckmässige Gestaltung des schlepperführerstandes *Schriftenreihe Landarbeit und Technik*, 1955, 20.

Elbel, E. R. Relationship between leg strength, leg endurance and other body measurements. *Journal of Applied Physiology*, 1949, **2**, 197.

Ely, J. H., Thomson, R. M., & Orlansky, J. Layout of workplaces. Report No. WADC-TR-56-171, 1956, Wright Air Development Center, WPAFB, Ohio.

Eriksen, C. W. Multidimensional stimulus differences and accuracy of discrimination. Report No. WADC-TR-54-165, 1954, Wright Air Development Center, WPAFB, Ohio.

Foxboro Company. Handwheel speed and accuracy of tracking. Report No. 3453, 1943, National Defense Research Committee, Office of Scientific Research and Development, Washington, D.C.

Harkins, W. H. The KB System: A modular control panel concept. *Human Factors*, 1966, **8**, 291.

Helson, H. Design of equipment and optimal human operation. *American Journal of Psychology*, 1949, **62**, 473.

Hertzberg, H. T. E. Unpublished data. 1954.

Hertzberg, H. T. E. Some contributions of applied physical anthropology to human engineering. *Annals of the New York Academy of Science*, 1955, **63**, 616.

Herbert, M. J. Speed and accuracy with which six linear arm movements can be visually positioned from two different control locations. Report No. 260, 1957, U.S. Army Medical Research Lab., Fort Knox, Ky.

Hick, W. E. Friction in manual controls with special reference to its effect on accuracy of corrective movements in conditions simulating jolting. Report No. APU-18, 1945, Applied Psychology Research Unit, Medical Research Council, Cambridge, England.

Hugh-Jones, P. The effect of limb position in seated subjects on their ability to utilize the maximum contractile force of the limb muscles. *Journal of Physiology*, 1947, **105**, 332.

Hunt, D. P. The coding of aircraft controls. Report No. WADC-TR-53-221, 1953, Wright Air Development Center, WPAFB, Ohio.

Hunt, D. P., & Craig, D. R. The relative discriminability of thirty-one differently shaped knobs. Report No. WADC-TR-54-108 1954, Wright Air Development Center, WPAFB, Ohio.

Jenkins, W. L. Design factors in knobs and levers for making settings on scales and scopes. Report No. WADC-TR-53-2, 1953, Wright Air Development Center, WPAFB, Ohio.

Jenkins, W. L. & Connor, M. B. Some design factors in making settings on a linear scales *Journal of Applied Psychology*, 1949, **33**, 395.

Jenkins, W. L. & Karr, A. C. The use of a joystick in making settings on a simulated scope face. *Journal of Applied Psychology*, 1954, **38**, 457.

Jenkins, W. L. & Olson, M. W. The use of levers in making settings on a linear scale. *Journal of Applied Psychology*, 1952, **36**, 269.

Jenkins, W. L., Maas, L. O., & Rigler, D. Influence of friction in making settings on a linear scale. *Journal of Applied Psychology*, 1950, **34**, 434.

Jenkins, W. L., Maas, L. O., & Olson, M. W. Influence of inertia in making settings on a linear scale. *Journal of Applied Psychology*, 1951, **35**, 208.

Knowles, W. B. & Sheridan, T. B. The "feel" of rotary controls: Friction and inertia. *Human Factors*, 1966, **8**, 209.

Konz, S. A. & Day, R. A. Design of controls using force as a criterion. *Human Factors*, 1966, **8**, 121.

Lehmann, G. Physiological basis of tractor design. *Ergonomics*, 1958, **1**, 197.

Lovinger, D. N. A Study of Controllers for an Astronaut Maneuvering Unit. In *National Conference on Space Maintenance and Extra Vehicular Activities*, sponsored by U.S. Air Force Aeropropulsion Labs., Research and Technology Division, Orlando, Florida. Mar. 13, 1966.

MIL-C-10812. U.S. Department of Defense, Washington, D.C.

MIL-STD 803A (USAF) U.S. Department of Defense, Washington, D.C. (Now superseded by MIL-STD 1472)

Müller, E. A. Der beste Handgriff und Stiel. *Arbeitsphysiologie*, 1935, **8**, 28.

Orlansky, J. The human factor in the design of stick and rudder controls for aircraft. Report No. SPECDEVCEN-151-1-8, 1948, U.S. Naval Training Devices Center, Port Washington, N.Y.

Orlansky, J. Psychological aspects of stick and rudder controls in aircraft. *Aeronautical Engineering Review*, 1949, **8**, 22.

Provins, K. A. A study of some factors affecting speed of cranking. Report No. 53/755, 1953, Royal Naval Personnel Research Committee, Medical Research Council, London, England.

Provins, K. A. Maximum forces exerted about the elbow and shoulder joints on each side separately and simultaneously. *Journal of Applied Physiology*, 1955, **7**, 390.

Reed, J. D. Factors influencing rotary performance. *Journal of Psychology*, 1949, **28**, 65.

Rockway, M. R. The effect of variations in control-display ratio and exponential time delay on tracking performance. Report No. WADC TR-54-618, 1954, Wright Air Development Center, WPAFB, Ohio.

Searle, L. V. & Taylor, F. V. Studies of tracking behavior: I. Rate and time characteristics of simple corrective movements. *Journal of Experimental Psychology*, 1948, **38**, 615.

Stump, N. E. Manipulability of rotary controls as a function of knob diameter and control orientation. Report No. WCRD-TN-53-12, 1953, Aero Medical Lab., Wright Air Development Center, WPAFB, Ohio.

U.S. Air Force Systems Command. Handbook of instructions for aerospace personnel subsystem design. Report No. AFSCM 80-3.

U.S. Army Materiel Command. Engineering design handbook: Maintainability guide for design. Report No. AMCP 706-134, August, 1967, U.S. AMC, Washington, D.C.

Vince, M. A. The intermittency of control movements and the psychological refractory period. *British Journal of Psychology*, 1940, **38**, 149.

Woodson, W. E. & Conover, D. W. *Human engineering guide for equipment designers.* (2nd Ed.) Berkeley: University of California Press, 1966.

Chapter 9

Design of Individual Workplaces*

This chapter provides principles and data for optimum workplace design. These include requirements for visual and physical access to equipment and displays, as well as the constraints that these requirements place on hardware design; a general approach to planning workplaces; and a survey of general and specific guidelines. Specific information about displays, controls, and body dimensions needed for workplace design is found in other chapters devoted to these topics. When multi-man workplaces are to be considered, the information of this chapter should be integrated with that presented in Chapter 10.

CONTENTS

* This chapter is based in part on Chapter 7 of the previous *Guide* and on original material prepared by Wesley E. Woodson. Maurice P. Ranc, Jr., and Donald W. Conover also made significant contributions to the chapter, which was reviewed by Howard W. Stoudt.

9. Design of Individual Workplaces

9.1 Approach to Workplace Design

Ideal workplace design will be compatible both with expected user and with system performance requirements. It will also reflect a knowledge of factors likely to affect equipment operations such as user capabilities and limitations. When design conceptualization occurs, other questions should be asked in addition to the obvious ones of, "Who will be the specific user of the work area?", or "How will the work area be used?" For example, "What are the specification requirements?" "Is proper attention paid to maintenance as well as operational functions?" and "How will layout design affect technical publications' content and training programs? and "How can a satsifactory layout be obtained with the least cost?"

The general design approach as outlined by Ely et al. (1956) and Thompson et al. (1958) provides the basis for workplace development. In this approach, the design is kept "tentative" until all parts have been defined. After establishing an explicit set of requirements based on performance functions, physical constraints, and specifications, the user population must be identified and the workplace designed for it. The following rules are useful:

1. Plan the whole, then the detail.
2. Plan the ideal, then the practical.
3. Plan the process and equipment around the system requirements.
4. Plan the layout around the process and equipment.
5. Plan the final enclosure around the layout.
6. Use the mockups to evaluate alternate layouts and to check final layout.

A design team receiving an assignment for workplace definition in a system complex must usually work from input data limited to certain generalities, specifications, and perhaps, initial task analyses and personnel/equipment data. As design progresses, more detailed information becomes available, but the designer has less freedom for workplace modification. Early consideration of system requirements thus becomes imperative and designers are cautioned to obtain all facts regarding operational mission, maintenance requirements, and physical and environmental constraints before starting layout conceptualization.

9.2 Layout of Workplaces

1. Define what the operator must be able to see:

a. Outside his work station (or vehicle).
b. Within his workplace (panel-mounted displays and controls).
c. Other people.
d. Other equipment (displays, recorder tapes, reels, status boards, etc.).

2. Consider what the operator must be able to hear:

a. Direct oral communications with adjacent operators.
b. Signals from loudspeakers, earphones.
c. Warning bells, sirens, etc.
d. Equipment operation, e.g., auditory sound unique to certain hardware systems.

3. Specify what the operator must reach and manipulate:

a. Hand and foot controls.
b. Latches.
c. Restraint harness, fasteners, adjustments.
d. Seat adjustments, optic adjustments, canopy/cover openings.
e. Emergency items, e.g., flashlight, survival gear, fuse, light bulb.

4. Determine body clearance:

a. Possibility of the operator bumping elbows, knees, head, etc., during both nominal and emergency exit and/or crash (ejection).

b. Possibility of inadvertent snagging or disturbance of controls or handles.

c. Grouping of related instrument/control system elements and interacting or related system groups (arrangement relative to each other).

d. Relationship of an operator workplace with other operator stations.

By following the general design approach and basic information requirements described above, designers can use the principles and practices described in the remainder of this chapter to develop workplace layouts for specific applications.

9.3 General Principles

When exact workplace configurations have yet to be determined, and it is necessary to establish overall space requirements for both operators and equipment to define room or van dimensions, suggested estimates indicate a minimum of 2.5 ft. by 2.5 ft. (6.25 sq. ft.) of floor space requirement for seated workers at a fixed station. Taking into account the vertical space needed, a total of about 30 $ft.^3$ should be provided (i.e., 4.8 ft. in vertical dimensions). Since greater area may be required, the figure is only a baseline for the typical 95th-percentile military operator. Human body dimensional ranges and definitions of percentile grouping are to be found in Chapter 11.

Figure 9–1 illustrates certain principles applying to workplace layout (Woodson and Conover, 1964).

1. To avoid unnecessary retraining and operator error because of established habit patterns, maintain relative placement of controls and displays for similar models or types of equipment.

2. Distribute workload as evenly as possible among hands and feet. Ordinarily, primary controls requiring precision should be given to the right hand. Emergency controls should be equally available to either hand.

3. Utilize anthropometric data within the 5th- to 95th-percentile range. (See Chapter 11.) When identifying the arm-reach envelope for location of controls, consider limits imposed by operators with shorter arms. For clearance requirements for head, knees, etc. choose data from larger members of the user population. For visual functions, accommodate the total range of potential eye positions, i.e., seat-to-eye height of both smaller as well as larger seated operators (consider adjustable seats where possible) and eye-height of standing operators.

4. Anticipate all safety hazards and required emergency actions in advance before starting to design.

9.3.1 Functional Factors

Two functional considerations in good workplace layout, visibility and clearance, are both related to an operator's anthropometric and biomechanical characteristics. Procedural efficacy, a third factor, is related to perception and reaction. A fourth factor is accessibility to displays, controls and work surfaces, and storage areas.

9.3.2 Visibility

The designer must consider primary and secondary visibility factors. "Out of the window observation" might define a primary visual task; a secondary task would be to monitor instruments or lights inside the workplace (such as a cockpit). The primary visual function normally establishes a workplace layout's principal orientation with other visual functions integrated into it.

9.3.3 Clearance

Clearance at various levels is important for access to and from the workplace, for ease in grasping and operating controls, for ease in adjusting the body properly to the visual-control task, and for isolation of the operator from physical discomfort or injury. All of these factors may be influenced to a great extent by the manner in which the operator is restrained and by the special clothing he may be required to wear to insure his safety and/or to provide life support. In establishing clearance requirements,

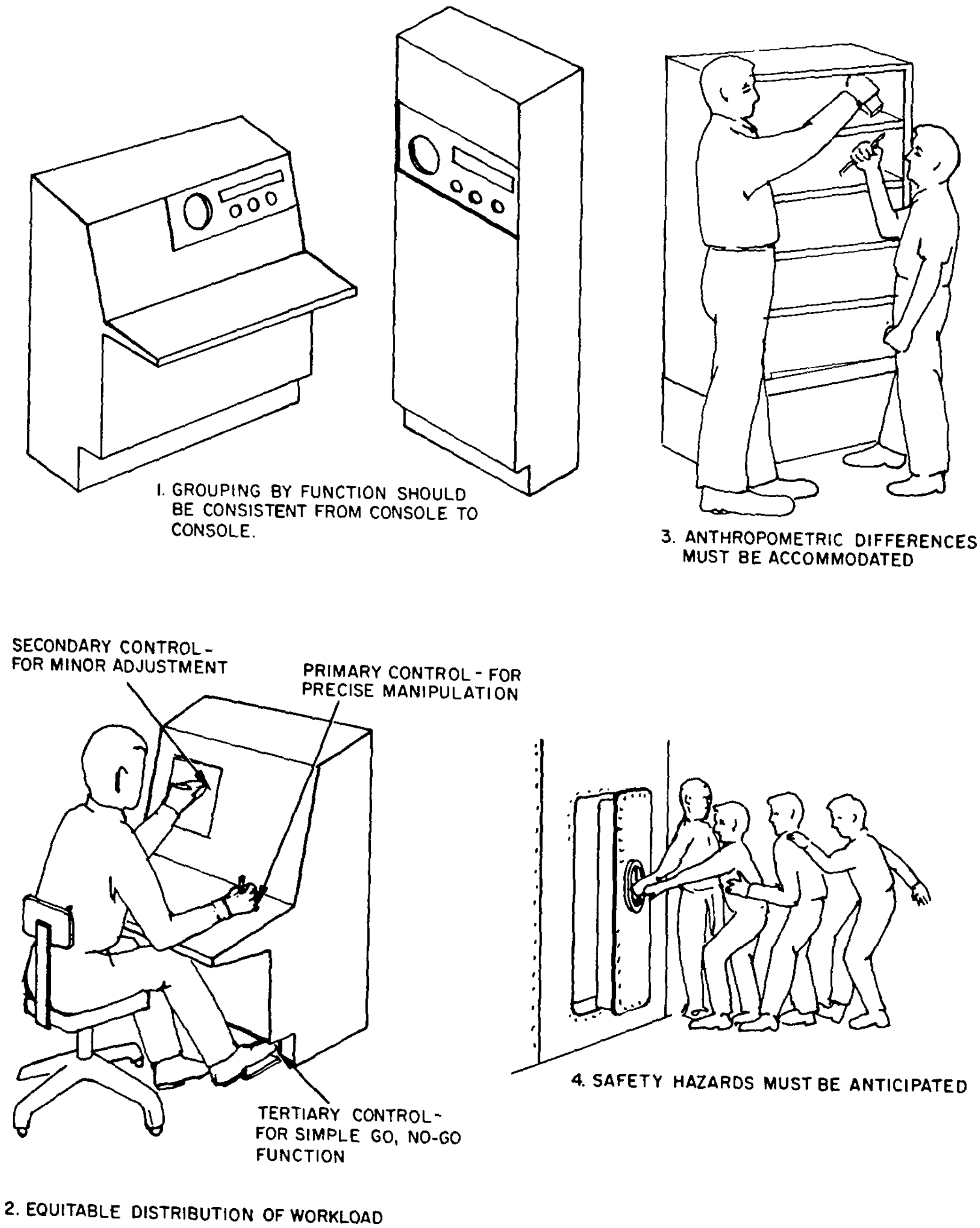


FIGURE 9–1. General workplace design principles (Woodson and Conover, 1964).

whether related to access, control manipulation, body position, or operator isolation, the designer must recognize the specific needs of the user.

9.3.4 Procedural Efficacy

Procedural efficacy is dependent upon logical arrangements of task elements to enhance operator performance and minimize error. Proper associations between controls and displays, logical grouping of instruments and controls by subsystem, and arrangement for following a natural procedural sequence, contribute to a coherent, compatible man–machine relationship.

9.3.5 Physiological Factors

Workplace design must maintain physiological systems within acceptable limits. In addition to recognizing gross personal hazards by providing crash protection, for example, designers should also be responsive to more subtle physiological stresses arising from simple design incongruities, i.e., (a) lack of postural control, (b) improper distribution of body weight, (c) cardiovascular restriction, and (d) fatigue-inducing activity. Physiological factors, environmental stresses, and design implications interact. Major relationships are summarized in Table 9-1 and illustrated in Figure 9–2.

TABLE 9–1. RELATIONSHIPS BETWEEN OPERATOR PHYSIOLOGICAL SYSTEMS, ENVIRONMENT, AND WORKPLACE LAYOUT

Considerations	Physiological systems affected		
	Musculo-skeletal	Cardio-vascular	Gastro-intestinal
Environments:			
Vibration	2	3	3
Oscillation	2	2	1
Acceleration	3	2	3
Impact	1	2	2
Noise	1	3	3
Workplace layout features:			
Improper postural support	1	2	2
Poor distribution of operator's body/limb weight	2	2	3
Awkward body or limb positions	1	3	3
Frequent requirement to use maximum reach or force	1	3	3

Rating criteria: 1 = Critical*, 2 = Important, 3 = Minor.

*Sound reasons can be presented for differential weighting of the above factors in terms of specific condition relationships; however, a good rule of thumb to follow is to recognize gastrointestinal and cardiovascular disturbance minimization should take precedence over musculoskeletal considerations, especially in systems requiring long-term operator exposure.

9.3.6 Psychological Factors

A primary psychological objective of workplace design should be to create user acceptance. An operator will be motivated if his workplace is (a) well organized, (b) convenient, (c) simple, (d) reliable, (e) safe, and (f) attractive. He will be frustrated if the workplace is completely disorganized, inconvenient, or unattractive. If an operator has difficulty getting into position, or seeing certain displays, or reaching and monitoring controls because of poor arrangement, his motivation to work will be reduced. Figure 9–3 illustrates the range of negative results from various workplace problems, all the way from discomfort to fatigue to sickness, injury, incapacitation, or even death.

In performing tasks, the operator's sensory, cognitive, and psychomotor attributes are influenced by the design of the workplace. For an operator to perceive and respond efficiently, he needs availability and compatibility of displays and of control output devices.

9.3.7 Dimensional Factors

Workplace dimensions should also be compatible with anthropometric characteristics of anticipated operator populations, for which basic data are presented in Chapter 11. A dynamic evaluation can be accomplished by mocking up designs in full scale using "real" people who represent size extremes and can conduct simulated operations. Such mockups, varying in sophistication from nonfunctioning cardboard and paper to functioning "breadboard" designs, will uncover problem areas seldom evident from scaled drawings of the workplace.

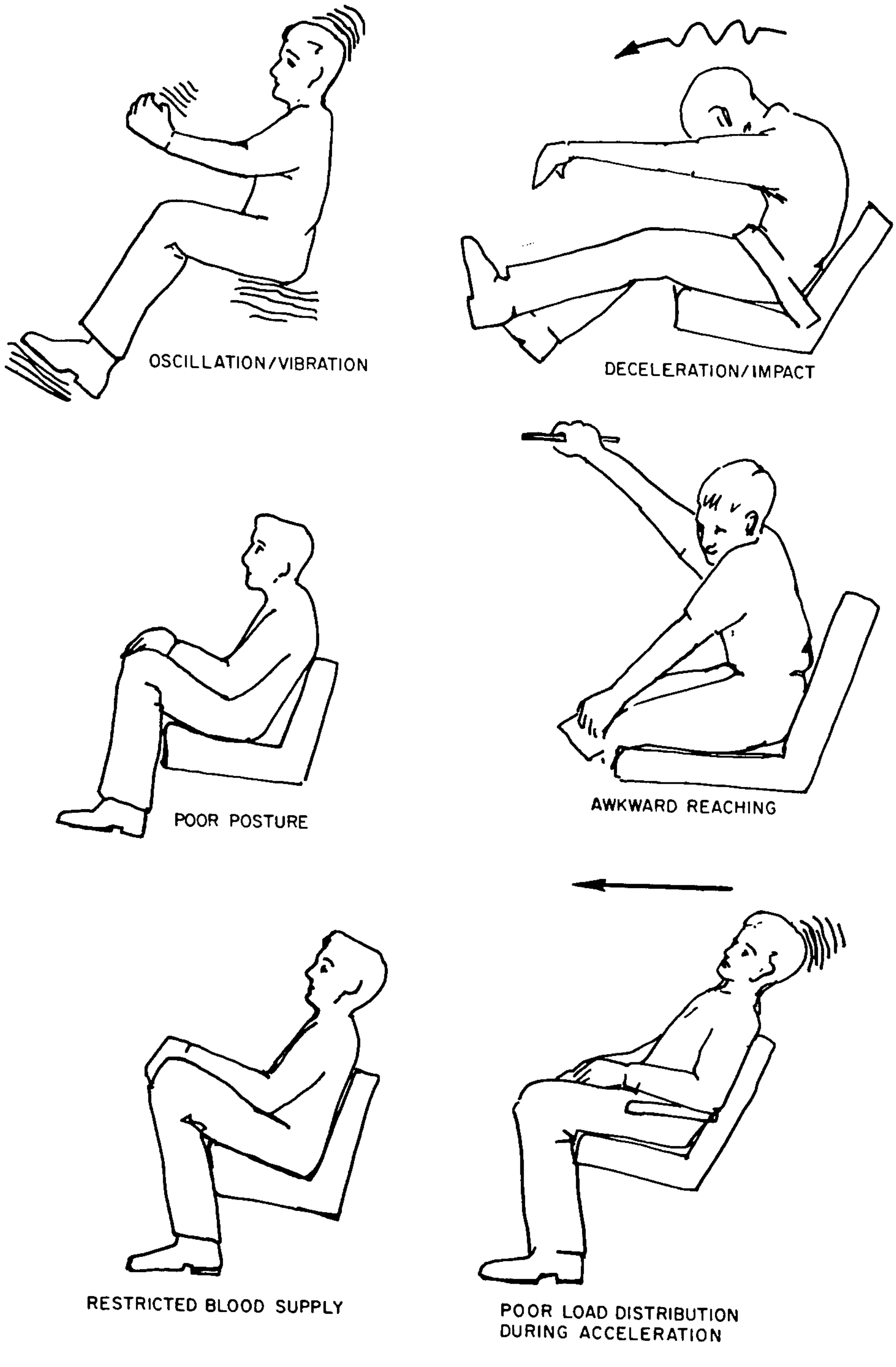


FIGURE 9–2. Typical physiological/environmental/workplace design problem areas.

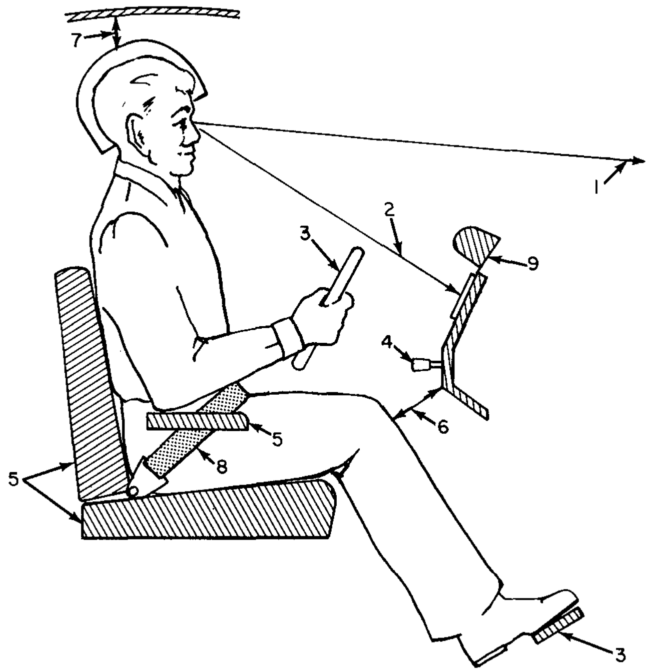


FIGURE 9–3. A checklist establishing a general priority for considering various workplace function requirements (Woodson, 1963).

9.3.8 Design Recommendations Based on Body Dimensions

1. Large operator dimensions should define clearance requirements; those of smaller operators should define reach requirements. When seating is adjustable, an extreme reach range can be accommodated. Both larger and smaller operator floor-to-eye, or seat-to-eye heights are important for design.

2. Restraint by a seat belt, shoulder harness, or a fixed viewing distance makes arm reach a critical factor, but even limited body motion causes these dimensions to be less critical. Other layout requirements can then assume higher priority.

3. Allowance should always be made for body dimensions, natural body slump, effects of bulky clothing, body excursions caused by vehicle oscillations, sudden deceleration, attitude changes, or weightlessness—and for dimensional alterations introduced by stooping, squatting, twisting, turning, or doubling up. (See Figure 9–4.)

9.3.9 Environmental Considerations

Vibration, noise, light, thermal radiation, pressure, etc., should be attenuated at the source or, when this cannot be accomplished, at the workplace. For example, proper orientation of a display panel may reduce the effects of glare from an ambient light source. Structural support for a hand or arm may alleviate vibration effects, thus improving the precision of manual control. Independent seat suspension, seat padding, and contoured seating reduce stress from road shock as well as fatigue from long duty periods in confined quarters. Even proper use of colors and color coding will counteract effects resulting from a crowded workplace. The designer should strive to eliminate or minimize the effects of environment on operator performance.

9.3.10 Priority of Design Considerations

Since designers must choose among competing priorities in workplace layout, the following guidelines are listed in the order of recommended priorities.

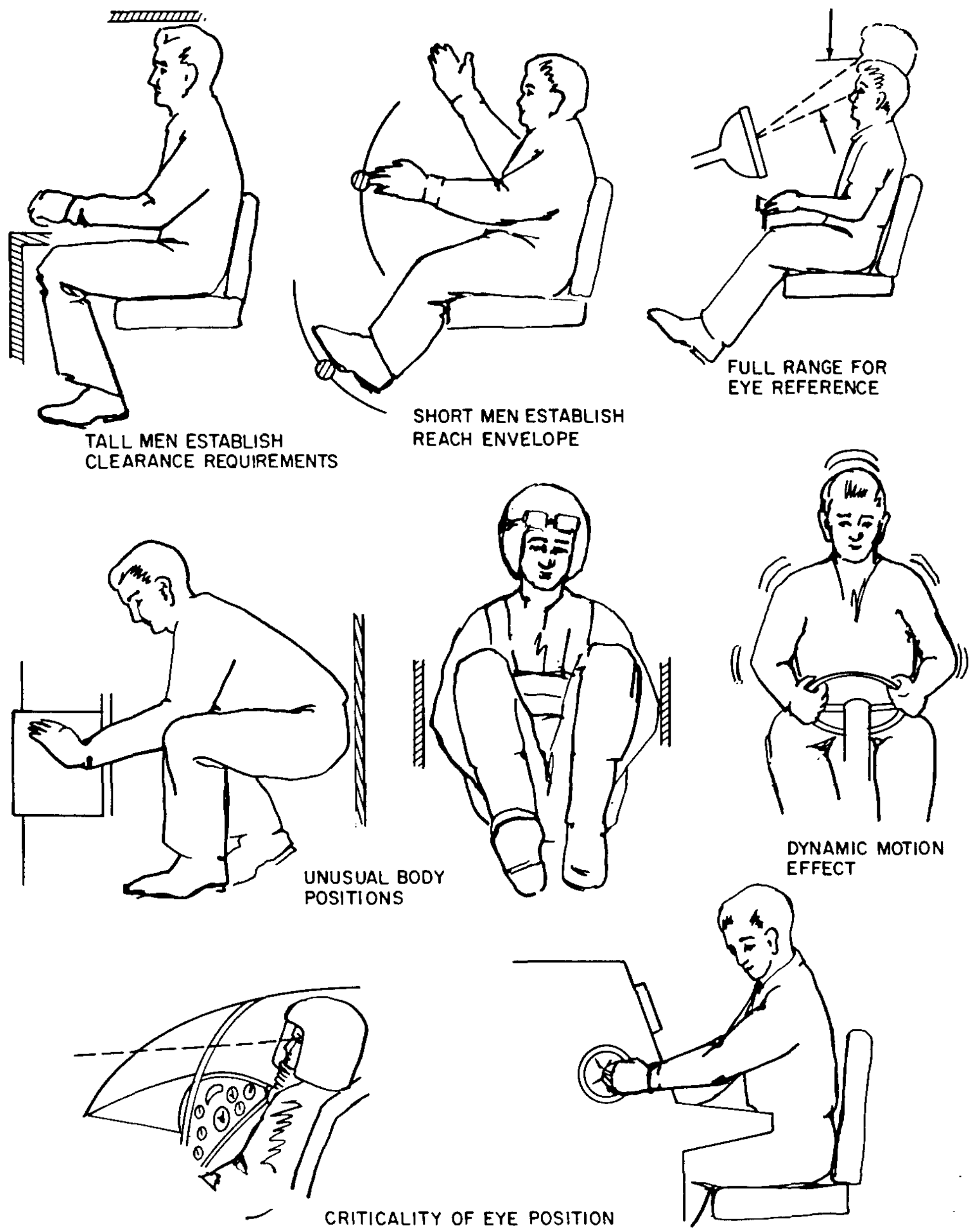


FIGURE 9-4. Influence of body dimensions on layout.

1. Primary visual tasks claim first priority. Whether it be to look out of a windshield, or to view a cathode ray tube display, eye position relative to task establishes the basic layout reference point.

2. Second priority should be given to placement of primary controls which interact with primary visual tasks (e.g., a steering control). Emergency controls should also be a second priority item. A seat reference may be the next design datum to identify. Primary and emergency control positions are generally related to this seat reference point.

3. Third priority should be given to control/display relationships. Controls should be near displays they affect, without causing the operator's hand to cover a display, and should have a direction of movement compatible with movements on the display that the control affects.

4. Fourth priority should be given to arrangement of workplace elements in anticipated sequence of operation, generally, from left to right and top to bottom.

5. Fifth priority should be given to convenient placement of workplace elements according to frequency of use.

6. Sixth priority should be given to consistency of layouts within the same system or other similar systems.

9.3.11 Compromise

Human body dimensions are least amenable to compromise. Man's physical limits for bending, stretching, and/or compressing are such that the machine must be made to adapt to the man rather than the converse. Behavior characteristics are somewhat more flexible. Man can adjust his sensory-motor behavior to some degree, and because of his ability to think and reason, he can utilize alternate procedures and make up for certain equipment inadequacies. However, the reader is cautioned to remember that as operator load increases due to task complexity, fatigue may reduce operator reliability; system performance could degrade at a critical time in a mission. Workplace layout should favor the man's physical and behavioral capability in all cases in which a likely error in human performance could affect the safety of the mission. The designer cannot assume that personnel selection and training will be a panacea for improper workplace considerations.

9.3.12 Integration

An approach to workplace layout should always be within the "total system" concept. Since a structure's physical aspects or a system's mechanical parts cannot be first defined and later modified to fit the man, hardware design should be resolved concurrently with human engineering design. In other words, when design insures structural integrity of a workplace, it should also solve problems of operator support and protection. The user must be given equal and timely consideration with other factors of design.

9.3.13 Safety

Safety for both operator and equipment should take top priority in workplace design. A complete layout conceptualization will include a hazard analysis. Projections and sharp corners are immediately obvious, but a layout mistake which could cause incorrect operator response is more difficult to recognize.

9.3.14 Standardization

The designer should investigate previous workplace layout solutions, particularly when they reflect the guides and specifications of published military requirements. Standardization among systems provides several important benefits including reduction in training time for the new system, less chance of operator error in transferring from one system to another, the obvious cost savings in development of new hardware, and reduced logistic support costs for the using activities. The designer should, however, recognize the dangers inherent in carrying over a poor design concept just to avoid the task of thoroughly analyzing the operator requirements and developing a proper solution.

As a corollary to standardization, commercially available components should be considered. For example, control and display panels can be designed to fit manufactured console and equipment racks conforming to workplace requirements and specifications. Panel widths normally available are 19 in., 24 in., and 30 in.

TABLE 9–2. ANTHROPOMETRIC DATA FOR THE DESIGN OF INSTRUMENT CONSOLES

Type of console	Maximum console height from standing surface A	Console depth at base B	Vertical dimension of panel, including sills C	Console panel angle—from vertical D	Minimum pencil-shelf depth E	Minimum writing surface depth—including pencil shelf F	Minimum knee clearance G	Foot support to seat‡ H	Seat adjust-ability I	Minimum thigh clearance at mid-point of "I" J	Writing surface height from standing surface K	Seat height at mid-point of "I" L	Maximum console panel breadth M
1. Sit-stand.	62.0	Opt.	26	15°	4	16	18	18	4	6.5	36.0	28.5	36
2. Sit (w/vision over top).	47.5* to 58.0	Opt.	22	15°	4	16	18	18	4	6.5	25.5 to 36.0	18.0 to 28.5	36
3. Sit (w/o vision over top).	51.5† to 62.0	Opt.	26	15°	4	16	18	18	4	6.5	25.5 to 36.0	18.0 to 28.5	36
4. Stand (w/vision over top).	62.0	Opt.	26	15°	4	16	----	----	----	----	36.0	----	36
5. Stand (w/o vision over top).	72.0	Opt.	36	15°	4	16	----	----	----	----	36.0	----	36

*"A" must never be more than 29.5 in. greater than "L."
†"A" must never be more than 33.5 in. greater than "L."
‡When seat-to-standing surface exceeds 18 in., a heel catch should be provided.
Note: Standard values for critical dimensions used in the design of instrument consoles for the seated and/or standing operator, with and without a requirement on the operator to maintain horizontal visual contact with other displays or test apparatus beyond the console. Design values for each console established to accommodate 95 plus percent of USAF population. (Anthropology Branch, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, 1963.)

9.4 Workplace Dimensions and General Layout

Representative workplace layouts should accommodate a 5th- to 95th-percentile user population. This range is usually specified for military systems equipment and can be considered an acceptable baseline. Other applications for a specific population or broader range, e.g., female electronics assembly personnel, or 1st- to 99th-percentile operators, are approached by utilizing the appropriate anthropometric data. (See Chapter 11.) Since each workplace should be developed for a specific user performing specific tasks, it cannot be overemphasized that the information of this section should only be used as a guide. Operator-related dimensional factors that influence workplace configuration are:

1. Eye position with respect to display area and/or field of view.
2. Reach envelope of arms and legs.
3. Manner and position of human body support.

Available consoles and equipment enclosures of commercial manufacturers should be surveyed before defining work places. Some of these standard or "off-the-shelf" units will have been developed in accordance with the dimensional constraints of a 5th- to 95th-percentile population. If compatible with other system and user requirements in regard to types and numbers of controls and displays needed, they can replace special-purpose work station enclosures, and save both design time and manufacturing cost. Ascertaining total control and display needs in workplaces is recommended before selecting enclosures.

9.4.1 Equipment Stations

Equipment stations are designed for seated or standing operations or for combined "sit-stand" operations. In some cases the decision as to the type of equipment station to use is defined by specification or is quite obvious; in other cases the designer may consider the tradeoffs between various types of equipment stations in making his decision. Vehicle stations and shared operation stations require particular attention when designing the equipment station. The designer should consider the relative advantages of each equipment station concept for the tasks to be performed using the following principles:

Seated Operator Stations

A seated position for equipment operators is advantageous for:

1. A high degree of body stability and equilibrium.
2. Long work periods.
3. Use of both feet for control actions.
4. Precise foot control actions.
5. Large force application or range of movement with foot controls.

Critical dimensional factors in developing the seated operator station include:

1. Proper eye position relative to the viewing tasks, either on the console or the surrounding environment.
2. Seat height, depth, and back angle with proper posture control.
3. Leg and knee clearance.
4. Hand and/or foot reach requirements for control actions.
5. A common eye position for large and small operators by means of an adjustable seat height.

Seated operator workplace dimensions are shown in Figures 9–5 and 9–6 and Table 9–2. When panel space requirements exceed 40 in. in width, a "wrap-around" console would place all controls within reach. Left and right segments should be positioned at an angle of 110° in front of the central segment. (See Figure 9–7).

Standing Operator Stations

A standing position for equipment operators is advantageous:

1. For mobility to reach and monitor controls and displays.
2. When precise manual control actions are *not* required.
3. When it is impossible to provide leg room for a seated operator.

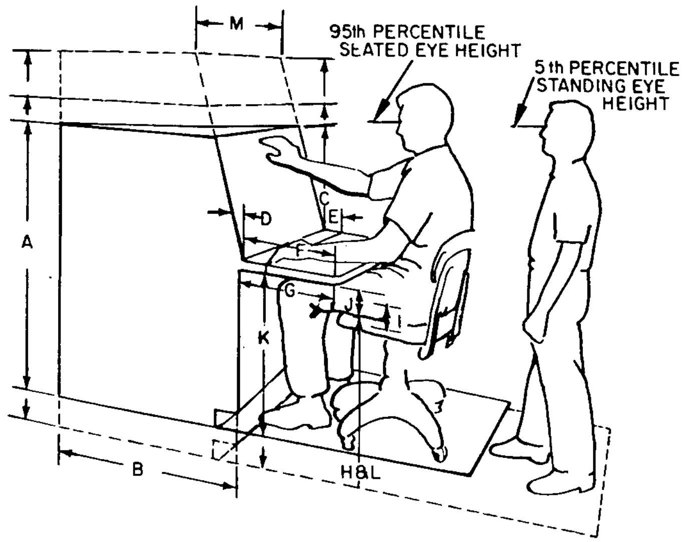


FIGURE 9–5. General workplace dimensions adapted to large and small men (adapted from Webb, 1964).

4. For a large functional panel area in conjunction with visual access to surrounding environments.

5. When foot control actions, other than simple go/no-go or on/off are *not* required.

Standing operator stations are not recommended for on-duty periods of extended length; for short duty periods, operators can minimize fatigue by moving around. A standing operator station design should insure control and display location within the smallest operator's reach and visual field. The requirement for a portable platform to extend the eye height and reach envelope for small operators using standing operator stations is considered a safety hazard and is not recommended as a workplace characteristic. Generalized workplace dimensions for a standing operator at an equipment console are shown in Figure 9–8.

Manual plotting on a rotating table or tactical plotting board is another common standing operation. Although plotting boards are generally vertical and drafting tables generally horizontal, there are instances where plotting angles in between these extremes may be desirable. Because of constraints imposed by plotting board configurations, operators cannot reach as far when the board is horizontal as when it is in various upright positions, as shown in Figure 9–9. These applications should be developed for the smallest potential user.

Sit-Stand Operator Stations

Combination sit-stand operator stations are recommended when the operator may be engaged in two types of tasks: one requiring a seated operation, the other a standing operation. For example, the operator may require the stability provided by seated operation for precise control actions and the mobility provided by standing operation for monitoring of large functional panel areas. The combination sit-stand operator station is also useful when the operator is required to be on duty for extended periods of time and would benefit by alternately sitting and standing to relieve muscular fatigue.

The sit-stand operator station provides a compromise position that gives the operator a high chair or stool by which he maintains his seated eye height approximately the same as his standing height. Common uses of this type of arrangement are illustrated in Figure 9–10.

A difficult arrangement is that in which a cathode ray tube display should be viewed with a minimum of parallax while sitting or standing. Variations of the illustrated workplace are possible. Seated and standing operator station

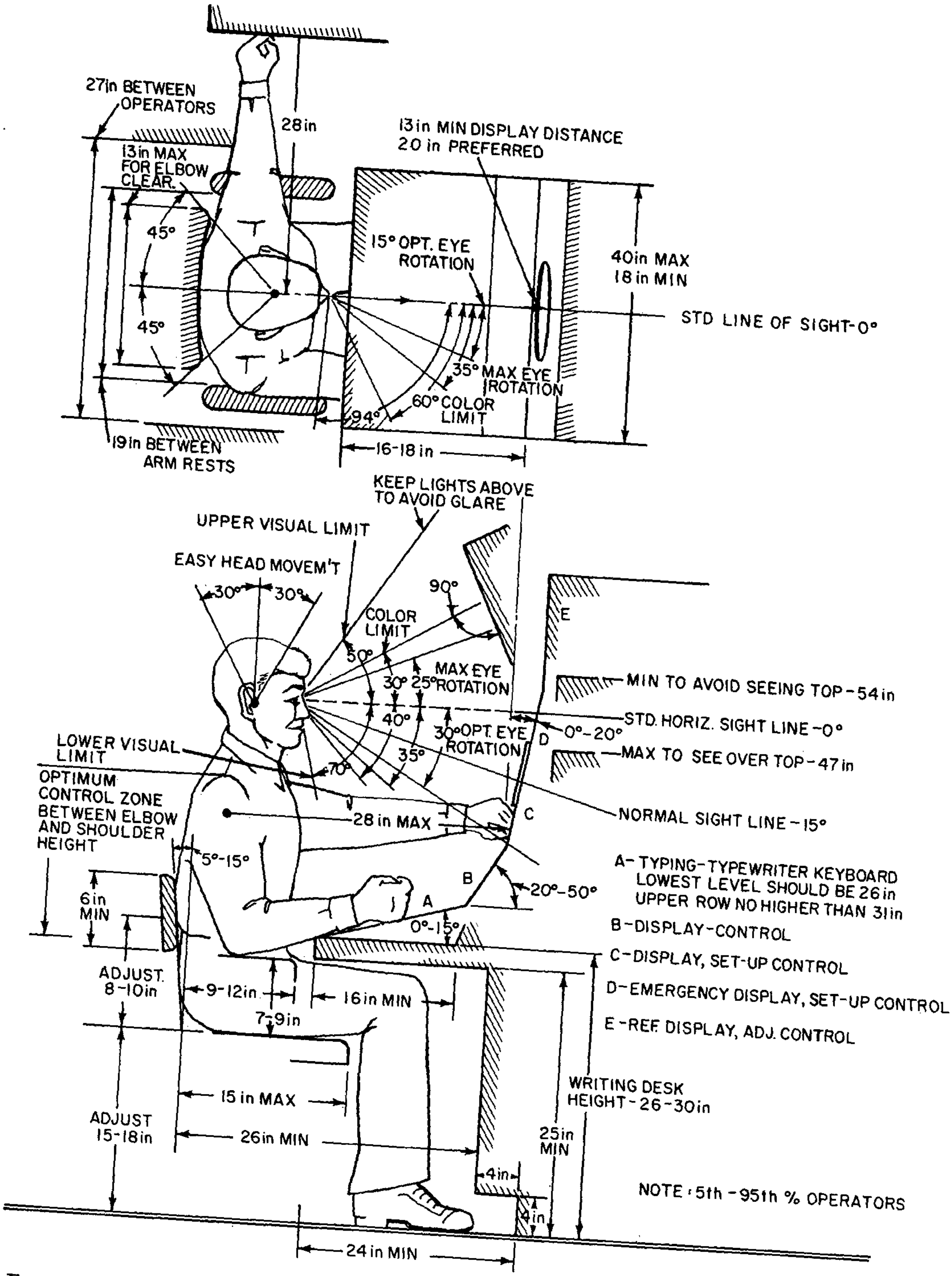


FIGURE 9-6. Suggested parameters for mockup of a seated operator console (after Dreyfuss, 1959; Kennedy and Bates; 1965, Woodson, 1964).

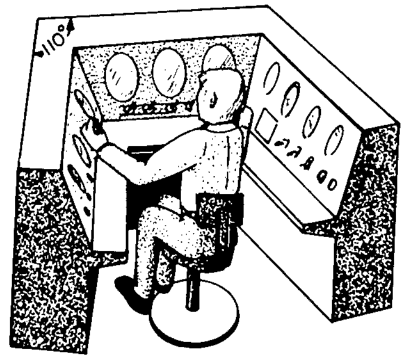


FIGURE 9–7. Horizontal wrap-around console.

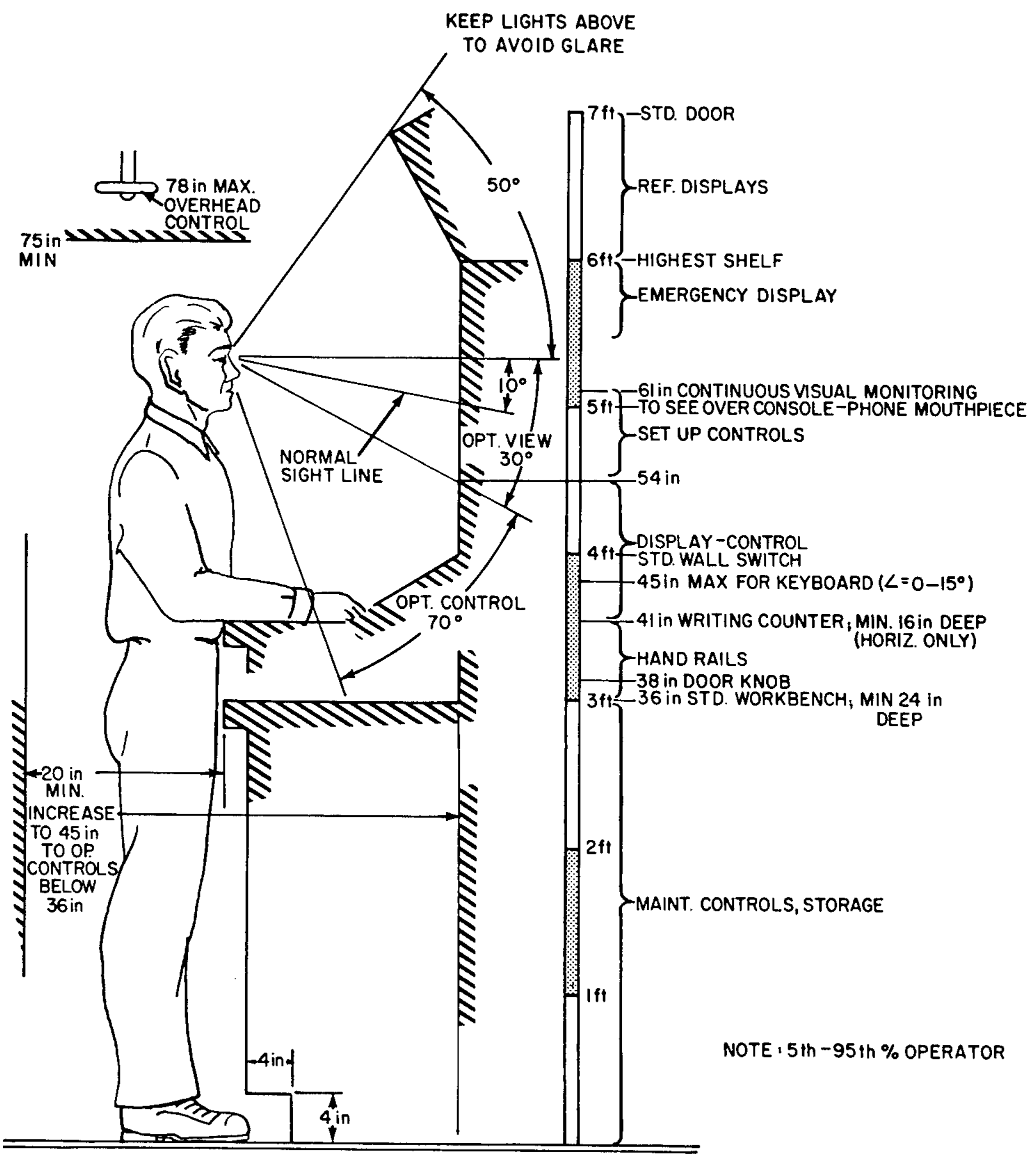


FIGURE 9–8. Suggested parameters for mockup of standing operator workplaces (Dreyfuss, 1959; Woodson and Conover, 1964).

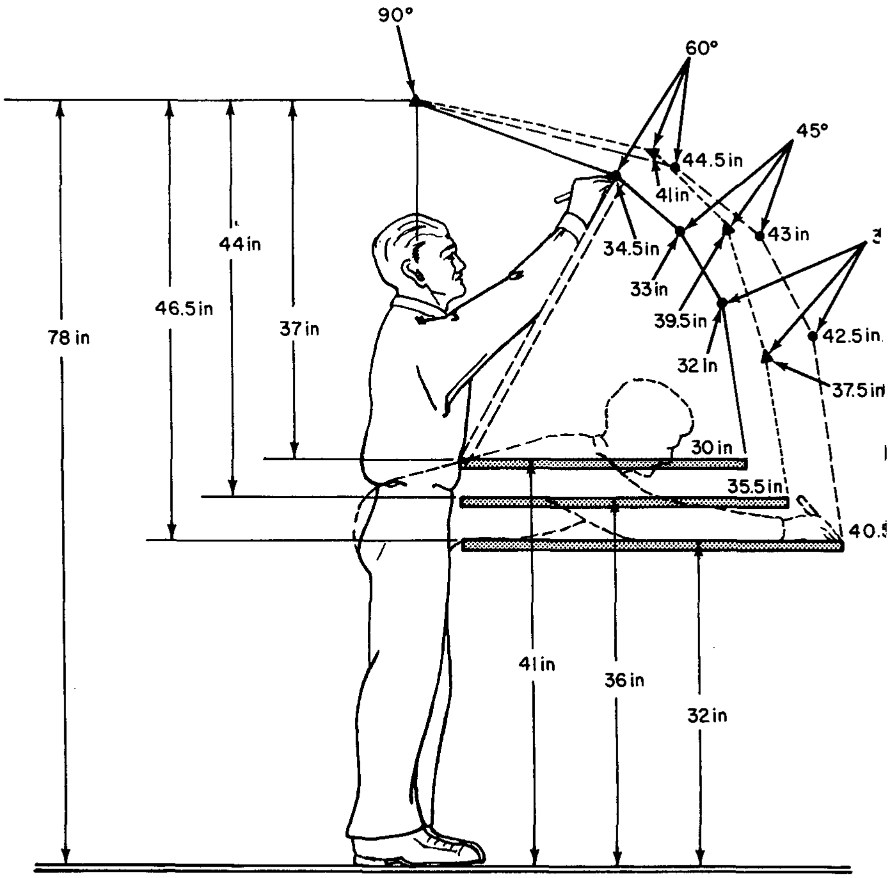


FIGURE 9-9. Drafting and plotting board dimensions (based on an approximate 5th-percentile man) (Woodson, 1954).

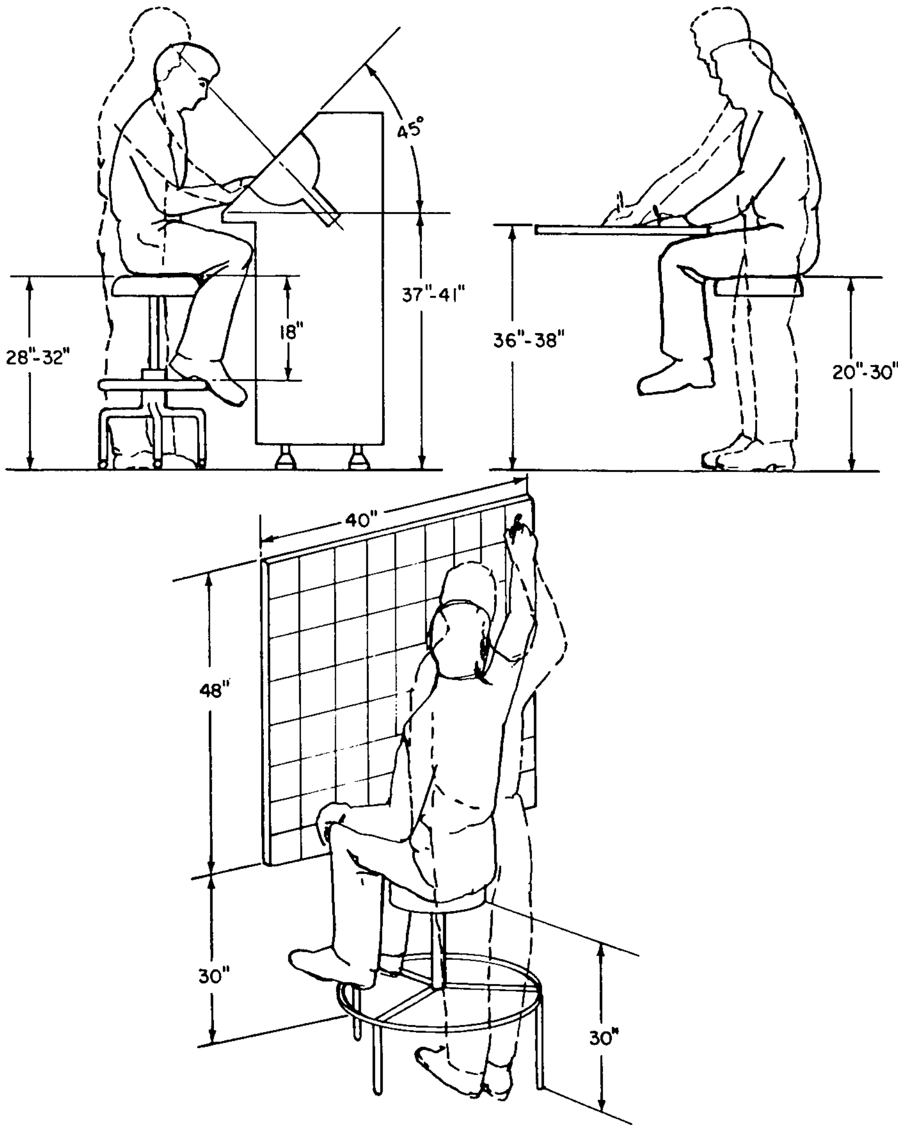


FIGURE 9–10. Sit-stand workplaces (Woodson, 1963; Woodson and Conover, 1964).

configurations previously discussed may be modified to provide for a common line of sight and adequate knee space.

An independently adjustable foot rest in the sit-stand layout is a necessity. The ratio of foot rest to seat height should remain constant, although the height of the seat may be adjustable.

9.4.2 Vehicle Stations

Vehicle workplace layout requires consideration of (a) proper eye position relative to the viewing task on the control/display panel or beyond the windshield, (b) seat height, depth, and back angle with proper posture control, (c) leg and knee clearance, and (d) hand and foot reach requirements for control actions. A suggested driver workplace configuration is shown in Figure 9–11.

9.4.3 Shared Operations

Although the subject of multi-manned system layout is covered more fully in the next chapter, it is important to mention briefly the impact of shared operations on the layout of the individual workplace. Dimensionally speaking, the primary operator should be given priority in terms of his operational efficiency. However, there are situations in which alternatives are available which will improve the utility of certain displays and controls for the secondary or "sharing" operator or supervisor. The two most common situations for sharing are: (a) the side-by-side arrangement, and (b) the operator-observer arrangement.

The primary constraints for side-by-side layouts are the minimum separation between operators and the practical arm reach limits for each. Shared foot controls are not recommended. It would be impractical to specify the entire envelope of dimensions possible, not knowing the specific problem at hand. However, the reader may develop such envelopes readily by utilizing the arm reach data in Chapter 11 plus the illustrative examples in Figure 9–12.

In the case of the operator-observer shared work station, the principal caution is to avoid any compromise of the primary operator's efficiency in adjusting a workplace to accommodate an observer. In many cases, this can be accomplished through the selection of the workplace concept. For example, a sit-stand layout lends itself well to secondary observer relationships since both operators are essentially at the same height. Seated operations also provide for easy "over-the-shoulder" observation of operator activities.

Specific recommendations pertaining to shared operation work stations are provided in Chapter 10.

9.5 Selection and Arrangement of Workplace Elements

Workplace elements are the pieces and parts—control/display panels, writing surfaces, seats, etc.—that are integrated to make up an individual workplace. Again, we must emphasize that the guidelines of this section are presented only for guidance purposes; the designer must define and consider the system and user requirements in defining the selection and arrangement of workplace elements for a particular workplace layout. The selection and arrangement of workplace elements must be pursued in a manner that considers each workplace element with respect to the operator's position and intended use.

9.5.1 Control/Display Panels

A control/display panel is made up of individual controls and displays defined by the operator's input and output requirements. The combining of these controls and displays is the major task of the workplace designer. A panel that cannot be used, or causes the operator to commit errors, will result in an unsuccessful mission or will degrade system performance.

Six considerations must be integrated in defining a control/display panel. Since a panel that fully complies with all guidelines related to all six areas is rare, compromises or trade-offs of preference are necessary. The considerations are (see also Chapters 3 and 8):

1. Visibility. The operator should see all displays from his normal working position, without excessive shifting of his head or body.

2. Grouping. Controls and displays should be arranged in functional and/or sequential

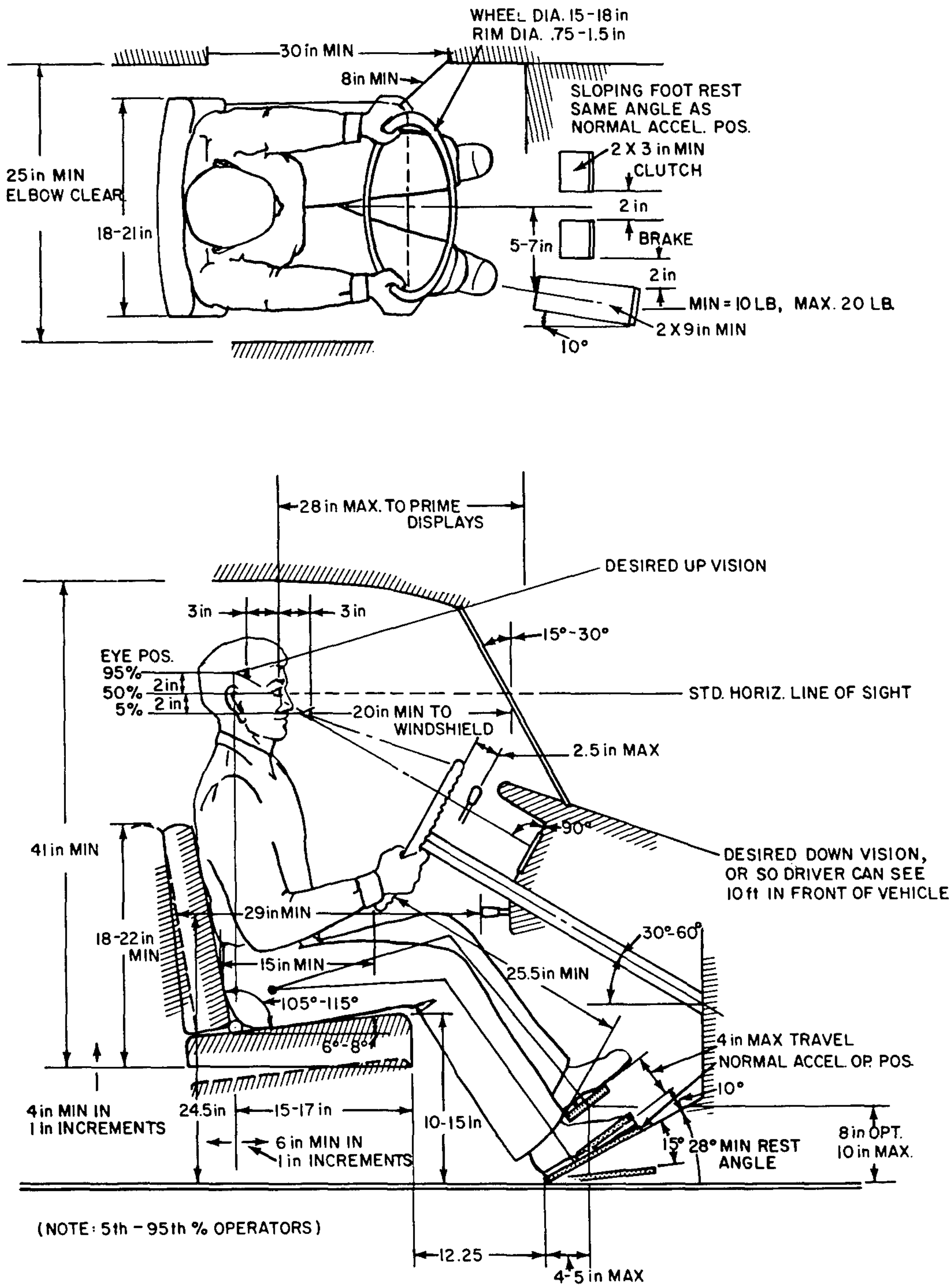


FIGURE 9-11. Suggested parameters for mockup of a road vehicle driver position (after Dreyfuss, 1959; Hedgcock and Chaillet, 1964; Woodson, 1954; McFarland and Mosley, 1954).

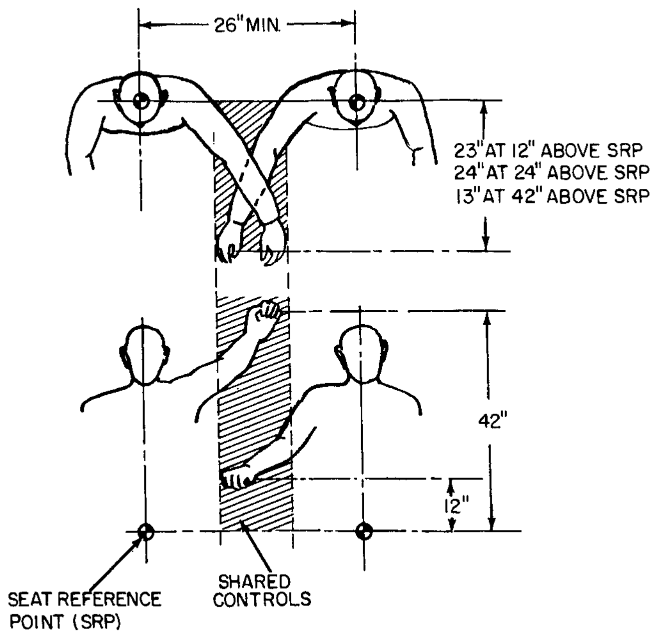


FIGURE 9–12. Example of planning for side-by-side operators.

groups to give logical patterns of movement or progressions.

3. Identification. The operator should find group or individual controls/displays rapidly without error.

4. Population stereotypes. Control and display arrangements should provide expected position and direction of movement relationships. These should be based on task and operator background and experience factors.

5. Clearance. Controls should be spaced far enough apart, or away from adjacent structures, to permit adequate grasp and manipulation through an entire motion range. If gloves are to be worn by the operator, larger controls and control spacings will be required.

Visibility

Preferred visual areas of control/display panels center around the operator's normal line of sight—approximately 10° down from horizontal.

Important principles for panel visual display categories are:

1. Warning Lights and Primary Displays. Including emergency or hazard indicators, critical monitoring displays and displays where color identification is critical. Use of these displays should not require excessive movement of the operator's head or eyes from normal line of sight.

2. Secondary Displays. These are frequently used operational displays. Use of these displays may require eye movement from the normal line of sight, but not head movement.

3. Auxiliary Displays. Including infrequently used displays such as console power indicators and maintenance displays. Use of these displays may require operator head and eye movement from normal line of sight.

Preferred panel location for each display category is shown in Figure 9–13. All warning displays (those indicating a present or potential system failure or personnel/equipment hazard) should be within 30° of normal line of sight or 45° for a sit-stand workplace.

In order to prevent interference with primary operator visual tasks, it is desirable to arrange displays and controls so that the displays are generally in the center of the panel (or upper portion) and controls are arranged in the lower section or about the periphery of the panel. If reach is not taxed, a horizontally oriented panel favors accessibility and display visibility, particularly for the seated operator.

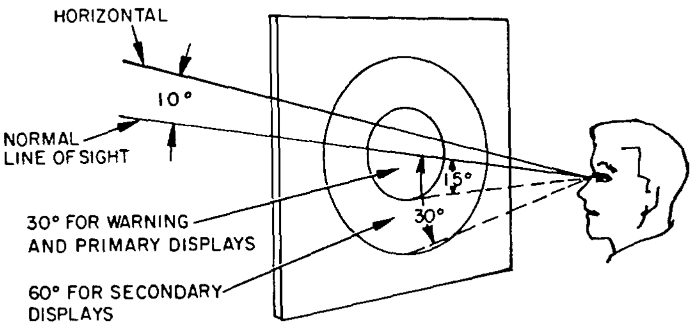


FIGURE 9–13. Preferred placement for visual displays.

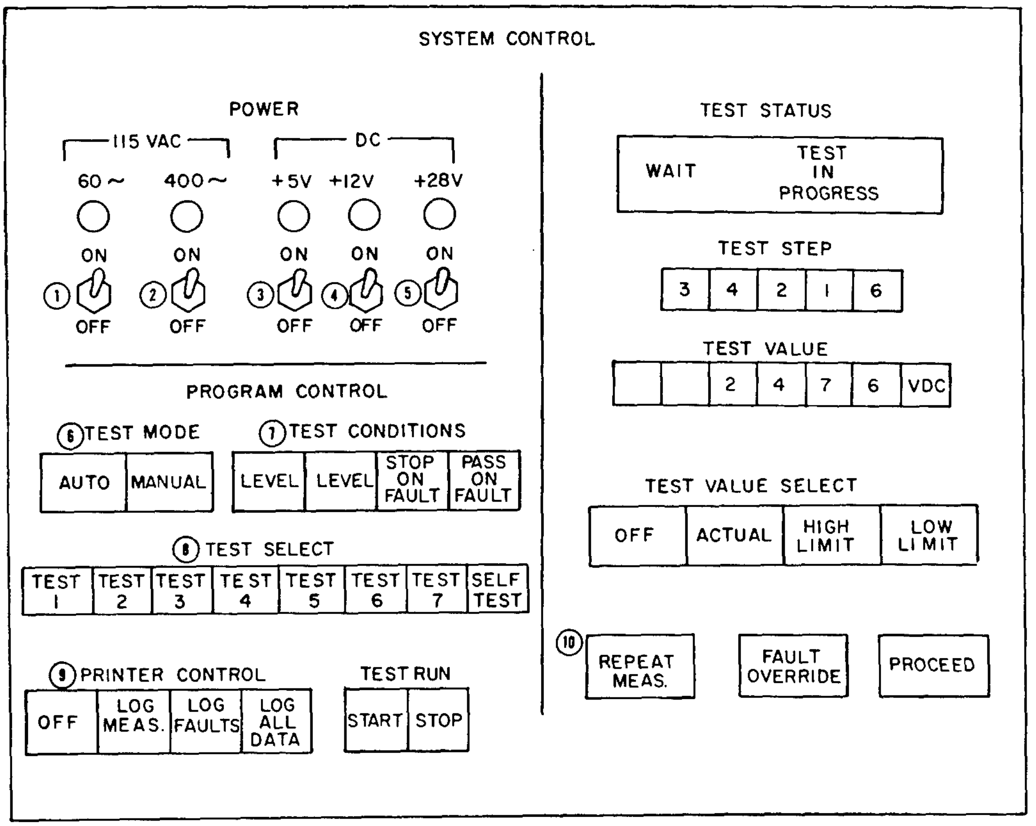


FIGURE 9–14. Examples of panels arranged according to sequential order of operation and function. Numbered callouts indicate the sequence of events.

Grouping of Controls and Displays

All displays and controls should be properly labeled. Even so, an operator is likely to read the wrong indicator, or lose valuable time hunting for an indicator or control, unless the panel is organized in a meaningful and logical manner. The grouping of controls and displays is the primary method of organizing panels. While grouping may not be feasible or desirable in all cases, it should be used whenever possible, especially as panel complexity increases. The various methods of grouping described below can be used in combination on control panel layout.

The two prime methods of grouping controls and displays are by function and by sequence of use. (See Figure 9–14.) In functional grouping, controls and displays related to one function are grouped together and segregated in one area of the panel; e.g., flight instruments, engine controls and displays, system power controls and displays, test program controls and displays, etc. In addition, sets of controls and displays within functional groups can also be grouped. For example, as shown in Figure 9–14, a functional group containing system power controls/displays can be further grouped by AC and DC functions; and the program control functional group can be further grouped by Test Mode, Test Select, Printer Control, etc.

The grouping of controls and displays by sequence of use is an aid in reducing operator errors of omission. When a typical sequence of control and monitoring can be defined, especially if performed frequently, controls and displays can be arranged to provide operator movement that does not require retracing or "skipping" around the panel. As a general rule, sequential grouping should provide for operator movements from left to right and from top to bottom of the panel. Controls and displays may be arranged sequentially for the overall panel or within a functional group.

In addition to grouping by function and sequence of use, other grouping techniques can be used to aid operator performance. One method is to group the most frequently used controls and displays in the center of the panel, consistent with other panel configuration requirements. Another useful technique, when large numbers of meters must be monitored, is to arrange all instrument pointers in a common direction for "normal" conditions so abnormal readings can be readily detected.

The "graphic panel" technique enables visualization and tracing of the flow of events or actions by a simplified system picture or schematic painted on the panel. Key controls and displays are mounted at appropriate points within the graphic. (See Figure 9–15.) This technique is particularly useful for liquid or gas transfer and other process control panels.

Because rapid response is important, special care should be taken to group warning displays and emergency controls together in convenient visible locations. Most government standards and guides provide specific recommendations for placement of these panel elements. In some cases where it may be impossible to follow these recommendations (such as in locating an ejection seat handle), the total sequence of events should be analyzed to establish warning, emergency control, and display locations compatible with the anticipated operator position and activity.

Identification of Controls and Displays

It is important to rapidly and reliably identify controls on a panel. In addition to labeling, shape and color coding should be considered. (See Chapters 3 and 8.) Groups of controls and displays can be more easily identified if they are obviously separated. Techniques for accentuating control/display groups are illustrated in Figure 9–16. Those commonly used are:

1. Contrasting color or shading between subpanel and basic panel.
2. Outlining with contrasting color, a line border around a group of items.
3. A panel relief.
4. An insert panel.
5. Alternating slopes and modular supplement.

Red surfaces or borders should be reserved for emergency controls and display areas.

Arrangement of labels for whole panels must be considered. General rules for panel labeling are:

1. Labels should be located consistently, i.e., never mix positions on the same panel (e.g., above one display, but below another).
2. Every console/rack, panel, functional group, control/display, and control position should be labeled, with labels graduated in size, increasing approximately 25% from smallest to largest in the following order: (a) control position, (b) control/display, (c) functional group, (d) panel, (e) equipment console or rack. For a normal 28-in. viewing distance, the smallest label should use approximately 1/8-in. characters.
3. Lettering should be oriented in a horizontal line.
4. Place nonfunctional labels (e.g., nameplate, manufacturer, part number, etc.) inconspicuously so they will not be confused with operating labels.
5. Avoid placing labels on curved surfaces.

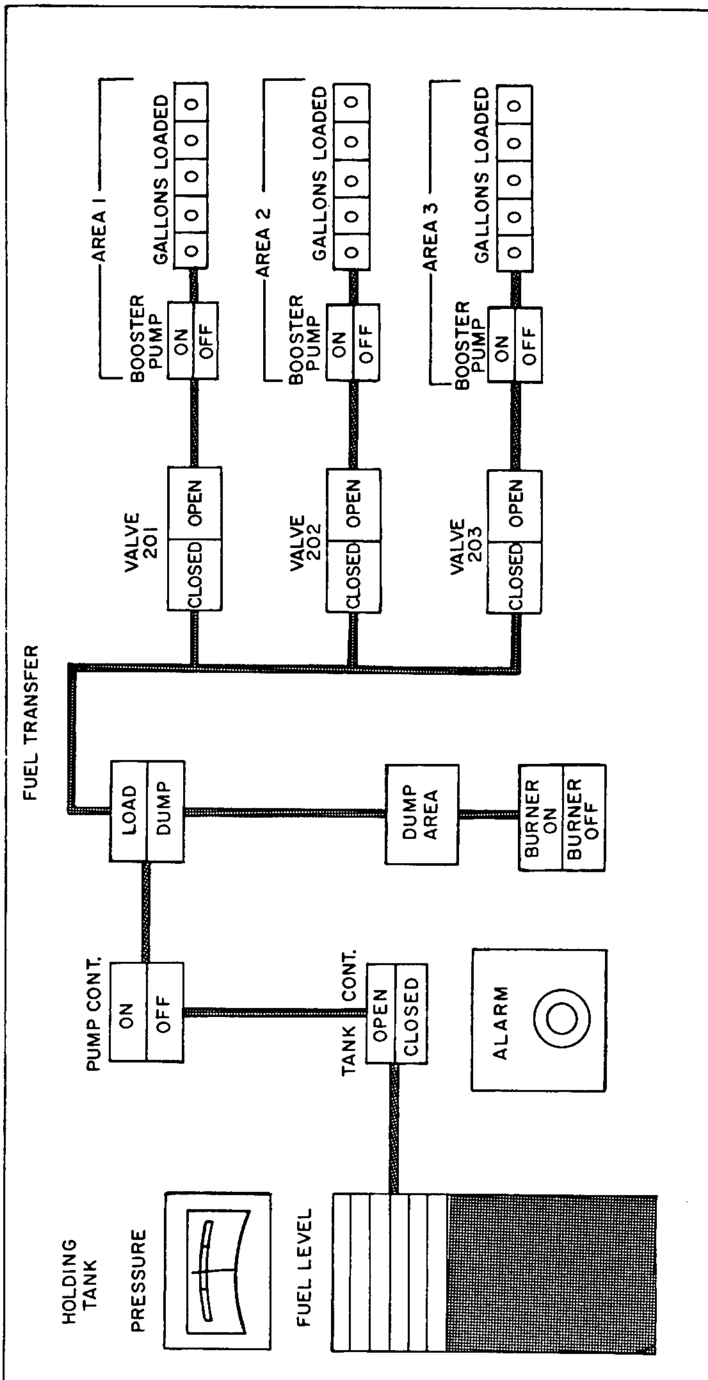


FIGURE 9–15. Example of graphic panel (Meister and Farr, 1965).

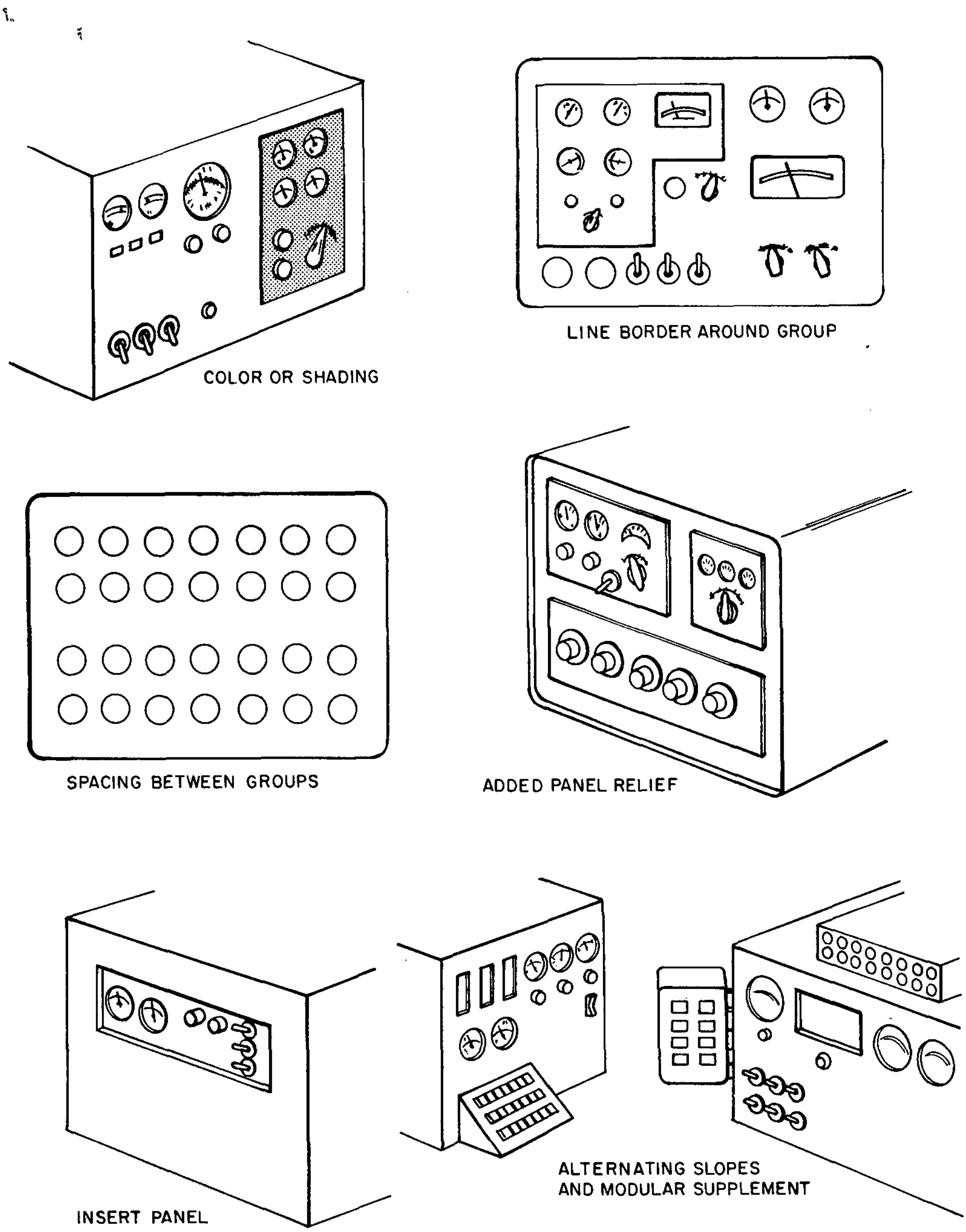


FIGURE 9–16. Methods for accentuating grouping and functional association.

In the event this is necessary, make sure the significant portion is directly visible.

6. Labels should appear upright. If the label must be placed on a moving part, and could be rotated to an upside-down position during operation, consider a second inverted label so identification can be read regardless of position.

9.5.2 Population Stereotype Considerations

Principles dictated by population stereotype, i.e., "what the operator expects," should be used in defining control/display panels. Some direction of motion expectancies are natural, i.e., pushing a throttle forward to increase forward speed, turning a wheel clockwise to turn right, etc. Certain direction of motion relationships have become traditional such as turning a faucet clockwise to shut off water. The designer should avoid control relationships either between controls and displays or between control and vehicle motion which imply wrong or unexpected direction of motion relationship.

When using a large number of controls and/or displays, their arrangement should aid in determining: (a) which control affects which display, (b) which control affects which equipment component, and (c) which equipment component each display describes. Note the following rules based on population stereotype considerations:

1. When a group of equipment components has the same function (e.g., engines of a multi-engine aircraft), positions of the related (associated) controls and displays depend on the direction the operator faces, relative to the normal direction of movement of the vehicle.

a. When the operator faces the direction of vehicle motion, controls and displays correspond exactly to engine positions.

b. If the operator faces to the side or rear relative to vehicle motion, controls and displays should be arranged as though the operator were still facing the normal direction of vehicle motion.

2. When a control is always associated with a specific display, it should be located near that display and below or to one side so that, in using the control, the operator will not obscure the displays. Large numbers of similar control display units arranged on the same panel should all maintain constant control display positional relationships. Common two-position switches, such as the toggle, should have a consistent directional position for "on." When two or more rows of displays must be associated with one row of controls, and vice versa, arrangements shown in Figure 9–17 should be considered.

3. When obvious mechanical motion relationships exist, controls for changing direction of mechanical motion should be oriented and labeled so that direction of motion coincides.

These rules being illustrative, the designer should, if at all possible, check with operational personnel to assure that violation of a population stereotype will not be incorporated in his workplace concept.

9.5.3 Clearance Considerations

Adequate spacing between controls or adjacent structures will help prevent inadvertent actuation and assure that controls can be operated without inconvenience or injury. In spacing, consider: (a) requirements for simultaneous or sequential use of controls, (b) the body member being used, (c) control size and amount of movement (displacement), (d) requirements for "blind" reaching (i.e., being unable to see the control), (e) effects on system performance of inadvertently using wrong controls, and (f) personal equipment that might hinder control manipulation (e.g., pressure suit, gloves, boots).

For "blind" reaching, separation of hand controls located forward of the operator should be at least 6 in.; for areas behind or above his shoulders on either side, control separations should be 12 in. (Fitts and Crannell, 1950).

To conserve panel space or aid in sequential operations, two or three knobs may be mounted on concentric shafts. The probability of accidental actuation increases, however, if either knob diameter or thickness is too large, too small, or differs by too small amounts.

Constraint Considerations

Constraints are factors which restrict the alternatives available to the designer. The designer must identify and consider all con-

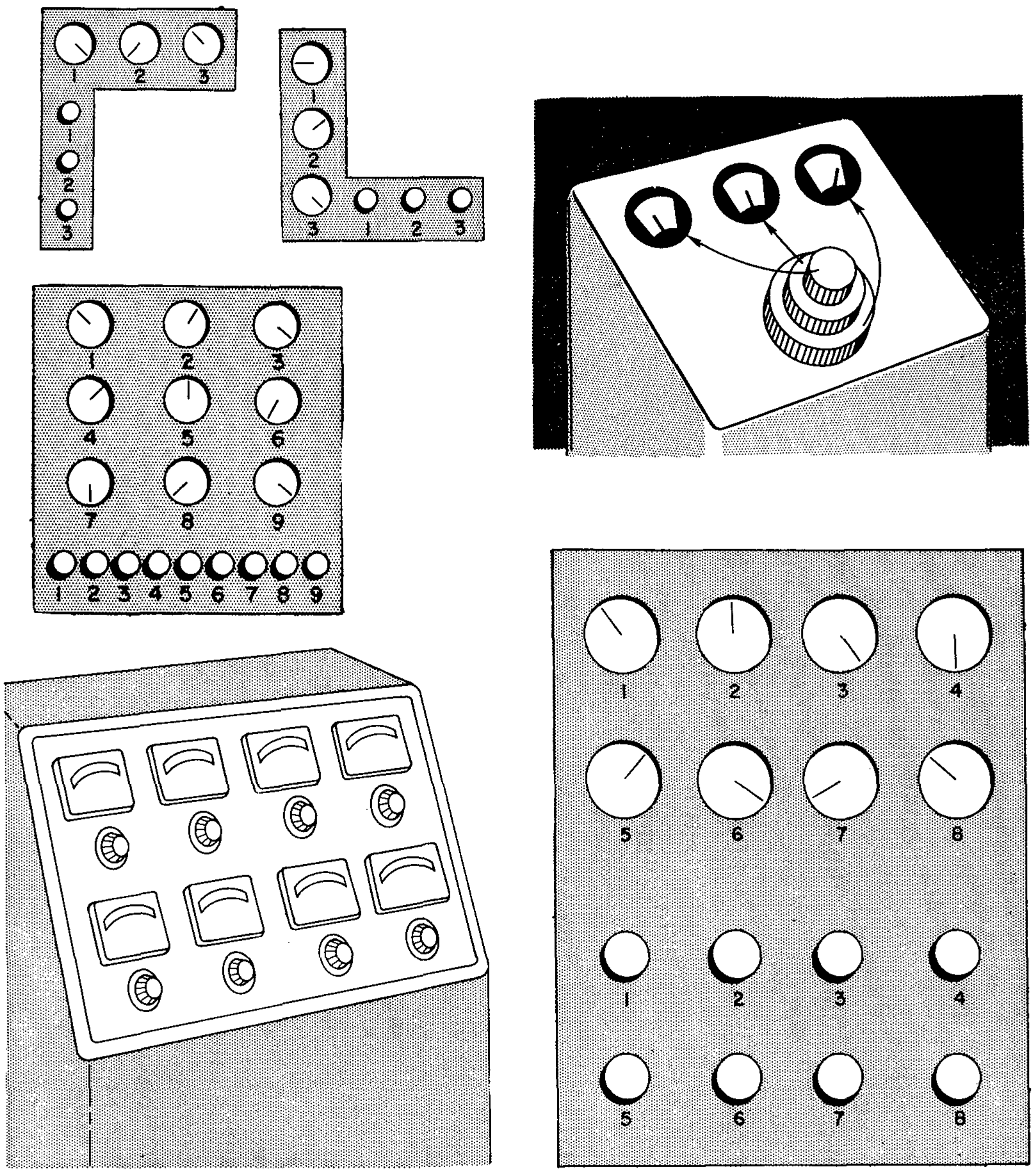


FIGURE 9-17. Recommended arrangements to maintain control display association.

straints imposed on the workplace under development or his design may be rendered unfeasible.

Major constraints include (Meister and Farr, 1965):

1. Restricted panel size which affects the number of controls and displays that can be included on the panel. Restricted panel size usually results in the loss of freedom to organize the panel using such devices as spacing of groups and graphics (flow diagrams). Often nomenclature is severely limited and abbreviated and controls/displays tend to be cramped, resulting in a panel configuration likely to be somewhat difficult to operate. One partial solution to restricted panel size is to judiciously combine input/output functions using switchlights, split-lens indicators, etc.

2. Standardization requirements, limiting the designer to particular control/display positions, types of controls/displays, and coding and nomenclature usage. These constraints are usually stated by specification or regulation and are not necessarily an undesirable limitation. The consistency imposed by standardization requirements (e.g., color coding in accordance with MIL-STD-1472 or nomenclature abbreviations in accordance with MIL-STD-12B) makes it easier for operators to use the panels.

3. Internal packaging demands which may limit the number and arrangement of controls/displays on a given panel due to the availability of space behind the panel. The bulkiness of stress beams and component mounting hardware as well as accessibility requirements may restrict the panel layout. (See Chapter 12.)

4. Many workplaces require the incorporation of "off-the-shelf" assemblies in their design, i.e., consoles include commercially available assemblies as modules or sub-panels. Common examples are standard instrumentation units (counters, oscilloscopes, digital volt meters, etc.) which are inserted as a unit into six-foot high rack-mounted chassis. Under these circumstances, modification of panel arrangement or labeling is difficult. Besides certain layout inadequacies in individual units, a console rack containing a number of off-the-shelf assemblies usually results in different panel faces (color, controls, display color coding, etc.) for each instrument (Meister and Farr, 1965).

If an off-the-shelf equipment panel face cannot be specified or modified, it can be standardized with other system equipment. Control knobs and indicator lenses can be changed and modular color patches can be applied to major functional areas. This simply involves painting around functionally related controls and displays with a color different than the panel itself, e.g., dark gray on a light gray panel face. Another technique standardizes labeling within system equipment by use of a thin overlay panel over the controls/displays of the off-the-shelf assembly, but allows freedom in labeling and functional area designation. (See Figure 9–18.)

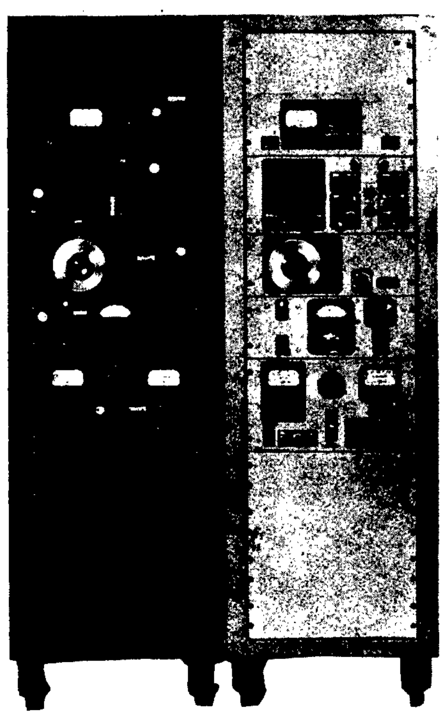


FIGURE 9–18. Off-the-shelf modification (Meister and Farr, 1965). By changing to properly human engineered control knobs and applying pads or lines to designate functional areas it is possible to considerably improve "off-the-shelf" panels. Integration of such equipment into a system provides standardization and ease of operator use.

9.5.4 Selection of Individual Controls and Displays

It is advisable to explore the range of controls and displays available in order to select devices which not only are effective but are least sensitive to critical placement in the development of the workplace. Some displays not mounted with line of sight can still be used effectively—while others are subject to parallax problems. Some controls can be operated satisfactorily even though their relative position to the operator is not convenient. Detailed discussions of controls and displays are presented in Chapters 3 and 8 of this *Guide*.

9.5.5 Visual Display Considerations

Visual display elements range from natural display of objects as seen from aircraft or vehicle windows or in a command-control complex, to specific hardware displays such as instruments, lighted indicators, signs, and status and map boards. Certain visual displays must be perpendicular to the line of sight in order to eliminate parallax and allow for accurate reading. This limits freedom in arrangement. If it is not possible to place all displays in optimum positions, tradeoffs will have to be made. Visual displays sensitive to parallax are:

1. Cathode ray tube (CRT) displays with engraved overlays which require alignment between target and engravings.
2. Scale and pointer instruments in which very precise pointer–scale alignment is required.
3. Stacked, edge-lighted digital readout displays in which a number plate at the rear of the stack may be obscured by the display case.
4. All instruments in which the instrument face is deeply inset below the instrument face cover glass (this may be true for meters, digital counters, and some types of CRT's).

Rules for arrangement of visual displays are generally quite obvious. They include placement of visual displays in front of the operator, as nearly perpendicular to his line of sight as possible, at a distance for adequate resolution of visual detail without causing fatigue and out of direct light which causes reflection and/or glare. (See Chapter 3.)

Certain other rules for arranging visual displays are less obvious. Interaction between instrument position within a workplace and an out-of-window viewing task is important; e.g., in piloting an aircraft or driving a vehicle, the operator must scan both the outside world and his instrument panel. Therefore, instruments most directly related to this shared visual task should be arranged so that eye and head movements are minimized. For example, arrangement of aircraft positional and directional instruments are one above the other, directly in front of the pilot, so that instrument indications and outside world cues are geometrically compatible and warning lights are within the scanning path.

Instrument scanning or viewing time should be minimized. By arranging instruments and displays so their "normalcy" patterns are alike, scanning an array for an abnormal indication can be done rapidly.

Typically, workplace layout problems are related to display surface reflections and glare. Shields or instruments set in below a panel surface help reduce glare caused by ambient illumination on a display face. Since displays must be located very close to the operator's normal line of sight to prevent partial masking, arrangement difficulties emerge and are aggravated if the display area is large. One effective solution is tilting glass covers so that incident light rays are not directed into the operator's eyes. Nonreflective coatings also reduce reflection.

Bright sunlight shining on a pilot's face or clothing causes light to be reflected in the instrument cover. If the cover is perpendicular to normal line of sight, this reflection will be bright enough to obscure instrument detail. Slight tilting of instruments "just off" the normal axis will reduce the problem.

Although not as severe as outdoor bright sunlight, self-reflection problems occur indoors on such displays as cathode ray tubes. Fatigue and loss of display information result when self reflection is close to the threshold of detectability. In addition to cover-tilting and nonreflective coating techniques, it is possible to alleviate reflection by selective filtering of ambient light and the use of filter glass over the cathode ray tube face. One popular technique is the "cross-polarized filter." Other systems utilize narrowband color filters in conjunction with colored lighting systems. (See Chapter 3.)

The topic of visual obstruction is also a consideration of workplace layout. Certain window structures create "blank spots" causing serious loss of visual information. Windshield support posts should either be placed where they will not obstruct vision, or designed so that the area blocked is small enough to "see around" the post. With road vehicles as much "down vision" as possible should be provided. For example, a cab-over-engine truck increases the driver's forward visibility better than a vehicle which has a long engine hood. The nose-high landing attitude of modern aircraft has forced the development of adjustable cowl assemblies which can be lowered for more downward visibility. A forward motion at a time when downward vision is most critical results in enough blurring of objects immediately in front to lose the benefits from seeing down. Minimum visual distance may be a function of the point just beyond which blurring occurs. Designing for operator vision to this point rather than to the nose of the aircraft relieves an otherwise unachievable requirement (Woodson and Conover, 1964; Havron, 1962).

9.5.6 Auditory Displays

Location and positioning of auditory displays is less critical than for visual displays. However, the direction from which a sound comes may be used as a cue for differentiating one signal from another. When air traffic control tower operators must hear multiple communication channels simultaneously, one means for maintaining channel identity is to route each channel through a separate loudspeaker. If these speakers are separated by 10°, the speaker can be identified.

Normally, a loudspeaker should be pointed at a listener at "head height" or slightly higher. In noisy environments, loudspeakers should be near listeners' ears. In this manner, the operator can adjust speaker gain so as not to contribute to overall noise level. Other audio devices such as microphones, intercoms, and telephones also should be located for maximum user convenience.

Voice operated (VOX) communications sets mounted on panels or headset units minimize interference with primary tasks. If "hardwired" communications equipment is used, connecting cords should not interfere with operator actions. If a device such as a telephone or handset may be used in conjunction with writing activities, it should be placed for left-handed operation.

9.5.7 Manual Controls

Manual controls that are used frequently should be in a comfortable position. General control placement guidelines are illustrated in Figure 9–19. Operators should not have to reach, contract their arms, or hold them in awkward positions for long periods. Controls occupying prime layout areas are joysticks, steering wheels, plotting or pantograph devices, and stylus-type data pick-off devices. Simple writing should be considered a primary control task if it is continuous. Supplementary controls, such as rotary and toggle switches, and adjustment controls should be located near displays they affect.

When the application of considerable force is required, controls should be located so that maximum use of combined sets of muscle groups plus support from a seat back or other structure is possible. Where quick operator response is important, locate controls "at the operator's fingertips."

Controls requiring motion excursion should be located so the range of control excursion remains within comfortable arm reach; the operator should not have to move his body to initiate or complete the control motion. Location of controls should not create a hazard nor an obstruction during an emergency, nor possible accidental actuation. A control operated in a jouncing, pitching, or other disturbing environment should be located so that the operator can steady his arm or hand, and/or so that the operator's hand remains in contact with the control. Linear controls should operate perpendicularly in the direction of G forces so the operator does not have to compensate for forces acting along the line of control. Rotary control movements are little affected by G forces interacting with limb mass.

Maximum distance of control from the operator will vary as a function of arm reach,

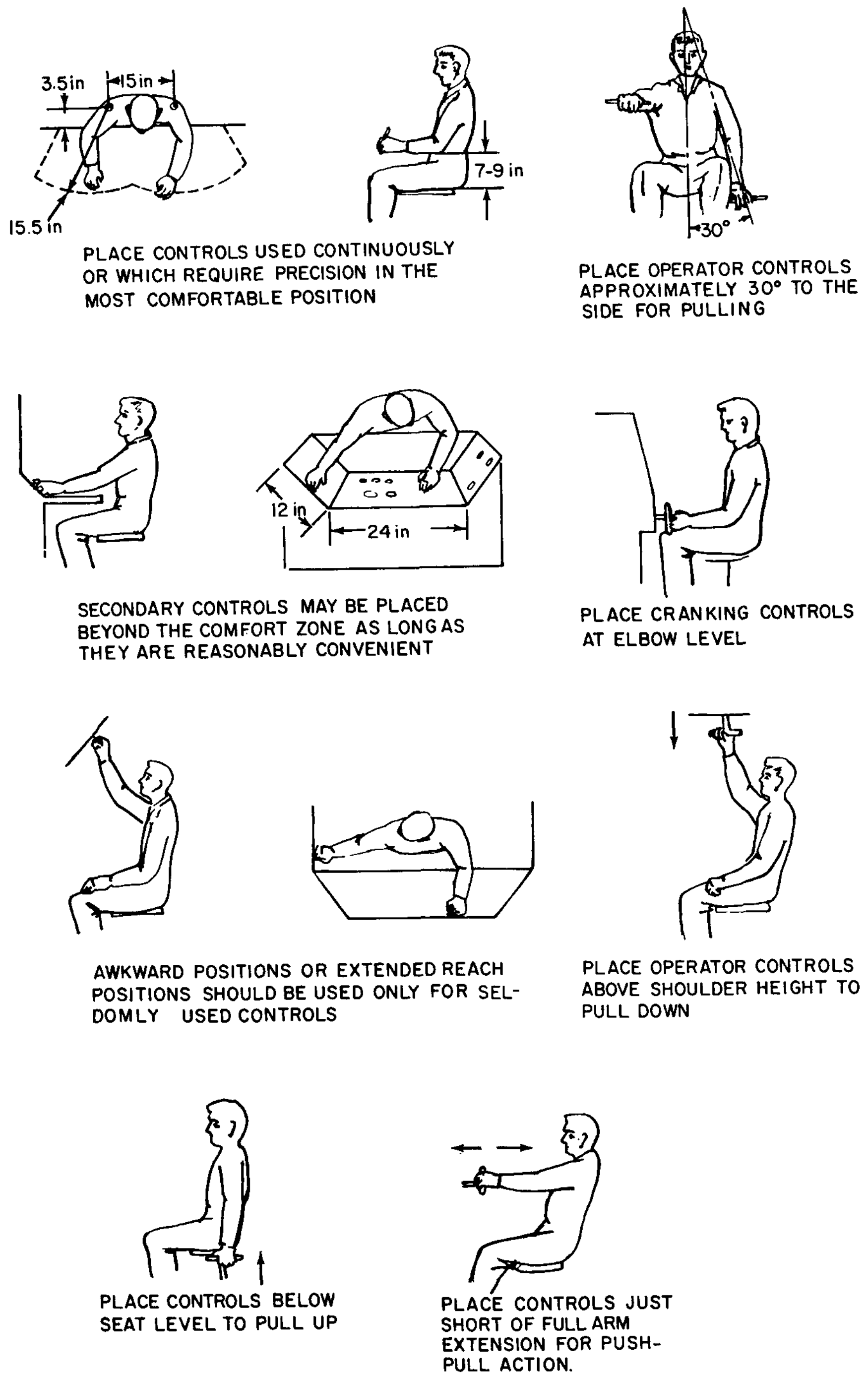


FIGURE 9–19. General control placement guidelines.

body movement, vertical and lateral location, and type, or shape of control. Ideally, all forward distances should be measured horizontally from the seat reference point (SRP). However, because backrest angle influences reach distance, recommendations are based on measurements from its inside surface at shoulder height:

1. Controls operated with the whole hand should be located approximately within 27 in. of the back of the operator's shoulder.

2. Finger-operated controls should be located approximately within 29 in. from the same shoulder reference point. Controls located near these maximum distances should not be ones frequently used, except for those requiring maximum pulling force.

Controls should be located where they can be handled with elbows at angles of 90° to 135°. The best angle for exerting force from a seated position is about 120° with the following variations (Briggs, 1955; Caldwell, 1959; Hugh-Jones, 1947; and Hunsicker, 1955):

1. Push-pull forces are strongest when controls are farthest from the operator, i.e., at or near full elbow extension.

2. Push-only forces are strongest at elbow angles of 150° to 160°.

3. Up-down forces are maximum at intermediate (120°) location.

4. For right-left movements, fore-aft control locations make little difference, although slightly greater forces can be applied when they are positioned close to the body.

5. For rotary movements, there is little difference at most fore-aft locations, although farthest points should be avoided.

The best fore-aft locations for controls are different for prone and seated operators. For prone operators (a) push-pull is greatest with elbow angles of 120° to 180°; (b) up-down is greatest close to the body, and (c) right-left movements make little difference, although they are somewhat stronger close to the body.

Both the speed and accuracy of visually controlled and manually positioned movements are greater when controls are as close as 7 in. They become progressively less as distance increases.

Foot controls operated while standing afford freedom of movement, but the operator is at a disadvantage since he must support himself with the other foot as shown in the upper portion of Figure 9–20. Conversely, a seated operator's range of movements for foot control are more limited, as shown in the lower portion of Figure 9–20. Depending upon force and precision, foot controls can be operated with either toe or heel satisfactorily. Some foot controls can be operated at the instep, although this position is less preferable.

9.5.8 Writing and Charting Surfaces

Factors to consider in writing or charting surface design are: (a) the nature of the task and type of manual aids to be used, (b) relationship to surface size, shape, and position, (c) inherent constraints of body position, arm reach, and eye position, (d) possibility of sharing work area. Guidelines for designing writing and charting surfaces are illustrated in Figure 9–21.

The size of writing material defines work as well as surface area required by arms and hands for precise control of writing, drawing, or plotting. A recommended minimum surface is 24 in. wide by 16 in. deep. Racks or other storage devices for writing and drawing aids should be provided.

Storage facilities on the work surface should make writing and drafting aids accessible, secure, and within easy reach of the operator. Complex storage techniques are not necessary; simple clips on a panel surface or even a standard clipboard are often satisfactory. Special devices such as VELCRO materials, used to prevent small items from floating away during weightlessness, may have certain applications.

If more writing space is required than is within arm's reach, arrangements allowing access from more than one side of an adjustable (tilt) work surface can be used. A work surface for drawing with a slight slope up to 10° is more comfortable; beyond this point pencils slide. If used infrequently, retractable writing surfaces are favored for crowded workplaces. Ideal positioning should be for both right- and left-handed people. If this is not possible, the right-hand rule should apply.

9.5.9 Workbenches

Workbench design (size, shape, and height above the floor) must reflect: (a) the charac-

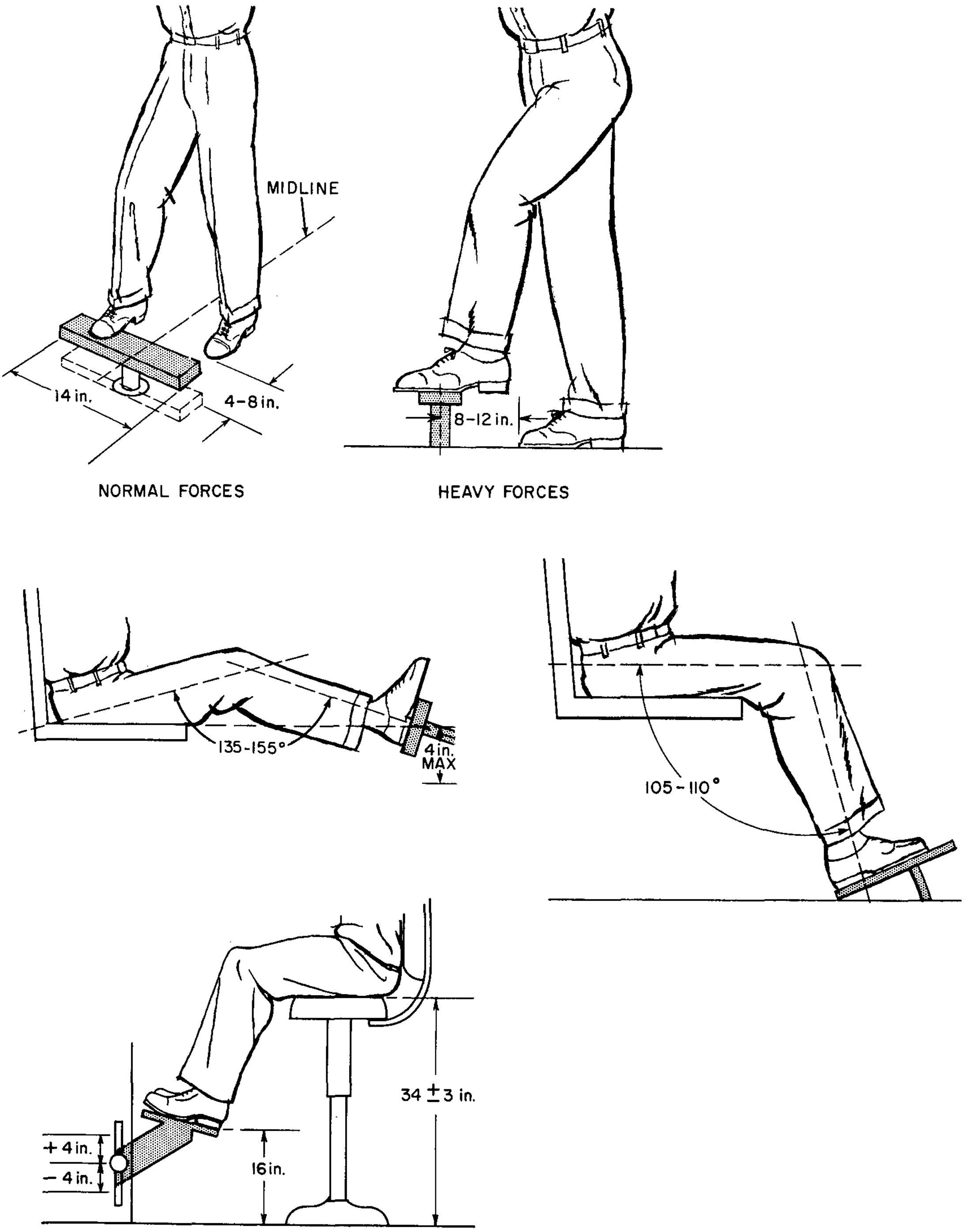


FIGURE 9-20. Foot control placement depends on force and comfort.

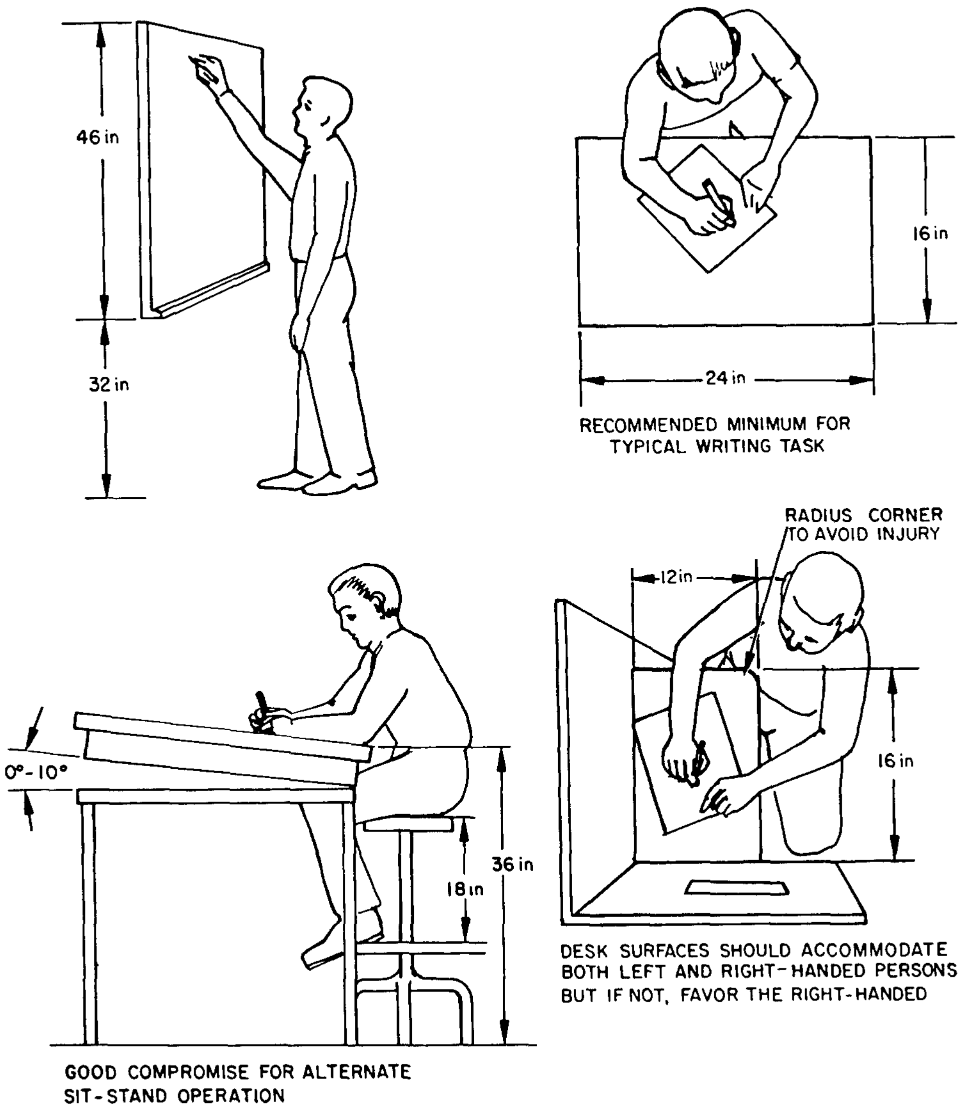


FIGURE 9–21. Suggestions for design of writing, drawing, charting and plotting work surface layout.

teristics of equipment placed on it, and (b) how the technician will perform his task. Size and weight of equipment should be considered in determining bench as well as final working height, since problems of lifting may be equally as important as ease of repair.

Important convenience factors in workbench design include: (a) electrical outlets in position for equipment testing and power tools, (b) storage shelves which place test equipment displays at proper viewing height and angle, (c) storage for spare parts and tools both temporary and semipermanent, and (d) maximum access to all sides of equipment being worked on.

Although workbench dimensions may be extrapolated from workplace data, the uniqueness of particular equipment repair tasks should be a principal determinant of workbench configuration. Safety factors to be considered include grounding, switches for equipment and tools, special eye protection, and mechanical and thermal guards. Principles for workbench design are illustrated in Figure 9–22.

9.5.10 Seating and Restraint

A properly designed seat contributes to efficiency and safety. It must provide: (a) accessibility to the task, (b) proper support, (c) security and protection, (d) accessibility, (e) comfort.

Although most tasks may be accomplished satisfactorily from a simple stationary seat, a range of adjustment may be necessary to place the operator in proper working position. Typical adjustments which are useful for this purpose include: (a) seat height, (b) rotation, (c) fore/aft movement, (d) seat/backrest angle, and (e) lateral movement. Of these, seat height and rotation most often affect ability to perform. Types and ranges of adjustable seats should be established at the same time other workplace dimensions are being developed; i.e., seat design should not be expected to account for all workplace variations.

For proper support, a seat must be designed to fit body dimensions, distribute weight to relieve pressure points, and support posture. Seating dimensions and considerations for typical seats are shown in Figure 9–23.

A seat should provide security from dynamic forces which tend to "unseat" a person. A "pitch-back," plus slightly contoured sides on the seat pan are means for securing against mild dynamic influences. For more violent motions, restraint systems such as lap and shoulder harnesses should be provided.

Special protective devices in operator seats should be a function of environment and expected use. Adding guards and rails will protect arms, legs, feet, and head from violent dynamic forces such as the wind blast in aircraft seat ejection. The designer should not overlook requirements for seat ingress and egress in workplace development. In tight quarters, a movable seat (e.g., with rotary or translational adjustments) could provide a solution.

Comfort may be provided by adding slight contouring to seat pan and back rest, and by application of soft padding. Padding amount is a function of material density, but should be sufficient to prevent the subject from compressing the material to its absolute limit; there should be some resiliency remaining.

9.5.11 Test Equipment and Tools

Functions of maintenance should be considered in initial workplace planning. Detailed information pertaining to maintenance is presented in Chapter 12. General layout requirements pertaining to workplace maintenance are:

1. Position test equipment at proper distance, height, and angle for viewing displays and operating controls.
2. Locate test equipment so that interconnecting leads will not interfere with the technician's viewing or manual activities.
3. Arrange test equipment, tools, and other elements of maintenance tasks to simplify the work flow and eliminate unnecessary motion.
4. Provide a convenient place to lay tools without having to place them on test or other equipments.
5. Provide convenient electrical outlets.
6. Provide adequate illumination, properly located.

General concepts for organization of test equipment and tools are shown in Figure 9–24.

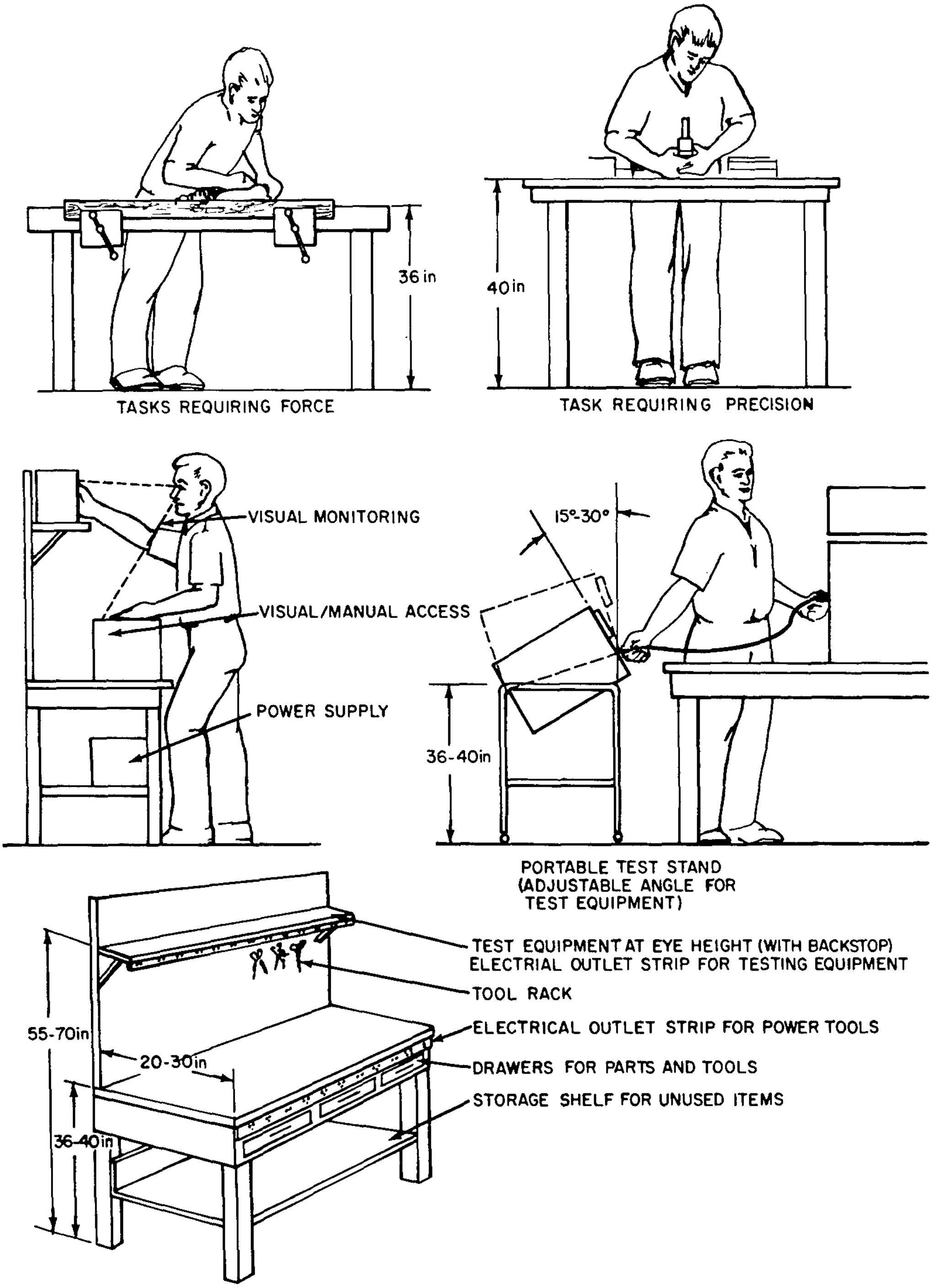


Figure 9–22. Recommendations for typical workbench layout and design.

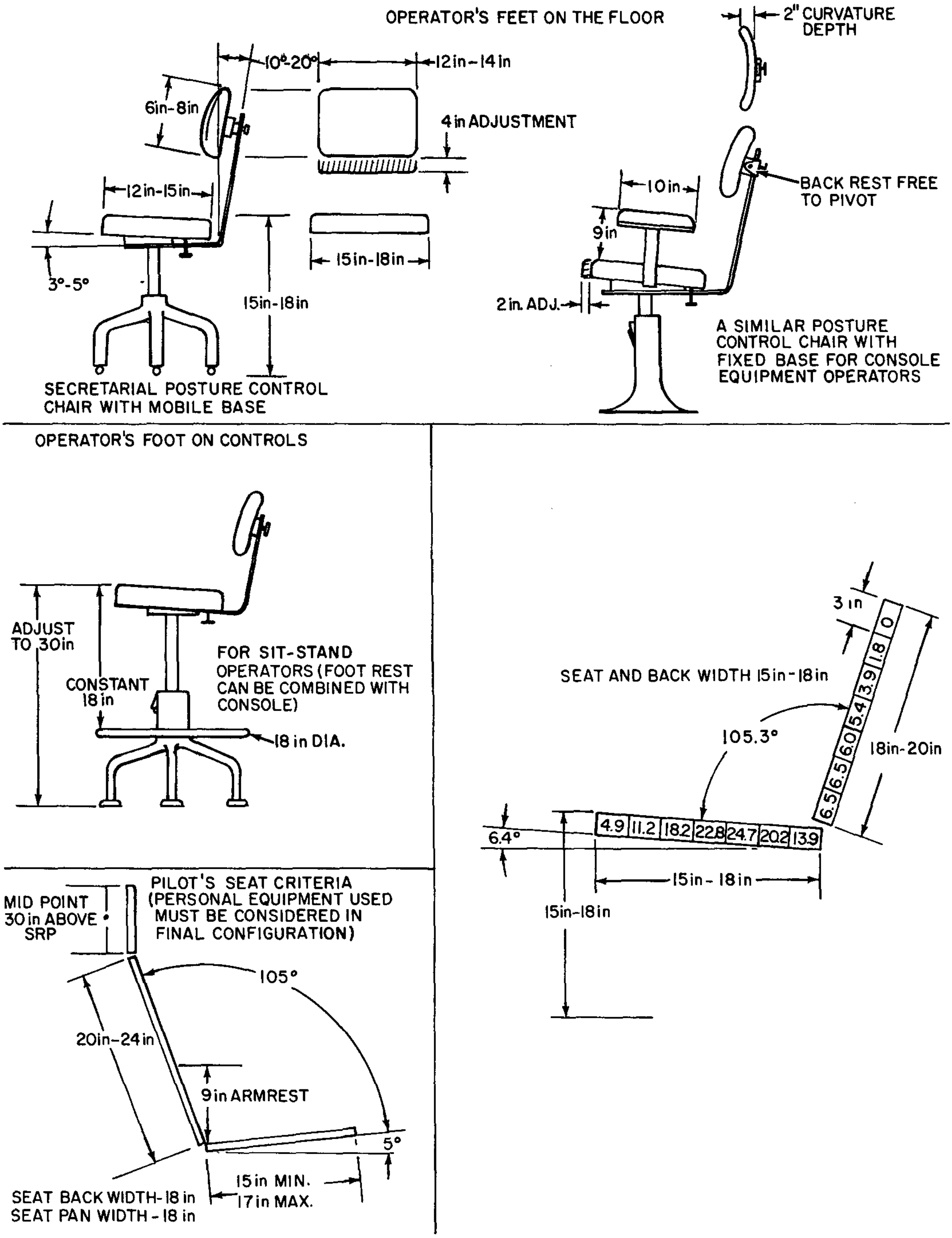


FIGURE 9–23. Typical dimensions for common seat configurations (Lay and Fisher, 1940).

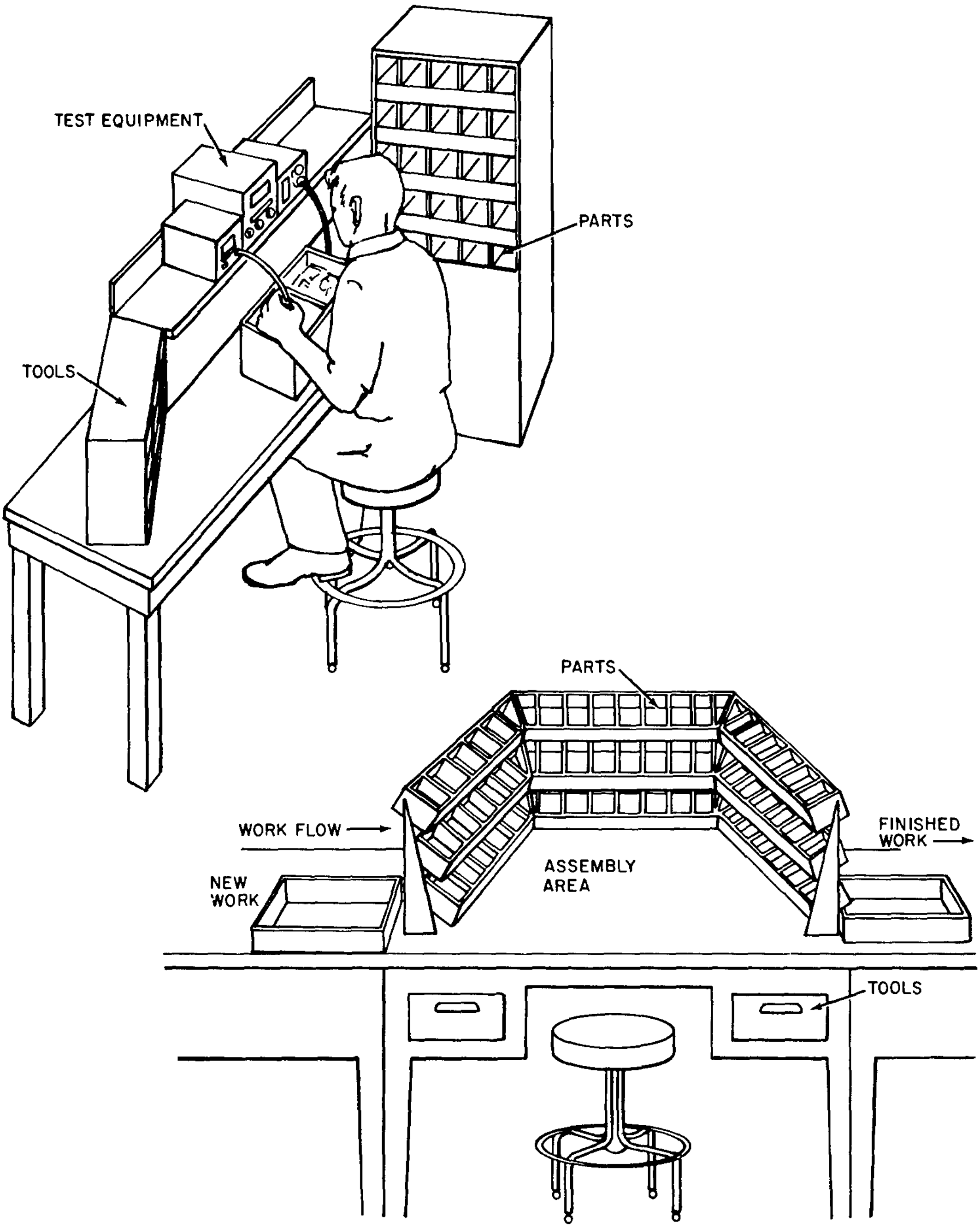


FIGURE 9–24. Concepts for layout of test equipment, parts, and tools.

Dimensional factors applicable to these concepts are the same as those that have been presented in the discussions of operational workplaces.

9.5.12 Storage

Individual workplaces should usually provide for storage. Such items as operational checklists, maintenance handbooks, special test equipment, tools, outdoor or protective clothing, and coffee cups are usually stored in individual workplace facilities.

The designer should avoid the tendency to create storage out of "what is left over." In planning storage, the following factors should be considered:

1. Size and shape of items to be stored.
2. Method of inserting and retrieving stored items.
3. Separation of items within the storage area so individual items can be retrieved quickly and easily.
4. Reach to the storage area.
5. Labeling of storage so items are easy to find and return to place.
6. Proper illumination.
7. Design of the storage area so that it does not collect dirt easily, and so that it can be cleaned out.

References

Briggs, S. J. A study in the design of work areas. Unpublished Ph. D. Dissertation, 1955, Purdue University, Lafayette, Ind.

Caldwell, L. S. The effect of elbow angle and back support height on the strength of horizontal push by the hand. Report No. 378, 1959, U.S. Army Medical Research Labs,. Fort Knox, Ky.

Caldwell, L. S. The effect of the spatial position of a control on the strength of six linear hand movements. Report No. 411, 1959, U.S. Army Medical Research Labs., Fort Knox, Ky.

Dreyfuss, H. *The measure of man.* New York: Whitney Library of Design, 1959.

Ely, J. H., Orlansky, J., & Thomson, R. M. Layout of workplaces. Report No. WADC–TR–56–171, 1956, Wright Air Development Center, WPAFB, Ohio.

Fitts, P. M., & Cranell, C. Location discrimination: II. Accuracy of reaching movements to twenty-four different areas. Report No. AF–TR–5833, 1950, Aero Medical Lab., Wright Air Development Center, WPAFB, Ohio.

Havron, M. D. Information available from natural cues during final approach and landing. Report No. HSR–RR–62/3–MK–X, 1962, Human Sciences Research, Inc., McLean, Va.

Hedgcock, R. F., & Chaillet, R. F. Human factors engineering standard for vehicle fighting compartments. U.S. Army HEL Std. No. S–2–64, 1964, U.S. Army Human Engineering Labs., Aberdeen Proving Ground, Md.

Hugh-Jones, P. The effect of limb position in seated subjects on their ability to utilize the maximum contractile force of the limb muscles. *Journal of Physiology,* 1947, **105,** 332.

Hunsicker, P. A. Arm strength at selected degrees of elbow flexion. Report No. WADC–TR–54–548, 1955, Wright Air Development Center, WPAFB, Ohio.

Kennedy, K. W., & Bates, C. Development of design standards for ground support consoles. Report No. AMRL–TR–65–163, 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Lay, W. E., & Fisher, L. C. Riding comfort and seat cushions. *Society of Automotive Engineers Journal,* 1940, **47** (5), 482–495.

McFarland, R. A., & Mosley, A. L. Human factors in highway transport safety. Harvard School of Public Health, Boston, Mass., 1954.

Meister, D., & Farr, D. E. Designer's guide for effective development of aerospace ground equipment control panels. Report No. AMRL–TR–66–21, 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Thomson, R. M., Jacobs, H. H., Covner, B. J.,

& Orlansky, J. Arrangement of groups of men and machines. ONR Report No. ACR–33, 1958, Office of Naval Research, Washington, D.C.

Webb, P. (Ed.) Bioastronautics data book. Report No. NASA–SP–3006, 1964, National Aeronautics and Space Administration, Washington, D.C.

Woodson, W. E. *Human engineering guide for equipment designers*. Berkeley, Calif.: University of California Press, 1954.

Woodson, W. E. Human engineering principles in console design. Un-numbered report, 1954, U.S. Navy Electronics Lab., San Diego, Calif.

Woodson, W. E. Human engineering design standards for spacecraft controls and displays. Report No. GDA63–0894–1, 1963, General Dynamics/Astronautics, San Diego, Calif.

Woodson, W. E., & Conover, D. W. *Human engineering guide for equipment designers* (2d ed.). Berkeley: University of California Press, 1964.

Chapter 10

Design of Multi-Man–Machine Work Areas

Robert M. Thomson

Operator performance and safety depend upon the way in which work areas and accesses are arranged and designed. While an adequate quantitative methodology has yet to be developed in support of the many necessary human engineering trade offs, qualitative assessments are helpful. These could include link analyses, adoption of anthropometric data in appropriate situations, and judgment based on past experience. Topics considered in this chapter are design tools and methods, followed by a discussion of various factors affecting workplaces areas, such as safety, mobility, equipment access, and visual, and communicative requirements. Finally, arrangement of groups of men within specific areas such as conference rooms, and layouts of traffic spaces, are considered.

CONTENTS

This chapter was reviewed by Harold P. Van Cott.
It is based in part on material appearing in Chapter 8 of the previous *Guide*.

10. Design of Multi-Man–Machine Work Areas

10.1 Approach to Design of Multi-Man–Machine Work Areas

One of the most important but difficult aspects of human engineering is the design of multi-man–machine work areas. In principle, the logic of the design process is simple. In practice, many complex factors constrain the freedom of the designer and force him to make compromises. Nevertheless, poor work area arrangment dramatically affects system performance as well as human efficiency and safety.

Before planning even a provisional work area, the designer needs information about each of the following:

1. The system's mission and the environment in which it will operate.
2. The system maintenance concept.
3. The necessary equipment items with their dimensions and functions.
4. Equipment interconnections.
5. The number and function of all personnel.

With this knowledge in hand, the designer will be able to:

1. Arrange the most important equipment and personnel functions within the space constraints of the work area using link analyses and other design tools.
2. Consider secondary equipment and support personnel, as well as facilities for storage, biological relief, rest, and recreation.
3. Adjust relationships between equipment and work locations and attempt to optimize viewing and talking links, insuring adequate space for movement of personnel and materials, for maintenance, and for normal and emergency access and egress.
4. Check adequacy of visual angles and lines of sight.
5. Test the provisional layout or portions of it by using reduced scale mockups, models, and other design tools.

In completing these steps, the designer should follow these general principles, as stated by Woodson (1965):

1. Plan first the whole, then the parts.
2. Plan the ideal, then compromise with the possible.
3. Plan the layout around the system and its requirements.
4. Plan the final enclosure around the layout.

These are simple-sounding prescriptions, but their implementation is difficult. Layout design tends to become "frozen" early, precluding extensive changes later. Expensive and time-consuming as prior testing is, it is essential because many interactive effects in a layout cannot be anticipated. Considering these drawbacks, the designer must be a skillful salesman as well as a careful technician—selling the importance of his plans and designs in terms of the costly consequences of human error or system failure.

10.2 Design Tools and Methods

A number of qualitative methods and tools can be of use in resolving design problems. These methods, described below, will help throughout the analysis and design process.

10.2.1 Mockups (Reduced-Scale)

A reduced-scale mockup (1 in. = 1 ft. is a generally convenient scale) is a valuable tool in the early stages of multi-man–machine layout design. When constructed with reasonable accuracy, tolerances can be measured and scaled to the equivalent of ±0.5 in. The reduced-scale mockup may be portable so that it can be inspected and manipulated easily. Modification can be quick and inexpensive. These features are generally lacking in a full-scale mockup.

10.2.2 Mockups (Full-Scale)

Two advantages gained only with a full-scale mockup are (a) a realistic concept of size, and (b) an accurate measure of the effects of human body articulation. Specifically, the effects of possible awkward body positions, obstructions to hand or tool access (such as through access portholes), etc. can be assessed only in this manner. For most purposes, cardboard or similar materials yield sufficient exactness. A full-scale mockup of only a portion of the total work area may be all that is needed. See Seminara and Tevis (1963) for a brief but inclusive survey of full-scale mockup uses and advantages.

10.2.3 Models

Reduced-scale models of particular aspects of a layout (built to a ratio as small as 1:16) are useful for the measurement of lines of sight, visual angles, and other geometric properties not easily observed from a two-dimensional blueprint or drawing.

It is desirable that the equipment in scale models be outfitted with transparent plastic skirts representing required clearances. This is particularly helpful in preventing possible infringement on these clearances which could occur when space constraints are severe and clearances are ignored. Vital clearances that should be considered are ventilation, cable, operator, and maintenance clearances. (See Figure 10–1.)

10.2.4 Operational Sequence Diagram (OSD)

The Operational Sequence Diagram (Brooks, 1960) is a good means for examining two significant aspects of compartment layout: (a) sequential interactions among men and equipment, and (b) critical time constraints on performance.

The OSD is a sequential record of each action performed by each element of the system. In its simplest form, this is a repetitive chain of reception of information, decision, action taken, and further information generated. This can be as fine grained as needed. Ordinarily, it will become progressively more detailed as specific information about required actions becomes available.

10.2.5 Correlation Chart

The correlation chart is a matrix for recording and examining relationships between personnel

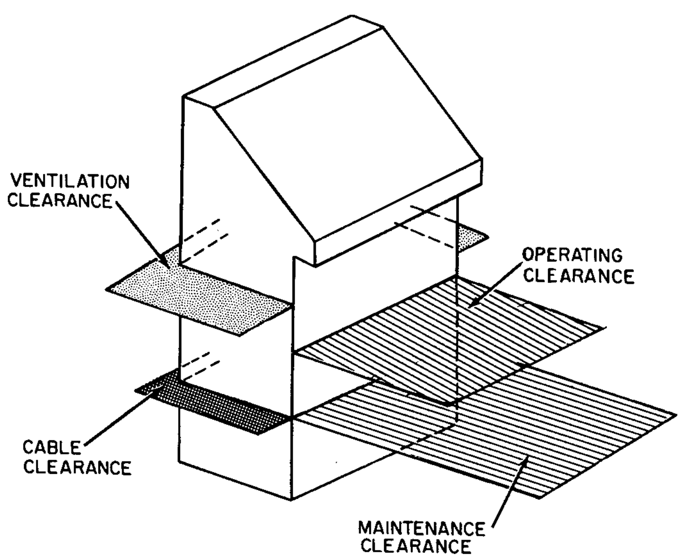


FIGURE 10–1. Transparent plastic skirt represents required clearances around scale model (Thomson et al., 1958).

and equipment. Personnel and Equipments are row and column headings, and their relationships are entered in each matrix cell. Also recorded are the type of relationship, its importance, and such information as special limitations governing relationships (e.g., maximum allowable separations due to cabling interconnections, direct voice communications in the midst of high ambient noise level, and so on). Information for the correlation matrix chart may be derived from an Operation Sequence Diagram.

10.2.6 Link Analysis

Link analysis is a technique which provides information needed to produce an acceptable arrangement of men and machines in a system. Before a link analysis can be performed, firm decisions must be made about the exact items of equipment to be used, the number of men who will operate them, and the functions that will be performed. The rationale behind the link analysis technique is that the "best arrangement" can be found only by optimizing different types of links (such as communication and movement) that are important in the particular system being designed.

The term "link" refers to any connection between a man and a machine or between one man and another. If, for example, one man must talk to another, this need is represented by a link between them. Similarly, if a man must see the display on a machine or operate a control on a machine, he has a link to the machine. Links include walking, talking, seeing, and movement of material and information. Ordinarily, links between machines can be neglected *if* the length of the link is of no importance in the system being considered.

The procedure for making a link analysis of a man–machine system is as follows:

Step 1. Draw a circle for each man in the system and label it with a code number for his particular function (e.g., "1" for radio operator, "2" for navigator, "3" for plotter, etc.).

Step 2. Draw a square for every item of equipment used by a human operator and label it with a code letter (e.g., "A" for radio, "B" for plotting board, "C" for compass, etc.). It makes little difference how the circles and squares are arranged at this point so long as there is some room between them.

Step 3. Draw connecting lines (links) between each man and any other man or men who have any direct interaction in the operation of the system.

Step 4. Draw connecting lines between each man and any machines with which he must interact.

Step 5. Redraw the resulting diagram, reducing to a minimum the number of crossing links in order to obtain the simplest possible arrangement.

For many systems, this procedure completes the link analysis, which will then show how those men and equipments belong together and how to separate those that have little relation to each other in the operation of the system. If this has been such a case, one would skip the next two steps and continue with Step 7.

Where the preceding steps yield many crossings that reveal conflicting requirements for the proximity of men and machines, it is necessary to evaluate the frequency of use, and the importance of each type of link. When this is true, proceed as follows.

Step 6. Evaluate each link by one or more of the following methods:

Where the *importance of each link* is the criterion, have an experienced person—someone who thoroughly knows the operation of the system—rank each link according to its relative importance by assigning low numbers to unimportant links and high numbers to important links.

Where frequency of use is the criterion, obtain data from the simulated or operational use of the system. Use these data to rank each link according to the amount of use it gets or the time it is in operation. Enter these rankings on the links.

Where *both* frequency of use and importance of a link must be considered, experienced observers can judge the relative weights to be given so as to assign a single composite value to each link.

Step 7. Redraw the diagram so that the links having the higher values are shorter than those having lower link values and reduce the number of crossing links. This is the optimum link diagram.

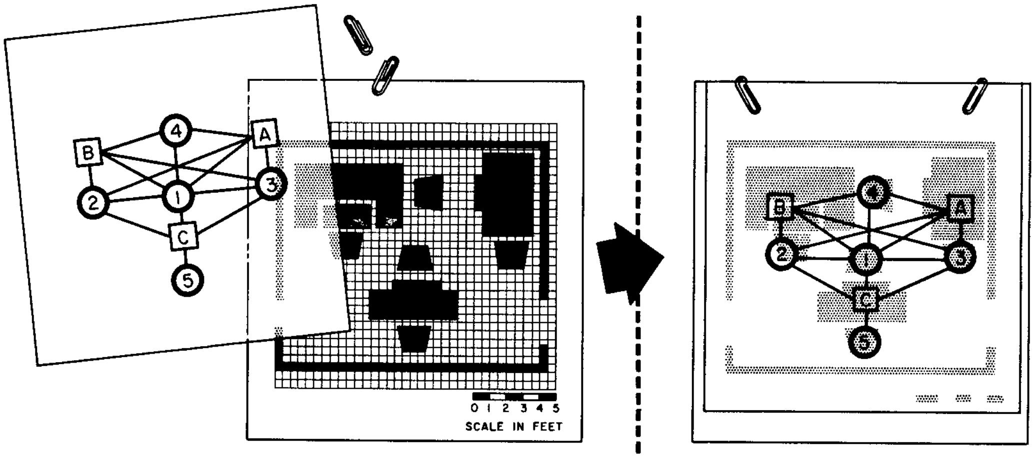


FIGURE 10–2. Link diagram superimposed on scale drawing.

Step 8. Redraw the link diagram, as necessary, to fit it into the available space or, preferably, design the space to suit the shape of the diagram.

Step 9. Confirm the final link analysis on a scale drawing of the actual positions of the men, machines, and spaces comprising the system. Such a drawing enables the designer to visualize the physical dimensions of the system and to discover difficulties that are not revealed by the link analysis. An example of a link diagram by itself, the layout and the link diagram, superimposed on the scale drawing of a Navy communications and control center, is shown in Figure 10–2. Models with computer solutions are now in use for a variety of these analytic methods.

10.3 Factors Affecting Workplace Area Design

10.3.1 Environmental Effects

The subject of environmental effects is one of considerable complexity and cannot be covered here. The guidelines recommended in this chapter assume a normal environment, i.e., unit gravity, air pressure near 15 p.s.i., and temperature 50° to 100° F.

10.3.2 Safety

Specific guides for designing for safety can be listed as follows:

1. All moving parts of machinery and transmission equipment, including pulleys, belts, gears, and blades should be provided with safety guards.

2. Self-locking or other foolproof devices should be incorporated on elevating stands and work platforms to prevent accidents or inadvertent collapse.

3. An anchor or outrigger should be employed on equipment stands with high centers of gravity.

4. Handrails should be affixed to platforms, stairs, and around floor openings, wherever personnel may fall from an elevation.

5. A safety bar chain or rail should be attached across stair or step openings on a platform.

6. Emergency doors and exits should be readily accessible and quick to open. The design should allow the door or hatch to be opened by a single motion of hand or foot. Provisions should be made for emergency exit from secure or classified areas.

7. Areas should be specifically identified where protective clothing, tools, or equipment, such as insulated shoes, nonsparking tools, gloves, or suits, are necessary.

8. "NO STEP" markings should be incorporated where applicable.

9. Illumination adequate for tasks should be provided in all work and work access areas. Work areas should be illuminated by at least 20 ft.–c., access areas by at least 8 ft.–c.

10. Provisions should be made for skid-

proof flooring and stair or step-tread coverings.

11. In the design of telescoping steps or ladders, clearance for fingers should be considered.

10.3.3 Human Body Size and Dynamics

Standards for work area size and shape will depend primarily on the minimum requirements of (a) body size, dimensions, and dynamics, (b) crew size (number), (c) equipment, hardware (geometry, volume), (d) space required for operating, (e) space required for maintaining, and (f) space required for storage.

The above factors determine the basic minimum volume and geometry requirements. In addition, important factors determining configuration and additional space requirements include (a) crew mobility; physical access to and from equipment and in and out of compartments for operation and maintenance, (b) visual requirements; direct lines of sight to other personnel, to equipment, and to the exterior (wind screens, etc.), and (c) auditory requirements; person-to-person voice communication.

Because of human adaptability, there may be pressure on the designer to "shade" the minimum standards recommended for individual and multiple man–machine systems. As a general rule, a waiver of a human engineering minimum requirement places a burden on the operator in the actual working environment which could be unforeseen "on paper." Minimum knee clearances tend to be further compromised by addition of electric and telephone cables, protruding fasteners and hinges, and other reductions to the planned dimensions which may be overlooked in detail in the design planning.

The human body dimensions and dynamics that most directly affect compartment size and shape are given in Chapter 11. The measurements presented in Chapter 11 describe for the most part the nude body. These are acceptable for lightly clothed persons, adding on 1.0 in. to standing dimensions for shoes. Heavy clothing or encumbering equipment have the following effects on nude dimensions:

1. An increase in such static dimensions as trunk thickness and thigh clearance.

2. A decrease in dynamic dimensions such as arc of arm and leg movement.

When special clothing or equipment is worn and no data on space tolerance are available, a mockup is recommended so that the information can be gotten empirically.

The following recommendations apply to the use of horizontal body dimensions in plan-view layouts:

1. Use the 95th-percentile when the size of the operator to be stationed at any particular equipment cannot be predicted in advance.

2. Add 4 in. to either side of lateral dimensions to allow for normal "elbow room" and to take into account activities such as rotary arm movements that require several pounds of torque.

Individual workplace dimensions are discussed in detail in Chapter 9.

Mobility

The size, shape, and design characteristics within a given compartment are affected by the size and mobility of the crew stationed there. Obviously, large crews require space, since mobility requires aisles, corridors, etc. to move about. Such qualitative differences as relationships between aisle and corridor location with respect to equipments or the protection of visual clearances also depend on crew size and mobility.

Crew size

A large crew imposes conditions which affect the following:

1. Communications. As crew size increases, unaided voice communications decrease in effectiveness and reliability. Either physical aids to voice communication or personnel mobility permitting face-to-face contact become necessary.

2. Physical access. Additional space must be provided for access from compartment entrance to and between the various work stations. For one or two operators, this can be reduced to a minimum (e.g., an aircraft cockpit). For accesses through which several persons must sometimes hurry, additional clearances should be provided.

3. Illumination. Illumination requirements become more complex for a large crew, especially where these requirements vary among operators. General lighting must be provided for the overall area, aisles, and exits. Individual illumination, filters, and partitions often must be provided for specific personnel. Electrical outlets for local plug-in illumination should be provided for all maintenance operations.

4. Acoustics. When several operators are in a compartment, special equipment may be needed to permit internal voice communication and to filter interfering and extraneous sounds. The use of common voice circuits brings about special problems.

5. Environment. As crew size increases the problems of providing adequate ventilation and air conditioning increase.

6. Space and equipment. When two or more men occupy the same compartment, savings are possible in space and in equipment through sharing of common functions.

7. Visual information. Visual information presented on a common display to three or more persons must be increased proportionally in size to the number of persons viewing the information up to about 12 persons.

Crew mobility

The nature of an operator's job determines his work station layout. Provision for freedom to move about means provision for uncluttered access to workspaces.

Standing allows mobility. For jobs that require mobility, the operator will generally stand or will alternate between sitting and standing. The effects of these positions are simple but important. For instance, the standing position is characterized by the following compartment-layout considerations:

1. The standing operator requires less front-to-rear operating room; the additional vertical room is usually provided at no penalty.

2. The standing operator is more mobile, even when confined to a small operating area, and can move out of another person's way. Thus, if traffic is not too heavy, the standing operator can be stationed in a traffic space or aisle. For example, aboard ship, controls and displays are mounted on the deck and bulkheads of the bridge. Thus, the space against the after bulkhead of the pilot house, between the bridge wings, is used for mounting displays to accommodate bridge walkers and to serve as a passageway. (See Figure 10–3.)

3. The standing operator does not require furniture.

Standing operators permit greater design latitude. The factors listed above permit design latitude in compartment layout as well as savings in total floor or deck space as compared with designs for seated crew members. The standing operator can be surrounded by controls and displays which are not all within his frontal vision or reach. (See Figure 10–4.)

Seating permits superior operation. Standing operators find some operations of pedal controls awkward; in moving vehicles they may have to expend considerable effort to maintain their footing.

Operators should not be required to stand continuously unless the job requires it. Preferably, the designer should allow the operator to stand or sit at will, especially where the job is continuous.

Limitations to seated operation. The seated position has the following drawbacks:

1. The control area may be limited, with movement from one position to another hindered.

2. Open space (not a part of the useful operating area) generally is required for access to the seat, although swiveling and (in-out) sliding seats yield access to minimum space requirements.

3. If seats are in a row, additional space is required for access to the inside seats.

Where operators are seated in close proximity, however, space will be saved by running an aisle directly in back of the seats. This aisle space can be used jointly for general traffic, individual access to operating stations, seat adjustments, and maintenance purposes.

Transient Personnel

Crew size may vary intermittently with transient personnel who are concerned with messages, maintenance, liaison, and support. Without adequate provision for such personnel,

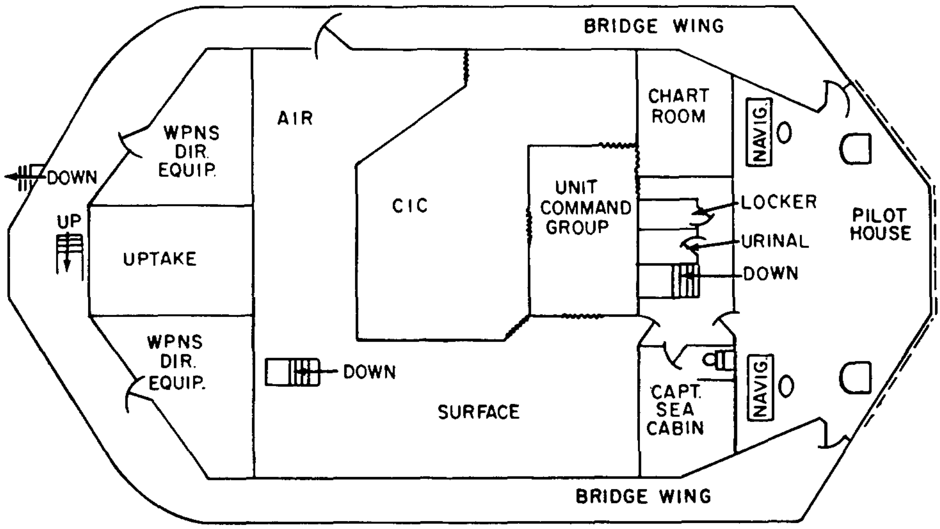


FIGURE 10–3. Mobility arrangement on bridge of ship (Thomson et al., 1958).

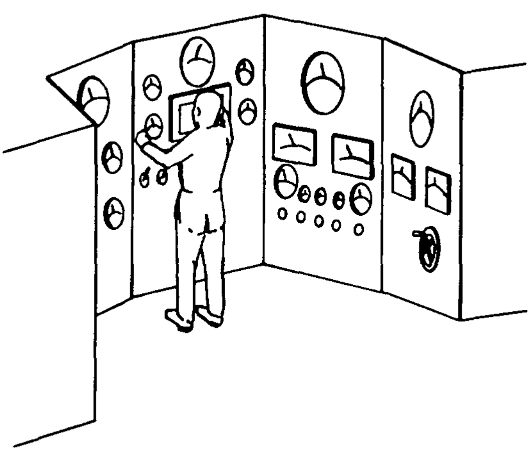


FIGURE 10–4. Standing operations permit greater design latitude.

they can hamper operation of the system. The following should be considered:

1. Access space (and adequate visual clearance) to important displays should be provided.
2. Space should be provided next to those persons who must receive and dispatch written or printed messages.
3. Aisle space to "operating" positions should allow for continuous traffic.
4. Visiting observers are generally allotted special permanent space set aside for this purpose.
5. Design criteria should provide for the most crowded condition (excluding special test trials, visiting "VIP" exercises, etc.) occurring in full system use.

10.3.4 Equipment

The space required for an equipment must include its gross dimensions as well as clearances necessary for its operation and maintenance. The operational requirements must be considered as primary and as inviolate as the physical space occupied by the equipment itself. Other space requirements might include:

1. Cables and Ventilation. Space becomes fixed once the cable or duct run location has been selected.
2. Maintenance. Space must be provided but can be shared with other functions, e.g., (a) operation of equipment (if the equipment is non-operational while maintenance is being performed in that area), and (b) maintenance of adjacent equipment. Conduct of maintenance must not infringe on an *adjacent* operating area.

In general, space can be saved when small rather than large or "consolized" equipment is used because small equipment is more versatile with respect to arrangement.

1. Small units can be ideally located relative to the operator—mounted at any height and facing in any direction.

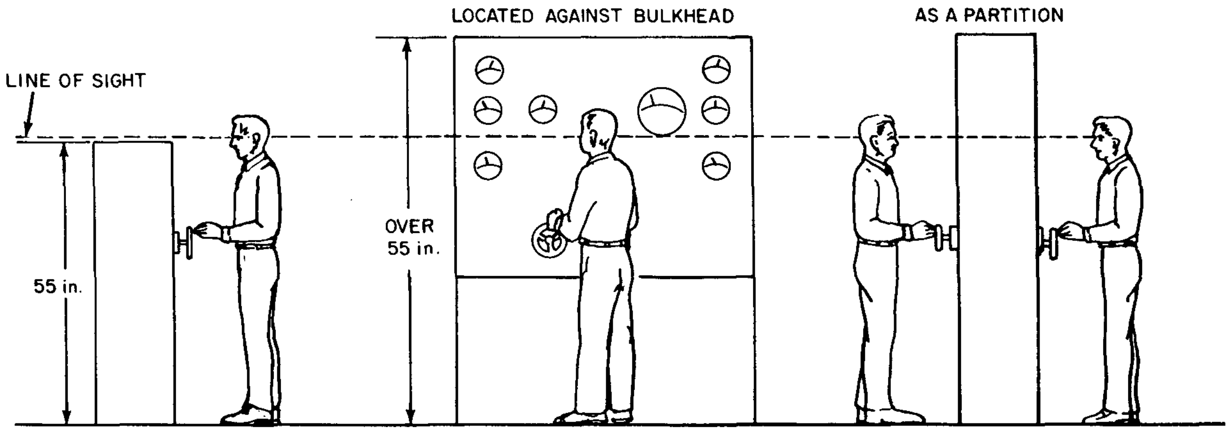


FIGURE 10–5. Equipment size and location affect operator visibility (Thomson et al., 1958).

2. Small equipment can be stacked or mounted overhead or on bulkheads.

3. When used infrequently, units can be mounted in low-priority areas (i.e., once located, they can later be rearranged).

4. Small equipment can be carried or moved with little effort or mechanical assistance.

Once a satisfactory arrangement has been thoroughly tested in operation, a console may be designed to house as many of the units as possible in order to assure certain advantages, e.g., a smooth and continuous operating surface, centralized ventilation, power supplies, test points, efficiency of edge lighting, intensity controls, etc.

Equipment height partly determines compartment layout, especially where clearances are needed for viewing the same display by more than one operator and for direct voice communication between operators. The maximum height of equipment over which something can be viewed will vary according to the location of the object to be viewed and position of the viewer. Before locating any equipment over 55 in. high (about 5 in. less than the standing eye height of the short operator), it should be determined that the extra height will not interfere with any operator's line of sight. Generally, equipment taller than this should be located against a wall or bulkhead or in an open space where it serves as a partition. (See Figure 10–5.)

Equipment Operation

The volume required for operating equipment can be added directly to that occupied by the equipment itself. The design of the equipment will determine, within close limits, the method of operation and, consequently, the required operating room, regardless of the location of the equipment with respect to other equipment within the compartment. If supervision is a required part of the operating routine, additional space must be added as a proper part of the normal operating space. If the operator is seated, the volume occupied by the seat together with its movement must be added to the operating area.

Access to Equipment

1. Bulkheads, brackets, and other units should not interfere with removal or opening of covers of units. Unobstructed clearance for opening or removal of covers should be provided.

2. Maintenance checks should be possible without the use of special rigs and harnesses.

3. Units should be laid out to allow a minimum of place-to-place movement by the operator during checkout.

4. Removal of any replaceable unit should involve opening a minimum number of covers or panels.

5. Units should be so located that no other equipment must be removed to gain access.

6. When one unit is placed behind another, the unit requiring most frequent access should be most accessible.

Maintainability. (See Chapter 12.) Electronic equipment should be designed so that as many maintenance operations as possible (especially unscheduled or corrective maintenance) can be performed in repair shops or some location remote from the operating area. This prevents the maintenance operator and his activities from interfering with the performance of the operators' duties and allows maintenance to be performed unhindered by equipment operators, under conditions of adequate light, visibility, cleanliness, and temperature, with all necessary test equipments and tools conveniently accessible.

Where possible, the design for access to replaceable or adjustable assemblies should include *each* of the following:

1. The use of hinged or removable chassis, permitting replacement of an entire electronic "drawer" (with further maintenance performed remotely).
2. Design of major units and assemblies with removable housings to make complete visual inspections possible.
3. Correlation of unit accessibility features with the accessibility requirements of the overall system.

Rests and stands on which removable units can be placed should be provided. These should incorporate provisions for test equipment tools and manuals, and if they must be used for maintenance, the rests and stands should be a part of the basic chasses.

The removal of any replaceable item should require opening of only one access, unless these are of the latched and hinged door type.

Visual access should be possible for all maintenance operations requiring visual inspection and control and particularly where electrical or mechanical hazards can be encountered within the access. The technician should never be required to work blindly.

Maintenance accesses should be located:

1. Only on equipment surfaces that will be accessible in normal installation.
2. To permit direct access and maximum convenience for job procedures.

Covers and shields should be designed, located, and mounted so that (a) they do not obscure or interfere with controls, displays, test points, or connections related to work within the access or enclosure; and (b) when in the open position they are provided with adequate stops and retainers to prevent them from swinging into or being dropped on fragile equipment or personnel.

Units should be removable along opening lines that are straight or only slightly curved.

Cases should be selected, designed, and mounted so that (a) maintenance portions of the equipment are fully exposed when the case is removed, and (b) the case itself can be stored temporarily in the aisle without totally blocking access to the equipment.

Design of Access Openings in Order of Relative Importance

1. Hinged doors, hoods, and caps permit easiest access. Such covers do require "swinging space," however, and may interfere with other operations. Where opening space is a problem, double-hinged or split doors might be feasible.
2. Sliding doors or caps are particularly useful where "swinging space" is limited. Small sliding caps are useful for small accesses that do not require a close seal.
3. Removable doors, plates, or caps require little space for opening and, once removed, do not interface with work space. However, to remove them conserves time and effort in searching, bending, or reaching.
4. Removable cover panels or sections are useful where access to the whole side of equipment is required. They discourage non-maintenance personnel from opening the access. They do not require "swinging space," but are easily damaged and awkward to handle.

Cable Routing

Interconnecting communication cables should either be routed to minimize the possibility of their use as handholds or steps or a protective guard should be placed over the cables. Extension cables should be planned, designed, and provided to:

1. Increase the efficiency and ease of maintenance.
2. Avoid removal of assemblies of components for testing.
3. Allow each functioning unit to be checked in a convenient place.

10.3.5 Visual Requirements

Visual requirements are necessary factors in determining compartment size and shape. (See Chapter 3.) Clearances for adequate vision to major displays and other personnel will affect the *arrangement* of equipment. When a group display must be viewed by three or more operators within a compartment, the degree of distortion that can be tolerated in viewing the display should be determined. Normally, the operators should be stationed so that their lines of sight to the display surface form an angle between 60° and 90° (never less than 45°). Where multiple operators must view a single display, the area containing the display and operators will be wedge-shaped (or an elongated rectangle). This situation will set at least one minimum dimension for the compartment. (See Figure 10–6.)

As a rule of thumb for a row of operators viewing a single display, the row of operators should be seated no closer than the width of the display and preferably not closer than 1.5 times the display width. For moving (animated) displays, operators should not be seated closer than two times the width of the display. It is, of course, desirable to be able to look "head-on" at a display. An oblique viewing angle introduces distortion of perspective and the chance for parallax error (Kinney et al., 1965).

The relationship of a row of observers to a single display requires a trade off between the width of the row and the width of the display. The following rule permits more accuracy than a single "rule of thumb." The minimum distance M between row and display is 0.87 (R+D) where R = width of row, D = width of display. Standard dimensions for all equipment locations can be determined by the operators at each end of the row.

In addition to visual links with a main display, clear lines of sight may be necessary between personnel for communication purposes and to auxiliary displays. This may require an area of unobstructed vision in several directions, with all equipment over 55 in. in height being eliminated from the work area in these directions. (See Figure 10–7.)

A single row arrangement of operators has these advantages:

1. The physical separation between operators may be small enough so that displays viewed by one operator can also be seen by another.
2. Side-by-side seating makes for better communication (by voice, writing, manual signal) than a column (tandem) arrangement.
3. A single set of controls may be operable by pairs of operators, which permits standby operation by either operator in the pair.

A single row arrangement of several operators who share a single display may be relatively inefficient, however, because to obtain an acceptable viewing angle, considerable space must be placed between the row and the shared display. For wider rows, the space between the row and the shared display becomes increasingly large. (See Figure 10–8.)

While a column arrangement alleviates the viewing angle problem, the minimum dimensions of shared displays are governed by the

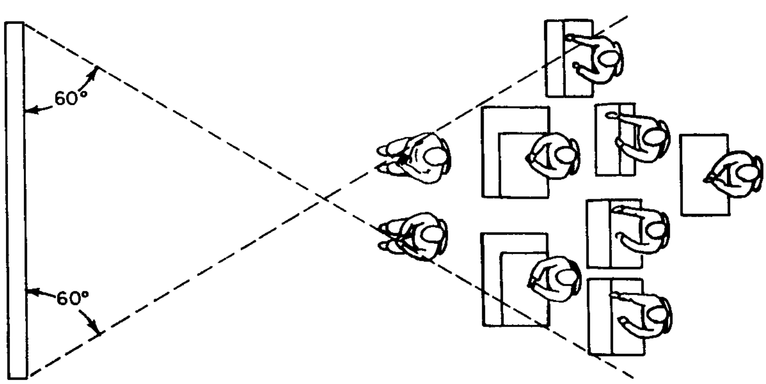


FIGURE 10–6. Area within which operators can view display.

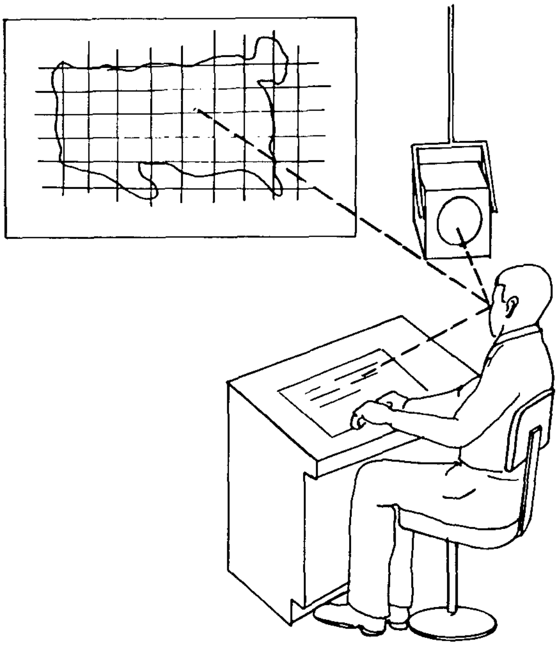


FIGURE 10–7. Unobstructed lines of sight are required to see several displays.

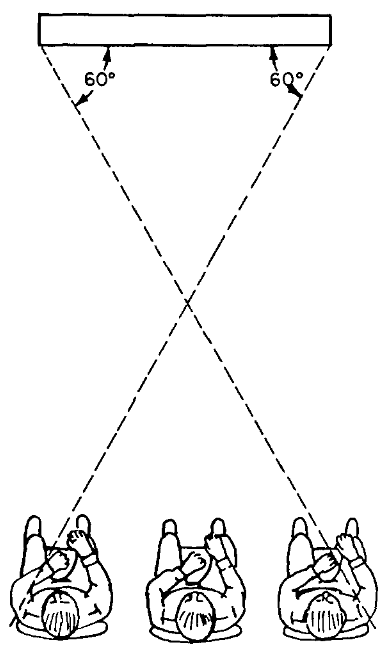


FIGURE 10–8. Arrangement required for an acceptable viewing angle.

most distant viewer in the column. Rows, on the other hand, provide approximately equal distances among all observers to the display. A column arrangement is most economical of space for operators who must share a single display, where there is adequate space vertically, and where a row is necessarily wide.

The designer should insure first that the visual demands on all personnel are above the minimum recommended standards, i.e., that none of these angles is less than 45° where the minimum distance M is unacceptably large. There are at least three possible solutions to this problem as illustrated in Figure 10–9:

1. Angle the outer edge of the display so that angles M′NA and NM′E are equated. This will permit moving the row closer to the display (by effectively reducing the D term in the expression 0.87 (R+D).
2. Provide a supplementary display for the operators seated at opposite ends of the row.
3. Form a second row of operators.

Arrangement of Columns of Operators

Special circumstances may require a columnar arrangement of operators. By offsetting displays to either side of a column, special column arrangements can be avoided (such as staggering). The amount of offset should be determined by the viewing angle of the operators at the ends of the column. If the end operator has a full view beyond the head and shoulders of the person directly in front of him, the view for the others will be automatically clear. It will be desirable to rotate the display a few degrees to give the front operator a better viewing angle. A good general rule is to equate the viewing angles for the front and rear operators; then operators in between will also have an acceptable viewing angle. (See Figure 10–10.)

Seated vs. Standing Operations

Seated operators. Because a columnar arrangement for standing operations often requires excessive headroom, a seating arrangement is preferred:

1. The seated operator's eye and head heights can be 12 to 15 in. lower than those of the standing operator.
2. With adjustable seats, the tall and short operators can use the same seat.
3. When operators are seated, the differential in height because of varying leg lengths is

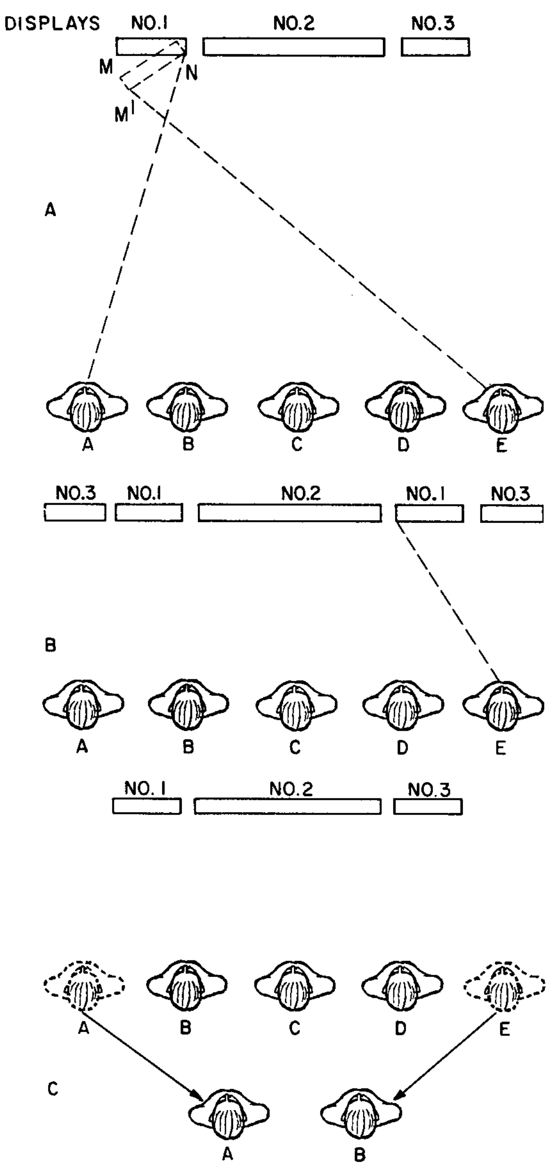


FIGURE 10–9. Three solutions are possible to achieve acceptable viewing by groups: A. Equate visual angles for end operators; B. Provide supplementary displays for end operators; and C. Form two operator rows.

eliminated, leaving only the differential resulting from torso variations; this saves nearly 7 in. of headroom allowance for each operator. (See Figure 10–11.)

Seated and standing operators. When a column of only a few operators is involved, a further reduction in overall headroom can be made by having both seated and standing operators.

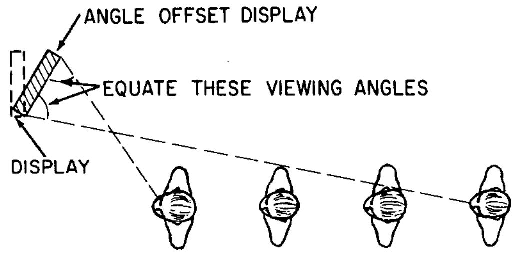


FIGURE 10–10. Display offsets improve viewing angles of operator column.

This arrangement can be provided in the following manner:

1. The front operator can sit in a seat with a low seat pan.
2. The next operator can sit in a standard seat.
3. The next operator can stand or sit on an elevated seat or stool. (See Figure 10–12.)

With adjustable seats and tiered platforms, any number of arrangements can be made. (See also Chapter 9, Section 4.1.)

Standing operators. Where a standing operator must see over the head of another operator in front of him, the following dimensions are involved (see Figure 10–13):

1. The height of the bottom edge of the display above the floor (A).
2. The distance of the front operator from the display (B).
3. The distance between operators (C).
4. Head height of the front operator above the floor whether he is standing or seated (D).
5. Eye height of the second operator (E).

With these dimensions, the seat position can be determined successively for each operator. For example, suppose the dimensions are as shown in Figure 10–14. By subtracting measurement E from D, (74.2 − 61.9 in.), it is evident that the second row must be elevated 12.3 in. At this elevation, the second operator would be able to see horizontally, but when the bottom of the display is lower than the front operator's head, an additional increment must be added. The following formula closely approximates this increment:

$$\frac{(D - A)C}{B}.$$

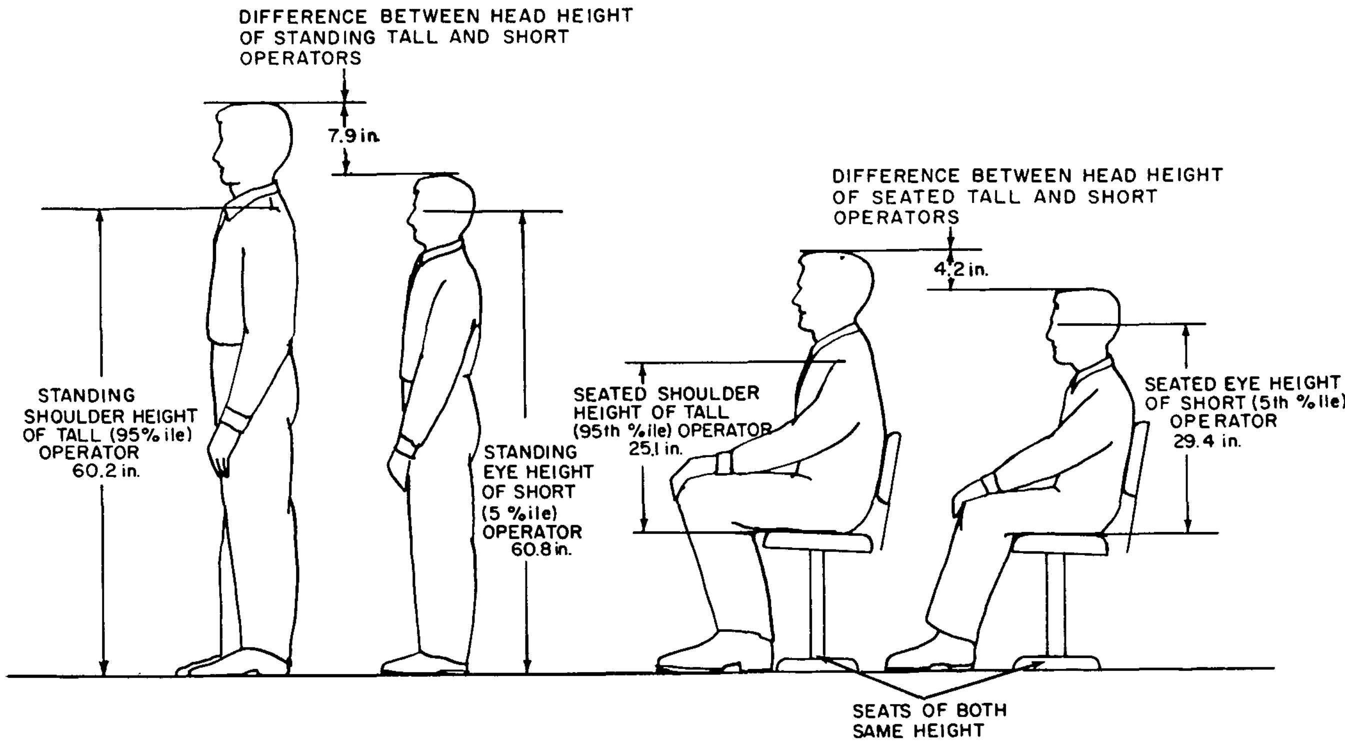


FIGURE 10–11. Seating can be used to reduce viewing problems caused by differences in operator heights (Thomson et al., 1958).

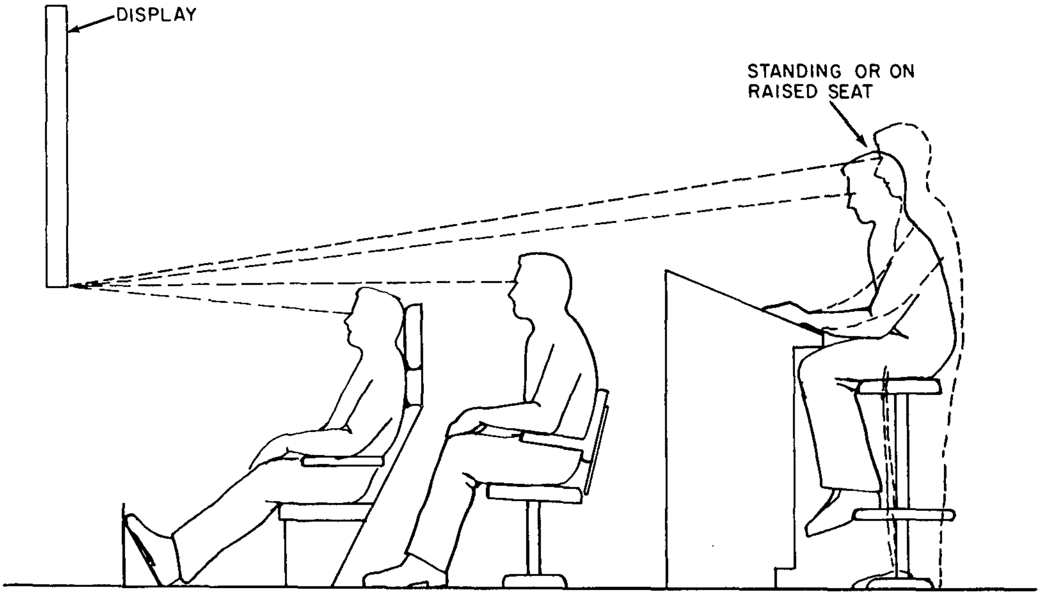


FIGURE 10–12. Differential seating permits all operators unobstructed line of sight to common display (Thomson et al., 1958).

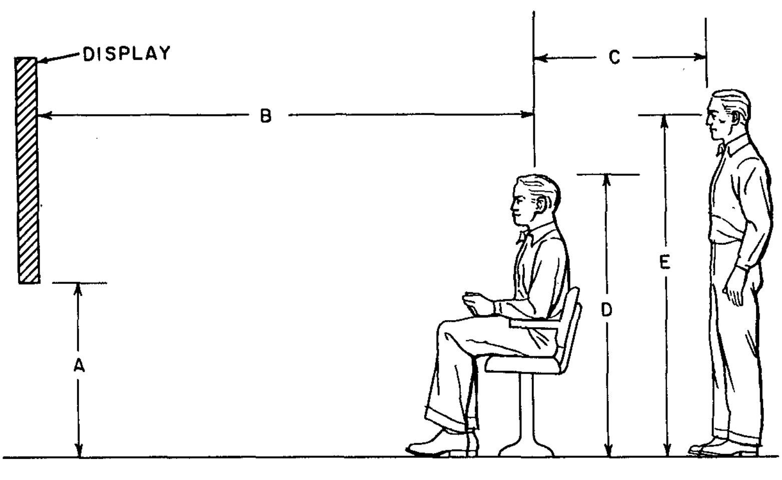


FIGURE 10–13. Dimensions used to compute positions of standing and seated operators.

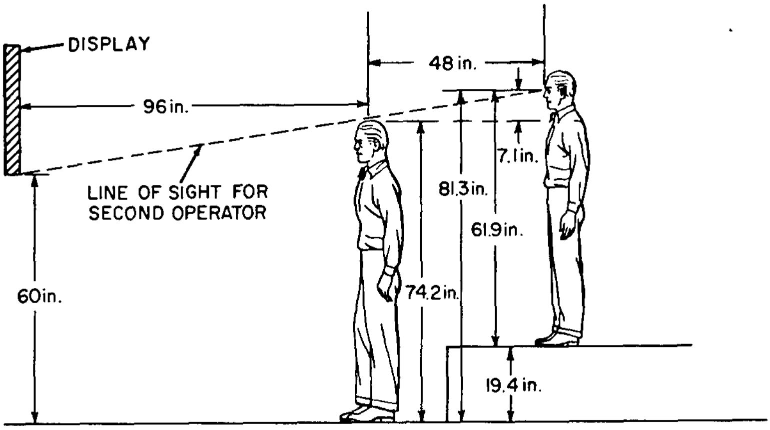


FIGURE 10–14. Examples of dimensions and use of platform.

In the above example, this increment would be (74.2 − 60 in.)(48)/96 = 7.1 in. Thus, the second row should be elevated 19.4 in. (12.3 + 7.1 in.) above the floor level. Note that when the bottom of the display is *higher* than the head of the front-row operator, the quantity (D − A) is negative, and the increment is *subtracted* rather than added. The identical process can be used for computing the platform elevation for each succeeding operator by calculating the required height with reference to the operator in the row in front.

This formula as given imposes *stringent* requirements. If tolerances can be relaxed to the extent that the operator is allowed to stretch or lean to either side to view obstructed portions of the display, smaller clearances are feasible. Reductions in tolerances are appropriate where information displayed changes infrequently, but dynamic information will be permanently lost if momentarily obstructed from view.

Figure 10–14 gives examples of measures involved in determining positions of standing operators and platform height.

10.3.6 Communication Requirements

Communication facilities require spatial planning to allow for the equipment itself, for room to operate it, and to have access to it.

Communication equipment should be located where there is maximum protection from damage during system operations or inadvertent operation by crew members.

The location of a communication set should not interfere with the normal range of movement of the crew (highly restricted in moving vehicles) or be a hazard.

The control panel of the communication set should be visible and readily accessible to all of its operators.

The location of antennas should minimize the possibility of RF burns to personnel.

Control boxes should be located for ease of access to all controls.

Visual Methods of Communication

Visual methods of communication have a great advantage over auditory communication with respect to the flexibility possible in amount and manner of presentation. Auditory communications are superior in most other respects: cost, size, portability, ruggedness, and physical and attention demands on the user. In current systems, visual displays tend to be "primary" equipment, while auditory displays are "secondary."

The distance between the viewer and visual displays determines the required size and brightness of the display for effective visibility. In general, visual information should be arranged within 30° of the viewer's normal line of sight. (See Chapter 3.)

As indicated, the space between viewers and displays must be unobstructed. If the displayed information changes often, it must be free from even momentary obstruction. Since this free space is a three-dimensional volume, the design problem requires a careful mockup verification.

Auditory Methods of Communication

There should be uninterrupted space between personnel who communicate directly by voice to insure adequate (without shouting) intelligibility—intelligibility is enhanced when the listener can see the talker. (See Chapter 5.) Where continuous communication is required, personnel should sit side-by-side or face-to-face. At the least, all personnel requiring direct voice communication should be within normal voice range from the speaker. Acceptable voice range depends on the prevailing noise level and the acoustical characteristics of the compartment.

Intercom units, handsets, voice tubes, plug-in headsets, etc. require little space and can be fitted between larger equipment items, mounted overhead, grouped on fiddleboards or bulkheads within the convenient reach of the operator and out of prime visual-manual operator and out of prime visual-manual operating areas. Communication equipment should always be placed where it is not exposed to damage or where it will cause accidents.

Message Distribution

Communications between compartments can be carried out by message distribution systems: teletypewriter, pneumatic tube, or messenger. Each method requires space for the equipment involved and for storage and delivery of message forms. Where messengers or runners are to be used, they will require access space to key personnel receiving messages. This will involve aisle width of 18 in. minimum for access and a clearance of 24 in. adjacent to the message delivery equipment. In addition, there are requirements for storage of duplicate messages, pneumatic tubes, and other materials.

10.3.7 Storage

Racks or bins should be designed so that stores can be stowed and easily removed either manually or (for heavy or bulky materials) by hoists or lifts. This should be an easy operation from rear and side sections of storage as well as from the front. In moving vehicles, stored materials should be protected from falling out or banging against each other when the vehicle is moving or undergoing sharp shock-vibration (e.g., aircraft encountering rough weather).

In moving vehicles, the items should be individually removable (such as ammunition rounds in armored vehicles), so that the release of one item will not allow other adjacent items to drop, roll, etc.

Bins or shelves should be designed so that the materials can be stored and removed without having to move other equipment.

Unobstructed work space should be provided for transferring items from storage to usage areas.

Items which are inflammable or subject to damage by leakage should be stored to receive reasonable protection from hot engines, exhaust components, and overhead pipes.

Drain holes should be arranged so that they will not be blocked by normal storage.

When signal or warning lights are part of the control unit, the box should be located in such a manner that the signal warning is within the responsible crewman's field of vision.

Storage hooks should be placed in the general area of each crew member for storing (audio) accessories.

Hooks should be located where they will be out of the normal path of crew members.

Items which would be prevented from operating by being exposed to a given set of climatic conditions (moisture, heat, dust, mildew) should receive special storage.

In military vehicles, floor storage boxes should not interfere with crew footing.

Stowed items within a fighting compartment should not interfere with the entrance, exit, escape, movement, or operations (such as gun recoil, loading) of crew personnel.

In vehicles, items to be stored should be capable of being stowed and unstowed by the 5th- through the 95th-percentile man, wearing gloves, without having to assume an unnatural position.

Where space is extremely limited, storage items should be tailored to the available configuration.

10.4 Arrangement of Groups of Men

Men must be arranged in groups when one or more of the following requirements exist:

1. More than one person must see the same equipment or display.
2. Face-to-face communication is desirable.
3. A single equipment requires several operators.
4. Two or more equipments have to be operated in close proximity to each other.
5. Supervision of several persons is required by one supervisor.
6. Available space does not permit physical separation.

10.4.1 Location of Groups

Small groups are located together to jointly share a common area (thereby forming a larger group) for one or more of the following reasons:

1. Material, information, equipment, or personnel is shared or flows from group to group.
2. A single supervisor directs several groups.
3. There is a need for coordination of group activities by several supervisors.
4. Facilities are shared by several groups.
5. A reduction of personnel or improvement of performance is achieved through joint operations.

To meet these requirements, plan the physical location and arrangement of groups in accordance with the following principles:

1. Locate closely related groups in a single building, on a single floor if posible.
2. Arrange groups according to a plan of work flow.
3. Integrate related activities (such as mechanical and nonmechanical, professional and nonprofessional) in accordance with actual requirements for interaction. Don't separate activities simply because work is different.
4. When noise or special lighting requirements are a problem, use partitions between related activities.
5. Where noise or lighting is not a problem, related activities may be separated by space rather than partitions (where space is available).
6. Provide privacy for supervisors.
7. Locate supervisors so that they can readily contact subordinates or subgroup supervisors.
8. Integrate supervisory locations of closely related subgroups.
9. Provide face-to-face access for related groups.
10. Integrate working areas that are closely related; do not separate arbitrarily by storage space, equipment, etc.
11. Give a central location to a group that has many contacts with other groups.

10.4.2 Examples of Arrangements

Illustrated here are some examples of desirable *group arrangements* of varying complexity applicable to a variety of situations. These are provided to serve as a source of ideas for the designer of man–machine systems. Each example is based on actual cases, and its principal features are described.

Ship machinery control station. The arrangement shown in Figure 10–15 (not to exact scale) has the following principal features:

1. Grouping of two cooperating teams: power control team and damage control team.
2. Central overview of all activities by responsible officer (engineering officer).
5. Adjacency of personnel with whom engineering officer must talk (monitoring officer and hull technician).
4. Grouping of power control team: electrical, engine, and boiler monitoring.
5. Arrangement of power control section to permit handling of normal cruising functions by two technicians.

Electronics development laboratory. The arrangement shown in Figure 10–16 (not drawn to exact scale) has the following principal features:

1. Central locations for people or areas having relatively equal contact with, or use by, most groups (chief project engineers, secretarial pool, first aid, building superintendent, and washrooms).
2. Relatively central locations for facilities used by a high percentage of the staff (dining and recreation areas, library, and conference room).

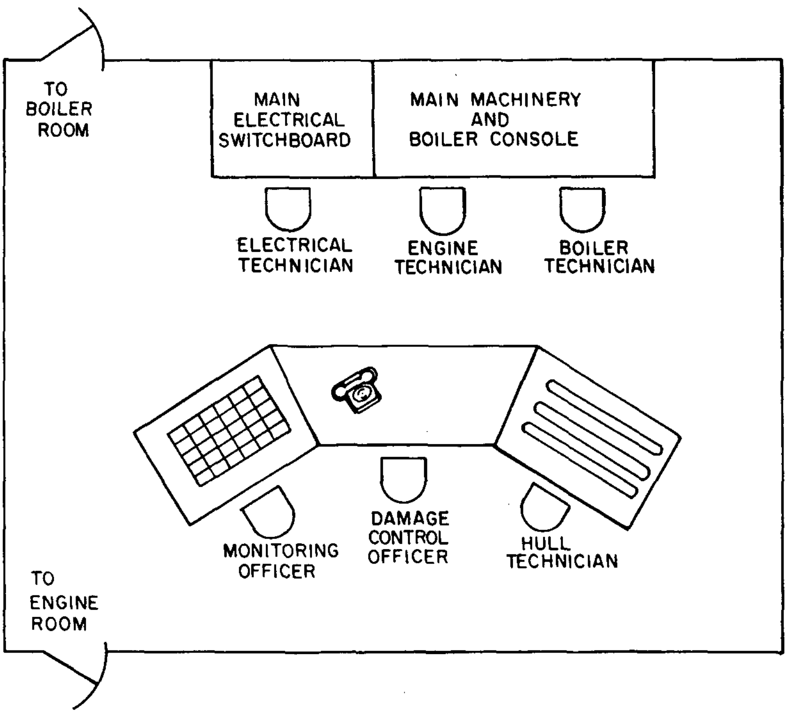


FIGURE 10–15. Arrangement of ship machinery control room.

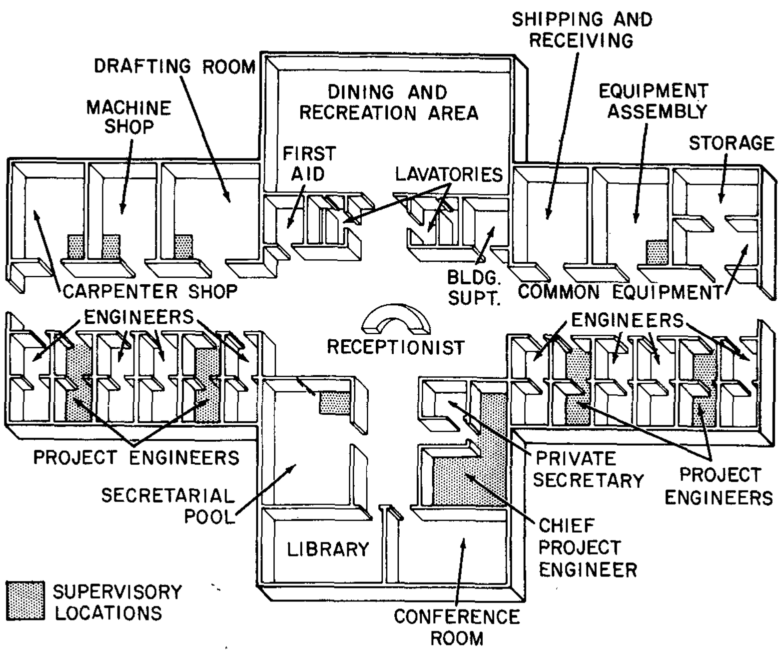


FIGURE 10–16. Electronic equipment laboratory arrangement.

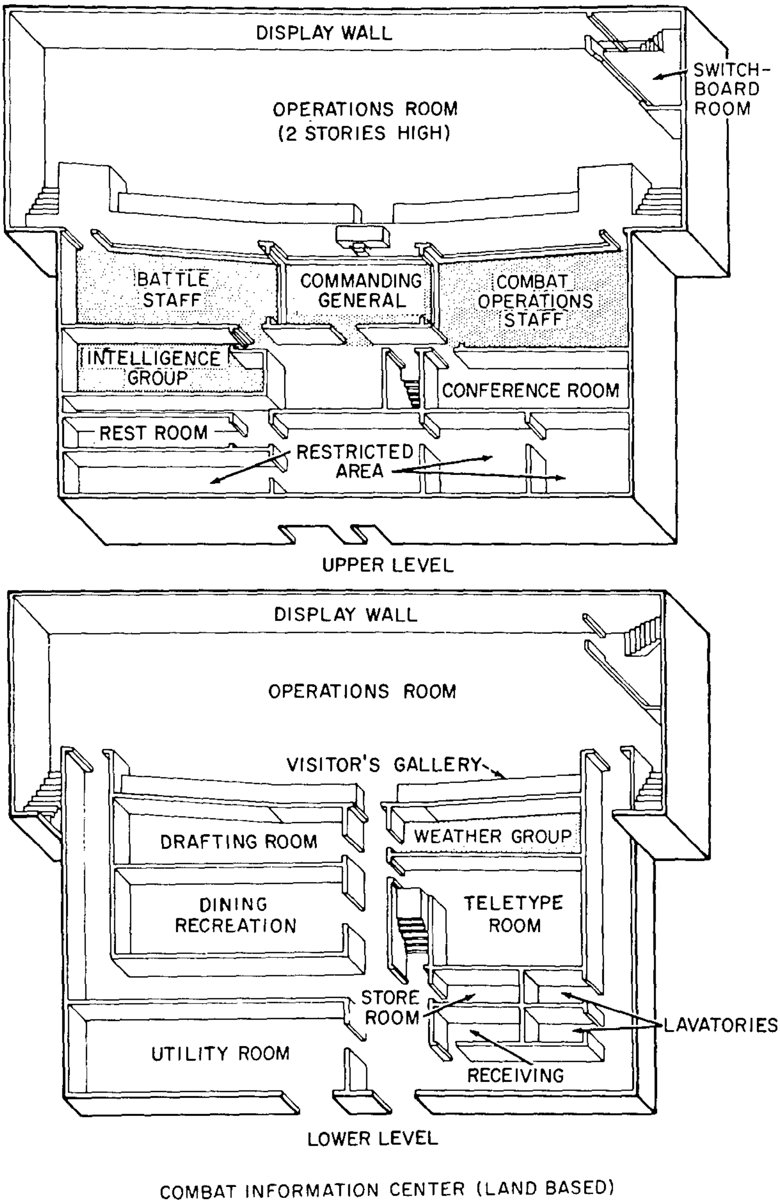


FIGURE 10–17. Land-based combat information center arrangement.

3. Interaction is facilitated by face-to-face access to group areas.
4. Supervisory locations are compartmented but integrated with those of subordinates.
5. The development staff has privacy with ease of access to shop facilities, to superiors, and to each other.
6. Isolation is minimized and counteracted by the location of entrances.
7. A large central area accommodates traffic peaks.

Land-based combat information center. The arrangement shown in Figure 10–17 (not drawn to exact scale) has the following principal features:

1. Central cluster of key personnel.
2. Location plus transparent partitions permit personnel in areas of the battle staff, commanding general, and combat operating staff to have both privacy and a clear view of information displayed in the operations room.
3. The conference-room location facilitates accessibility by the three primary user groups:

battle staff, commanding general, and combat-operations staff.

4. Location plus transparent partitions permit the switchboard operator to search visually for personnel who are not at their stations.

Data processing system. The arrangement shown in Figure 10–18 (not drawn to scale) has the following principal features:

1. Supervisory personnel are accessible to their own groups as well as to each other.
2. The juxtaposition of areas facilitates the flow of work and results in relatively short intergroup distances.
3. Input-output and control groups are shielded from noise.
4. System supervisor is centrally located.
5. A conference room for intra- and inter-system meetings is located near the personnel most likely to use it—the system supervisor and the program group.

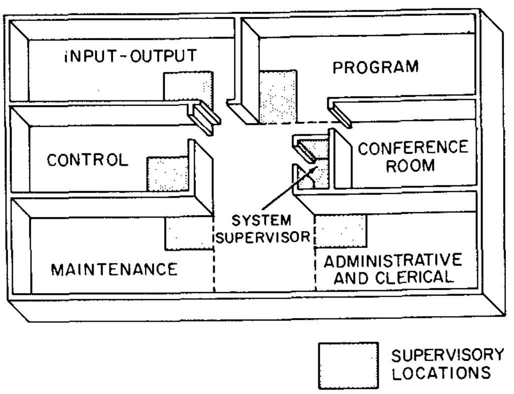


FIGURE 10–18. Land-based command post arrangement.

Aircraft combat information center. The arrangement of land-based, long-range radar aircraft shown in Figure 10–19 (not drawn to scale) has the following principal features:

1. The central supervisory location makes it readily accessible to all other groups.
2. The location of the communications area permits visual and verbal contact with the officer in charge.
3. The standby and monitoring facility is convenient to either the search-track group or officer in charge.
4. Equipment required in the work area is placed to one side so as not to break up the working groups.
5. Maintenance and storage areas are readily accessible to the working area.
6. The dining-recreation area and the lavatory are physically separated from, but convenient to, the work area.
7. The rest area is isolated from the principal work areas.

10.5 Space as a Design Constraint

The previous sections have listed a number of individual factors bearing on work area design. This section presents design recommendations and principles for specific types of facilities, compartments, and compartment elements where space constrains design.

The presence or absence of constraints on available spaces is a chief factor governing design approach. In unconstrained design, equipment and personnel can be arranged in

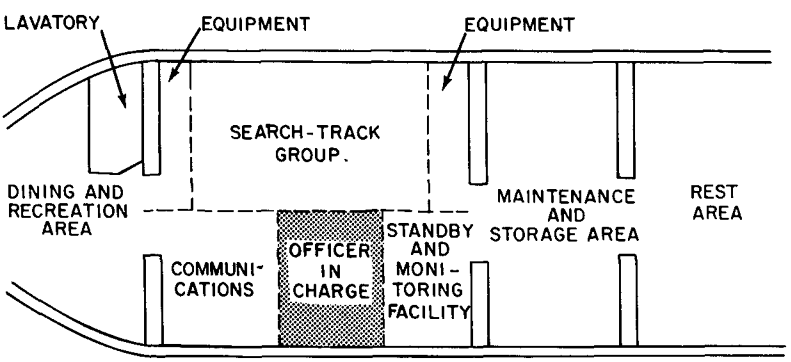


FIGURE 10–19. Airborne command post configuration.

three dimensions without encroaching on human engineering standards.

1. All display characteristics can be within normal visual range.

2. All controls can be within normal reach limitations with respect to distance, relative direction, and operating time.

3. Aisles and access spaces to equipments and operators will be adequate.

Where space is constrained, a satisfactory solution to all arrangement relationships is possible only in simple arrangements where the number of personnel and equipments involved are few in number and the space loading factor is low: i.e., the ratio between (*number of equipments*) + (*personnel*) divided by (*total available space*) is relatively small.

Usually there is no solution which will not violate one or more human engineering minimum recommendations. The principal problem here is to attempt to minimize the seriousness of these encroachments; personnel–equipment "links" are potentially the most troublesome to handle.

Space constraints come in two general forms and each calls for different design goals:

Fixed area or volume as a goal. Here the designer is required to "package" equipment and personnel within fixed limits. His goal is to reduce human engineering deficiencies without imposing other penalties on system effectiveness. The real problem is to determine the loss of effectiveness attached to each failure to meet the human engineering standard.

Minimization of area or volume as a goal. Here the aim is to fit the system into the smallest envelope possible. There is no fixed limit to such a solution. In most man–machine systems, and in moving systems in particular, performance effectiveness is highly correlated with the space (volume) and weight requirements. In a combat vehicle design, for example a space or weight reduction enables a choice of one or more of: (a) power and fuel reduction, (b) larger payload, (c) more personnel, (d) better roadability, (e) smaller silhouette. Trade offs between these factors may not be straightforward. For example, a 2% decrease in volume may permit a 10% increase in payload, which will give an estimated 50% increase in combat effectiveness or tactical advantage. To minimize the volume of a workspace, the designer should first determine possible trade off choices. He can then specify the layout problem, and go to the system designer, the equipment designer, or the human engineer for specific value estimates. On the basis of this information he can then design a compromise arrangement which gives the highest overall system value.

10.6 Conference Rooms and Auditoriums

There are functional distinctions between conference rooms, classrooms, and auditoriums:

Conference (*or seminar*). An arrangement of a small group seated, in a closed loop, facing inward and toward one another, in which frequent participation of each member is expected.

Class. An arrangement of a group facing generally in the direction of the group leader or teacher and his presentation. Group participation is controlled by the leader who presents most, if not all, of the information. Some flexibility (movement of chairs, desks, tables) is possible.

Auditorium. The distinction between auditorium and classroom is essentially that of even larger numbers and less audience participation, the requirement for artificial aids in order to communicate adequately, and no flexibility in rearrangement of facilities.

A conference grouping in which all participate actively probably should not include more than 8 to 10 persons. Five persons are the ideal number for a "conference team" it has been found. Beyond this number, free exchange of ideas and the functioning of the group as a unit almost invariably break down. Side conversation is indulged in and information is no longer unanimously shared. (See Figure 10–20.)

The face-to-face, closed-loop seating arrangement is desirable for enhancing individual participation up to a limit of around 25 to 30 persons. Seating in a conference room should be around-the-table arrangement. This arrangement is more satisfactory for encouraging discussion than the speaker-audience arrangement found in classrooms and auditoriums. The table also provides a large work surface for material to be examined at close range by all participants.

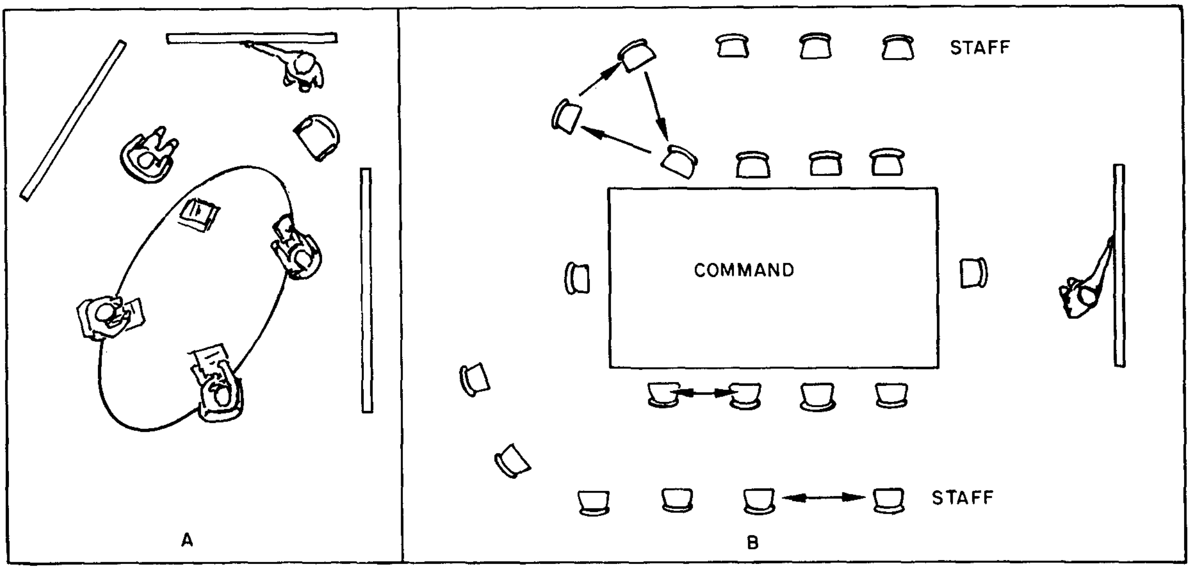


FIGURE 10–20. Conference room arrangements. (A) Five persons; and (B) Twenty-one persons (command and staff).

10.6.1 Conference Rooms

Equipment Arrangement

A properly equipped conference room should have one or more optical projectors (for slides and moving pictures), a screen, a blackboard, and a display board. The viewing angle between the line of sight and information on a screen or wall should be at least 60° if possible, but never less than 45°; for moving pictures, the larger value should be used. Recommendations on the design of visual details to ensure adequate legibility and visibility are discussed in Chapter 3.

Although a microphone and loudspeaker should be available for an unusually large number of personnel in the conference room, or for a lecturer who will make a long briefing, the room size and acoustical properties should make voice amplification unnecessary. If each person partially faces the other people in the room, acoustics are better. No conferee should be more than to 4 to 5 ft. from a microphone.

Seating Arrangement

Maximum seating of 25 to 30 persons assumes that the principals will be limited to 10 persons with others participating as advisors or observers. These latter personnel should be seated in a second outer ring surrounding the principals seated at the conference table.

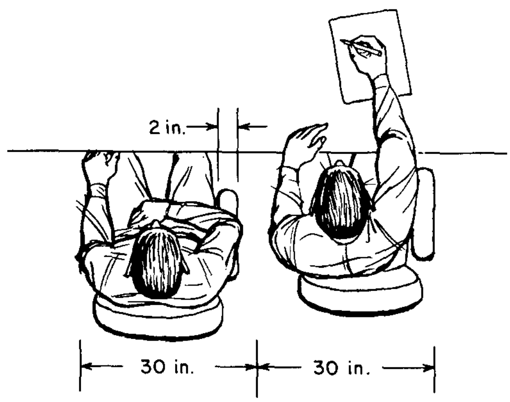


FIGURE 10–21. Member side-by-side conference dimensions.

The distance recommended as the minimum lateral allowance for men seated at a table is 30 in. per person. This clearance permits the use of armchairs with 2-in. wide armrests. (See Figure 10–21.) When several persons are seated at a table in a row roughly perpendicular to a visual presentation, the lateral distance between personnel becomes critical and the required visual clearance will depend on:

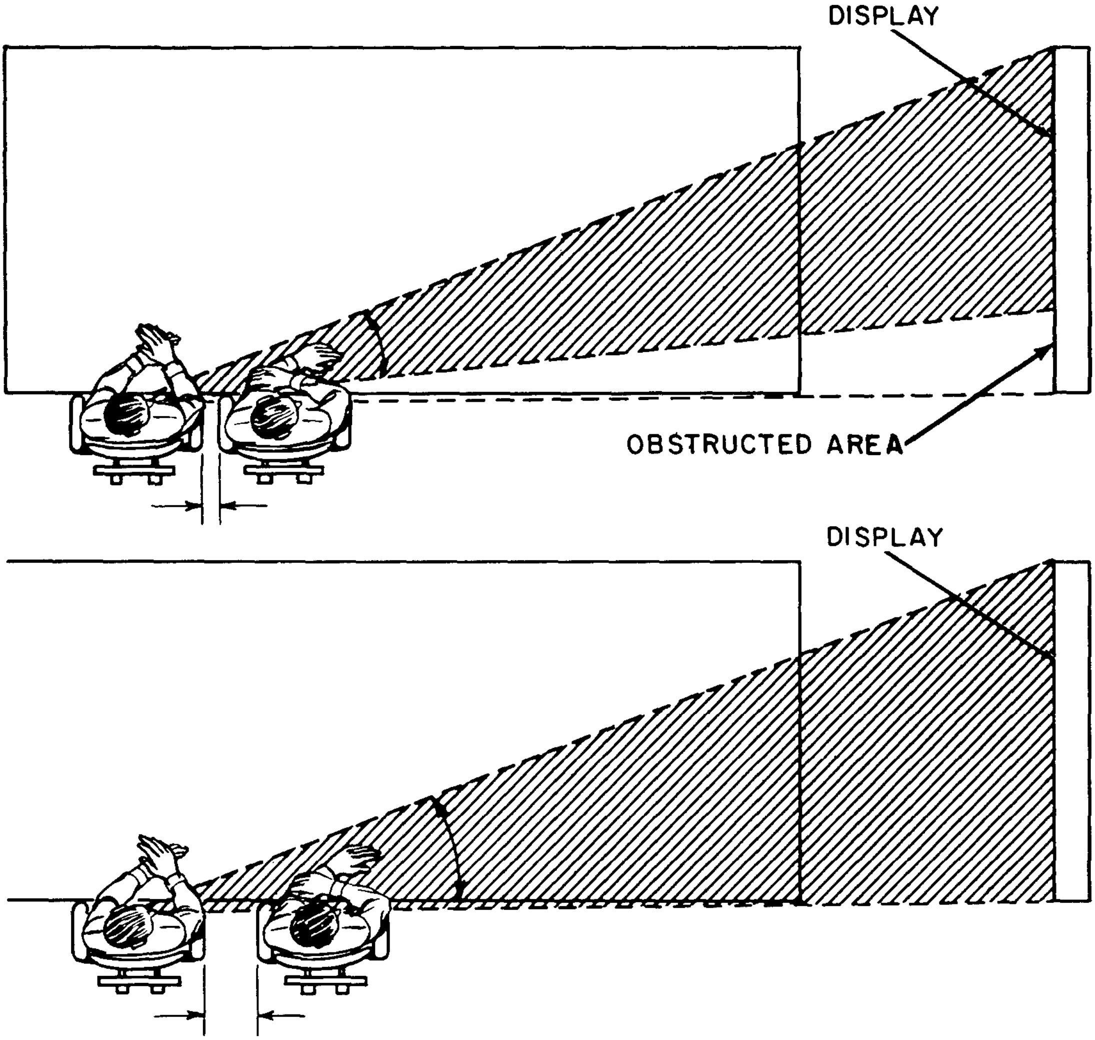


FIGURE 10–22. Increasing distance between viewers increases viewable area of displays.

1. The width of the display in relation to the width of the table.

2. The lateral distance between each person. (See Figure 10–22.)

Although persons nearest the display can shift backward to get out each others' way, this interferes with their view. In a prolonged conference, where a part of the program is to be given over to individual presentations, the conference table can be replaced by narrow table-workbenches (see Figure 10–23) to allow all participants a head-on natural view of the presentation.

Aisles and Exits

A minimum clearance of 4 ft. between the table edge and the nearest wall will provide a 2-ft. aisle along the wall when people are seated at the table. An additional clearance of 12 in. will provide more freedom of movement for seated people. (See Figure 10–24.) With a 6-ft. minimum clearance between table and "front" wall, this wall can be used for displays that can be seen by everyone in the room. This "front" wall should be one of the two narrower walls in a rectangular room to minimize extreme viewing angles for persons seated toward the front against the walls.

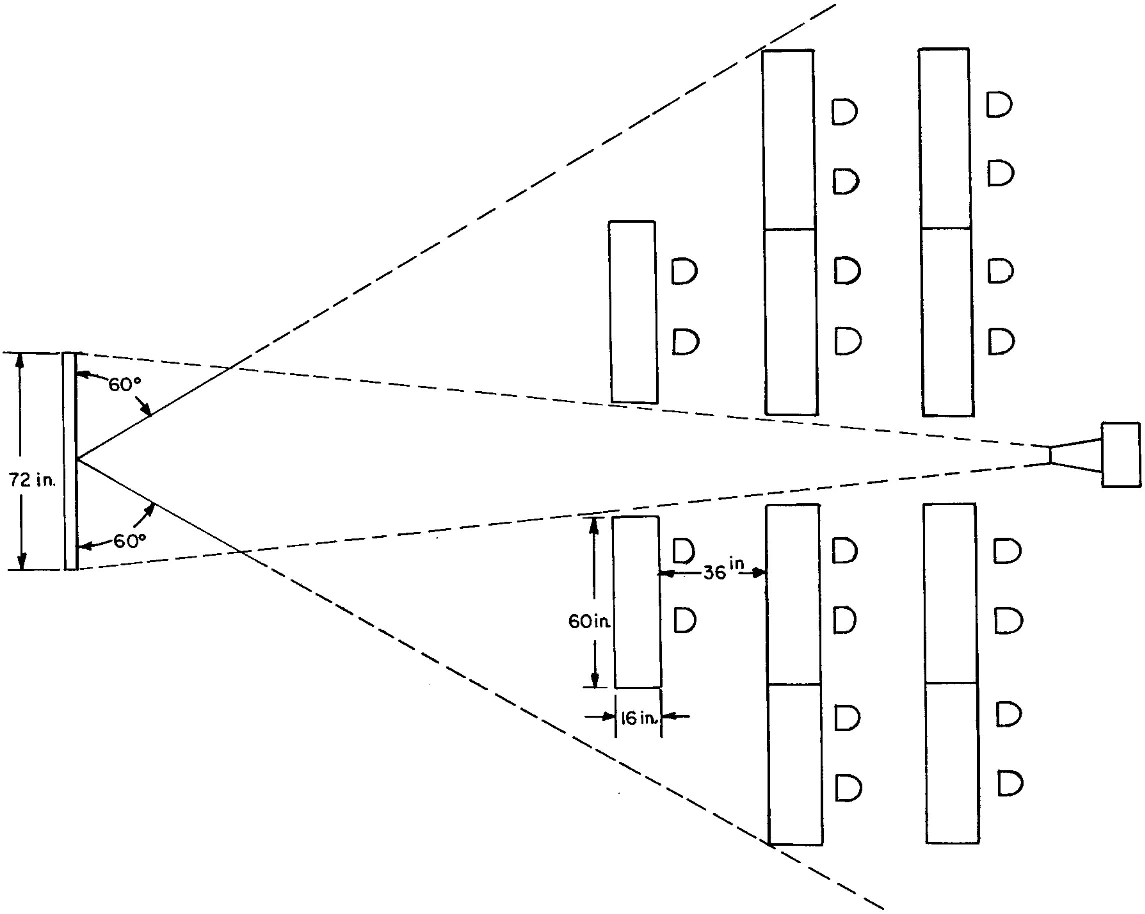


FIGURE 10–23. Use of narrow tables in conference room permits natural head-on view by all participants.

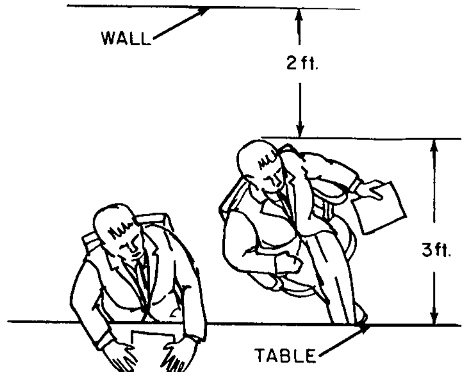


FIGURE 10–24. Clearance required around conference table.

Exits should be located in a rear corner of the conference room in order to provide maximum wall space for large displays, to minimize the clearances needed for exit space, and to lessen disturbance from persons entering or leaving the room while the conference is in progress.

Where a large conference room can be partitioned, exits should be located side-by-side, one on either side of the partition. If it is desirable to guard the entrance to these rooms, the layout suggested in Figure 10–25 provides the following additional features:

1. Receptionist to check security, assure privacy, take messages, etc.
2. Lobby area where necessary conversations can be held.
3. Telephone outside of meeting rooms.

Accessible storage space should be provided for overcoats, packages, luggage, etc., so that conferees need bring into the conference area only essentials.

10.6.2 Classrooms

Classrooms are best set up using narrow (16- to 20-in. deep) tables. In this way, more persons can participate in the listener role. (See Figure 10–26.)

Tables have proven to be highly versatile and efficient, lacking only one advantage of desks: they do not provide out-of-the-way storage for the books and notes of one group while another group uses the same room. For economy, it is possible to form the tables into into a conference "ring" so that bulky conference tables do not have to be set up and stored.

The principal difference between a conference room and a classroom is the use to which it is put. Except for arrangement of furniture, identical facilities can be used for either purpose.

10.6.3 Auditoriums

Viewing. To view a display, chart, or movie screen, the ideal shape of an auditorium is a fan or a truncated wedge. (See Figure 10–27.) Acceptable viewing angles are not less than 60° in the horizontal plane. (See Figure 10–28.)

If the vertical angle exceeds 10°, either up or down, viewers will tend to slump or lean forward in their seats. Ideally, the screen should be normal (90°) from the observers' line of sight. In practice, this ideal cannot be realized, and an effective compromise is shown in Figure 10–29. The most serious distortion results from combined horizontal and vertical (oblique) deviations from normal viewing angles.

For recreational movies, these limits can be widened because the viewer rapidly adapts to distortion. Where undistorted vision is important, these guidelines should be observed. Where the auditorium is to be used for both purposes (e.g., showing of training films and briefings during the day and used as a recreational theater in the evening) the more lenient restrictions should be used with the front several rows held vacant where good visual angles are desired. When viewing a "live" presentation (speaker, panel, etc.) these seating recommendations can be relaxed since the viewing angle restrictions are no longer pertinent. (See Schlanger, 1942–1944.)

Seat type and arrangement. Auditorium seats should have a narrow silhouette and be of the "spring-up" type. Where the seats are divided by a single armrest, a 24- to 26-in. seat width is recommended. (See Figure 10-30). A 28-in.-wide seat is recommended where two armrests are provided for each individual seat. The recommended separation (front-to-rear) between seats is 40 in. A fold-up writing table, similar to those provided on commercial airlines, should be provided for each seat, together with a rack or bin for temporary storage of papers. Since seating will be staggered, folding writing tables and paper receptacles should be mounted across the backs of the row of seats immediately forward.

Although the recommended separation between seats in adjacent rows is 40 in., increased seating capacity may be a necessity. In this case 32 in. is the absolute minimum. To save ceiling height, floor elevation can be specifically structured with a reverse incline for the first few rows. (See Figure 10–31.) The advantage of the reverse incline is that it minimizes floor slope or pitch. However, with fewer than 22 rows a reverse incline is impractical.

It is seldom necessary to arrange seating so that a person must see directly over the head

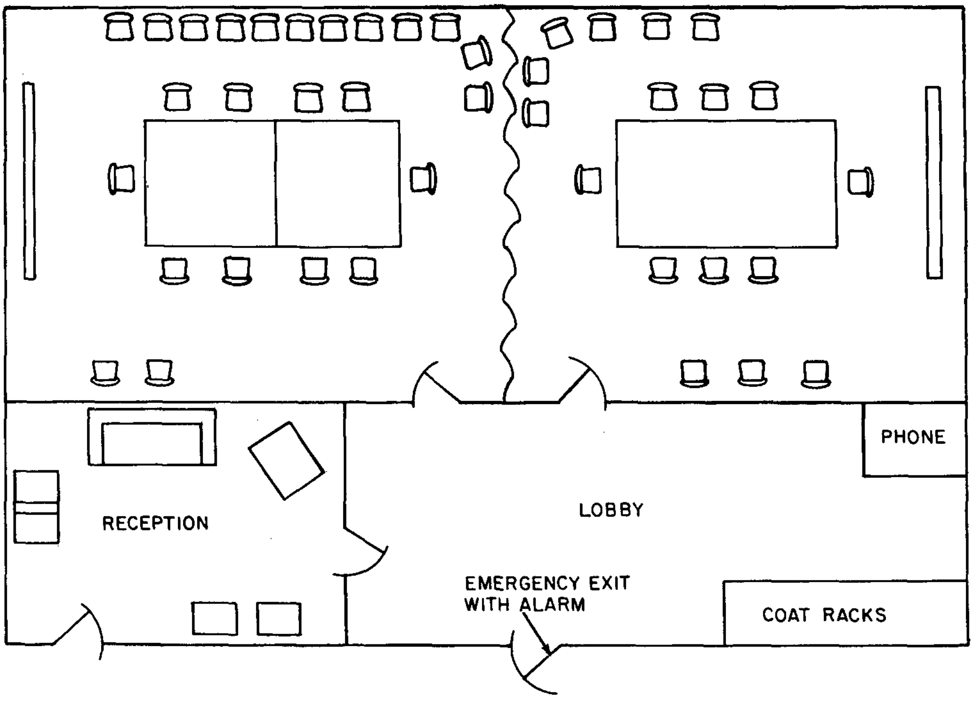


FIGURE 10-25. Use of movable partitions for secured conference facility.

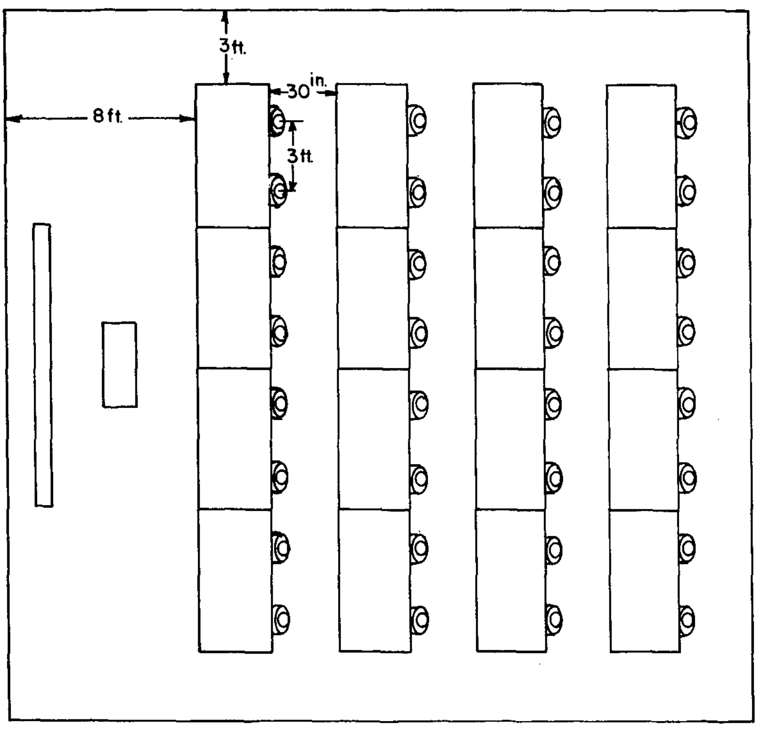


FIGURE 10–26. Classroom arrangement.

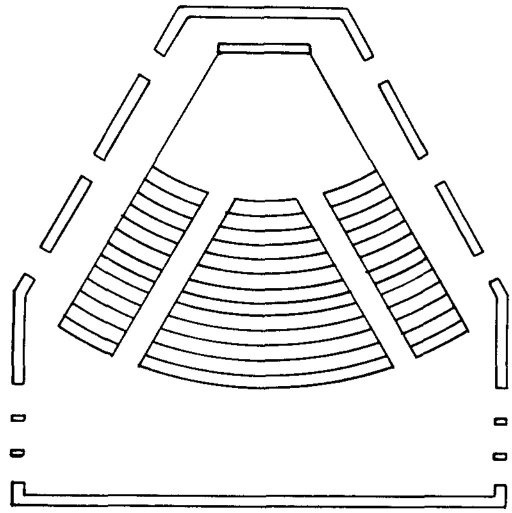


FIGURE 10–27. Auditorium seating arrangement.

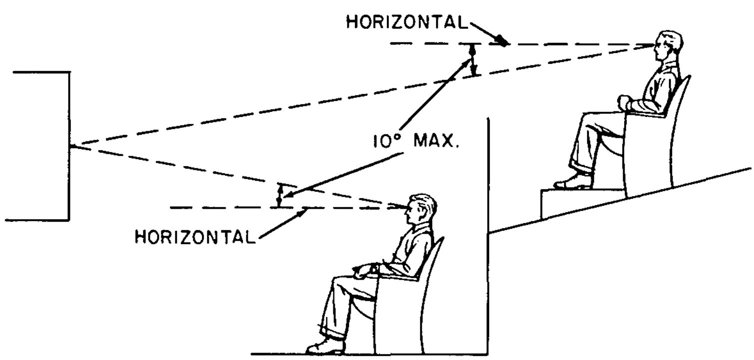


FIGURE 10–28. Vertical viewing angle.

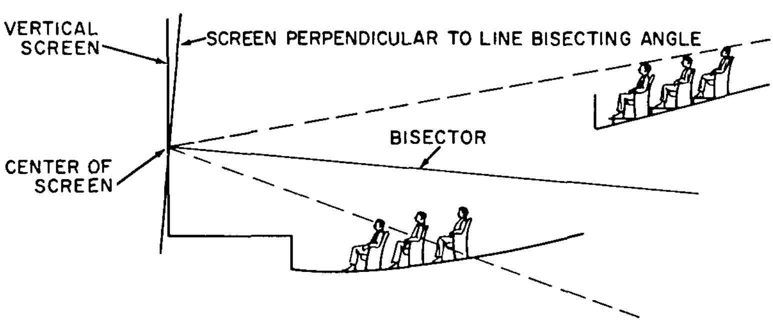


FIGURE 10–29. Screen position should be normal for the line of sight for all observers.

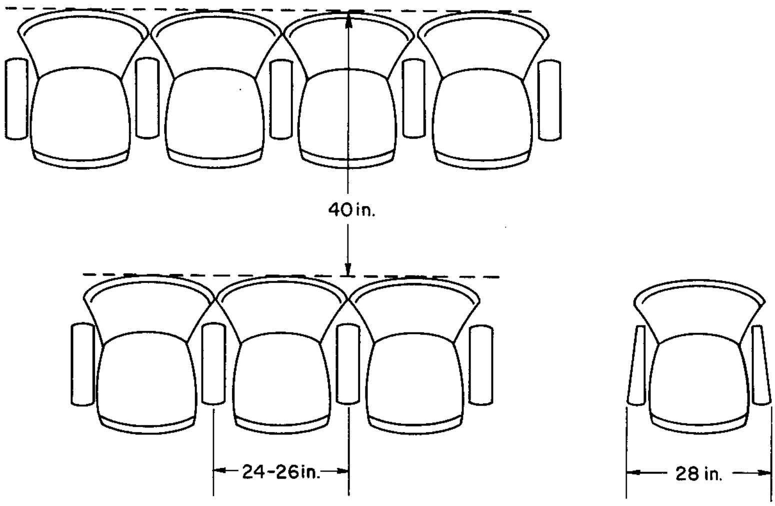


FIGURE 10–30. Row to row armrest dimensions.

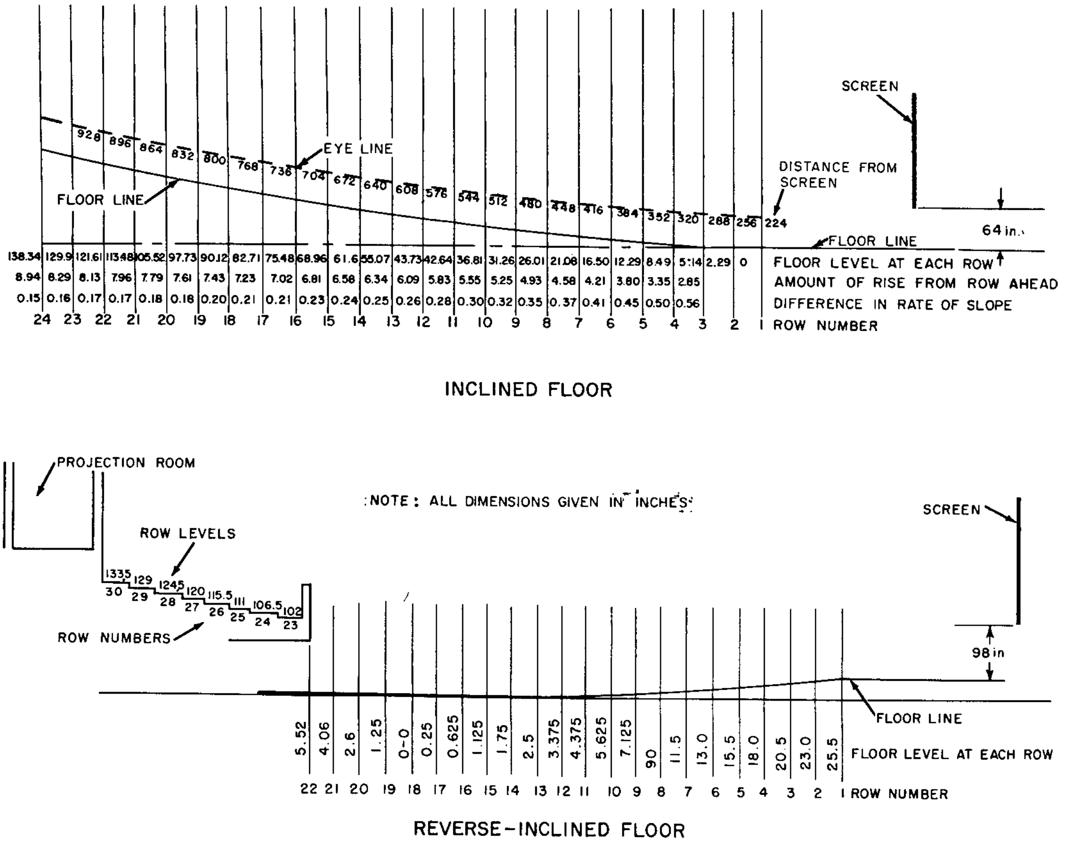


FIGURE 10–31. Dimensions for inclined and reverse-inclined floor auditorium.

of another person seated in front. Few installations can provide the required ceiling height for such a steep floor slope. In practice, seats usually can be staggered or offset so that a person can look over the shoulders and between the heads of those in front. The only restriction to such a staggered arrangement is a possible requirement that aisle seats be precisely aligned from front-to-rear.

In arranging rows of seats to insure proper visual clearance, each seat should be offset by a head's width (8 to 10 in.) so that each person can obtain a full view of the display. A staggered seating arrangement for maximum visual clearance should permit a short man (sitting *eye* height of 29.4 in. plus seat pan height) to see the bottom of the display over the *shoulder* height of a tall man in front of him (sitting *shoulder* height of 25 in. plus seat pan height). The horizontal angle of the field of view can be increased by either a greater lateral separation between seats or a greater separation between rows. For further details on optimizing the horizontal field of view, see Ramsey and Sleeper (1965).

On a level floor, the eye height of the seated short man is about 4 in. above the shoulder height of the seated tall man. Thus, when the bottom of the display is *above* eye level, e.g., on the main floor of an auditorium, the short man has no difficulty in seeing the bottom of the display. But when the bottom of the display or viewed object is *below* the level of the eye, as in a balcony, the angle between the eye of the viewer, the lower part of the display, and the floor should not be less than 4°. (See Figure 10–32.) While this does not permit the short man to see completely over the head of any person two rows in front, the obstruction at that distance is small (less than 6°) and covers only a portion of the bottom of the object being viewed. Any slope greater than 8° from the horizontal is unsatisfactorily steep for a main floor.

Aisles and exits. The following general recommendations will help provide maximum safety and speed under emergency conditions: (a) allow adequate clearance between rows, and use the narrow-profile, spring-up seats as recommended; (b) provide main aisles along either side of the auditorium; and (c) place exit doors along the aisles. (See Figure 10–33.) For large auditoriums, a center aisle also should be provided: (a) exit doors should push open from the inside; (b) aisles that run into narrow corridors so that the two form a "T" should be avoided' especially those where right-angle turns allow persons to get jammed into a corner. (See Figure 10–34.)

Balconies. The design of balconies is dependent on what is to be viewed. When a large central display is the subject, the desirable viewing distance from the screen is a function of image size and sound quality. To keep the vertical viewing angle close to 90°, the floor pitch should remain fairly small. (See Figure 10–35.) When an auditorium is to be used predominantly for "live" presentations, it is important to minimize the distance between the audience and stage. This can be accomplished by bringing the front of the balcony close to the stage and by increasing its pitch. The use of a reverse-pitch main floor permits the balcony to be lowered closer to the main floor.

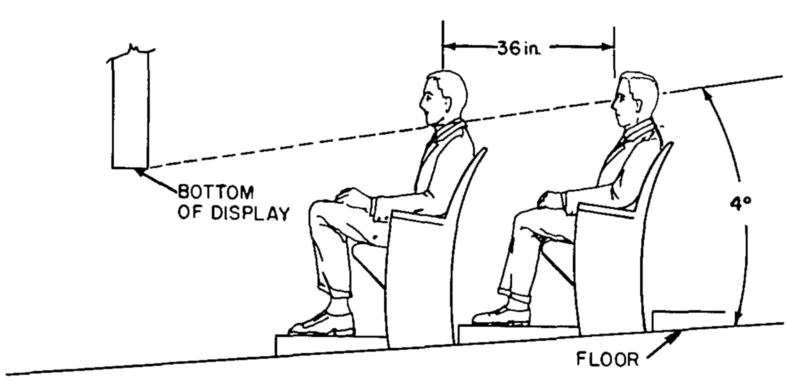


FIGURE 10–32. Display viewing on inclined floor.

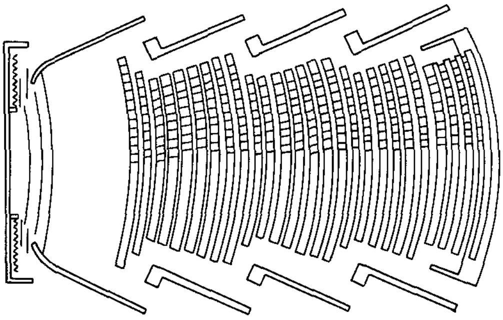


FIGURE 10-33. Auditorium arrangement for emergency exit (Thomson et al., 1958).

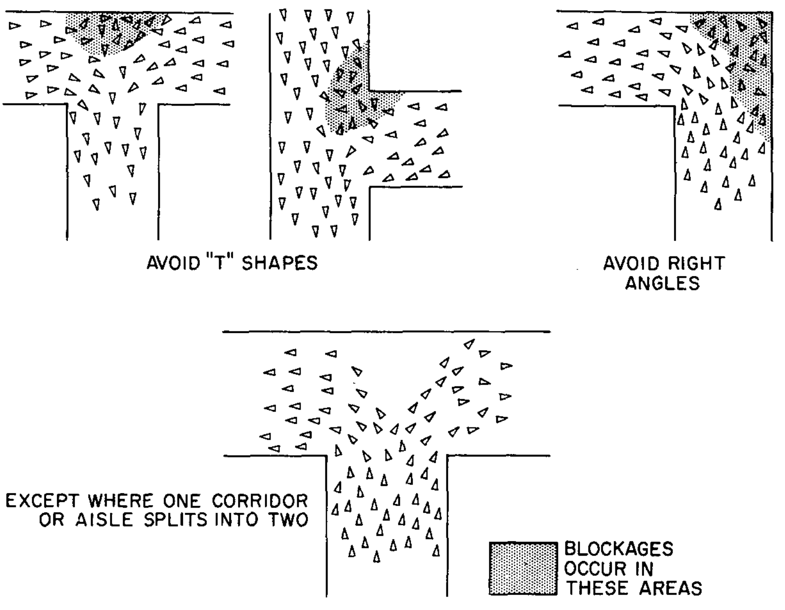


FIGURE 10-34. Aisle designs for mass emergency exit.

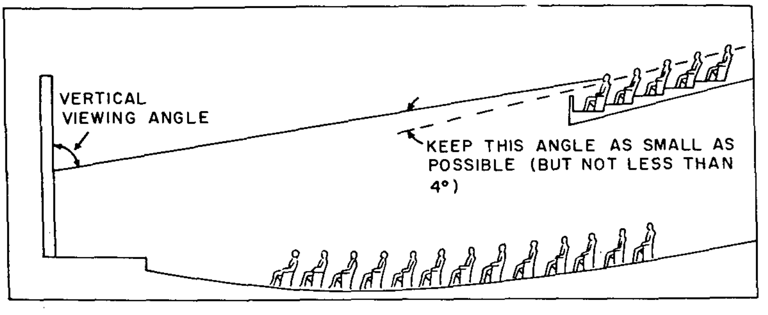


FIGURE 10-35. Reverse pitch floor minimizes balcony height above main floor (Thomson et al., 1958).

10.7 Layout of Traffic Spaces

In multi-man–machine work areas, it is necessary to provide access spaces, passageways, and stairs. These spaces generally are not a part of the working area except in special cases where:

1. Standing personnel can satisfactorily operate in the traffic space and move out of the way of transient personnel.
2. Personnel are fixed in one location throughout the duration of a mission and require access and egress only at the start and close of the mission.

Therefore, the arrangement of working areas must be given priority over work accesses. If a choice is required, adequate operating room at the expense of space for traffic can usually be justified.

An architect or industrial engineer designs space for traffic in terms of the type, amount, and weight of traffic with consideration for the frequency of use, routing, and speed. Discussion of these factors can be found in standard architectural and industrial engineering sources on plant layout. The present discussion is limited to the human-engineering aspects of access or traffic space design.

Aisles and corridors

An aisle is necessary where an appreciable amount of traffic flows within a compartment during normal operations. When there is moderate traffic, such as when personnel rotate at the end of a shift, traffic can make its way through whatever path is available (where there are no special requirements for safety). For civilian and military personnel wearing normal clothing, an unobstructed aisle of 24-in. width (never less than 20-in.) is sufficient to avoid brushing against equipment and inadvertently activating switches or knocking unattached gear to the floor. In any case, the arrangement of operators, displays, and equipment together with the establishment of clear lines of sight should be considered.

Figures 10–36 and 10–37 show recommended widths for aisles and corridors. An aisle can be

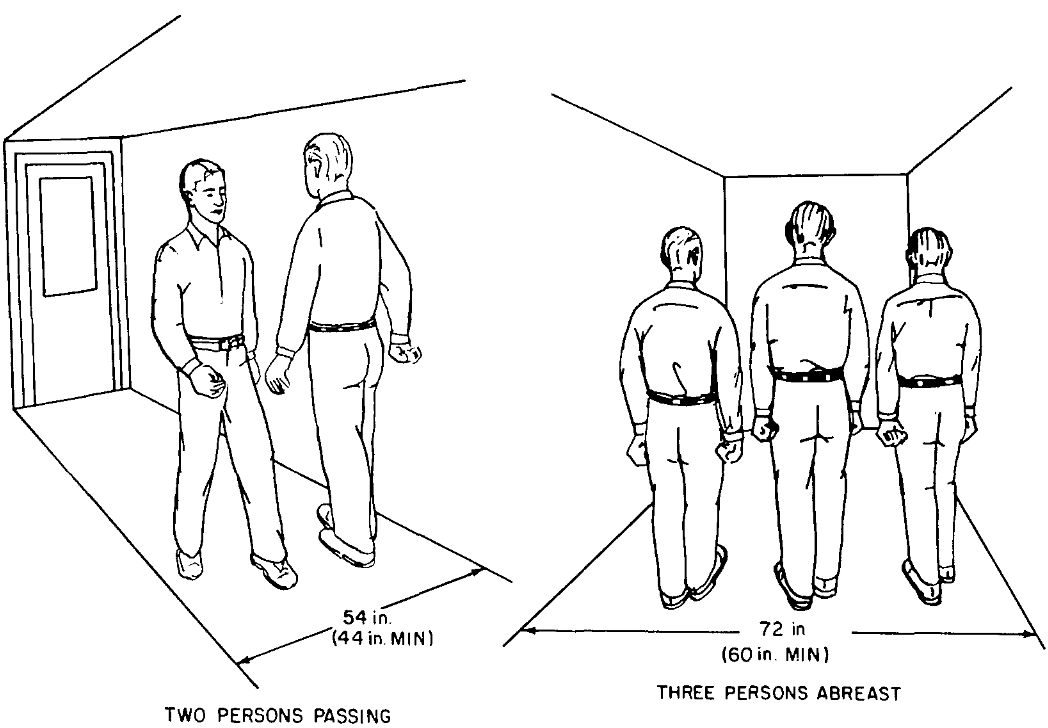


FIGURE 10–36. Aisle widths for two- and three-person flow.

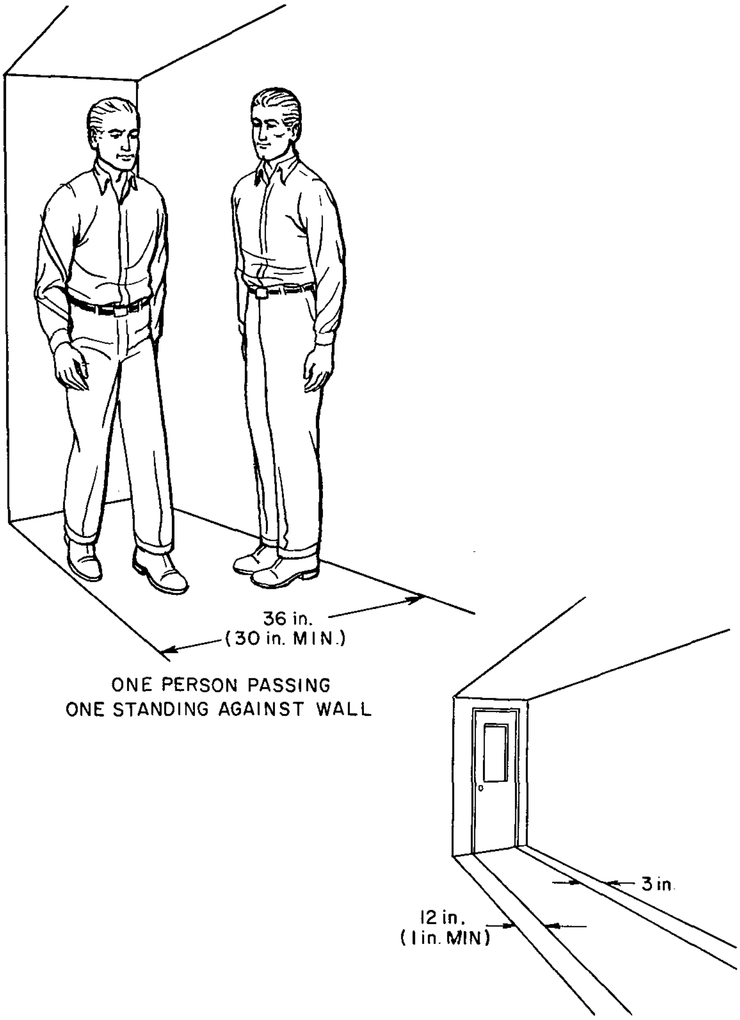


FIGURE 10–37. Aisle and door arrangements.

narrower than a corridor because it handles less traffic and it usually is convenient for one person to wait or stand aside while another passes. However, corridors must accommodate: (a) the continuous passage of large equipment with inflexible dimensions, and (b) free passage for all personnel. Figure 10–38 shows recommended locations of aisles with respect to exits.

The following additional recommendations should be observed in the design of aisles and corridors.

1. Avoid blind corners.

2. Locate paths for minimum distances, using flow charts or diagrams to indicate where the densest traffic will be.

3. Mark traffic guides (aisle limits, arrows, etc.) on floors, walls, or ceilings.

4. Make intersections converge at 90°; this arrangement will minimize lost floor space.

5. Doors should never open into corridors. Occasionally there is no alternative (such as with a small utility closet). In such cases, use sliding or folding doors.

6. Keep aisles clear; do not allow equipment or structural-support columns to protrude into any aisle.

7. Avoid locating an aisle against a blank wall because this will permit access from only one side.

8. Avoid one-way traffic in aisles. (It is practically unenforceable.) One-way traffic in corridors is feasible, but generally is not desirable.

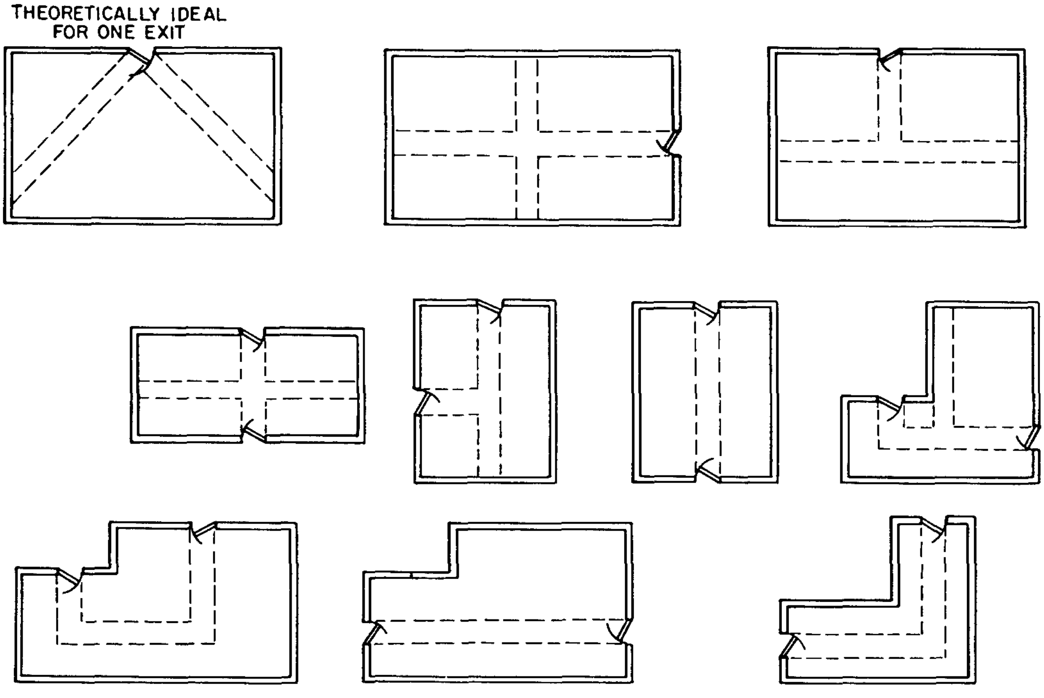


FIGURE 10–38. Recommended location of aisles in relation to exits.

Catwalks and Tunnels

Catwalks and tunnels are specialized facilities for handling traffic, and are recommended only where space or environmental conditions preclude normal corridors. The floor of a catwalk should have a non-skid surface. The stairway to the catwalk should approach the catwalk at right angles. Guard rails, to which wire mesh or other protective screening is attached to the lower half, should be provided.

Fully enclosed walkways or tunnels that allow the user to walk erect or nearly erect may be shaped in a way to approximate minimum space and weight requirements. The lateral measurements of such a passageway (see Figure 10–39) will have to be increased for personnel wearing or carrying bulky equipment. The minimum 12-in. floor width requires walking "cat-fashion" (steps aligned fore-and-aft). (See Woodson, 1965.) When used for emergency evacuation of personnel or where a load must be carried, a 16-in. width is preferable. If the enclosed walkway is in a moving vehicle, extra space should be provided for a handrail.

Sometimes the only feasible passageway is one in which the user must stoop, crawl, slide, or ride (on a pallet, trolley, or truck). Here the clearance from top to bottom of crawl floor and ceiling should be no less than 25 in. The dimensions for clearance in tunnels, (See Figure 10–40) are 25 in. for shoulder clearance allowing for clothing and freedom of movement, 16 in. for body clearance in the first case (upper figure), and 22 in. for body-plus-elbow room (lower figure). These latter dimensions are measured from the top of the personnel carrier, and are not the recommended diameters of the tunnels. The tunnel cross section may be elliptical or rectangular. All of these dimensions are based on the 95th-percentile man. Any allowance required for equipment or heavy clothing should be added to them. Refer to Chapter 11 for more detailed analysis of the appropriate increments to be added for certain kinds of bulky clothing.

10.7.1 Doorways and Hatches

The dimensions recommended below for doorways and hatches will accommodate over 99% of the population wearing light to medium clothing, e.g., shoes, trousers, jacket, and cap. These dimensions may be decreased if less than 99% of the population is to be accommodated

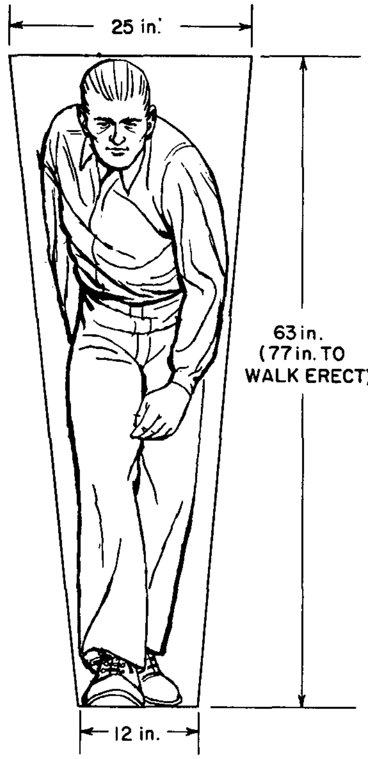
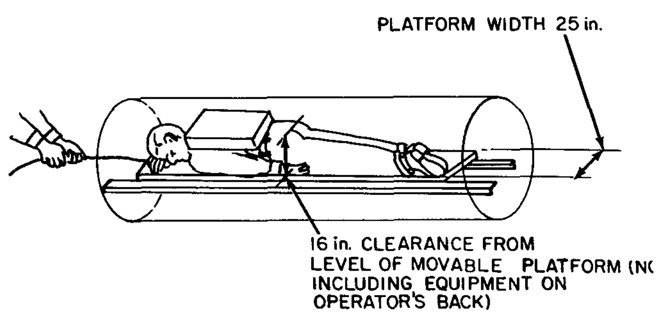


FIGURE 10-39. Minimum tunnel dimensions.



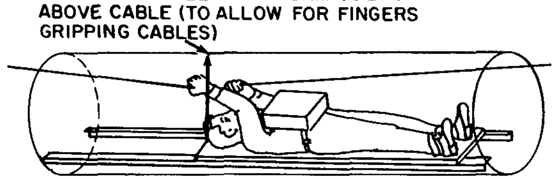


FIGURE 10-40. Minimum crawl tunnel dimensions.

(see Chapter 11 on Anthropometry), or if the users can stoop, bend, or otherwise adapt to inadequate openings. On the other hand, dimensions must be increased to allow for bulkier or specialized clothing and equipment. For most purposes, these dimensions should be considered minimum, with additional space to be supplied when possible.

Aisles passing through the center of a space are more efficient than those passing along its periphery. Doorways are best approached "head on" rather than from the left or right sides. Except for conference rooms and private offices, the location of doorways in a corner of the room is restrictive of traffic flow, tending to force it along walls or bulkheads.

Hinged Doors

The dimensions shown in Figure 10–41 are more than adequate for access to and from a compartment for one person at a time. Larger dimensions will be required for moving bulky equipment.

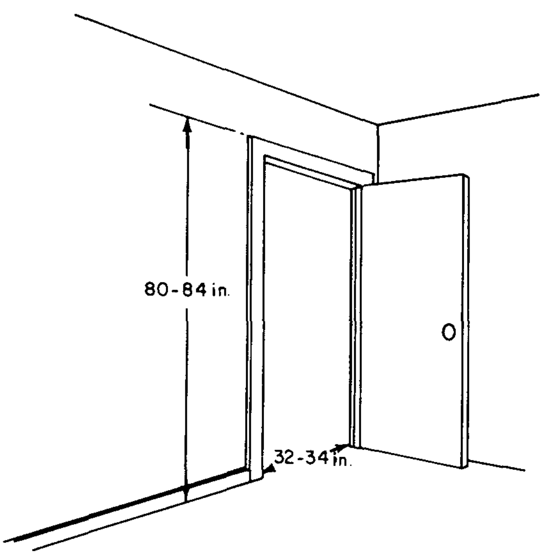


FIGURE 10–41. Converted door dimensions for one person at a time.

There is no general preference for doors opening either from the left or from the right. For small rooms, such as private offices, the door is located most conveniently in a corner of the room. For larger rooms which house several persons, doorways should be located in the near-center of the wall or partition.

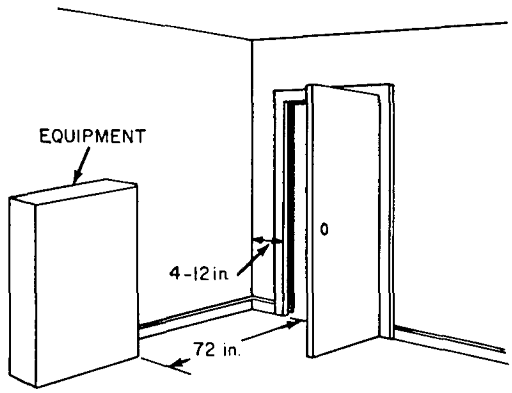


FIGURE 10–42. Relation of equipment to door movement.

Hinged doors should never open outward into a corridor. Hinged doors connecting two rooms should be removed where possible. If this is not feasible, then:

1. Entrances should be protected to keep occupants of the room from inadvertently being struck by the opening door. A row of file cabinets with backs facing the doorway is effective. A space of 6 in. should be provided between the moving door edge and any equipment to permit a person approaching the door to move out of its way (keep hands and limbs from being scraped) should it open suddenly.

2. A desk adjacent to the door should not face the aisle leading to the door, since persons standing in front of the desk are endangered by the opening door.

Doorsills are not recommended except where specifically required for engineering reasons. Ventilation control can usually be maintained by a flexible strip attached to the door bottom. Clearance of at least 4 in., and preferably 12 in., between the door and wall is recommended where a door opens next to a wall. When a door is located in a corner, equipment should be located no closer than 5 ft. from it, as shown in Figure 10–42.

Sliding and Folding Doors

Horizontal or vertical (overhead) sliding doors are useful for cramped spaces, but they are easily jammed if subjected to blast, collision, etc. They should not be the only exit from a work area. Both folding doors and sliding doors (unless automatically activated) are slow and cumbersome to operate and are generally unsatisfactory for other than light traffic (or heavy but sporadic traffic, such as changeover of personnel shifts). Where large vehicles or pieces of equipment must regularly move into and out of compartments, a sliding door is recommended. For very large sliding doors, (hangars, heavy equipment maintenance buildings), a separate hinged door may be built into the sliding door for personnel use.

Swinging Doors

Swinging doors (spring-return to closed position) are hazardous and should not be used except where the operator is unable to manually open and close the door. This does not refer to doors fitted with a standard door closer-and-check, and opening in one direction only. Where their use is unavoidable, swinging doors should open in one direction only and be used in pairs, e.g., an "In" and an "Out" door. From the standpoint of pedestrian safety, they should: (a) be separated by a door post, (b) have openings or windows for visibility of oncoming traffic, and (c) be mounted "hinges in" (adjacent) to avoid interference. When vehicles or equipment wider than a single door must pass through it, the door or doors, must be mounted "hinges out" with no intermediate post.

Pressure sensitive pads may be used to open swinging exits and entrances. The exit and entrance must be conspicuously labeled, and the areas leading to each should be separated by a railing. The pressure sensitive pad must actuate the door opening mechanism sufficiently in advance of a rapidly moving pedestrian so that he will not collide with the door. The doors themselves should be of transparent glass so the user can see persons on the other side who might be stuck.

Revolving doors.

Revolving or rotary doors are not recommended for use.

Glass doors and floor-to-ceiling windows

When transparent glass is the principal material employed in a door or a large floor-to-ceiling window, the glass area should be labeled, marked, partly milked, or otherwise broken into a pattern so that a person will not mistake the glass area for an unobstructed opening. Many injuries have occurred when persons have walked into a glass door or window that was not properly marked.

10.7.2 Open Doorways and Archways

Open doorways and archways (the terms are used here synonymously) refer to any unobstructed opening in the wall (bulkhead) between adjoining compartments or between compartment and corridor. The dimensions given in Figure 10–43 apply to any such opening, and will permit the passage of two persons in opposite directions simultaneously. The 54-in. width is recommended if space and structural strength permit.

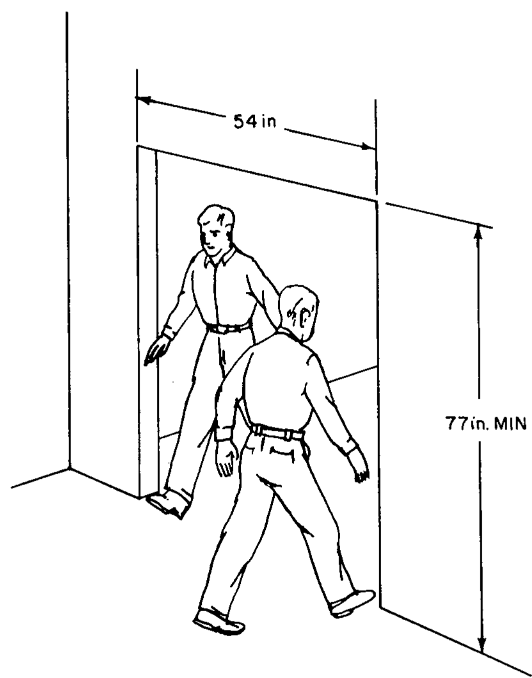


FIGURE 10–43. Archway dimension for two-person flow.

10.7.3 Escape Hatches and Emergency Exits

Standard emergency exits for public buildings, including location, construction, markings, lighting, etc., are adequately handled in local public building codes and in architectural design guides, and must so conform with these building codes. The design of public emergency exits is therefore not discussed further here. Escape hatches and watertight doors and armored hatches (see next section) refer to special design requirements for use in military vehicles such as aircraft, tanks, submarines, etc.

The proper dimensions for escape hatches will depend on structural requirements. Obviously, the larger the opening the better from the standpoint of the "escape," and the worse from the standpoint of structural integrity. Other prime design considerations include the escapee's equipment and clothing, and the environment he will enter. Alternative uses for the hatch must also be considered, e.g., for on- and off-loading of equipment and supplies including the overall dimensions of these objects. Figure 10–44 shows escape-hatch sizes for low-speed aircraft which do not require special protection for the ejected personnel (White et al., 1952). For the design of ejection capsules, see Chaffee (1960). The top- and side-escape hatch dimensions are appropriate for other vehicles. These dimensions include provisions for a parachute, but not for bulky clothing. It should be possible to open the hatch with a single motion of the hand or foot. When a handle is used to open the latch, the required activation force should not exceed 30 lbs.

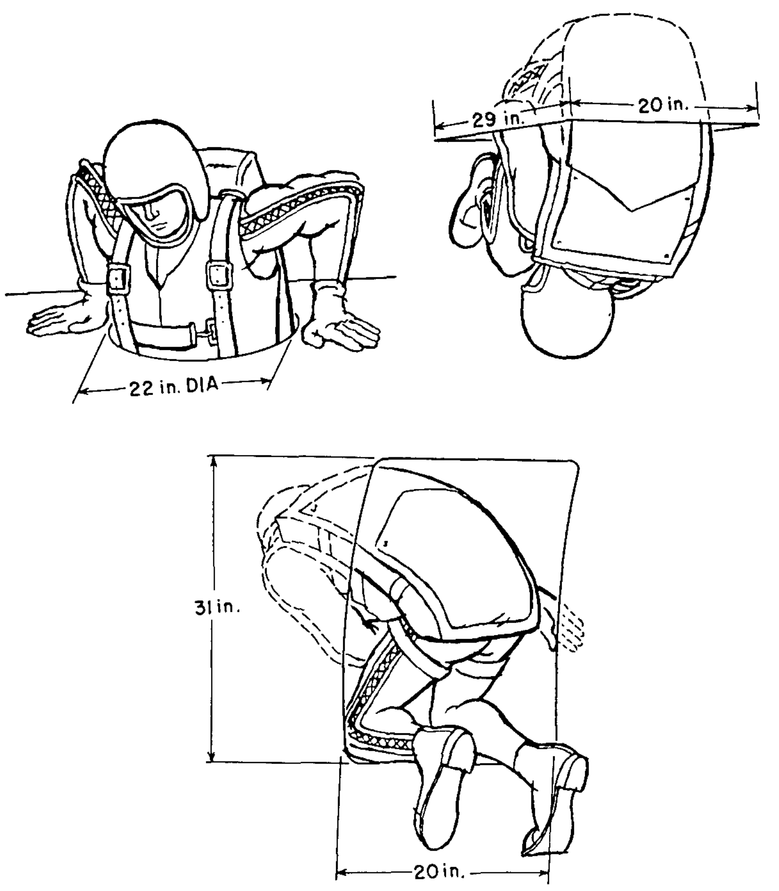


FIGURE 10–44. Escape hatch dimensions.

10.7.4 Watertight Doors and Armored Hatches

Heavily constructed hatches must be as small as possible to reduce weight and preserve the structural integrity of the bulkhead or deck in which they are to be mounted. Figure 10–45 shows recommended and minimum dimensions for bulkhead-mounted hatches. The 76-in. height permits the helmet-wearing 95th-percentile man to remain erect. If men must pass through a bulkhead-mounted hatch carrying heavy loads, the risk of muscular strain is less for stepping over a high coaming than for stooping excessively. For such situations, a 68-in. minimum is recommended for the top of the hatch, with the height of the coaming being 10 in. (and not over 14 in.). Bulkhead-mounted hatches should be designed for the range of the population that will use the hatch and not just the average man. In any event, the coaming should not be higher than 20 in. (at least 10 in. below the crotch height of the 5th-percentile man).

Horizontal, deck-mounted opening armored hatches (battle tanks, armored decks which raise and lower by hand) have particular maximum weight restrictions which depend on whether the hatch must be raised with:

1. One arm (about 40 lbs. force can be exerted);
2. Rigid arm with lift provided by torso (60 lbs. force); or
3. Two hands (80 lbs.).

If more than one person can simultaneously apply force, these forces can be additive provided the positions of the lifting personnel are not out-of-balance and strain-producing. In these cases, the actual action required should *always* be tested in a full-scale mockup which *fully duplicates* the intended production arrangement in dimensions and weight. This will avoid difficult and harmful situations in which the operator is required to exert a force from a strained position. Hundreds of thousands of military man-hours are lost annually because ruptures, hernias, or torn and strained ligaments or muscles have resulted from poor workplace and work-area design. Maximum forces to be applied are discussed in Chapter 11. The guidelines given should never be substituted for a live test in a faithfully reproduced design mock-up, however.

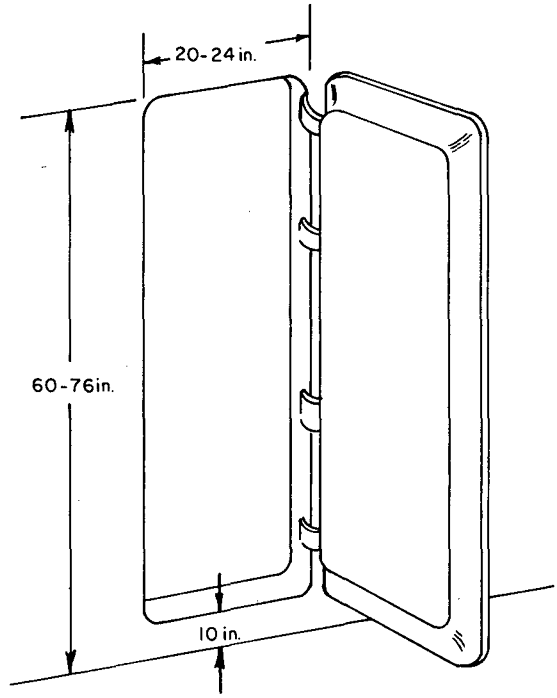


FIGURE 10–45. Armored door dimensions (Thomson et al., 1958).

Figure 10–46 shows mimimum and recommended dimensions for deck-mounted (horizontal) hatches. The actual depth of the hatch depends on the angle X of the ladder leading up to the hatch. The greater the angle X, the greater must be the depth of the hatch to provide head clearance. A usable rule of thumb is: hatch depth equals 76 (tangent X) in. The minimum depth is 24 in. Greater clearance must be added to this minimum when personnel will wear encumbering clothing, equipment, or harnesses. (See Figure 10–47.) For angled ladders, the minimum recommended vertical distances between the lower front edge of the hatch and ladder tread immediately below this point are shown in Figure 10–48.

10.7.5 Ladders, Stairs, Ramps, and Poles

Ladders

Ladders should be used where the desired rise from the horizontal is at an angle of 50° or more, or where a stairway is not practicable.

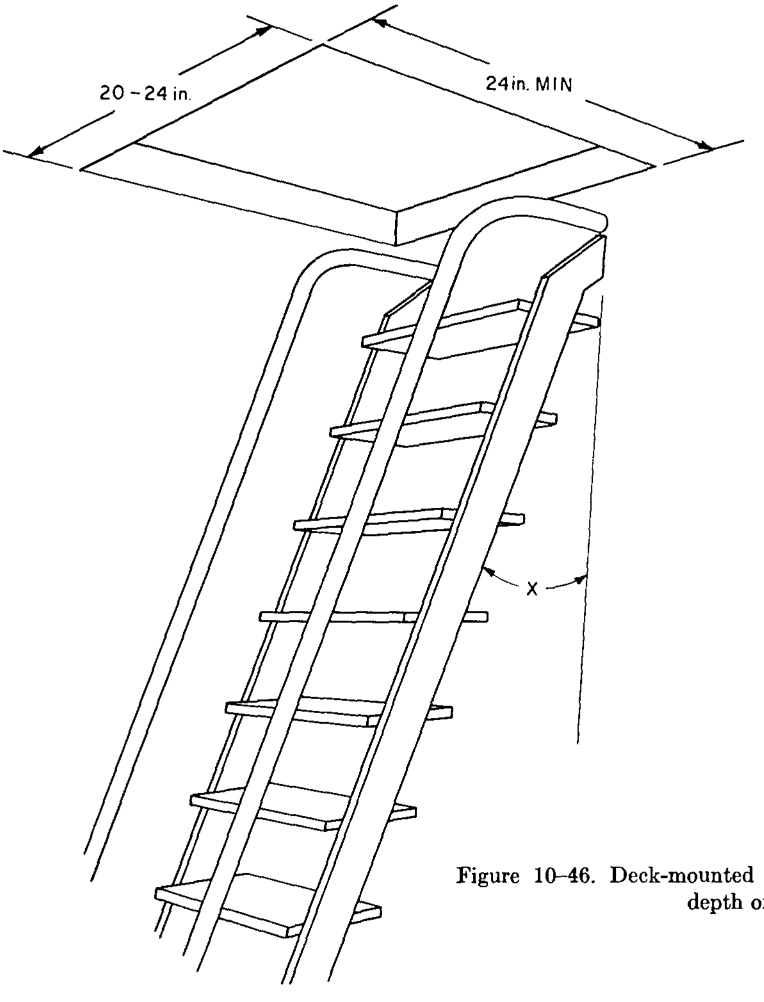


Figure 10–46. Deck-mounted hatch dimensions; as angle X increases, depth of hatch must increase.

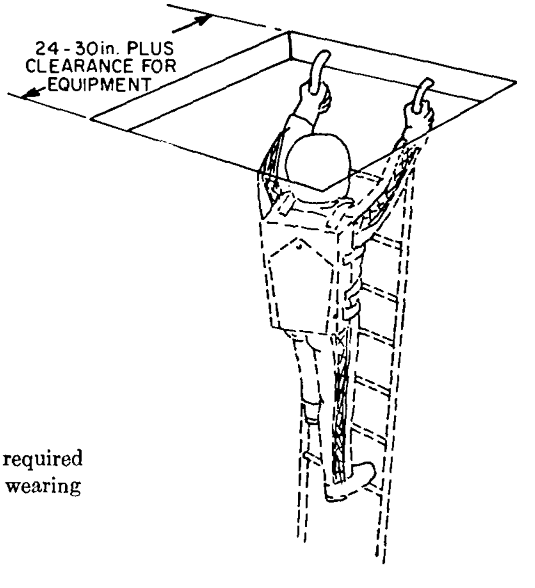


Figure 10–47. Added clearance in width required for vertical hatch use by persons wearing equipment.

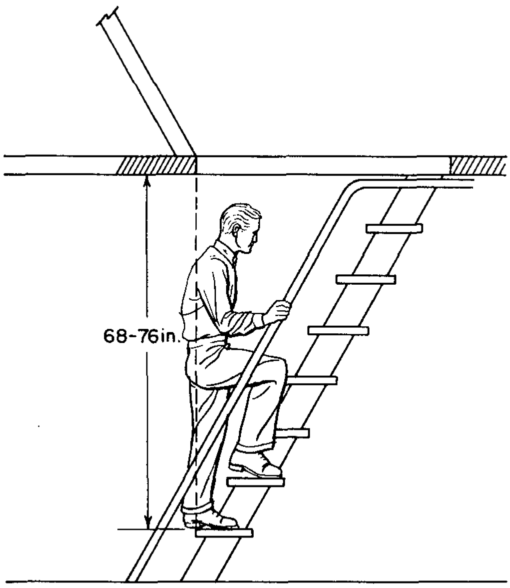


FIGURE 10–48. Minimum hatch edge to ladder tread distance.

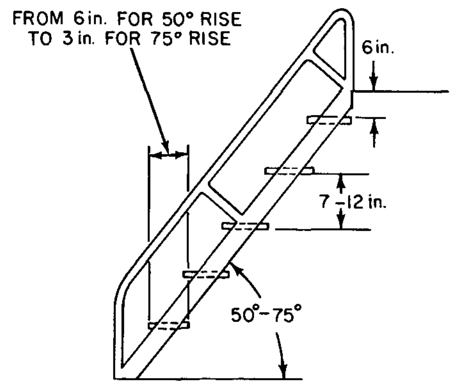


FIGURE 10–49. Ship's ladder dimensions.

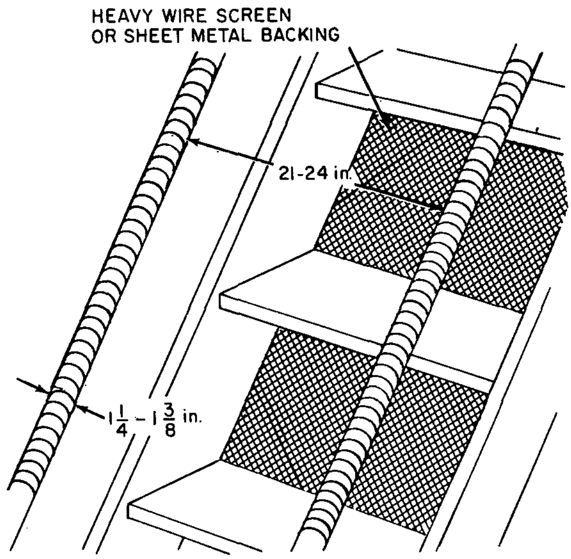


FIGURE 10–50. Handrail arrangements.

The round rung on the vertical ladder is necessary to provide a handgrip. Non-vertical ladders should have flat horizontal treads (as opposed to round rungs) and handrails. The most familiar example of this type is the ship's ladder, which usually rises at an angle of 68° from the horizontal (60° is a preferable angle), with a clearance for only one person. If simultaneous two-way traffic is desired, separate up-and-down ladders are provided, with a maximum tilt angle of 60°, preferably with a double handrail in the center.

Figure 10–49 shows recommended dimensions for this type of ladder. The optimum height between treads is 8.5 to 9 in. Treads should be open (without risers) and provided with non-skid surfacing. Depth of tread depends upon the angle of the ladder. As a rule, the rear of each tread should overlap the front of the tread immediately above, varying from 1 in. for a 70° ladder to 3 in. for a 50° ladder. Although portions of the shoe may extend beyond this point, this design will be in contact with the weight-bearing portion of the shoe sole. Metal screening should be fastened to the underside of the ladder to prevent the foot from slipping through. When two or more flights of such ladders are located one above the other, solid metal sheeting instead of screening will protect those on the lower ladder from falling dirt particles, etc. Handrails with a diameter of 1½ to 1⅜ in. and a spacing of 21 to 24 in. (see Figure 10–50) on both sides of the ladder should be covered with a nonslip surfacing.

For vertical ladders, round rungs are used to provide both hand grips and foot supports (for inclines between 75° and 90°). Figure 10–51 shows the recommended dimensions of such a ladder.

The optimum height between treads is from 11 to 12 in. If ladders are used to provide more or less permanent access to several levels, they should be offset at each level and protected by guardrails around the opening at the top of each ladder. (See Figure 10–52.)

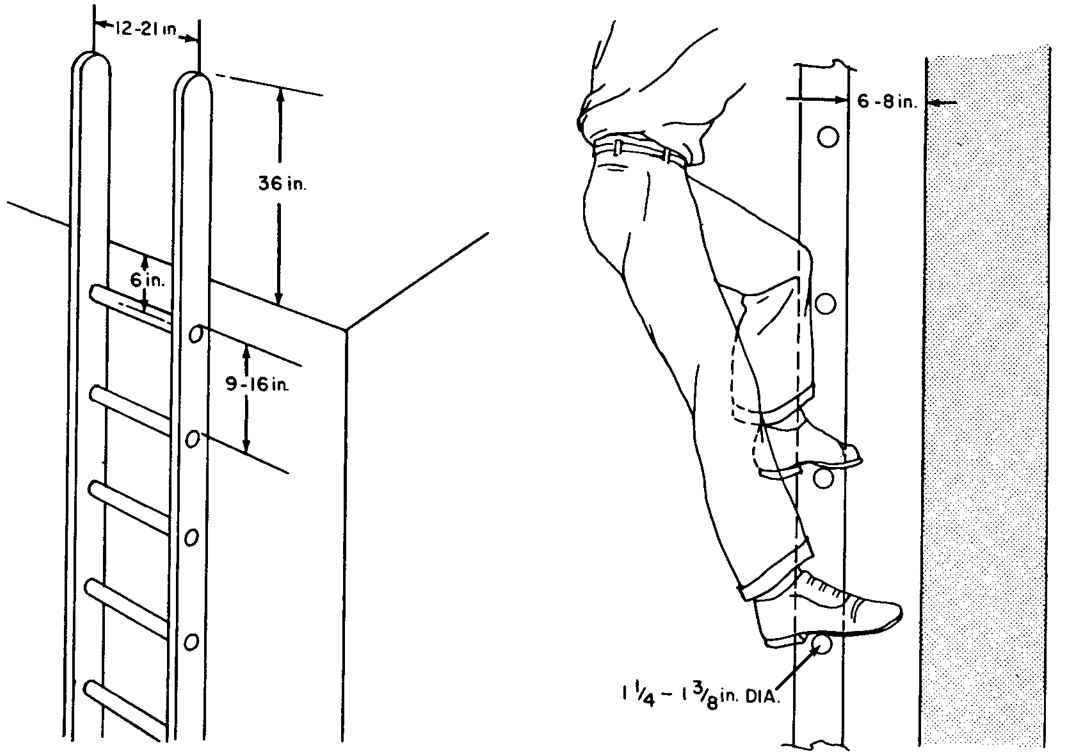


FIGURE 10-51. Vertical ladder design.

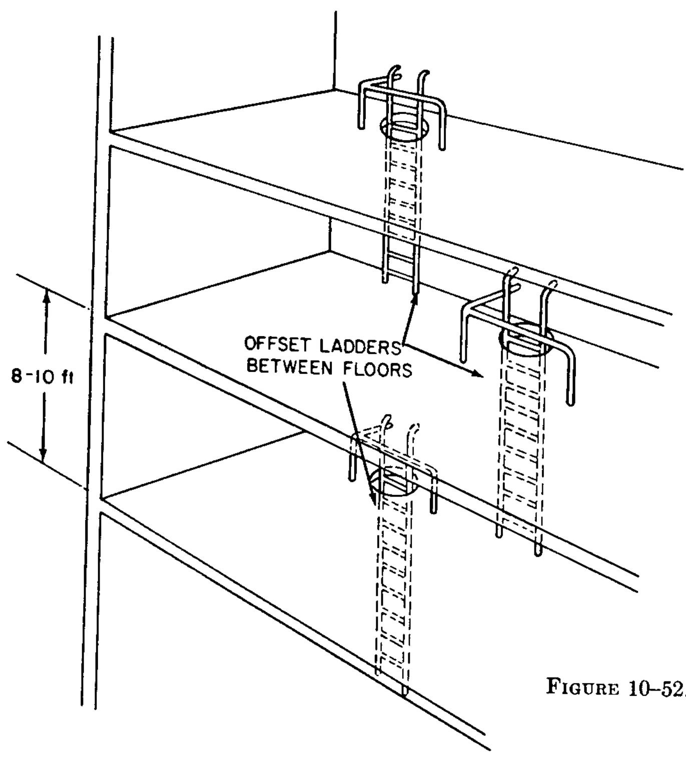


FIGURE 10-52. Offset of ladder between floors is recommended.

Stairs

Stairs should rise from the horizontal at an angle of between 20° and 35°. (See Figure 10–53.) This rise angle automatically determines the ratio of riser height to tread depth, but the minimum riser height should be 5 in. and the maximum 8 in. The optimum tread depth is 9.5 to 10.5 in. plus a 1 to 1.5 in. overhang. (See Figure 10–54.) These dimensions provide depth such that, in descending the stairs, the ball of the foot does not extend beyond the front edge of the tread, and the heel comfortably clears the overhang of the step above.

Long continuous flights of stairs should be avoided. Where space permits, landings should be provided every 10 to 12 treads. In addition, stairs enclosed by walls should have a handrail on at least one side. Recommended height of handrails is shown in Figure 10-55. The width of stairs (between handrails or between wall and handrail) should be as shown in Figure 10–56.

For open stairways and landings, a guardrail should be provided halfway between the handrails and treads. In addition, screen guards should be provided between the guardrail and floor for landings where the stairway is at right angles to the landing. (See Figure 10–57.)

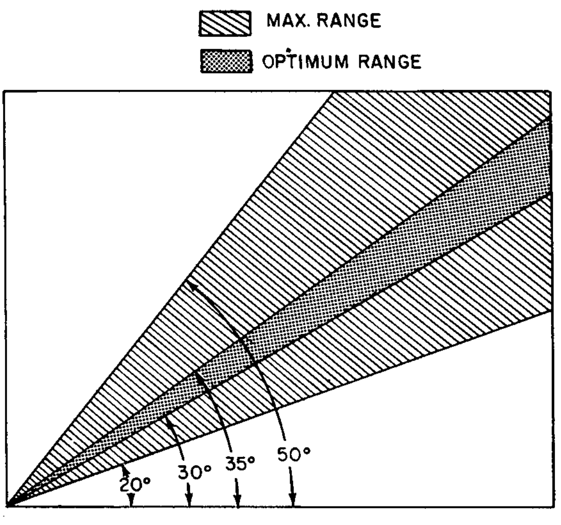


FIGURE 10–53. Stair rise angles (Thomson et al., 1958).

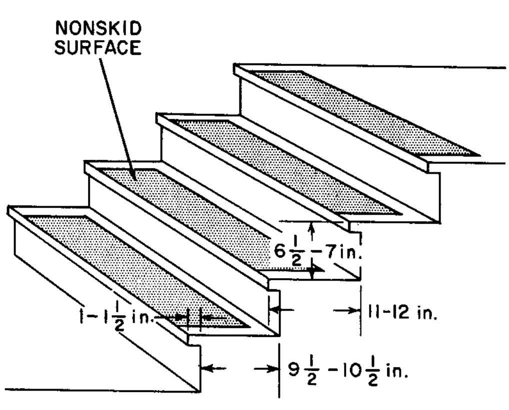


FIGURE 10–54. Stair tread dimensions.

Ramps

Ramps or inclines should be used for grades under 20° where rolling stock must be moved between different levels. For pedestrian traffic only, a stairway is more efficient from the standpoint of space, safety, and speed.

Ramps with a small incline do possess one advantage for pedestrian traffic: they allow elderly persons, or persons in poor physical condition, to expend their energy slowly and to avoid the abrupt raising of the knee required in climbing stairs; the user may shuffle at whatever step length he chooses. In designing for a military population, however, the requirement to provide this type of facility should be clearly justified. When a ramp is to be used for pedestrian traffic, cleats should be provided for slopes of over 15°. Maximum ramp slope may not exceed 20°. For outdoor ramps with slopes in excess of 15°, a non-skid surfacing must be used, and where liquids are likely to be spilled, a similar surface should be applied to indoor ramps as well. Indoor ramps (e.g., the ramps used in the Pentagon building) with slopes of 10° of less may be surfaced with standard materials. Distance between cleats should be 20 in. for slopes of 15°, decreasing as the slope increases to a separation of 14 in. for a maximum slope of 20°. When the ramp is for pedestrian use only, cleats should extend from handrail to handrail at right angles to the slope. (See Figure 10–58.)

Where a smooth (but nonskid) surface or runway for small wheeled vehicles is needed in conjunction with a passage for pedestrians, it should be located in the center of the ramp with the cleated portions on the outside next to the handrail. (See Figure 10–59.)

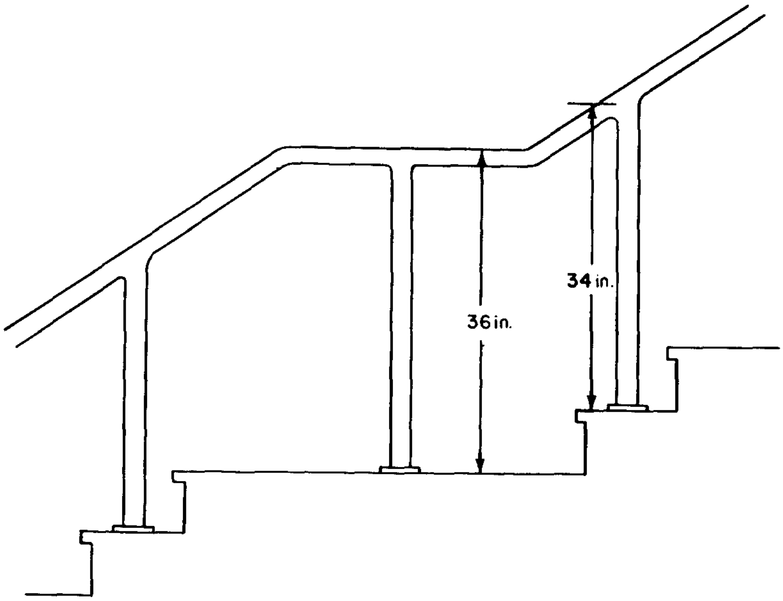


FIGURE 10-55. Recommended handrail heights.

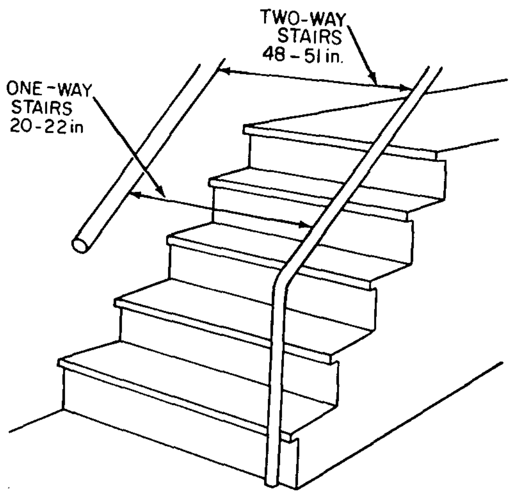


FIGURE 10-56. Recommended stair widths between handrails.

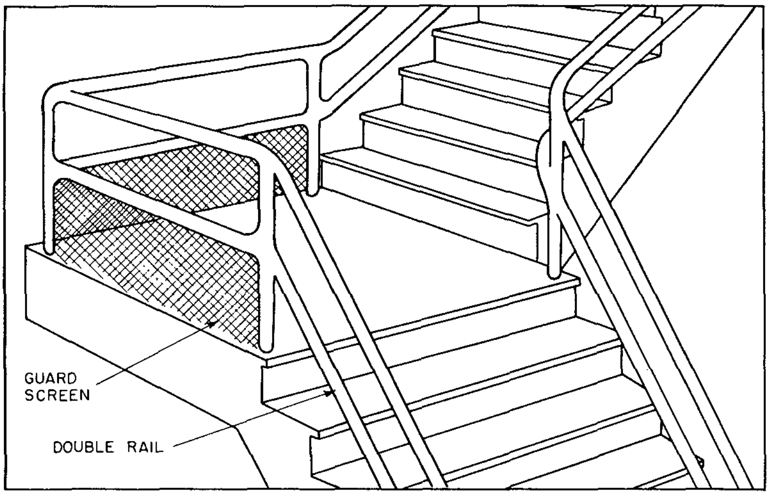


FIGURE 10–57. Use screen guards and guard rail when stair flights are at right angles.

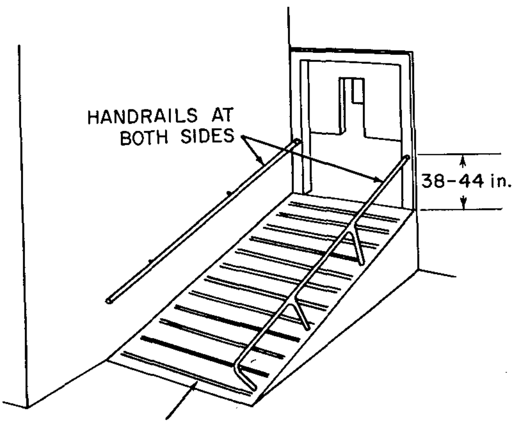


FIGURE 10–58. Ramp design for pedestrian use.

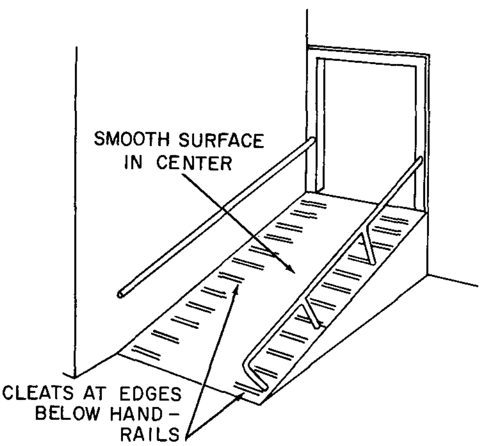


FIGURE 10–59. Ramp design for pedestrian and vehicle use.

Circular Stairs

If they must be used, the space between the handrail and steps of a circular stair or ladder should be enclosed with a metal screen. However, circular stairs are inherently hazardous and are not recommended because:

1. The ratio of tread width to riser height varies continuously across a step.
2. If persons approach in opposite directions, the inside person is forced to step on a very narrow tread. This situation is particularly dangerous if he is descending.

Poles

Vertical poles are used as a means of rapid access from one floor downward to the next. They permit a person in a "ready alert" situation to be transported quickly to another location in a ready-to-operate condition. Since the small possible increment of time saved in comparison to use of stairways is outweighed by the prevalence of accidents, the use of poles is not normally recommended. A survey of fire stations indicates that accidents are frequent. Moreover, fire station architects are discontinuing use of multi-story construction, with the "ready room" (lounge, cots, writing tables) now located adjacent to and on the same level as equipment. This latter practice is recommended here for military and other forms of construction.

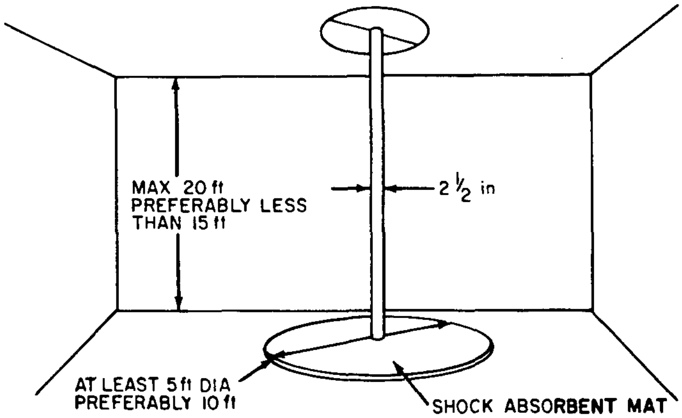


FIGURE 10–60. Pole design recommendations.

If poles must be used, the following features are recommended (see Figure 10–60): pole diameter 2.5 in.; vertical drop limited to 20 ft. and preferably not over 15 ft.; bottom of pole surrounded by shock-absorbent pad not less than 5 ft. in diameter and preferably 10 ft. in diameter to protect the person who falls after landing.

The entry hole in the floor above should be between 3.5 to 4 ft. in diameter, circular, with the pole located in the center. The entry hole should be covered by a divided cover which is springloaded to open downward automatically and remain open upon release of the holding latch.

10.7.6 Escalators, Elevators, and Lifts

Escalators and elevators have the disadvantage of becoming inoperative because of mechanical and power failure, although escalators can still be used as stairways when they become inoperative. Elevators and escalators have the advantage, however, in that they can transport rapidly large numbers of personnel between multiple levels, particularly where the distance to be transported is several stories, and the personnel are heavily loaded.

Escalators

The recommended rate of travel for an escalator is from 120 to 138 ft/min, although rates from 90 to 180 ft/min can be found. When entering an escalator running over 120 ft/min, users will pause to judge their footing at each entrance point. These pauses slow traffic to an extent which offsets the travel time gained by further advancing the escalator rate of travel and increases the hazard. Moreover, when a person is loaded with heavy equipment, a high travel rate may upset his balance when entering the escalator.

The preferred angle of ascent for escalators is 30°, with angles up to 45° satisfactory. In addition to providing support, the moving handrail paces the user's entering speed with that of the escalator steps. The handrail must, of course, move at the same speed as the steps. The customary smooth solid black handrail should have conspicuous white (high-contrast) markings (bars, diamonds) at least every 18 in. to make its movement apparent. This will permit entering users to gauge its speed and not mistake it for a stationary railing. (See Figure 10–61.)

At the exit point, the escalator guard walls should also extend from 5–6 ft. to permit the person to become accustomed to walking on a level and stationary surface before encountering pedestrian traffic. Thirty inches is the minimum width of escalator stairs, while 36 in. is a preferable width. To permit maximum traffic, a 52-in. tread width will allow persons to ride side-by-side.

Entries and exits should be illuminated at a level at least 20% greater than that of the surround. Illumination of the floor at these points should be maintained at this same level by concealed lighting fixtures located at waist level or below. Exits and entrances should be of non-

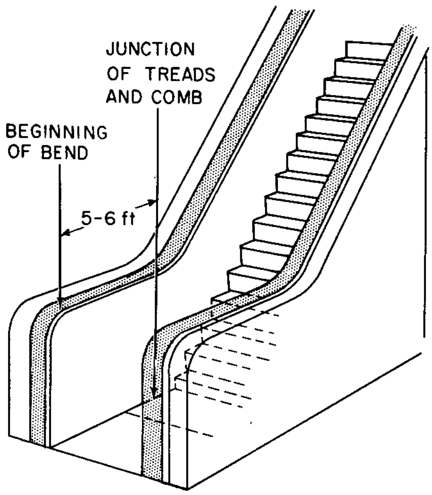


FIGURE 10–61. Recommended escalator features.

skid material both for firm footing and to prevent glare. Eight to 10 ft.-c. is adequate for the remaining areas of the escalator and surround.

Clearly marked emergency power-off panels which will stop the escalator should be located at the head and foot of each escalator flight. The power switch should be located under a lid which will automatically sound an alarm when raised. Restarting the escalator should be possible only by a special key or control which can be operated by authorized personnel.

Elevators

The requirement for military personnel to use elevators in standard military constructions is rare. Elevators should be used for transporting:

1. Heavy loads of materials between levels.
2. A limited number of personnel over distances of three or more levels.
3. Physically disabled personnel.

An elevator is the fastest means of vertical transportation between levels. However, the capacity is limited, and the speed is dependent on the availability of the elevator. For one to two stories, a stairway is faster; for movement of personnel in quantity, an escalator is preferable for three or more levels. In military situations where enemy action might make an elevator nonfunctional, elevators should not be relied upon. Escalators still can be used as ordinary stairs in such an emergency. The use of elevators for transporting military personnel between levels should be restricted to specialized requirements where the high cost, space, maintenance, unreliability, and low capacity penalties are justified.

Man-Lifts

A man-lift is a specialized and hazardous means of rapid transport of individuals between one or more levels. It is useful where a person must move continuously between levels, where space and cost do not permit installation of an escalator or bank of elevators, and where the continuous use of stairs is time-consuming and fatiguing. The potential hazards are so apparent that the casual user employs extreme caution when riding a man-lift; this probably accounts for the relatively low accident rate. However, all current commercial units examined are unsatisfactory from the human engineering standpoint. The principal difficulty is the continuous moving belt and pulley arrangement. This requires an "UP" and a "DOWN" set of handgrips which are "cupped" in opposite directions to prevent confusion. This ingenious method of making the best of a bad situation is still unsatisfactory. The man-lift is especially hazardous when a breakdown or other stoppage occurs with a rider suspended between floors. There may be no readily safe means available to him to get off of the lift.

The emergency stop provisions are also unsatisfactory. The most prevalent means is a cord running the length of the man-lift and adjacent to it. This requires the rider to remember to have a free hand next to the emergency cord at all times.

References

Brooks, F. A., Jr. Operational sequence diagrams. *IRE Transactions on Human Factors in Electronics*, March 1960.

Chaffee, J. W. Anthropometric considerations for escape capsule design. Convair Report, Contract No. AF 33(600)–36200. 18 January 1960.

Churchill, A. V., & Lewis, R. E. F. Ships control spaces: machinery control headquarters. Report No. DMRL 562, October 1964, Defense Research Board, Canada.

Kinney, G., Manning, S., & Smith, L. Further research on the effect of viewing angle and symbol size on reading ease. Report No. ESD–TDR–64–633, January 1965, Electronics System Division, U.S. Air Force, Hanscom Field, Bedford, Mass.

Ramsey, C. G., & Sleeper, H. R. *Architectural graphic standards* (6th ed.), New York: John Wiley & Sons, 1965.

Schlanger, B. How to design a better motion picture theatre. *Better Theatres*, 1942–1944 (series).

Seminara, J. L., & Tevis, J. M. Mockups: plain and fancy. *Machine Design*, 20 June 1963.

Thomson, R. M., Jacobs, H. H., Covner, B. J., & Orlansky, J. Arrangement of groups of men and machines, ONR Report No. ACR-33, 1958, Office of Naval Research, Washington, D.C.

White, C. B., Johnson, P. J., & Hertzberg, H. T. E. Review of escape hatch sizes for bailout and ditching. Technical Note No. WCRD-52-81, September 1952, Wright Air Development Center, WPAFB, Ohio.

Woodson, W. E. *Human engineering guide for equipment designers* (3rd ed.). Berkeley, California: University of California Press, 1965.

Chapter 11

Engineering Anthropology

H. T. E. Hertzberg

Human Engineering Division
Aerospace Medical Research Laboratory
Wright-Patterson AFB, Ohio

Anthropometry is the technology of measuring various human physical traits, primarily such factors as size, mobility, and strength. Engineering anthropology is the effort to apply such data to equipment, workplace, and clothing design to enhance the efficiency, safety, and comfort of the operator. Information in this chapter covers body dimensions, ranges of motions for body extremities, and muscle-strength capabilities for various groups of people within the United States, and for other national populations of interest to the equipment designer. The dimensions presented are some that relate most directly to the characteristics of equipment and workplaces used by people and that affect their performance, safety, and comfort.

CONTENTS

This chapter is revised from material which appeared in Chapter 11 of the previous *Guide*. This chapter was reviewed by Lee Caldwell, E. Ralph Dusek, and Robert M. White. Dr.-Ing. E. K. H. Kroemer prepared most of the rough draft of the Section on Muscle Strength and Muscle Power (11.7).

11. Engineering Anthropology

The human body, in its structure and mechanical function, occupies a central place in man-machine design. Failure to provide a few inches, which might be critical for the operator, can jeopardize performance, operator safety, and machine reliability. With proper forethought, these critical inches usually can be provided without compromising the design. By using reliable anthropometric data it is possible to establish proper sizes of equipment involving human use.

The engineer should design to accommodate the body dimensions of the population that will be using the equipment. Normally, an attempt is made to achieve universal operability. This being an ideal goal, compromises to 98% or even 90%—but no less—might be necessary, but nearly universal operability is desirable because many situations where machinery is being operated require human interchangeability. Any equipment, no matter how well it has been engineered, can be destroyed or abused by an uncomfortable or inefficient operator. If this is a result of some oversight of the designer, the failure might better be termed "design error" than "operator error." As the supply of qualified operators for a complex machine may be limited, their selection and training is easier if machinery fits most operators. Fitting most operators usually requires adjustability.

Adjustability must be considered in design because equipment built according to any one arbitrary set of dimensions can seldom accommodate the entire range of body sizes in the user population. Adjustability is especially important for products intended for a world market, due to the great diversity in human body size around the world. For efficient design of adjustability, reliable anthropometric data are essential, and some of such data are presented in this chapter.

Yet, currently available anthropometric data from various sources may not be comparable because of differences in measuring techniques. Faulty technique can cause apparent size differences that do not exist, or mask those that actually do. To achieve collective standards of measuring technique and terminology for worldwide use, a convention of anthropologists was called and their technical choices were reported (Hertzberg, 1968a). The terminology in this chapter has been modified to conform to the standards chosen by that group. For example, "measurement" is used to denote the act of measuring, and "dimension" or "variable" to designate the technical entity measured. The words "data" or "values" denote the numerical results of measuring the entity.

11.1 Human Variability

There are wide differences in body dimensions between different national or ethnic groups, and even within them. Caucasians, Negroes, and Orientals differ from one another in height, weight, body proportion, and in body physique. (See Newman and White, 1951; Ishii, 1957; and Hertzberg, 1968b.) For example, stature varies from about 4.5 ft. among Congo pygmies to more than 6.5 ft. among certain African tribes, among whom individuals of 7 ft. are not unknown. Various groups of northwest European derivation also are among the world's tallest people. The population of the United States, a national group of mixed racial origins, is among the larger-sized members of the human species. Even within the same national population, there is a wide range of size and proportions. For example, steel workers, truck drivers, and athletes are usually larger and more muscular than bookkeepers, college professors, and artists. Usually there are size differences between successive generations, and within a given generation individuals tend to change size as they age from young adulthood.

Because of human variability, anthropometric data for design use must be perceived as a range of values rather than an average. The use of the average alone is an error. (See Section 11.1.4.) Hence every machine should include the adjustability needed to accommodate the range of size of its operators. The same forethought is needed for maintenance. If a machine like the automobile, the tractor, or similar equipment, is to enter the world market, the designer should have a reasonable knowledge of the world range of size of potential operators and repair men. As a step in that direction, this chapter presents some data on men and women, as available, for a variety of national groups around the world. Adjustability of equipment is a keynote of design.

11.1.1 Historical Variability

Changes in human body size have been occurring from prehistoric times, with a worldwide trend noted in recent centuries toward an increase in height. Size increases in recent decades have invalidated anthropometric surveys of the 19th and early 20th centuries. A review of previous surveys shows an increasing average stature (Gould, 1869; Baxter, 1875; Davenport and Love, 1921). American soldiers of World War II averaged 0.7 in. taller and 13 lb. heavier than those of World War I. (See Davenport and Love, 1921; and Newman and White, 1951.) Comparisons between fathers and sons among a group of Harvard men showed the sons to be significantly taller and heavier than their fathers at the same age (Bowles, 1932). The general trend appears to be continuing in the American population (Karpinos, 1961, Newman, 1963), and even among native-born sons of immigrant fathers (Damon, 1965). Table 11–1 below shows the preliminary results of a recent unpublished survey of Air Force flying personnel taken in 1967 compared with data obtained in 1950 (Hertzberg et al., 1954). These military data substantiate the civilian trends toward increased stature and weight. Although improved nutrition has been cited as instrumental in explaining the increase, other factors, including heterosis, may also contribute (Dahlberg, 1942; Hulse, 1957). Heterosis is the increase in size that results from interbreeding between different physical types.

11.1.2 Variability in U.S. Military Groups

Selection criteria among the Air Force, Army, Navy, and Marines are not uniform, and men of different body sizes, geographical origins, and education choose different services; hence the several services exhibit considerable differences in the body size range of their personnel. This variation is so marked that any subgroups using standardized equipment—such as truck drivers or aviators—should be measured to determine their range of sizes, and from this the limits of design for them.

Most of the body size data presented in this chapter for U.S. military groups are now 10 to 25 years old, so do not reflect the considerable change in body size that anthropologists and others have observed since World War II. This change caused the Department of Defense to order new military anthropometric surveys in

TABLE 11–1. COMPARISON OF WEIGHT, STATURE, AND SITTING HEIGHT OF USAF FLYING PERSONNEL MEASURED IN 1967 AND 1950

Dimension	Year	N	Mean	SD	Mean	SD
Weight	1967*	2420	173.60 lb	21.44	78.74 kg	9.72
	1950†	4063	163.66 lb	20.86	74.23 kg	4.42
Increase	----	----	9.94 lb	----	4.51 kg	----
Stature	1967*	2420	69.82 in	2.44	177.34 cm	6.19
	1950†	4063	69.11 in	2.44	175.54 cm	6.19
Increase	----	----	0.71 in	----	1.80 cm	----
Sitting height	1967*	2420	36.69 in	1.25	93.18 cm	3.18
	1950†	4063	35.94 in	1.29	91.28 cm	3.27
Increase	----	----	0.75 in	----	1.90 cm	----

* Clauser et al., unpublished data, report in preparation.
† Hertzberg et al. (1954).

the armed services. Between 1965 and 1967, over 19,000 men in the Air Force, Army, Marines and Navy were measured. Although no reports of the new data were available when the tables in this chapter were compiled, the following unpublished data on U.S. Air Force flying personnnel are presented by permission of Aerospace Medical Research Laboratory to give designers some idea of the trend in USAF body size since 1950 (Table 11–1). Changes in percentile values between the two samples are found in Table 11–2.

These preliminary results show that the percentage of larger men in the Air Force has inincreased since 1950. If the 1967 values are assessed on the 1950 percentile scale, it can be seen that the 95th percentile of the 1967 data has shifted upward to roughly the 98th percentile of the 1950 scale for the three variables of weight, stature, and sitting height. Approximately the same shift upward is apparent at the lower end of the distributions. Similar shifts are very probable for the other services.

Although these increases are appreciable, they are not necessarily severely damaging to a design planned on 1950 data. If no compensatory changes can be made in the design—for example, an increase of 0.8-in. adjustability of a pilot's seat, at either the top or the bottom of its vertical travel—this growth in 1967 body size means that the distance that formerly was adequate for 95% of the 1950 sample will now accommodate only the 85th percentile of the sitting height of the 1967 population. Whether or not this reduction may be acceptable depends on many factors which the designer must reconcile.

11.1.3 Variability Among Different National Groups

It is commonly known that there is a considerable range in size between individuals within any given national group. After the group has been measured, this range can be accurately expressed statistically by a mean value and its standard deviation. Tables 11–9, 11–10 and 11–15 present the means of stature, weight, and sitting height, and percentiles of each, for numerous national groups and subgroups around the world. (See Sections 11.3.4. and 11.4.2.)

It is not so commonly known that there are also important differences in body proportion among groups, especially between widely separated racial stocks. Table 11–3, summarized from a recent review of this situation (Hertzberg, 1968b), emphasizes some of the proportional variations found between such disparate groups as Mediterraneans, Orientals, and USAF flying personnel. By dividing the range of values for dimensions of the 1950 USAF sample into quartiles, a yardstick is established by which differences between this and other groups can be assessed. Taking the 75th percentile of the USAF group as a reference percentage, the percentages of the other groups that show the same values as the USAF groups can be computed. Any group of similar body size and proportion to the USAF group will also show a percentage of about 75 for each dimension. Table 11–3 shows that, in these samples, about 93% of the Mediterranean men, and 99% of these Oriental men, do not exceed the value of the USAF 75th percentile.

TABLE 11–2. PERCENTILES OF WEIGHT, STATURE AND SITTING HEIGHT OF USAF FLYING PERSONNEL: 1950 VERSUS 1967

	Year	N	Percentiles 1st	5th	50th	95th	99th
Weight (lbs.)	1967*	2420	127.6	140.2	172.4	210.8	227.7
	1950†	4063	123.1	132.5	161.9	200.8	215.9
Increase	----	----	4.5	7.7	10.3	10.0	11.6
Stature (in.)	1967*	----	64.3	65.9	69.8	73.9	75.6
	1950†	----	63.5	65.2	69.1	73.1	74.9
Increase	----	----	.8	.7	.7	.8	.7
Sitting height (in.)	1967*	----	33.9	34.7	36.7	38.8	39.6
	1950†	----	32.9	33.8	36.0	38.0	38.9
Increase	----	----	1.0	.9	.7	.8	.7

* Hertzberg et al. (1954).
† Clauser et al., unpublished data; report in preparation.

TABLE 11–3. COMPARISON, BETWEEN DIFFERENT NATIONAL GROUPS, OF PERCENTAGES OF SELECTED BODY DIMENSIONS WHOSE VALUES DO NOT EXCEED THE 75TH PERCENTILES (3RD QUARTILE) FOR USAF FLYING PERSONNEL*

	Italian AF**	Greek AF**	Turkish AF**	South Korean AF†	Japanese AF‡
Stature	93	93	96	99	99
Weight	88	95	97	98	99
Sitting height	89	86	88	81	87
Buttock-knee length	91	96	97	100	----
Gluteal furrow height	93	96	97	100	----
Foot length	82	85	87	100	----
Foot breadth	32	39	39	73	----
Hand length	78	84	83	97	95
Hand breadth at metacarpale	69	79	86	98	83
Head circumference	89	95	98	94	91

* Hertzberg et al. (1954).
** Hertzberg et al. (1963).
† Kay et al. (1961).
‡ Oshima et al. (1962).

In most of these dimensions, chosen because they are indicators of body proportion, the comparison groups are seen to be distinctly smaller than the USAF sample; but the variation in their percentages shows that their proportions differ from both the USAF and each other. For example, among the Orientals, torsoes are longer and legs are shorter for a given stature; and Mediterranean and Oriental hands differ appreciably from USAF hands, while their feet are shorter and much wider than those of the USAF.

Such data re-emphasize the requirement that designers must provide adjustability to accommodate for differences in human body size and proportion. For more specific information concerning these and other national groups, the reader is referred to the sources at the bottom of each table.

11.1.4 The Fallacy of the "Average Man"

Designing to fit the "average man" is a serious error. By definition, 50% of any group might suffer from a design sized to the 50th percentile, and this could have serious consequences. For example, the smaller 50% will be unable to reach a control just fitted to the average or 50th-percentile operator.

A further fallacy of the "average man" concept is that it ignores the fact that no one is average in all respects, and few people in several. To demonstrate this for any sample, if one takes the middle third of any dimension, then takes the middle third of that third with respect to an unrelated dimension, and then repeats the process for a third independent dimension, one obtains $\frac{1}{3} \times \frac{1}{3} \times \frac{1}{3} = \frac{1}{27}$, or less than 4% of the sample, who will be "average" in three traits. This has been found to be true for stature, weight, and chest circumference in spite of the fact that these dimensions are correlated (Daniels, 1952). Even fewer persons, about 1%, will be average in four traits, and the fraction becomes vanishingly small as the number of traits included increases further.

The "average man" concept is equally unfit for muscle strength and other biomechanical data. For example, in the design of an ejection seat, if trigger force requirements were set to the strength capabilities of the average or 50th-percentile pilot, then the weaker 50% of pilots would be unable to escape (Hertzberg, 1955, 1960; Damon et al., 1966).

11.2 Factors Affecting Body Size

11.2.1 Age and Sex

All body dimensions increase, though somewhat irregularly, from birth to the late teens or early twenties. The age at which growth is complete varies with the individual person and the dimension. Table 11–4 indicates that full growth is normally attained by the age of 20 in males and 17 in females. Sometime around age 60 height decreases, but current evidence is inadequate to establish the time when this begins, its extent, or its influence on other body lengths. Weight and its correlated body breadths,

TABLE 11–4. HEIGHT AND WEIGHT OF WHITE MALE AND FEMALE AMERICANS AT DIFFERENT AGES

	Male				Female			
	Height (in.)		Weight (lb)		Height (in.)		Weight (lb)	
Age (yr)	Mean	S.D.	Mean	S.D.	Mean	S.D.	Mean	S.D.
1	29.7	1.1	23	3	29.3	1.0	21	3
2	34.5	1.2	28	3	34.1	1.2	27	3
3	37.8	1.3	32	3	37.5	1.4	31	4
4	40.8	1.9	37	5	40.6	1.6	36	5
5	43.7	2.0	42	5	43.8	1.7	41	5
6	46.1	2.1	47	6	45.7	1.9	45	5
7	48.2	2.2	54	7	47.9	2.0	50	7
8	50.4	2.3	60	8	50.3	2.2	58	11
9	52.8	2.4	66	8	52.1	2.3	64	11
10	54.5	2.5	73	10	54.6	2.5	72	14
11	56.8	2.6	82	11	57.1	2.6	82	18
12	58.3	2.9	87	12	59.6	2.7	93	18
13	60.7	3.2	99	13	61.4	2.6	102	18
14	63.6	3.2	113	15	62.8	2.5	112	19
15	66.3	3.1	128	16	63.4	2.4	117	20
16	67.7	2.8	137	16	63.9	2.2	120	21
17	68.3	2.6	143	19	64.1	2.2	122	19
18	68.5	2.6	149	20	64.1	2.3	123	17
19	68.6	2.6	153	21	64.1	2.3	124	17
20–24	68.7	2.6	158	23	64.0	2.4	125	19
25–29	68.7	2.6	163	24	63.7	2.5	127	21
30–34	68.5	2.6	165	25	63.6	2.4	130	24
35–39	68.4	2.6	166	25	63.4	2.4	136	25
40–49	68.0	2.6	167	25	63.2	2.4	142	27
50–59	67.3	2.6	165	25	62.8	2.4	148	28
60–69	66.8	2.4	162	24	62.2	2.4	146	28
70–79	66.5	2.2	157	24	61.8	2.2	144	27
80–89	66.1	2.2	151	24				

Stoudt et al. (1960).

depths, and circumferences, continue to increase significantly through middle age and then decrease in old age.

Men are larger than women at any given percentile for most body dimensions, but the extent of the difference varies considerably. Women are consistently larger only in hip-to-hip breadth, hip circumference, and thigh circumference. Men's legs are not only absolutely longer than women's, but also are longer relative to standing and sitting height.

11.2.2 Body Position

Many human dimensions vary with posture or body position. In order to standardize and compare, the anthropometrist usually requires specific, erect positions rarely assumed by people at work or at rest. Because few people normally stand or sit completely erect, "normal" standing height, sitting height, and eye height involve "slump," and are thus significantly less than when measured with the body erect (about 0.75 in. less for standing and 1.75 in. less for sitting heights). Standing height is shorter than prone or supine length. Buttock breadth and waist depth are larger in the seated than in the standing position. Most dynamic dimensions are altered by body movement: thus, maximum arm reach with free movement of the shoulder or trunk is much greater than with the shoulder and trunk restrained.

11.2.3 Body Build

Differences in size and sex are obvious causes for physical variation. Less obvious but sometimes critical are differences in body build. A man's body type certainly affects his choice of athletic event, and probably influences his choice of livelihood. For example, among white men, aged 18 to 40, truck drivers differ from research workers, and USAF fighter pilots differ from bomber pilots. It is also becoming common knowledge that among Olympic athletes dash men differ considerably from distance runners, and that track men vary greatly from athletes in the field events. Similar differences can be seen among other athletic classifications, such as basketball players versus football players. Such physical differences are fundamental. Body size, of course, is one striking difference between individuals, but there are metrical differences also in form and proportions of the body parts. Thus, a large man is not necessarily simply a scaled-up facsimile of a small man.

Although Hippocrates, and many later physicians, tried to classify human bodies according to verbally defined types, only two systems need be mentioned here. One, called "somatotyping" (Sheldon et al., 1940), conceives of relative degrees of fatness, muscularity, and thinness or fragility of the body as numbers on a scale from 1 to 7, rather than as verbal types. Another system, called "body composition" (Behnke and Siri, 1957; Brozek and Behnke, 1963) utilizes anthropometric dimensions, and skinfold thicknesses at selected body locations, to estimate the percentages of fat, water, protein and mineral in the body. Both methods, or some derivative of each, have been used extensively by anthropologists, medical men, physiologists, and physical educationists.

11.2.4 Clothing and Equipment

Specific environmental needs of an operator often involve special-purpose or protective equipment added to the nude body. Unusual heat, cold, altitude, and pressure are all examples of special environments. Escape hatches that do not permit passage of a flyer wearing a parachute, gun-charging handles in aircraft that cannot be operated by a gloved hand, and turrets that provide insufficient space for gunners wearing helmets and oxygen masks are results of poor design. Thus, the designer should acquaint himself with the environmental conditions and personal equipment with which operators of the machines will have to cope.

These machines should be designed to accommodate body sizes near both extremes, whether lightly clad or encumbered with complete special-purpose outfits. (See Figure 11–1.)

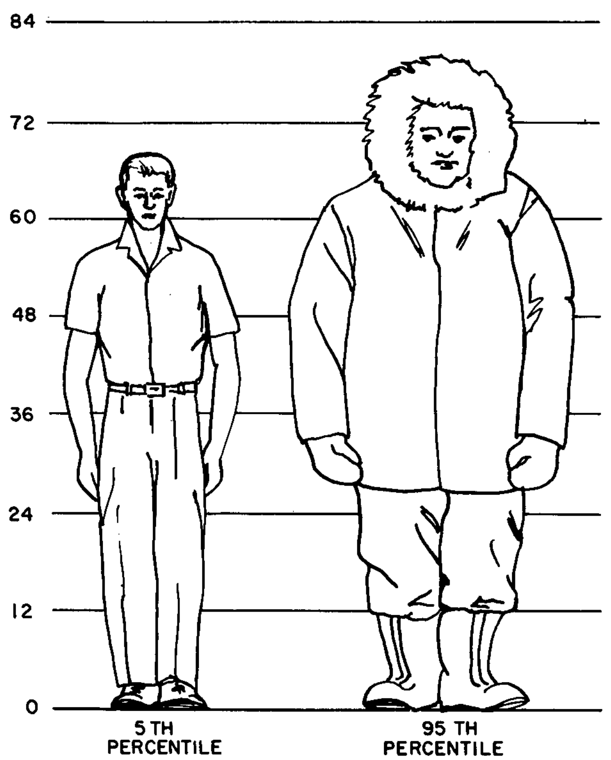


FIGURE 11–1. Comparison of normally clothed 5th-percentile civilian with 95th-percentile soldier in arctic clothing.

11.2.5 Effects of Clothing on Performance

Wearing heavy clothing affects performance. For example, gloves reduce tactility markedly, and pressurized gloves reduce it severely. With 17 subjects wearing pressure suits, Walk (1964) measured tactile performance on the Purdue Pegboard Dexterity Test under three conditions: (a) barehanded, (b) gloved but unpressurized, and (c) gloved and pressurized to 2.5 p.s.i. Accepting the barehanded results as standard (100%) performance, gloved but unpressurized performance declined to 65%, and pressurized gloved performance fell to 35% of standard. Part of the same trend is evident in performance tests involving the actuation of knobs, toggle switches, and pushbuttons (Sharp, 1964). In tests of the pressurized condition in comparison to the unpressurized for several pressure suits, it was found that the pressurized condition invariably increased actuation time.

11.3 Anthropometric Data

11.3.1 Data

In the collection of reliable anthropometric data, three criteria must be satisfied:

1. Measuring techniques must be specified and standard. This is the only way whereby data from different groups can be compared, and test subjects can be accurately located as percentiles of a user group. It makes a great deal of difference, as has been noted previously, whether the body is measured with or without clothing or is in the erect or slumped position, or at which place on the body the dimensions are taken.

2. The groups measured should be representative of equipment users. Neglecting to consider age, sex, race, occupation, and such factors as geography or socioeconomic status may lead to nonrepresentative sampling. Americans living in the Pacific Northwest average 0.6 in. taller and 3 lb. heavier than New Englanders; college men are larger than noncollege men. Military personnel, highly selected by age, health, height, and weight standards, cannot adequately represent the general population. Within a single military service, different specialists are physically distinct.

3. Samples should be large enough to yield required statistical reliability. The term *reliability* is used here to denote the extent to which results are reproducible from sample to sample. The degree of required reliability is dependent on the ultimate use of the data. For example, where the data will be used to establish preliminary workspace layout or for other applications where "ballpark" dimensions are needed, samples of 50 to 100 people may be acceptable. Larger samples are needed for dimensions with a wide range, such as weight, than for range-restricted dimensions such as those of the head, hand, or foot. Much larger samples are required when reliable measures are needed of the extremes of a distribution.

In evaluating equipment, the procedures listed below should be followed:

1. Obtain percentile data on the range of size of the intended operators.

2. Select and measure a small group of subjects to ensure that they represent the percentile range.

3. Dress the subjects in the full set of standard clothing and personal equipment that might have to be worn while operating the equipment.

4. Have the subjects operate the equipment, performing all motions necessary. Subjects should remain in the operating position for as many hours as standard operation requires, to become aware of those faults in the whole environment that appear only with time.

5. Note any shortcomings in design, especially the degradations in comfort, efficiency, vision, and safety in relation to the percentiles of the operator population, and use this information as a guide to redesign. (For a fuller discussion of these points, see Damon et al., 1966.)

11.3.2 Locating Sources of Data

The anthropometric data presented in this chapter are heavily weighted toward military populations, largely because soldiers are the most easily obtained subjects, but also because governments are realizing the impressive benefits of anthropometric data in improving efficiency of operation and economy of procurement. Some data on civilians have been listed, when available, including a sample from Eastern Europe. Few anthropometric surveys for engineering data have been conducted in the free world on civilian populations, numerically the most important group in any country. Nearly all of the surveys that have been made were focused on medical or biological descriptions (like color of skin, hair, and eyes, and other nonmetrical traits).

One important exception is the National Health Survey of U.S. adults (Stoudt et al., 1965). In this publication, weight and 11 other anthropometric dimensions were summarized for 3091 men and 3581 women aged 18 to 79, who were selected on a scientifically random, nationwide basis.

Another recent study also throws some light on civilian male body size, even though it is on air-traffic-control trainees—a rather specialized sample. Performed by the Federal Aviation Agency (Snow and Snyder, 1965), it contains data for 64 dimensions on 678 men ranging from 21 to 50 years of age.

Other data applicable to the U.S. civilian population are available from a classic study of railway travelers (Hooton, 1945), but the subjects were predominantly New Englanders, and the eight dimensions measured were chosen for their relevance to the design of railway-coach seats. Because of the special measuring chair used, these dimensions are not always directly comparable to those taken using standard techniques. A detailed anthropometric study of women (O'Brien and Shelton, 1941) was excellent for its intended purpose—garment and pattern construction—but it consists mostly of body heights, circumferences, and skin-surface distances, only a few of which are useful in solving engineering problems.

If the desired anthropometric information is not contained in this chapter or these reference sources, the designer is advised to consult active anthropometry groups. Several of these groups are indicated in Table 11–5.

TABLE 11–5. SOME SOURCES OF ANTHROPOMETRIC DATA

Name	Location
A. Primary: laboratories of anthropometry (which specialize in anthropometric research as well as gather a library of data)	
Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base.	Dayton, Ohio.
U.S. Army Natick Laboratories.	Natick, Mass.
Anthropology project, Webb Associates.	Yellow Springs, Ohio.
Department of Human Anatomy, University of Newcastle.	Newcastle-on-Tyne, England.
Centre d'Anthropologie Appliquée, Université de Paris.	Paris, France.
Department of Anthropology, Harvard University.	Cambridge, Mass.
School of Public Health, Harvard University.	Boston, Mass.
B. Secondary: repositories where anthropometric data may be found (where actual anthropometric services may or may not be obtainable)	
U.S. Naval Training Devices Center.	Orlando, Fla.
Aerospace Crew Equipment Laboratory, Naval Air Engineering Center.	Philadelphia, Pa.
Human Engineering Laboratory.	Aberdeen Proving Ground, Maryland.
Guggenheim Center for Aviation and Safety, Harvard University.	Boston, Mass.
Institute for Psychological Research, Tufts University.	Medford, Mass.
Biotechnology Laboratory, University of California, Los Angeles.	Los Angeles, Calif.
Furniture Institute Research Association.	Stevenage, Hertfordshire, England.
Unit for Research on Human Performance in Industry, Welsh College of Advanced Technology.	Cardiff, Wales.
Department of Ergonomics and Cybernetics, Loughborough College of Technology.	Leicestershire, England.
Institute of Engineering Production, University of Birmingham.	Birmingham, England.
Bureau Internationale du Travail.	Geneva, Switzerland.

11.3.3 Understanding Anthropometric Data Presentations

Anthropometric data for engineering use are best presented in percentiles. Percentile tables provide a faster and more convenient means of comprehending the dimensional range to be accommodated, and of locating the percentile equivalent of a measured value, than does the curve of normal distribution. Extreme values represent chance occurrences, which for practical purposes should be disregarded in designing equipment. Removing 1% at both ends of the range will eliminate most of these extreme values and leave a range covering 98% of the population. For some dimensions and equipment, the range from 1st to 99th percentile can be accommodated easily; for others only the spread from the 5th to 95th percentile is feasible. Some discussion of the rationale for such choices will be found in Randall et al., 1946; Coakley et al., 1953; Hertzberg, 1955; McCormick, 1964; and Damon et al., 1965. The designer should attempt to accommodate at least 90% of the population as a minimum, and strive for 98% or more, if possible.

Percentiles can serve the design engineer in the following ways:

1. They afford a basis for estimating the proportion of a group accommodated or inconvenienced by any specific design.
2. They permit selection and accurate use of test subjects. Any body dimension, value, or physical test score of a subject can be readily located.
3. They aid in the selection of operators. If the equipment imposes any size limitation, misfits can be avoided by their elimination before or after an established cutoff point. Of course, the cutoff point could be established in measurement units (in., lb., etc.) without using percentiles, but percentiles indicate the proportion of potential operators rejected. If the proportion rejected is too large, redesign is in order. For example, during World War II, the Air Force considered raising gunners' height and weight limits to 73 in. and 180 lb., but it was found that, even at the existing limits of 70 in. and 170 lb., a significant percentage of gunners (30% for one turret, 40% for another) had trouble operating the turrets, and larger men could not even enter the turrets. This information made it obvious that a redesign of the turrets was needed.

How to Find Percentiles

Percentiles can be obtained graphically from normal-probability paper for normally distributed groups, i.e., those following the bell-shaped or normal curve. On normal-probability paper, a normal distribution is a straight line that is defined by two points. (See Figure 11–2.) Thus a graph can be drawn permitting any desired percentile to be read from the paper, if any two percentiles or any two of the following values are known:

1. A single percentile.
2. An average, whether mean or median (50th percentile).
3. The standard deviation (see below).

Percentiles can be obtained arithmetically from the standard deviation if the mean also is known. The standard deviation (S.D.) is a measure of dispersion, variation, or scatter about an average. Thus, the average (or 50th percentile) ± 1 S.D. includes 68% of the measured group, ± 2 S.D. includes 95% of the group, and ± 3 S.D. includes 99.7% of the group (see Figure 11–2). The S.D. can be approximated by subtracting the value corresponding to the 16th percentile from that corresponding to the 50th percentile.

Percentiles computed from a known standard deviation provide an alternative to the graphic technique. With the factors in Table 11–6, the S.D., and the value of either the mean or median, percentiles can be computed as in the following examples:

1. To obtain the 95th percentile when the mean is 35.1 in. and the S.D. = 1.5 in., find the factor corresponding to the 95th percentile in Table 11–6, multiply the S.D. by it ($1.5 \times 1.645 = 2.5$ in.), and add the result to the mean: $35.1 + 2.5 = 37.6$ in. This is the 95th percentile.
2. To obtain the 5th percentile, given the same starting information, proceed as before, with the same numbers, but subtract instead of add: $35.1 - 2.5 = 32.6$ in. This is the 5th percentile.

11.3.4 Application of Anthropometric Data in the Design Situation

The correct procedure for designing equipment to accommodate human dimensions, using the data presented in subsequent sections, is as follows:

1. Determine the body dimensions important in the design (e.g., sitting height as a basic factor in seat-to-roof distance in automobiles or truck cabs).
2. Define the population to use the equipment. This establishes the dimensional range that needs to be considered (e.g., U.S. military, U.S. civilians, worldwide populations, etc.).
3. Select the percentage of the population to be accommodated (e.g., 90%, 95%, 98%, or whatever is relevant to the problem).

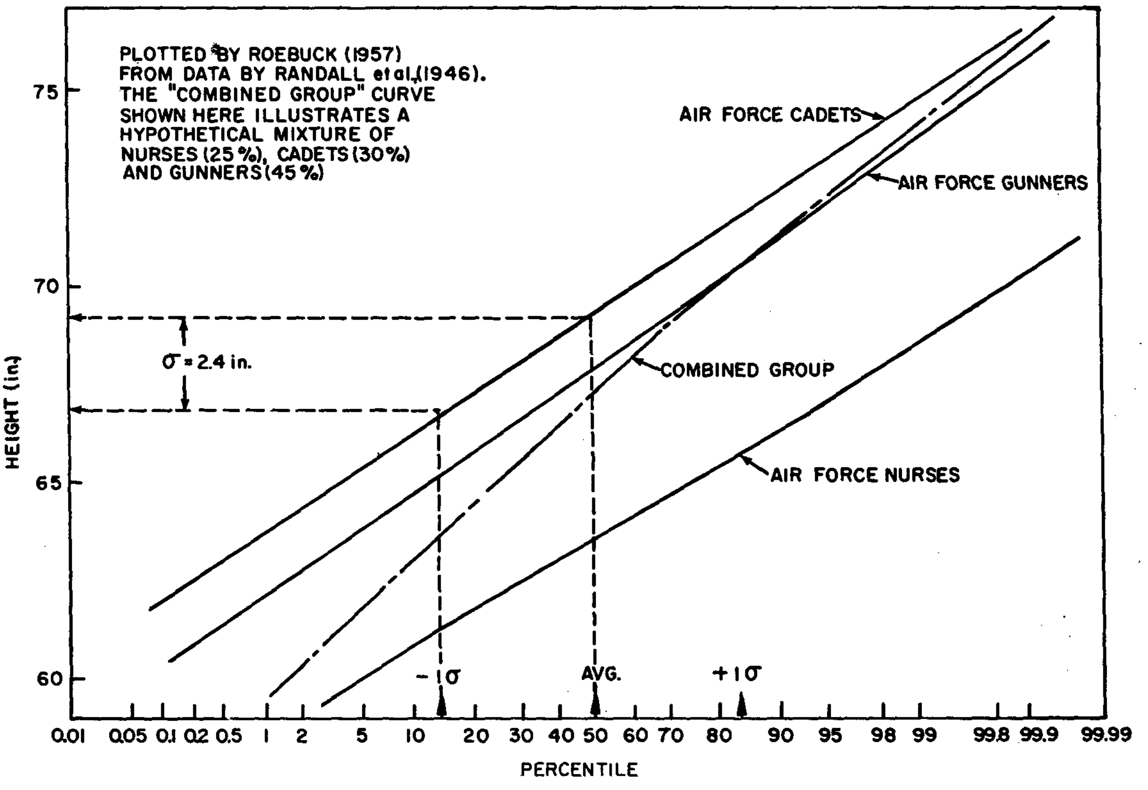


FIGURE 11–2. Comparison of heights among Air Force cadets, Air Force gunners, Air Force nurses, and the combined group.

TABLE 11–6. FACTORS FOR COMPUTING PERCENTILES FROM THE STANDARD DEVIATION

Percentiles		Factor	Percentiles		Factor
0.5	99.5	2.576	15	85	1.036
1	99	2.326	20	80	.842
2.5	97.5	1.960	25	75	.674
5	95	1.645	30	70	.524
10	90	1.282			

Roebuck (1957).

4. Extract the value of the selected percentage from the appropriate dimension table located in Structural Body Dimensions, Section 11.4; then check the text for qualifying information.

5. Determine the type of clothing and personal equipment that will be worn (e.g. civilian clothing or winter flying gear) and add the relevant clothing increments (see Tables 11–7 and 11–8) to the values indicated in the dimension table.

The following tables (Tables 11–9 and 11–10) present data on the major American military and civilian groups for which the engineer may be designing, and also for a number of foreign populations. Military populations are usually measured nude, while civilian populations are measured wearing light clothing, preferably without shoes. In the world market, necessary allowances for the range of relevant body dimensions should be made. The blank areas in Table 11–9 indicate lack of data for the dimension listed. These and other tables are not to be considered encyclopedic; they are rather a selected cross section to approximate the range of body size variability among men. In the selection of these samples, when a choice existed, larger samples have taken precedence over smaller ones; mature groups have usually been preferred to adolescent groups.

11.4 Structural Body Dimensions

Two kinds of anthropometric dimensions. "structural" and "functional," are related t the practical problems of design engineering. (These types were called "static" and "dynamic" in the first edition of this *Guide*; the new terms were chosen, as more descriptive of the body and its action, in accordance with the recommendations of the standardization conference previously mentioned; Hertzberg, 1968a). Structural dimensions, which are taken with the body of the subjects in fixed, standardized positions, are easily obtained and used in equipment design. (See Section 11.4.) Functional dimensions, which are taken with the body in various working positions, are usually more complex and difficult to measure. (See Section 11.5.)

TABLE 11–7. INCREMENTS ADDED TO NUDE BODY DIMENSIONS BY VARIOUS ARMY CLOTHING

Dimension	Increment (in.)*		
	Standard uniform†	Plus blouse or field jacket	Plus blouse or field jacket and overcoat‡
Stature	2.65	2.65	2.65
Sitting height	1.39	1.43	1.61
Weight (pounds)	9.4	11.8	18.6
Head length	3.5	3.5	3.5
Head breadth	2.8	2.8	2.8
Eye height, sitting	.04	.08	.16
Shoulder height, sitting	.16	.58	.92
Shoulder-elbow length	.14	.50	.94
Shoulder breadth	.24	.88	1.52
Chest depth	.41	.96	1.80
Elbow-to-elbow breadth	.56	1.04	1.84
Waist depth	.94	1.18	1.95
Hip breadth	.56	.76	1.08
Hip breadth, sitting	.56	.76	1.08
Buttock-knee length	.20	.30	.54
Knee height, sitting	1.32	1.32	1.44
Knee-to-knee breadth	.48	.48	.72
Foot length	1.6	1.6	1.6
Foot breadth	.20	.20	.20

Roberts et al. (1945).
* Except as noted.
† Underwear, shirt, and trousers or fatigues, shoes and socks, and steel helmet and liner.
‡ See Roberts et al., 1945, for additional increments to be added for combat suit, overcoat, wool cap and gloves.
Note: See also White, Kobrick, and Zimmer, 1964, for Arctic clothing dimensions.

TABLE 11–8. INCREMENTS ADDED TO NUDE BODY DIMENSIONS BY VARIOUS AIR FORCE CLOTHING

Dimension	Increment (in.)*			
	Winter flying gear†	T–1 Partial-pressure suit‡	T–5 Partial-pressure suit§	A/P 22–2 Full pressure suit¶
Stature	1.9	2.0	3.3	----
Sitting height	1.6	----	2.1	40.8
Weight (pounds)	20.0	----	----	----
Head length	0.4	----	----	----
Head breadth	0.4	----	----	----
Eye height, sitting	0.4	----	----	----
Shoulder height, sitting	0.6	----	----	----
Shoulder elbow length	0.3	----	----	----
Shoulder breadth	1.3	6.0	0.4	----
Chest breadth	0.6	2.5	----	17.0
Chest depth	1.4	4.5	0.8	----
Elbow-to-elbow breadth	4.4	11.0	----	----
Waist depth	1.4	5.0	----	----
Hip breadth	1.3	----	----	----
Hip breadth, sitting	1.7	5.5	2.9	----
Buttock-knee length	0.5	2.0	----	----
Hand length	0.3	----	----	----
Hand breadth	0.4	----	----	----
Knee height, sitting	1.8	----	----	----
Knee-to-knee breadth	2.5	9.5	----	----
Foot length	2.7	----	----	----
Foot breadth	1.2	----	----	----
Helmet breadth	----	----	----	10.5
Lower torso breadth	----	----	----	14.7
Elbow-rest height	----	----	----	13.4
Maximum thigh height, sitting	----	----	----	29.7
Maximum knee height, sitting	----	----	----	27.0

* Except as noted.

† Underwear, shirt and trousers or fatigues, boots and socks, jacket, helmet, and gloves (Damon, 1943).

‡ Inflated pressure suit, deflated ventilation suit, MD-1 antiexposure suit, MD-3A liner, and long cotton underwear (no boots).

§ Uninflated pressure suit, K-1 pressure helmet, and boots (USAF, 1953b).

¶Alexander, Garrett and Flannery (1969).

TABLE 11–9. MEANS OF WEIGHT, STATURE, AND SITTING HEIGHT OF NUDE MALES AMONG VARIOUS WORLD POPULATIONS

Population	Group	N	Weight Kg	SD	Lb	SD	Stature CM	SD	IN	SD	Sitting height CM	SD	IN	SD
Africa [1]	Dinka Nilotes	279	58.4	5.92	128.3	13.0	181.3	6.12	71.5	2.41	----	----	----	----
Australia [2]	Mixed Army rec.	3580	†72.6	----	†159.7	----	174.0	5.67	68.5	2.23	*90.5	----	*35.6	----
Belgium [3]	Recruits 20-yr.	17018	64.1	7.8	141.1	17.7	171.5	6.10	67.5	2.40	----	----	----	----
Belgium [4]	Flying personnel	2450	66.5	6.9	146.2	15.2	173.9	5.79	68.4	2.28	----	----	----	----
Bulgaria [5]	Civilians aged 26	114	67.7	8.13	148.9	17.9	169.8	5.89	66.9	2.32	89.4	3.68	35.2	1.45
Canada [6]	RCAF pilots	314	76.6	9.91	168.5	21.8	176.5	5.89	69.5	2.32	91.5†	3.35†	36.0†	1.30†
Canada [7]	Civilians 30–40	400	76.4	----	168.0	----	172.9	----	68.1	----	----	----	----	----
Ceylon [8]	Sinhalese	635	----	----	----	----	160.4	6.02	63.1	2.37	81.5	3.45	32.1	1.36
China (Taiwan) [9]	Young men	31	52.6	4.44	115.7	9.77	167.6	5.22	65.9	2.06	90.4	2.92	35.6	1.15
China (Taiwan) [10]	Military 20–30	1049	55.7	----	122.3	----	164.5	----	64.8	----	----	----	----	----
France [11]	Pilots, cadets	7084	----	----	----	----	171.3	5.81	67.4	2.29	----	----	----	----
France [11]	Pilots, cadets	1000	65.8	7.03	143.0	15.5	----	----	----	----	----	----	----	----
France [12]	Recruits, 20-yr.	234	65.4	7.17	143.9	15.8	169.5	6.05	66.7	2.38	----	----	----	----
East Germany [13]	Civilians 30–39	1651	72.0	----	158.4	----	171.0	----	67.4	----	----	----	----	----
East Germany [14]	Civilians 30–39	12200	----	----	----	----	170.9	7.19	68.3	1.83	----	----	----	----
West Germany [15]	Recruits 20–yr.	316202	68.5	----	150.5	----	175.2	6.6	69.0	1.68	----	----	----	----
West Germany [15]	Recruits 20–yr.	378133	68.4	8.9	150.4	19.6	----	----	----	----	----	----	----	----
Greece [16]	Mixed military	1084	67.0	7.62	147.4	16.8	170.5	5.88	67.1	2.31	90.3	3.01	35.6	1.19
Great Britain [17]	RAF bomb personnel	11772	66.1	7.06	145.1	15.5	173.2	6.2	68.2	2.44	----	----	----	----
Great Britain [18]	RAF Pilots	4357	72.1	9.27	158.6	20.4	177.3	6.15	69.8	2.42	93.3	3.3	36.7	1.30
India [19]	Civilians 19–60	499	50.9	8.00	123.0	17.6	163.0	6.10	64.3	2.40	85.1	3.04	33.5	1.20
India [20]	Civilians mixed	3774	----	----	----	----	163.9	5.57	64.5	2.19	----	----	----	----
Italy [16]	Mixed military	1358	70.3	8.43	154.7	18.6	170.6	6.23	67.2	2.45	89.7	3.20	35.3	1.26
Israel [21]	Civilians 40–44	3339	72.2	11.0	158.8	24.2	168.4	6.50	66.2	2.56	89.2	3.7	35.1	1.46
Japan [22]	JASDF pilots	239	61.1	5.86	133.2	12.9	166.9	4.80	65.7	1.89	90.8	2.62	35.8	1.03
South Korea [23]	ROKAF pilots	264	62.9	6.53	138.2	14.4	168.7	4.61	6.64	1.81	90.8	2.81	35.8	1.11
Mexico [24]	Rural	5217	----	----	----	----	160.2	----	6.31	----	83.8	----	32.3	----
Mexico [24]	Mixed rural	1502	54.1	----	119.0	----	----	----	----	----	----	----	----	----
Norway [25]	Recruits 20-yr.	5765	70.1	7.46	154.2	16.4	177.5	6.03	69.9	2.37	92.7	3.47	36.5	1.37
Thailand [26]	Mixed military	2950	56.3	10.3	123.8	22.7	163.4	5.3	64.4	2.09	86.4	3.1	34.0	1.22
Turkey [16]	Mixed military	915	64.6	8.23	142.1	18.1	169.3	5.76	66.7	2.27	89.7	3.15	35.3	1.24

TABLE 11–9. MEANS OF WEIGHT, STATURE, AND SITTING HEIGHT OF NUDE MALES AMONG VARIOUS WORLD POPULATIONS—Continued

Population	Group	N	Weight Kg	Weight SD	Weight Lb	Weight SD	Stature CM	Stature SD	Stature IN	Stature SD	Sitting height CM	Sitting height SD	Sitting height IN	Sitting height SD
USA:														
Air Force— 1950 [27]	Flyers	4063	74.4	9.27	163.7	20.9	175.5	6.19	69.1	2.44	91.3	3.27	35.9	1.29
Army [28]	Separatees	24449	70.3	9.33	154.8	20.5	174.0	6.38	68.5	2.51	90.9	3.4	35.8	1.34
Army [29]	Pilots	500	75.4	8.59	165.8	18.9	176.5	5.72	69.5	2.25	90.4	3.22	35.6	1.27
Navy— 1964 [30]	Pilots	1549	77.7	8.66	171.4	19.09	177.6	5.91	69.9	2.33	92.1	3.16	36.3	1.25
Nat'l. Health Survey [31]	Total males 17–79	3091	76.4	----	168.0	----	173.2	----	68.2	----	90.4	----	35.6	----
FAA [32]	Tower trainees 21–50	678	73.4	10.0	161.8	22.1	176.7	6.35	69.6	2.50	92.0	3.27	36.2	1.29
Vietnam [33]	Mixed military	2129	51.1	6.0	112.4	13.2	160.5	5.5	63.2	2.16	85.0	3.3	33.5	1.30

*Computed from published means and standard deviations.
†Taken by USAF technique.
1 Roberts and Bainbridge, 1963.
2 Aird et al., 1958.
3 Martin, 1958.
4 Evrard, 1954.
5 Bulgarian Academy of Sciences, 1965.
6 RCAF, 1963.
7 Pett and Ogilvie, 1957.
8 Stoudt and Marett, 1961.
9 Chen et al., 1963.
10 Crawley et al., 1956.
11 Ducros, 1955.
12 Monod et Pineau, 1958.
13 Ries, 1963.
14 Kroemer, 1964.
15 Institut der Wehrmedicinalstatistic, 1964.
16 Hertzberg et al., 1963.
17 Samuel and Smith, 1965.
18 Morant, 1951.
19 Sen, 1964.
20 Guha, 1935.
21 Lippert, 1965.
22 Oshima et al., 1962.
23 Kay, 1961.
24 Comas, 1943.
25 Udjus, 1964.
26 White, 1964a.
27 Hertzberg et al., 1954.
28 Newman and White, 1951.
29 White, 1961.
30 Gifford et al., 1965.
31 Stoudt et al., 1965.
32 Snow and Snyder, 1965.
33 White, 1964b.

Note: Since the compilation of this table (cut-off date, 1 March 1967) several sources have come to hand too late to be incorporated. These are: France: Olivier and Coblentz (1965). South Korea: Hart et al. (1966). U.S.A. Hathaway and Foard (1960). Clauser et al. (in preparation; 1969). Latin America: Dobbins and Kindick (1967).

Table 11–10. Stature (nude males); world population samples

Population	Group	N	Percentiles 1 cm	1 in.	5 cm	5 in.	Means cm	Means in.	95 cm	95 in.	99 cm	99 in.	SD cm	SD in.
MILITARY														
Australia* [1]	Mixed army	3580	160.0	63.0	164.1	64.7	174.0	68.5	184.1	72.5	188.2	74.2	5.67	2.76
Belgium* [2]	Recruits, 20-yr.	17018	156.5	61.6	160.8	63.3	171.5	67.5	182.2	71.8	186.5	73.5	6.1	2.4
Belgium* [3]	Flyers, 17–50	2450	159.3	62.7	163.6	64.4	173.9	68.5	184.2	72.6	188.5	74.2	5.79	2.28
Canada [4]	RCAF pilots	314	164.2	64.7	167.8	66.1	177.4	69.8	188.0	74.0	190.2	74.9	6.1	2.4
France* [5]	Pilots, Cadets	7084	157.2	61.8	161.1	63.5	171.0	67.4	181.1	71.4	185.6	73.1	5.81	2.3
France* [6]	Recruits 20-yr.	234	154.5	60.8	159.0	62.6	169.5	66.7	180.2	71.0	184.7	72.7	6.05	2.38
West Germany* [7]	Recruits 20-yr.	316202	158.8	62.5	163.6	64.5	175.2	69.0	187.0	73.6	191.8	75.5	6.6	2.6
Greece [8]	Mixed military	1084	157.5	62.0	160.9	63.4	170.5	67.2	180.3	71.1	184.8	72.7	5.88	2.32
Great Britain* [9]	Bomber personnel	11772	157.7	62.1	162.2	63.8	173.2	68.3	184.0	72.4	188.5	74.2	6.2	2.44
Great Britain* [10]	RAF pilots	4357	161.8	63.7	166.3	65.5	177.3	69.8	188.1	74.0	192.6	75.8	6.15	2.42
Italy [8]	Mixed military	1358	157.0	61.8	160.2	63.1	170.7	67.2	180.7	71.2	185.6	73.0	6.23	2.45
Japan [11]	JASDF pilots	239	157.3	61.9	159.4	62.8	166.7	65.7	175.0	68.8	180.4	71.1	4.8	1.89
South Korea [12]	ROKAF pilots	264	157.6	62.1	159.6	62.9	168.7	66.4	173.7	68.4	177.9	70.0	4.6	1.81
Norway*[13]	Recruits 20-yr.	5765	162.3	64.0	166.7	65.7	177.5	70.0	188.1	74.1	192.5	76.8	6.03	2.38
Thailand [14]	Mixed military	2950	151.5	59.7	155.0	61.1	163.5	64.4	172.0	67.8	176.0	69.3	5.3	2.08
Turkey [8]	Mixed military	915	157.8	62.2	160.6	63.2	169.0	65.5	179.2	70.6	182.4	71.6	5.73	2.26
USA:														
Air Force—														
'50 [15]	USAF flyers	4063	161.3	63.5	165.5	65.2	175.6	69.1	185.8	73.1	190.3	74.9	6.19	2.44
Army*[16]	Separatees	24449	157.8	62.2	162.4	64.0	173.8	68.4	185.0	72.8	189.7	74.7	6.3	2.48
Army [17]	Pilots	500	163.4	64.4	167.0	65.8	176.1	69.4	186.1	73.3	190.0	74.8	5.7	2.25
Navy—'64 [18]	Pilots	1549	165.3	65.1	168.2	66.2	177.5	69.9	187.7	73.9	191.3	75.3	5.91	2.33
Vietnam [19]	Mixed military	2129	148.1	58.3	151.6	59.7	160.4	63.2	169.6	66.8	173.0	68.2	5.5	2.16
CIVILIAN														
Africa*[20]	Dinka Nilotes	279	166.3	65.5	170.2	67.0	181.7	71.5	191.7	75.5	195.5	77.0	6.1	2.41
Bulgaria*[21]	Civilians 26-yr.	114	155.0	61.1	159.3	62.8	169.8	66.8	180.2	71.0	184.5	72.7	5.89	2.32
Ceylon*[22]	Sinhalese	635	145.6	57.3	150.0	59.1	160.4	63.2	171.1	67.4	175.5	69.2	6.02	2.37
China														
(Taiwan)*[23]	Young men	31	154.4	60.8	158.2	62.3	167.6	66.0	177.0	69.7	180.8	71.2	5.22	2.05
East Germany [24]	Civilians 30–39	12200	153.2	60.3	158.4	62.4	170.9	67.3	183.9	72.4	189.1	74.5	7.19	2.83
India [25]	Civilians 19–60	499	147.8	58.2	152.2	60.0	163.0	64.2	173.9	68.5	178.3	70.3	6.10	2.41
India*[26]	Mixed	3774	150.1	59.5	154.1	60.7	163.9	64.6	174.0	68.5	178.0	70.1	5.57	2.79
Israel*[27]	Aged 40–44	3339	152.1	59.9	156.9	61.8	168.4	66.3	180.0	70.9	184.8	72.7	6.5	2.56
USA: FAA [28]	Tower trainees (21–50)	678	*160.8	*63.3	166.4	65.5	176.6	69.5	187.0	73.6	*192.6	*78.5	6.35	2.5
National Health Survey [29]	Total males (18–79)	3091	156.7	61.7	161.5	63.6	173.5	68.3	185.0	72.8	189.4	74.6	----	----

* Data plotted from published means and standard deviations.
[1] Aird et al. (1958).
[2] Martin (1958).
[3] Evrard (1954).
[4] RCAF (1963).
[5] Ducros (1955).
[6] Monod et Pineau (1958).
[7] Institut der Wehrmedincinalstatistic (1964).
[8] Hertzberg et al. (1963).
[9] Samuel and Smith (1965) (after Morant).
[10] Morant (1951).
[11] Oshima et al. (1962).
[12] Kay (1961).
[13] Udjus (1964).
[14] White (1964a).
[15] Hertzberg et al. (1954).
[16] Newman and White (1951).
[17] White (1961).
[18] Gifford et al. (1965).
[19] White (1964b).
[20] Roberts and Bainbridge (1963).
[21] Bulgarian Academy of Sciences (1965).
[22] Stoudt and Marett (1961).
[23] Chen et al. (1963).
[24] Kroemer (1964).
[25] Sen (1964).
[26] Guha (1935).
[27] Lippert (1965).
[28] Snow and Snyder (1965).
[29] Stoudt et al. (1965).

Each of the following paragraphs deals with a single body dimension related to some aspect of equipment or workspace design. The basic data are given for males, as they constitute the bulk of the labor force. The discussion includes a brief description of the dimension and of the subject's body position while being measured; also correction factors for females, and for clothing, personal equipment, and other relevant variables.

Some features of the military data in this article must be specifically noted for their proper interpretation.

1. All measurements were made on nude subjects, standing erect unless otherwise indicated.

2. Some groups, notably the military, have been selected according to upper and lower limits of height and weight, thus eliminating the extremely tall, short, stocky, and thin persons who appear in the general population.

3. The column headed "50th percentile" contains, in some cases, the average or arithmetic mean. Ordinarily there is little or no practical difference between these two measures of central tendency. Weight is one exception, however; it is a physical trait that does not show a normal distribution. Because of its skewed distribution, the mean exceeds the median, or 50th percentile, by about 2 lb.

4. Some of the percentiles presented in the following tables have been read from graphs plotted on normal probability paper. This technique has been used where original frequency distributions were not available, and percentiles could be obtained in no other way. The graphic technique is accurate, however, only to the extent that the dimensional values are normally distributed. Although most body dimensions approximate the normal distribution closely enough to justify using the graphic technique, an occasional group deviates because of the biased or nonrandom selection of subjects or variations in measuring techniques. The percentile data for weight, derived in part from graphs (Table 11–17), therefore contain some inherent uncertainty, but its magnitude is far smaller than the uncertainty of having no weight percentile at all to accompany the data for stature.

5. Many of the design recommendations in the following paragraphs are based on the dimensions of Air Force personnel because, anthropometrically, Air Force personnel have been described in the greatest detail for design purposes. It may be repeated here that Air Force personnel, like all other U.S. military groups, have been selected from the general population according to standards that differ with each service. Also, the military samples were measured at different times in a period of rapid body-size change. The tables in this chapter show that military data are different from civilian data; no group can adequately represent the other. Therefore the designer must select his data with care, and should build in enough adjustability to accommodate operators in any potential user group.

6. "Large-people" design dimensions, i.e., those that provide clearances to accommodate the large members of the population, are based on some value between the 95th and 99th percentiles. "Small-people" design dimensions, which must accommodate the small members of the population, are based on some value between the 1st and 5th percentiles. Adjustment design dimensions, which must accommodate both large and small persons, are usually based on the middle 90% (5th to 95th percentiles) or 95% or 98% of the populations involved, or some other combination of values which may be chosen for special reasons.

11.4.1 Stature

This is the vertical distance from the floor to the top of the head. The subject stands erect and looks straight ahead. (See Figure 11–3.) The data are given in Tables 11–11 through 11–14.

For women, subtract 4.5 in. from the respective male values. For clothing increments, add 1.0 in. for men's shoes, 1.3 in. for military boots, up to 3.0 in. for women's shoes, roughly 1.0 in. for civilian caps, 1.4 in. for steel helmets, and up to 3.5 in. for flying helmets.

11.4.2 Sitting Height

This is the vertical distance from the sitting surface to the top of the head. The subject sits erect, looking straight ahead, with his knees at right angles. (See Figure 11–4.) The data are given in Tables 11–15 and 11–16.

FIGURE 11-3. Stature.

TABLE 11-11. STATURE: NUDE U.S. MALE MILITARY PERSONNEL

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
Air Force personnel*	63.5	65.2	69.1	73.1	74.9	2.44
Pilots, multi-engine	64.4	65.9	69.4	73.3	74.9	2.31
Pilots, fighter	63.8	65.2	68.8	72.6	74.2	2.24
Cadets	63.6	65.2	69.2	73.1	74.7	2.45
Bombardiers	63.5	65.2	69.1	73.0	74.5	2.32
Navigators	63.5	65.2	69.2	73.3	75.0	2.46
Observers	63.8	65.4	69.1	72.8	74.2	2.44
Flight engineers	63.1	64.8	69.0	73.2	75.0	2.51
Gunners	62.4	64.2	68.3	72.2	73.7	2.43
Radio operators	63.0	64.6	68.3	71.8	73.2	2.37
Basic trainees†	62.5	64.2	68.6	72.7	74.7	2.61
Army personnel:						
Inductees less than 20 years old‡	62.4	64.3	68.7	73.1	74.9	2.66
Inductees more than 20 years old‡	62.7	64.6	69.0	73.4	75.2	2.65
Separatees, white§	62.7	64.3	68.5	72.6	74.5	2.52
Separatees, Negro¶	62.3	64.0	68.0	72.2	74.0	2.58
Marine Corps personnel**	64.4	66.1	69.7	73.5	74.5	2.18
Recruits††	63.0	64.6	68.6	72.5	74.1	2.40
Navy personnel‡‡	64.1	65.7	69.7	73.5	75.1	2.34
Recruits, 18 years old††	62.8	64.5	68.5	72.6	74.2	2.50
Recruits, 17-25 years old§§	62.9	64.6	68.6	72.7	74.4	2.48
Enlisted men, general¶¶	63.2	64.8	69.5	73.5	75.5	2.48
Pilots, aircraft***	65.1	66.2	69.9	73.9	75.3	2.33
Cadets, aviation¶¶	65.1	66.6	70.1	73.8	75.2	----
Army aviators†††	64.4	65.8	69.4	73.3	74.8	----

*Hertzberg et al. (1954) (except as noted).
†Daniels et al. (1953b).
‡Damon (1957).
§Newman and White (1951).
¶ USA (1946).
** USMC (1949).
††USN (1949b).
‡‡King et al. (1947).
§§Gibbons et al. (1953).
¶¶USN (1955) (unpublished data).
***Gifford et al. (1965).
†††White (1961).

TABLE 11–12. STATURE: NUDE U.S. FEMALE MILITARY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel:						
WAF basic trainees*	59.3	60.3	64.0	68.2	69.9	2.34
Pilots†	60.8	61.7	64.9	68.3	70.0	----
Flight nurses†	59.0	60.2	63.5	67.7	69.3	----
Army personnel‡	58.4	59.9	63.9	68.0	69.7	2.42
WAC enlisted women§	58.3	60.0	63.9	68.0	69.6	2.40
WAC officers§	59.2	61.0	64.5	68.9	70.6	2.40
Nurses§	58.7	60.4	64.1	68.3	70.0	2.40

* Daniels et al. (1953a).
† Randall et al. (1946).
‡ Randall and Munro (1949).
§ Randall (1947).

TABLE 11–13. STATURE: NUDE MALE CIVILIAN SAMPLES (U.S., CANADA)

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Railroad travelers†	*62.5	*64.5	*69.0	*73.8	*75.6	----
Truck and bus drivers‡	63.0	64.6	68.4	72.5	74.1	----
Airline pilots§	64.4	66.0	70.0	73.9	75.6	2.40
Industrial workers¶	*64.4	*66.1	*70.3	*74.4	*76.2	2.46
College students:**	62.5	64.4	68.7	73.1	74.9	2.68
Eastern 18 years old††	64.5	66.1	69.9	73.8	75.4	2.38
Eastern 19 years old‡‡	65.0	66.5	70.2	74.0	75.5	2.30
Midwest, 18 years old§§	63.2	65.0	69.1	73.3	75.0	2.60
Midwest, 18–22 years old¶¶	64.2	65.9	70.0	74.1	75.8	2.49
Draft registrants***						
18–19 years old	62.0	63.8	68.0	72.3	74.1	2.61
20–24 years old	62.1	63.9	68.2	72.4	74.2	2.60
25–29 years old	61.9	63.7	68.1	72.4	74.2	2.63
30–34 years old	61.7	63.5	67.8	72.1	73.9	2.66
35–37 years old	61.3	63.2	67.6	72.0	73.8	2.64
Civilian men†††	61.7	63.6	68.3	72.8	74.6	----
Canadians‡‡‡						
18–19 years old	62.4	64.1	68.2	72.1	73.7	----
20–24 years old	62.0	63.8	68.3	72.5	74.3	----
25–29 years old	60.6	62.9	68.3	74.0	76.2	----
30–34 years old	61.5	63.4	68.1	72.8	74.8	----
35–44 years old	60.5	62.7	67.6	72.6	74.7	----
45–54 years old	59.7	61.8	66.8	72.0	74.1	----
55–64 years old	58.4	60.6	66.0	71.3	73.6	----
More than 64 years old	58.6	60.6	65.1	69.8	71.8	----

*Including shoes (subtract 1 in. for nude height).
†Hooton (1945).
‡McFarland et al. (1958).
§McCormick (1947).
¶Tyroler (1958).
**Diehl (1933a).
††Bowles (1932).
‡‡Heath (1945).
§§Damon (1955).
¶¶Elbel (1954).
***Karpinos (1958).
†††Stoudt et al. (1965).
‡‡‡Pett and Ogilvie (1957).

TABLE 11-14. STATURE: NUDE FEMALE CIVILIAN SAMPLES (U.S., CANADA)

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Railroad travelers†	59.1	60.8	64.9	69.1	70.8	----
Working women‡	58.1	59.7	63.6	67.5	69.2	2.43
City women:						
Eastern, white	56.5	58.4	62.8	67.3	69.2	2.69
Eastern, Negro	57.3	59.0	63.1	67.3	69.0	2.53
College students§	58.5	60.0	63.8	67.6	69.2	2.33
Eastern¶	59.8	61.2	64.8	68.3	69.7	2.15
Midwest**	58.8	60.5	64.4	68.4	70.0	2.36
Canadians††						
18–19 years old	57.4	59.0	62.7	66.3	67.8	----
20–24 years old	57.2	58.8	62.9	66.8	68.4	----
25–29 years old	56.9	58.5	62.6	66.5	68.2	----
30–34 years old	57.0	58.7	62.6	66.5	68.2	----
35–44 years old	56.6	58.2	62.4	66.4	68.0	----
45–54 years old	56.6	58.1	61.8	65.4	67.0	----
55–64 years old	55.9	57.4	61.0	64.9	66.4	----
More than 64 years old	54.8	56.4	60.6	64.7	66.4	----
General‡‡	57.4	59.1	63.2	67.2	68.8	2.48
Civilian women§§	57.1	59.0	62.9	67.1	68.8	----

*Including shoes (subtract 2 in. for nude height).
†Hooton (1945).
‡Bayer and Gray (1934).
§Diehl (1933b).
¶Bowles (1932).
**Donelson et al. (1940).
††Pett and Ogilvie (1957).
‡‡ O'Brien and Shelton (1941).
§§Stoudt et al. (1965).

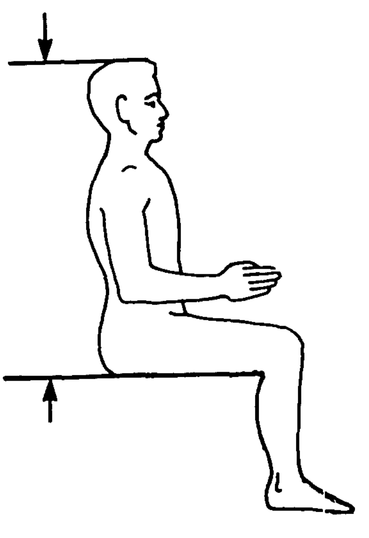


FIGURE 11-4. Sitting height.

TABLE 11–15. SITTING HEIGHT: NUDE MALE POPULATION SAMPLES

Population	Group	N	1 cm	1 in.	5 cm	5 in.	Means cm	Means in.	95 cm	95 in.	99 cm	99 in.	SD cm	SD in.
MILITARY														
Canada [1]	RCAF pilots	314	*84.8	*33.3	*85.8	*33.8	*91.3	*36.0	*96.3	*37.9	*99.2	*39.1	*3.68	*1.30
Greece [2]	Mixed military	1084	83.3	32.8	85.4	33.6	90.3	35.6	95.2	37.5	97.0	38.2	3.01	1.19
Great Britain [3]	RAF pilots	4357	32.8	33.4	87.1	34.3	92.7	36.5	98.0	38.6	100.2	39.5	3.3	1.30
Italy [2]	Mixed military	1358	82.0	32.3	84.2	33.2	89.7	35.3	94.8	37.3	97.1	38.2	3.2	1.26
Japan [4]	JASDF pilots	239	84.5	33.2	86.2	33.9	90.7	35.7	95.3	37.5	97.5	38.4	2.62	1.03
South Korea [5]	ROKAF pilots	264	83.3	32.8	86.1	33.9	90.7	35.7	95.3	37.5	98.2	38.6	2.81	1.11
**Norway [6]	Recruits 20-year	5765	84.0	33.2	86.5	34.2	92.7	36.5	98.8	38.9	101.3	39.9	3.47	1.37
Thailand [7]	Mixed military	2950	79.5	31.3	81.5	32.1	86.5	34.1	91.5	36.0	93.5	36.8	3.1	1.22
Turkey [2]	Mixed military	915	83.1	32.7	84.8	33.4	89.7	35.3	95.1	37.4	97.3	38.3	3.15	1.24
USA:														
Air Force—														
'50 [8]	Flyers	4063	83.5	32.9	85.8	33.8	91.4	36.0	96.6	38.0	98.9	38.9	3.27	1.29
Army** [9]	Separates	24449	82.5	32.5	85.1	33.5	90.9	35.8	97.0	38.2	97.0	39.2	3.4	1.34
Army [10]	Pilots	500	82.5	32.5	85.1	33.5	90.4	35.6	95.7	37.7	98.3	38.7	3.22	1.27
Navy—														
'64 [11]	Pilots	1549	85.0	33.4	87.0	34.2	92.1	36.3	97.4	38.4	100.0	39.4	3.16	1.25
Vietnam [12]	Mixed military	2129	77.5	30.5	79.6	31.3	85.0	33.5	90.5	35.6	92.5	36.4	3.3	1.30
CIVILIAN														
**Bulgaria [13]	26-year	114	80.5	31.7	83.1	32.7	89.4	35.2	96.0	37.8	98.5	38.8	3.68	1.45
**Ceylon [14]	Sinhalese	635	73.2	28.8	75.7	29.8	81.5	32.1	87.6	34.5	90.2	35.5	3.45	1.36
**China (Taiwan) [15]	Young men	31	83.7	32.8	85.3	33.6	90.4	35.6	95.5	37.6	97.7	38.5	2.92	1.15
**India [16]	Ages 19–60	499	77.6	30.6	80.0	31.5	85.1	33.5	90.4	35.6	92.7	36.5	3.04	1.19
**Israel [17]	Ages 40–44	3339	80.5	31.7	84.3	33.2	89.2	35.1	94.3	37.1	96.3	37.9	3.7	1.46
USA														
FAA [18]	Tower trainees (21–50)	678	83.8**	33.0**	86.2	33.9	92.0	36.2	97.5	38.4	100.1**	39.4**	3.27	1.29
Drivers [19]	Truck, bus	360	83.5	32.8	86.1	33.9	92.7	36.5	99.5	39.2	101.5	40.2	3.99	1.57
Nat'l. Health Survey [20]	Total males (18–79)	3091	81.0	31.9	84.3	33.2	90.6	35.7	96.4	38.0	98.8	38.9	----	----

* As taken by USAF technique.
** Plotted from published means and standard deviations.
[1] RCAF (1963).
[2] Hertzberg et al. (1963).
[3] Morant (1951).
[4] Oshima et al. (1962).
[5] Kay (1961).
[6] Udjus (1964).
[7] White (1964a.)
[8] Hertzberg et al. (1954).
[9] Newman and White (1951).
[10] White (1961).
[11] Gifford et al. (1965).
[12] White (1964b).
[13] Bulgarian Academy of Sciences (1965).
[14] Stoudt and Marett (1961).
[15] Chen et al. (1963.)
[16] Sen (1964).
[17] Lippert (1965).
[18] Snow and Snyder (1965).
[19] McFarland et al. (1958).
[20] Stoudt et al. (1965).

TABLE 11–16. SITTING HEIGHT: U.S. NUDE FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel*						
Pilots	31.8	32.4	34.1	35.8	36.3	----
Flight nurses	31.1	31.9	33.7	35.7	36.6	----
Working women†	30.9	31.7	33.7	35.7	36.5	1.15
College students:						
Eastern‡	31.6	32.4	34.2	36.0	36.7	1.10
Southern§	31.0	31.7	33.6	35.4	36.2	1.06
Civilian women¶	29.5	30.9	33.4	35.7	36.6	----

*Randall et al. (1946).
†Bayer and Gray (1934).
‡Bowles (1932).
§Carter (1932).
¶Stoudt et al. (1965).

For American women, subtract 2.2 in. from the male values. Add 0.2 to 0.3 in. for heavy clothing under the buttocks, add roughly 1.0 in. for civilian caps, 1.4 in. for steel helmets, and up to 3.5 in. for flying helmets.

11.4.3 Weight

The data are given in Tables 11–17, 11–18, 11–19, 11–20, 11–21.

For U.S. women, subtract 25–35 lb. from the male values. Add 5 lb. for men's light clothing, 3.5 lb. for women's light clothing, and up to 23 lb. or more for military winter clothing. (See also Tables 11–9 and 11–10.)

11.4.4 Maximum Body Depth

For this, dimension, using a lateral photograph of the subject, measure the maximum horizontal distance between the vertical lines tangent to the most anterior and posterior points on the trunk. The anterior points are on the chest or abdomen; the posterior points are in the shoulder or buttock region. The subject stands erect with his arms at his sides. (See Figure 11–5.) The data are given in Table 11–22.

11.4.5 Maximum Body Breadth

For this dimension, measure the maximum breadth of the body, including arms, as the subject stands erect with his arms hanging relaxed at his sides. (See Figure 11–6.) The data are given in Table 11–22.

11.4.6 Head Length

This is the dimension between the most anterior point on the head (between the brow ridges) and the most posterior point, wherever found (usually in the mid-sagittal plane; see Figure 11–7). The data are given in Tables 11–23, 11–24, and 11–25.

For U.S. women, subtract 0.5 in. from the male values. Clothing adds varying amounts, depending on the type of headgear; flying helmets add about 3.5 in.

11.4.7 Head Breadth

This dimension is the maximum horizontal head breadth above the ears. The location of this dimension is highly variable. The data are given in Tables 11–26 and 11–27.

For U.S. women, subtract 0.3 in. from the male values. Clothing increments vary, depending on the type of headgear; add 3.5 in. for steel helmets and 4.3 in. or more for flying helmets. (See Figure 11–8.)

11.4.8 Interpupillary Distance

This is the horizontal distance between the centers of the pupils when the subject looks straight ahead. (See Figure 11–9.) The data are given in Table 11–28.

11.4.9 Eye Height

This is the vertical distance from the floor to the inner corner of the eye. The subject stands erect and looks straight ahead. (See Figure 11–10.) The data are given in Table 11–28.

TABLE 11–17. WEIGHT: WORLD POPULATION SAMPLES (NUDE MALES)

Population	Group	N	Percentiles 1 kg	1 lb	5 kg	5 lb	Mean kg	Mean lb	95 kg	95 lb	99 kg	99 lb	SD kg	SD lb
MILITARY														
Belgium*[1]	Flying personnel	2450	49.1	108.0	54.2	119.1	66.5	146.2	78.5	172.7	83.4	183.5	6.9	15.2
Canada[2]	RCAF pilots	314	56.2	123.4	59.7	131.4	76.6	167.8	76.3	202.9	92.1	233.5	9.91	21.8
France*[3]	Pilots, cadets	1000	48.2	106.0	53.3	117.2	65.8	144.8	78.3	172.0	83.4	183.3	7.03	15.5
France*[4]	Recruits 20-year	234	47.2	103.9	52.4	115.1	65.4	143.8	77.8	171.3	83.0	182.5	7.17	15.8
West Germany*[5]	Recruits 20-year	378133	46.1	101.3	52.5	115.5	68.4	150.4	84.1	185.0	90.5	199.0	8.9	19.6
Greece[6]	Mixed military	1084	52.2	114.8	55.8	122.9	67.0	147.4	80.5	177.0	86.7	190.8	7.62	16.3
Great Britain*[7]	Bomber personnel	11772	48.6	106.9	53.7	118.1	66.1	145.1	78.8	173.3	83.9	184.5	7.06	15.5
Great Britain*[8]	RAF pilots	4357	49.0	107.7	55.8	122.8	72.1	158.6	88.6	195.0	95.4	209.9	9.27	20.4
Italy[6]	Mixed military	1358	53.9	118.7	57.6	126.8	70.3	154.7	85.0	187.0	93.1	204.8	8.43	18.6
Japan[9]	JASDF pilots	239	50.3	110.7	52.5	115.5	61.1	133.2	71.8	158.0	79.9	175.8	5.86	12.9
South Korea[10]	ROKAF pilots	264	50.2	110.3	53.1	116.8	62.9	138.2	76.5	168.5	82.5	181.7	6.53	14.4
Norway*[11]	Recruits 20-year	5765	51.5	113.2	57.0	125.3	70.1	154.2	83.3	183.2	88.8	195.6	7.46	16.4
Thailand[12]	Mixed military	2950	45.5	100.1	48.0	105.5	56.3	123.8	67.0	147.3	73.0	160.6	10.3	22.7
Turkey[6]	Mixed military	915	48.2	106.0	52.5	115.5	64.6	142.1	79.2	174.1	87.6	192.7	8.23	18.1
USA:														
Air Force—														
—'50[13]	Flyers	4063	56.0	123.1	60.3	132.5	74.4	163.7	94.5	200.8	98.1	215.9	9.27	20.9
Army*[14]	Separatees	24449	47.1	103.5	54.0	118.8	70.3	154.8	87.0	191.3	93.8	203.2	9.33	20.5
Army[15]	Pilots	500	56.2	123.6	61.8	135.9	75.4	165.8	89.8	199.7	96.6	212.5	8.59	18.9
Navy— '64[16]	Pilots	1549	58.7	129.3	63.7	140.3	77.7	171.4	92.3	203.6	100.2	220.9	8.66	19.1
Vietnam[17]	Mixed military	2129	39.8	87.5	42.4	93.1	51.1	112.4	61.5	135.2	70.0	154.0	6.0	13.2
CIVILIAN														
Africa*[18]	Dinka Nilotes	279	42.5	93.5	47.7	105.0	58.3	128.3	68.8	151.3	73.2	161.0	5.92	13.0
Bulgaria*[19]	Civilians age 26	114	47.3	104.1	53.3	117.3	67.7	148.9	82.0	180.4	88.0	193.5	8.13	17.9
China (Taiwan)*[20]	Young men	31	42.0	92.4	45.1	97.0	52.6	115.7	60.7	133.5	63.8	140.5	4.44	9.77
India*[21]	19–60	499	30.7	67.5	36.7	80.7	50.9	123.0	65.0	143.0	71.0	156.1	8.00	17.6
Israel*[22]	40 years	3339	44.9	98.8	52.9	116.3	72.2	158.8	91.8	202.0	99.8	208.3	11.0	24.2
USA: FAA[23]	Tower trainees (21–50)	678	†48.5	*106.6	58.0	127.9	73.4	161.8	90.3	199.0	*98.7	*217.0	10.0	22.1
National Health Survey[24]	Total males (18–79)	3091	50.9	112.0	57.3	126.0	76.3	168.0	98.6	217.0	109.5	241.0	----	----

*Estimated from graphs plotted from published means and standard deviations.
[1] Evrard (1954.)
[2] RCAF (1963).
[3] Ducros (1955).
[4] Monod et Pineau (1958).
[5] Institut der Wehrmedicinalstatistic (1964).
[6] Hertzberg et al. (1963).
[7] Samuel and Smith (after Morant; 1965).
[8] Morant (1951).
[9] Oshima et al. (1962).
[10] Kay (1961).
[11] Udjus (1965).
[12] White (1964a).
[13] Hertzberg et al. (1954).
[14] Newman and White (1951).
[15] White (1961).
[16] Gifford et al. (1965).
[17] White (1964b).
[18] Roberts and Bainbridge (1963).
[19] Bulgarian Academy of Sciences (1965).
[20] Chen et al. (1963).
[21] Sen (1964).
[22] Lippert (1965).
[23] Snow and Snyder (1965).
[24] Stoudt et al. (1965).

TABLE 11–18. WEIGHT: NUDE MALE MILITARY SAMPLES

Population	Percentiles (lb)					
	1st	5th	50th	95th	99th	S.D.
Air Force personnel†	123	133	162	201	216	20.9
Pilots, multi-engine	123	----	166	----	217	20.5
Pilots, fighter	123	----	159	----	225	20.7
Cadets	123	----	159	----	199	17.4
Bombardiers	126	----	169	----	211	20.6
Navigators	125	----	165	----	214	20.6
Observers	113	----	166	----	217	22.4
Flight engineers	124	----	166	----	222	23.3
Gunners	121	----	158	----	214	21.3
Radio operators	115	----	157	----	199	19.0
Basic trainees‡	109	118	145	186	208	21.0
Army personnel						
Inductees less than 20 years old§	----	(111)	159	(206)	----	29.4
Inductees more than 20 years old§	----	(122)	162	(202)	----	23.9
Separatees, white¶	114	124	153	192	215	20.6
Separatees, Negro**	----	(120)	152	(183)	----	19.2
Marine Corps personnel††	130	139	170	212	228	22.4
Recruits‡‡	----	(112)	143	(174)	----	18.7
Navy personnel§§	----	(126)	162	(197)	----	21.5
Recruits, 18 yr old‡‡	----	(110)	140	(171)	----	18.5
Recruits, 17–25 yr old¶¶	----	(119)	152	(185)	----	20.6
Enlisted men, general***	----	132	160	197	----	19.9
Pilots, aircraft†††	129	140	171	203	221	19.1
Cadets, aviation***	----	135	166	196	----	----
Army aviators‡‡‡	124	136	167	200	213	----

*Percentiles in parentheses were computed from the 50th percentile using the S.D. Because of the skewed distribution of weight, these values might differ somewhat from the true values and should be used with caution.
†Hertzberg et al. (1954).
‡Daniels et al. (1953b).
§Damon (1957).
¶Newman and White (1951).
**USA (1946).
††USMC (1949).
‡‡USN (1949a).
§§King et al. (1947).
¶¶Gibbons et al. (1953).
***USN (1955).
†††Gifford et al. (1965).
‡‡‡White (1961).

TABLE 11–19. WEIGHT: U.S. NUDE FEMALE MILITARY SAMPLES

Population	Percentiles* (lb)					
	1st	5th	50th	95th	99th	S.D.
Air Force personnel:						
WAF basic trainees†	95	102	122	148	162	14.5
Pilots‡	102	106	129	155	169	----
Flight nurses‡	104	107	122	135	143	----
Army personnel§	97	105	129	170	192	20.0
WAC enlisted women¶	----	(97)	130	(163)	----	20.6
WAC officers¶	----	(105)	132	(158)	----	16.1
Nurses¶	----	(95)	129	(162)	----	20.2

*Percentiles in parentheses were computed from the 50th percentile using the S.D. Because of the skewed distribution of weight, these values might differ somewhat from the true values and should be used with caution.
†Daniels et al. (1953a).
‡Randall et al. (1946).
§Randall and Munro (1949).
¶Randall (1947).

TABLE 11-20. WEIGHT: NUDE FEMALE CIVILIAN SAMPLES (U.S., CANADA)

Population	Percentiles* (lb)					S.D.
	1st	5th	50th	95th	99th	
Railroad travelers†	----	104‡	133‡	179‡	----	----
Working women§	----	(110)	136	(163)	----	16.0
City women:						
Eastern, white	95	108	140	200	229	27.2
Eastern, Negro	85	104	143	193	210	34.5
College students¶	----	(94)	121	(149)	----	17.1
Eastern**	----	(101)	125	(149)	----	15.2
Midwest††	----	(99)	126	(154)	----	16.9
Canadians‡‡						
18-19 years old	----	----	120	----	----	----
20-24 years old	----	----	122	----	----	----
25-29 years old	----	----	123	----	----	----
30-34 years old	----	----	126	----	----	----
35-44 years old	----	----	132	----	----	----
45-54 years old	----	----	142	----	----	----
55-64 years old	----	----	145	----	----	----
More than 64 years old	----	----	136	----	----	----
General§§	91	100	129	184	213	26.0
Civilian women¶¶	93	104	137	199	236	----

*Percentiles in parentheses were computed from the 50th percentile using the S.D. Because of the skewed distribution of weight, these values might differ somewhat from the true values and should be used with caution.
†Hooton (1945).
‡Including shoes and indoor clothing (subtract 3 or 4 lb. for nude weight).
§Bayer and Gray (1934).
¶Diehl (1933b).
**Bowles (1932).
††Donelson et al. (1940).
‡‡Pett and Ogilvie (1957).
§§O'Brien and Shelton (1941).
¶¶Stoudt et al. (1965).

TABLE 11-21. WEIGHT: NUDE MALE CIVILIAN SAMPLES (U.S., CANADA)

Population	Percentiles* (lb)					S.D.
	1st	5th	50th	95th	99th	
Railroad travelers [1]	----	**132	**167	**218	----	----
Truck and bus drivers [2]	----	129	164	213	247	----
Airline pilots [3]	----	(134)	168	(201)	----	20.3
Industrial workers [4]	----	**(130)	**170	**(210)	----	24.5
College students [5]	----	(112)	142	(172)	----	18.1
Eastern, 18 years old [6]	----	(122)	150	(178)	----	17.2
Eastern, 19 years old [7]	----	(132)	159	(187)	----	16.2
Midwest, 18 years old [8]	----	(115)	148	(180)	----	19.7
Midwest, 18-22 years old [9]	----	(118)	156	(195)	----	23.5
Draft registrants: [10]						
18-19 years old	----	(106)	141	(176)	----	21.1
20-24 years old	----	(109)	146	(183)	----	22.4
25-29 years old	----	(110)	151	(192)	----	24.8
30-34 years old	----	(110)	153	(195)	----	25.8
35-37 years old	----	(111)	154	(197)	----	26.1
Civilian men [11]	112	126	166	217	241	----
Canadians [12]						
18-19 years old	----	----	140	----	----	----
20-24 years old	----	----	151	----	----	----
25-29 years old	----	----	157	----	----	----
30-34 years old	----	----	168	----	----	----
35-44 years old	----	----	165	----	----	----
45-54 years old	----	----	161	----	----	----
55-64 years old	----	----	159	----	----	----
More than 64 years old	----	----	156	----	----	----

*Percentiles in parentheses were computed from the 50th percentile using the S.D. Because of the skewed distribution of weight, these values might differ somewhat from the true values and should be used with caution.
**Including shoes and indoor clothing (subtract 5 or 6 lb. for nude weight).
[1] Hooton (1945).
[2] McFarland et al. (1958).
[3] McCormick (1947).
[4] Tyroler (1958).
[5] Diehl (1933a).
[6] Bowles (1932).
[7] Heath (1945).
[8] Damon (1955).
[9] Elbel (1954).
[10] Karpinos (1958).
[11] Stoudt et al. (1965).
[12] Pett and Ogilvie (1957).

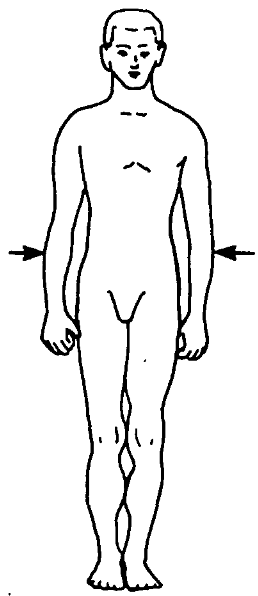


FIGURE 11-5. Maximum body depth.

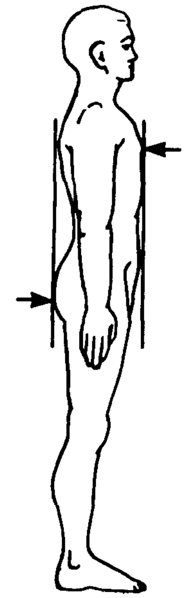


FIGURE 11-6. Maximum body breadth.

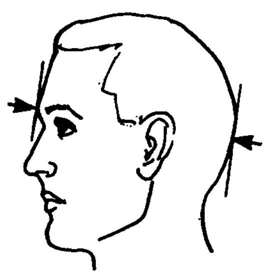


FIGURE 11-7. Head length.

TABLE 11-22. MAXIMUM BODY DEPTH AND BREADTH OF MALE U.S. AIR FORCE PERSONNEL AND COLLEGE STUDENTS

Dimension	Percentiles (in.)			S.D.
	5th	50th	95th	
Body depth	10.1	11.5	13.0	0.88
Body breadth	18.8	20.9	22.8	1.19

Hertzberg et al. (1956).

TABLE 11-23. HEAD LENGTH OF FOREIGN MALE MILITARY SAMPLES

Population	Personnel	Percentiles (in.)						S.D.
		N	1	5	50	95	99	
Canada [1]	Pilots	314	6.9	7.05	7.6	8.15	8.43	0.31
Greece [2]	Mixed military	1084	6.85	7.0	7.45	7.9	8.04	.27
Italy [2]	Mixed military	1358	7.0	7.16	7.6	8.04	8.2	.26
Japan [3]	JASDF pilots	236	6.58	6.78	7.37	7.84	7.95	.31
South Korea [4]	ROKAF pilots	264	6.65	6.85	7.25	7.88	7.8	.26
Thailand [5]	Mixed military	2950	6.42	6.63	7.0	7.45	7.65	.24
Turkey [2]	Mixed military	915	6.7	6.9	7.33	7.75	7.8	.26
Vietnam [6]	Mixed military	2129	6.5	6.7	7.23	7.56	7.75	.28

[1] RCAF Anthropometric Survey (1963).
[2] Hertzberg et al. (1963).
[3] Oshima et al. (1962).
[4] Kay (1961).
[5] White (1964a).
[6] White (1964b).

TABLE 11–24. HEAD LENGTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	7.2	7.3	7.7	8.2	8.3	0.25
Cadets [2]	7.2	7.4	7.8	8.2	8.4	.26
Basic trainees [3]	7.0	7.2	7.6	8.1	8.3	.28
Army separatees [4]	7.0	7.2	7.7	8.1	8.3	.28
Army aviators [5]	7.2	7.3	7.8	8.2	8.5	.27
College students:						
Eastern [6]	7.1	7.3	7.7	8.2	8.4	.27
Midwest [6]	7.1	7.3	7.7	8.2	8.4	.27
Naval aviators [7]	7.2	7.4	7.8	8.2	8.4	.26

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Newman and White (1951).
[5] White (1961).
[6] Damon (1955).
[7] Gifford et al. (1965).

TABLE 11–25. HEAD LENGTH OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES (U.S.)

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
WAF basic trainees [1]	6.1	6.4	6.9	7.3	7.5	0.30
WAC personnel and Army nurses [2]	6.7	6.8	7.2	7.7	7.8	.26
Working women [3]	6.8	7.0	7.4	7.7	7.9	.23
College students:						
Eastern [4]	6.8	7.0	7.4	7.7	7.9	.24
Southern [4]	6.8	7.0	7.4	7.8	7.9	.23

[1] Daniels et al. (1953a).
[2] Randall and Munro (1949).
[3] Bayer and Gray (1934).
[4] Carter (1932).

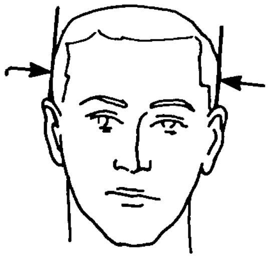


FIGURE 11-8. Head breadth.

TABLE 11-26. HEAD BREADTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	5.6	5.7	6.1	6.4	6.6	0.20
Cadets [2]	5.6	5.7	6.1	6.4	6.6	.21
Basic trainees [3]	5.4	5.6	5.9	6.3	6.5	.23
Army separatees [4]	5.4	5.6	6.0	6.4	6.6	.23
Army aviators [5]	5.6	5.7	6.1	6.5	6.8	.21
College students:						
Eastern [6]	5.5	5.7	6.0	6.4	6.5	.22
Midwest [6]	5.6	5.8	6.1	6.5	6.6	.20
Naval aviators [7]	5.6	5.8	6.1	6.5	6.6	.21

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Newman and White (1951).
[5] White (1961).
[6] Bowles (1932).
[7] Gifford et al. (1965).

TABLE 11-27. HEAD BREADTH OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
WAF basic trainees [1]	5.3	5.4	5.7	6.1	6.2	0.20
WAC personnel and Army nurses [2]	5.2	5.4	5.7	6.1	6.2	.22
Working women [3]	5.5	5.6	5.9	6.1	6.3	.17
College students:						
Eastern [4]	5.4	5.5	5.8	6.2	6.3	.20
Southern [4]	5.4	5.5	5.8	6.1	6.2	.18

[1] Daniels (1953a).
[2] Randall and Munro (1949).
[3] Bayer and Gray (1934).
[4] Carter (1932).

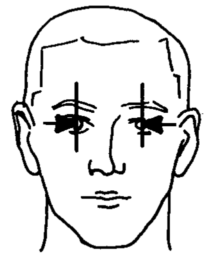


FIGURE 11–9. Interpupillary distance.

TABLE 11–28. INTERPUPILLARY DISTANCE AND EYE HEIGHT OF MALE USAF PERSONNEL

Dimension	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Interpupillary distance	2.19	2.27	2.49	2.74	2.84	0.14
Eye height	59.2	60.8	64.7	68.6	70.3	2.38

Hertzberg et al. (1954).

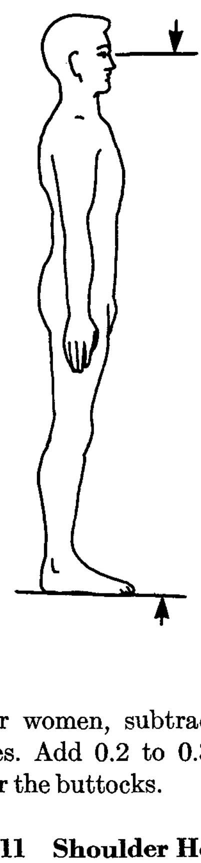


FIGURE 11–10. Eye height.

For U.S. women, subtract 4.5 in. from the male values. For clothing, add 1.0 in. for men's shoes, 1.3 in. for military boots, and up to 3.0 in. for women's shoes.

11.4.10 Eye Height, Sitting

This is the vertical distance from the sitting surface to the inner corner of the eye. The subject sits erect and looks straight ahead. (See Figure 11–11.) The data are given in Table 11–29.

For women, subtract 2.0 in. from the male values. Add 0.2 to 0.3 in. for heavy clothing under the buttocks.

11.4.11 Shoulder Height

This is the vertical distance from the floor to the uppermost point on the lateral edge of the shoulder with the subject standing erect. (See Figure 11–12.) The data are given in Table 11–30.

For U.S. women, subtract 4.0 in. from the

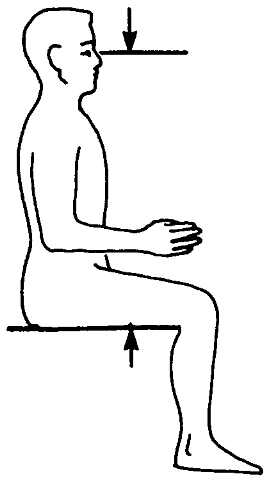


FIGURE 11–11. Eye height, sitting.

TABLE 11–29. EYE HEIGHT, SITTING, OF MALE AND FEMALE U.S. MILITARY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Male flight personnel [1]	28.5	29.4	31.5	33.5	34.4	1.27
Female pilots [2]	27.9	28.5	30.0	31.6	32.4	----
Female flight nurses [2]	26.3	27.3	29.3	31.1	32.2	----
Army aviators [3]	28.1	28.8	30.9	33.1	34.5	1.28
Naval aviators [4]	28.8	29.7	31.5	33.6	34.5	1.18

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] White (1961).
[4] Gifford et al. (1965).

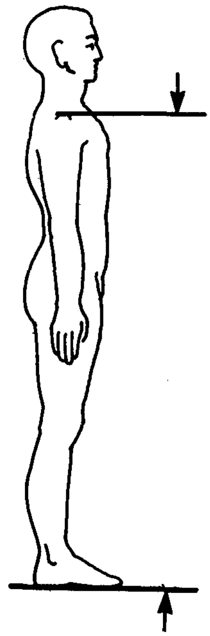


FIGURE 11–12. Shoulder height.

TABLE 11–30. SHOULDER HEIGHT OF MALE AND FEMALE USAF PERSONNEL

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Male flight personnel [1]	51.2	52.8	56.6	60.2	61.9	2.28
Male basic trainees [2]	50.3	52.0	55.9	59.9	61.8	2.41
Female basic trainees [3]	46.9	48.2	51.9	55.4	57.3	2.18

[1] Hertzberg et al. (1954).
[2] Daniels et al. (1953a).
[3] Daniels et al. (1953b).

male values. For clothing, add 1.0 in. for men's shoes, 1.3 in. for military boots, up to 3.0 for women's shoes, 0.2 in. for light clothing, and 0.9 in. or more for heavy clothing.

11.4.12 Shoulder Height, Sitting

This is the vertical distance from the sitting surface to the uppermost point on the lateral edge of the shoulder with the subject sitting erect. (See Figure 11–13.) The data are given in Table 11–31. Another 1.5 in. should be added for vertical distance from the highest point between the shoulder and neck—the more inclusive dimension. For women, subtract 2.0 in. from the male values. For light clothing, add 0.2 in. and, for heavy clothing, about 1.0 in.

11.4.13 Shoulder-Elbow Length

This is the vertical distance from the uppermost point on the lateral edge of the shoulder to the bottom of the elbow. The subject sits erect with his upper arm vertical at his side and the forearm making a right angle with it. (See Figure 11–14.) The data are given in Tables 11–32 and 11–33.

Another 1.5 in. should be added for vertical distance from the point of measurement to the highest point between the shoulder and neck—the more inclusive dimension.

For U.S. women, subtract 1.0 in. from the male values. Add 0.2 in. for light clothing and 1.0 in. or more for heavy clothing.

11.4.14 Arm Reach

This is the horizontal distance from the posterior surface of the right shoulder to the tip of the extended middle finger. The subject stands erect with heels, buttocks, and shoulders against the wall and the right arm and hand extended forward horizontally to their maximum length. (See Fig. 11–15.) The data are given in Tables 11–34 and 11–35.

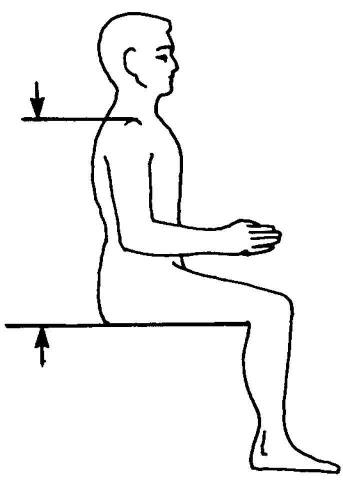


FIGURE 11–13. Shoulder height, sitting.

TABLE 11–31. SHOULDER HEIGHT, SITTING, OF MALE AND FEMALE U.S. MILITARY PERSONNEL

Population	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
Male flight personnel [1]	20.6	21.3	23.3	25.1	25.8	1.14
Female pilots[2]	21.8	22.4	23.8	25.2	25.9	----
Female flight nurses [2]	20.4	21.1	23.1	24.8	25.9	----
Naval aviators [3]	21.5	22.0	23.8	25.5	26.4	1.06

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Gifford et al. (1965).

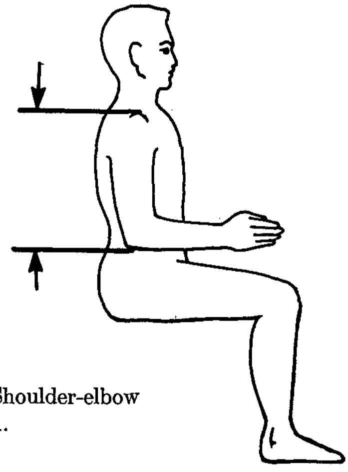


FIGURE 11–14. Shoulder-elbow length.

TABLE 11–32. SHOULDER-ELBOW LENGTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	12.8	13.2	14.3	15.4	15.9	0.69
Cadets [2]	13.2	13.6	14.7	15.8	16.3	----
Gunners [2]	12.9	13.3	14.5	15.6	16.1	----
Army personnel:						
Army aviators [3]	13.4	13.9	15.0	16.1	16.5	.70
Separatees, white [4]	12.3	12.9	14.3	15.6	16.3	.81
Separatees, Negro [5]	12.4	13.0	14.3	15.6	16.1	.80
Truck and bus drivers [6]	13.3	13.8	14.8	15.9	16.3	.81
College students [7]	12.8	13.3	14.5	15.7	16.1	.66
Naval aviators [8]	13.0	13.4	14.5	15.6	16.1	.67

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] White (1961).
[4] Newman and White (1951).
[5] USA (1946).
[6] McFarland et al. (1958).
[7] Bowles (1932).
[8] Gifford et al. (1965).

TABLE 11–33. SHOULDER-ELBOW LENGTH OF U.S. FEMALE MILITARY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel: [1]						
Pilots	12.3	12.7	13.7	14.7	15.2	----
Flight nurses	12.3	12.7	13.6	14.8	15.3	----
Army personnel [2]	11.3	11.9	13.1	14.3	14.9	0.74

[1] Randall et al. (1946).
[2] Randall and Munro (1949).

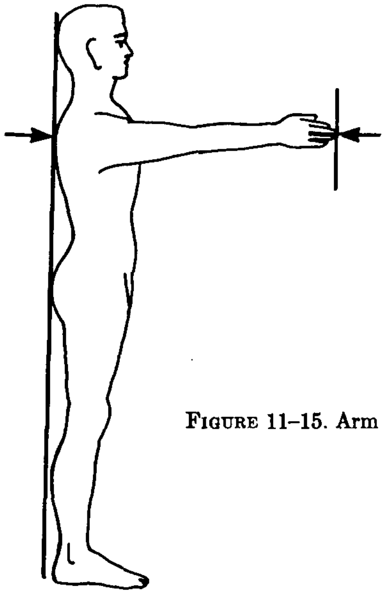


FIGURE 11-15. Arm reach.

TABLE 11-34. ARM REACH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	30.9	31.9	34.6	37.3	38.6	1.70
Cadets [2]	31.6	32.7	35.2	37.8	38.8	----
Gunners [2]	30.9	31.9	34.8	37.4	38.6	----
Navy personnel [3]	30.0	31.1	33.7	36.3	37.4	1.57
Enlisted men [4]	31.6	32.7	35.7	38.2	39.5	1.70
Cadets, aviation [4]	31.7	32.8	35.4	38.1	39.2	----
Truck and bus drivers [5]	31.9	32.9	35.7	38.4	39.5	----
Army aviators [6]	32.3	33.5	36.0	38.5	39.6	1.47

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] King et al. (1947).
[4] USN (1955).
[5] McFarland et al. (1958).
[6] White (1961).

TABLE 11-35. ARM REACH OF FEMALE USAF PERSONNEL

Population	Percentiles (in.) 1st	5th	50th	95th	99th
Pilots	29.2	29.7	31.8	34.1	34.9
Flight nurses	27.9	28.7	31.0	33.5	34.4

Randall et al. (1946).

TABLE 11-36. SHOULDER BREADTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	15.9	16.5	17.9	19.4	20.1	0.91
Cadets [2]	16.1	16.7	18.0	19.3	19.9	----
Gunners [2]	16.0	16.5	17.7	19.0	19.5	----
Army personnel:						
Aviators [9]	16.4	16.8	18.2	20.0	20.5	.88
Separatees, white [3]	15.8	16.4	17.9	19.6	20.6	.99
Separatees, Negro [4]	15.8	16.4	17.9	19.4	20.0	.89
Navy personnel [5]	15.1	15.8	17.6	19.4	20.2	1.09
Railroad travelers [6]	*15.7	*16.4	*17.6	*19.2	*19.8	----
Truck and bus drivers [7]	16.2	16.9	18.3	19.9	20.5	----
College students [8]	15.1	15.7	17.2	18.7	19.3	.86
Naval aviators [10]	16.6	17.3	18.8	20.3	20.9	.91

*Including light clothing (subtract 0.3 in. for nude dimension).
[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] USA (1946).
[5] King et al. (1947).
[6] Hooton (1945).
[7] McFarland et al. (1958).
[8] Bowles (1932).
[9] White (1961).
[10] Gifford et al. (1965).

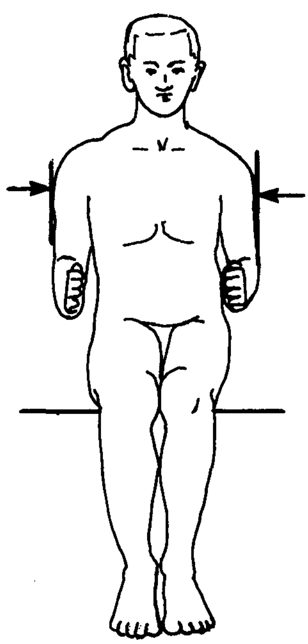


FIGURE 11-16. Shoulder breath.

For U.S. women, subtract 3.5 in. from the male values. Add 0.3 in. for light clothing, 0.2 in. for light gloves, and 0.3 in. for heavy clothing and gloves. For fingertip manipulation of controls subtract 0.5 in. for flip and 1.0 in. for push. For manipulation by the thumb and forefinger, subtract 3.0 in. (See data for Vertical Reach Height in Table 11-85.) For grasping by the whole hand, subtract 5.0 in. (See also the reach data as determined functionally in special measuring machines, Section 11.5.2.)

TABLE 11-37. SHOULDER BREADTH OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)				
	1st	5th	50th	95th	99th
Air Force personnel: [1]					
Pilots	14.3	14.9	16.1	17.6	18.0
Flight nurses	14.1	14.5	15.7	16.8	17.2
Railroad travelers [2]	*13.7	*14.4	*15.7	*17.6	*18.2

*Including light clothing (subtract 0.3 in. for nude dimension).
[1] Randall et al. (1946).
[2] Hooton (1945).

11.4.15 Shoulder Breadth

This dimension is the maximum horizontal distance across the deltoid muscles. The subject sits erect with his upper arms touching his sides and his forearms extended horizontally (see Figure 11-16). The data are given in Tables 11-36 and 11-37.

For U.S. women, subtract 2.0 in. from the tabular values. Add 0.3 in. for light clothing, 1.5 in. for heavy clothing, 0.4 in. for the partial-pressure suit uninflated, and 6.0 in. for the inflated unit.

11.4.16 Chest Depth

This is the horizontal distance from the front to the back of the chest at nipple level (on women, at the level where the 4th rib meets the breastbone). The subject stands erect and breathes normally. (See Figure 11-17.) The data are given in Tables 11-38 and 11-39.

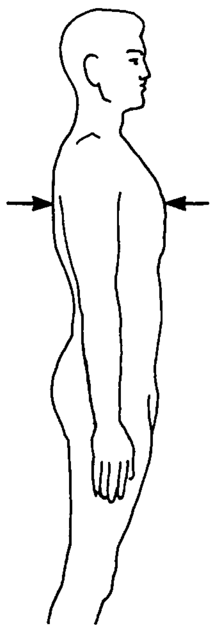


FIGURE 11–17. Chest depth.

TABLE 11–38. CHEST DEPTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	7.6	8.0	9.0	10.4	11.1	0.75
Cadets [2]	6.8	7.2	8.2	9.3	9.7	----
Gunners [2]	6.7	7.1	8.2	9.2	9.6	----
Army separatees [3]	6.7	7.2	8.3	9.6	10.5	.75
Aviators [4]	7.4	7.9	8.9	10.4	11.0	.79
Truck and bus drivers [5]	7.1	7.6	8.9	10.5	11.1	----
College students:						
Eastern [6]	6.5	6.9	7.9	8.9	9.3	.55
Midwest [7]	6.4	6.9	8.0	9.2	9.7	.71
Naval aviators [8]	7.8	8.3	9.4	10.6	11.1	.71

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] White (1961).
[5] McFarland et al. (1958).
[6] Heath (1945).
[7] Damon (1955).
[8] Gifford et al. (1965).

TABLE 11–39. CHEST DEPTH OF U.S. FEMALE COLLEGE STUDENTS

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Eastern [1]	5.8	6.3	7.4	8.6	9.0	0.68
Midwest [2]	6.0	6.4	7.3	8.2	8.6	.56

[1] Carter (1932).
[2] Donelson et al. (1940).

TABLE 11–40. CHEST BREADTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	10.4	10.8	12.0	13.4	14.1	0.80
Cadets [2]	9.8	10.3	11.3	12.4	12.8	----
Gunners [2]	9.7	10.1	11.1	12.1	12.5	----
Basic trainees [3]	9.7	10.2	11.4	13.0	14.3	.91
Army separatees [4]	9.3	10.0	11.1	12.4	13.2	.77
Truck and bus drivers [5]	9.6	10.2	11.8	13.5	13.9	----
College students:						
Eastern [6]	9.9	10.4	11.5	12.7	13.1	.67
Midwest [7]	9.3	9.9	11.1	12.4	12.9	.79

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Newman and White (1951).
[5] McFarland et al. (1958).
[6] Heath (1945).
[7] Damon (1955).

TABLE 11–41. CHEST BREADTH OF U.S. FEMALE AIR FORCE PERSONNEL AND COLLEGE STUDENTS

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
WAF basic trainees [1]	8.9	9.1	9.9	10.9	11.3	0.55
College students:						
Eastern [2]	8.3	8.7	9.7	10.7	11.1	.59
Midwest [3]	8.6	9.0	10.1	11.1	11.5	.64

[1] Daniels et al. (1953a).
[2] Carter (1932).
[3] Donelson et al. (1940).

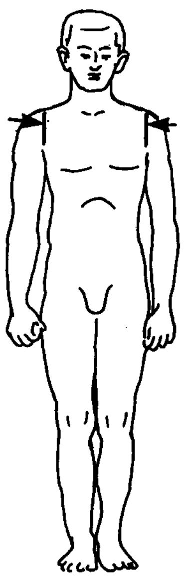


FIGURE 11–18. Chest breadth.

Add 0.5 in. for light clothing, 2 in. for heavy clothing, and 4.5 in. for inflated partial-pressure suits.

11.4.17 Chest Breadth

This is the horizontal lateral distance across the chest at nipple level (on women, at the level where the 4th rib meets the breastbone). The subject stands erect, breathes normally, and has his arms hanging naturally at his sides. (See Figure 11–18.) The data are given in Tables 11–40 and 11–41.

For U.S. women, subtract 1.5 in. from the above values. Add 0.3 in. for light clothing, 0.6 in. for heavy clothing, and 2.5 in. for inflated partial-pressure suits.

11.4.18 Waist Depth

This is the horizontal distance between the back and abdomen at the level of the greatest lateral indentation of the waist (if this is not apparent, at the level at which the belt is worn). The subject stands erect with his abdomen relaxed. (See Figure 11–19.) The data are given in Table 11–42.

Add 1.0 in. for light clothing, 2.5 in. for heavy clothing, and 5.0 in. for inflated partial-pressure suits. Add 0.1 in. for the sitting dimension.

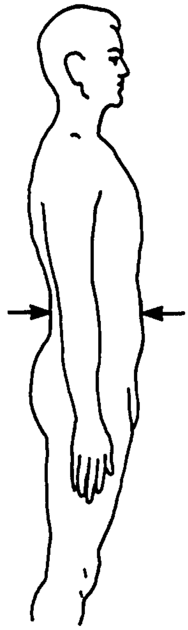


FIGURE 11–19. Waist depth.

TABLE 11–42. WAIST DEPTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	6.3	6.7	7.9	9.5	10.3	0.88
Cadets [2]	6.7	7.2	8.2	9.3	9.8	----
Gunners [2]	6.7	7.2	8.2	9.3	9.8	----
Army separatees [3]	7.5	7.9	9.0	10.5	11.5	.81
Truck and bus drivers [4]	7.3	7.9	9.5	12.1	13.1	----
Naval aviators [5]	6.9	7.3	8.5	9.8	10.5	.76

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] McFarland et al. (1958).
[5] Gifford et al. (1965).

11.4.19 Elbow Height

This is the vertical distance from the floor to the depression at the elbow formed where the bones of the upper arm and forearm meet. The subject stands erect with his arms hanging naturally at his sides. (See Figure 11–20.) The data are given in Table 11–43.

11.4.20 Elbow Height, Sitting

This is the vertical distance from the sitting surface to the bottom of the right elbow. The subject sits erect with his upper right arm vertical at his side and his forearm at a right angle to the upper arm. (See Figure 11–21.) The data are given in Table 11–44.

Informal tests show that these values (from Stoudt et al., 1965) can be reduced by about 1.5 in. for the relaxed sitting position ("slump").

For U.S. women, use the same values as given above. Clothing makes no difference because that under the buttocks is balanced by that under the elbow.

11.4.21 Forearm-Hand Length

This is the horizontal distance from the tip of the right elbow to the tip of the longest finger. The subject sits erect with his upper right arm vertical at his side and his forearm, hand, and fingers extended horizontally. (See Figure 11–22.) The data are in Table 11–45.

Add 0.2 in. for light clothing without gloves, 0.2 in. for light gloves, and 0.8–1.0 in. for heavy clothing and gloves.

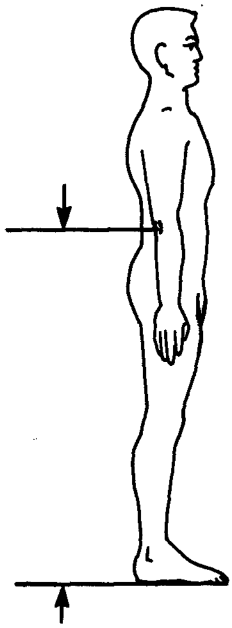


FIGURE 11-20. Elbow height.

TABLE 11-43. ELBOW HEIGHT OF MALE U.S. MILITARY PERSONNEL

Posture	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Standing, USAF [1]	39.5	40.6	43.5	46.4	47.7	1.77
Sitting, USAF [1]	6.6	7.4	9.1	10.8	11.5	1.04
Naval aviators [2]	7.0	7.6	9.3	10.9	11.7	.99

[1] Hertzberg et al. (1954). [2] Gifford et al. (1965).

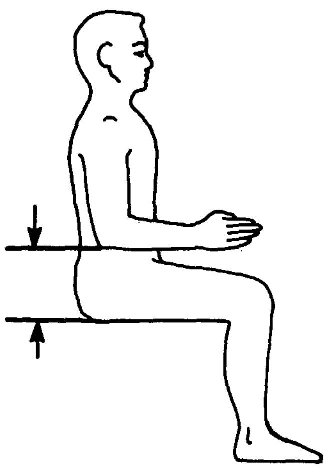


FIGURE 11-21. Elbow height, sitting.

TABLE 11-44. ELBOW HEIGHT, SITTING, OF U.S. CIVILIANS

Sex	Percentiles (in.) 1st	5th	50th	95th	99th
Male	6.3	7.4	9.5	11.6	12.5
Female	6.1	7.1	9.2	11.0	11.9

Stoudt et al. (1965).

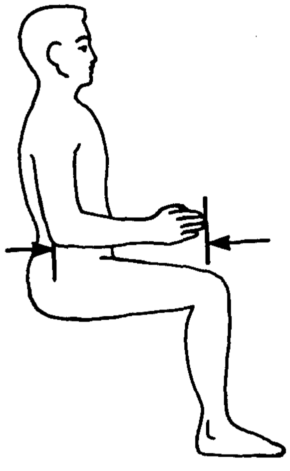


FIGURE 11–22. Forearm-hand length.

TABLE 11–45. FOREARM-HAND LENGTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	17.0	17.6	18.9	20.2	20.7	0.81
Army personnel:						
Aviators [2]	16.1	17.6	19.1	20.4	21.5	.86
Separatees, white [3]	16.6	17.3	18.7	20.1	20.8	.88
Separatees, Negro [4]	17.3	18.0	19.6	21.4	22.1	.94
Truck and bus drivers [5]	16.7	17.3	18.8	20.2	20.8	----
College students [6]	17.0	17.6	18.9	20.2	20.7	.75
Naval aviators [7]	17.5	17.9	19.1	20.4	20.8	.75

[1] Hertzbertg et al. (1954).
[2] White (1961).
[3] Newman and White (1951).
[4] USA (1946).
[5] McFarland et al. (1958).
[6] Bowles (1932).
[7] Gifford et al. (1965).

For fingertip manipulation of controls, subtract 0.5 in. for flip and 1.0 in. for push. For manipulation by the thumb and forefinger, subtract 3.0 in. For grasp by the whole hand, subtract 5.0 in.

11.4.22 Elbow-to-Elbow Breadth

This dimension is the horizontal distance between the lateral surfaces of the elbows. The subject sits erect with his upper arms vertical and lightly touching his sides, and his forearms extended horizontally. (See Figure 11–23.) The data are given in Tables 11–46 and 11–47.

For women, subtract 2 in. from the above values. Add 0.5 in. for light clothing and 4.5 in. for heavy clothing.

11.4.23 Hip Breadth

This dimension is the maximum horizontal distance across the hips. The subject stands erect with his heels together. (See Figure 11–24.) The data are given in Table 11–48.

For U.S. women, add 0.5 in. to the above values. Add 0.5 in. for light clothing and 1.5 in. or more for heavy clothing.

11.4.24 Hip Breadth, Sitting

This is the maximum horizontal distance across the hips when seated. The subject sits erect with his knees at right angles and his knees and heels together. (See Figure 11–25.) The data are given in Tables 11–49 and 11–50.

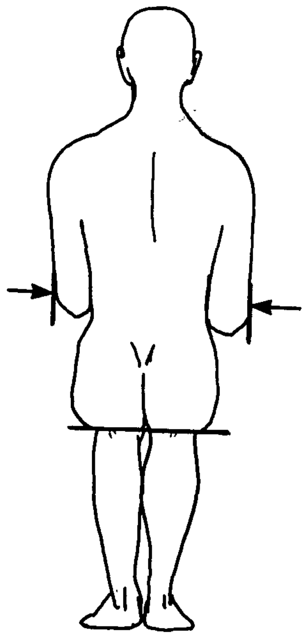


FIGURE 11–23. Elbow-to-elbow breadth.

TABLE 11–46. ELBOW-TO-ELBOW BREADTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	14.5	15.2	17.2	19.8	20.9	1.42
Cadets [2]	14.4	15.1	16.7	18.4	19.1	----
Gunners [2]	13.9	14.6	16.4	18.2	18.9	----
Army personnel:						
Separatees, white [3]	14.4	15.3	17.4	20.3	21.8	1.54
Separatees, Negro [4]	14.4	15.1	16.9	19.3	20.4	1.28
Truck and bus drivers [5]	13.8	14.9	17.5	20.7	22.2	----

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] USA (1946).
[5] McFarland et al. (1958).

TABLE 11–47. ELBOW-TO-ELBOW BREADTH OF U.S. FEMALE AIR FORCE PERSONNEL

Population	Percentiles (in.)				
	1st	5th	50th	95th	99th
Pilots	12.8	13.3	15.1	17.1	18.5
Flight nurses	13.0	13.5	14.9	16.7	17.3

Randall et al. (1946).

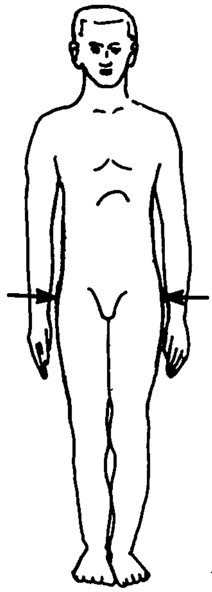


FIGURE 11–24. Hip breadth.

TABLE 11–48. HIP BREADTH OF MALE AND FEMALE U.S. AIR FORCE AND NAVY PERSONNEL AND COLLEGE STUDENTS

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Male USAF flight personnel[1]	11.3	12.1	13.2	14.4	15.2	0.73
Male basic trainees [2]	11.5	12.1	13.3	15.0	15.0	.94
Naval aviators [3]	12.1	12.6	13.8	14.9	15.4	.70
Male college students [4]	11.4	11.8	13.0	14.2	14.7	.67
Female basic trainees [5]	12.2	12.5	13.5	15.4	16.9	.95

[1] Hertzberg et al. (1954).
[2] Daniels et al. (1953b).
[3] Gifford et al. (1965).
[4] Bowles (1932).
[5] Daniels et al. (1953a).

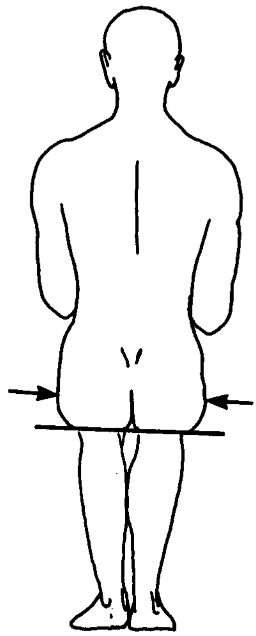


FIGURE 11-25. Hip breadth, sitting.

TABLE 11-49. HIP BREADTH, SITTING, OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	12.2	12.7	13.9	15.4	16.2	0.87
Cadets [2]	12.6	13.1	14.2	15.5	15.9	----
Gunners [2]	12.1	12.7	13.8	15.1	15.5	----
Army personnel:						
Separatees, white [3]	12.2	12.7	13.9	15.5	16.7	.90
Separatees, Negro [4]	11.6	12.1	13.4	15.0	15.8	.84
Aviators [5]	12.4	12.8	14.2	15.7	16.3	.87
Navy personnel: [6]						
Enlisted men	12.4	13.0	14.8	16.4	17.2	1.05
Cadets, aviation	13.4	14.0	15.4	16.8	17.3	----
Aviators [7]	12.7	13.1	14.5	15.9	16.6	.85
Railroad travelers [8]	*12.9	*13.7	*15.3	*17.4	*18.1	----
Truck and bus drivers [9]	12.4	13.2	14.5	16.3	16.8	----
Civilian men [10]	11.5	12.2	14.0	15.9	17.0	----

*Including light clothing (subtract 0.5 in. for nude dimension).
[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] USA (1946).
[5] White (1961).
[6] USN (1955).
[7] Gifford et al. (1965).
[8] Hooton (1945).
[9] McFarland et al. (1958).
[10] Stoudt et al. (1965).

TABLE 11-50. HIP BREADTH, SITTING, OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)				
	1st	5th	50th	95th	99th
Air Force personnel:[1]					
Pilots	13.0	13.5	15.0	16.9	18.1
Flight nurses	13.1	13.5	15.1	16.6	17.1
Railroad travelers [2]	*12.2	*13.1	*14.6	*17.2	*17.8
Civilian women [3]	11.7	12.3	14.3	17.1	18.8

*Including light clothing (subtract 0.5 in. for nude dimension).
[1] Randall et al. (1946).
[2] Hooton (1945).
[3] Stoudt et al. (1965).

For U.S. women, add 1.0 in. to the values given above. Add 0.5 in. for light clothing, 2.0 in. for heavy clothing, and, for partial-pressure suits, 3.0 in. when uninflated and 5.5 in. when inflated.

11.4.25 Buttock-Leg Length

This is the horizontal distance from the most posterior point on the buttocks to the base of the heel. The subject sits erect with his legs as far forward as possible on a horizontal surface. (See Figure 11–26.) The data are given in Table 11–51.

This dimension is involved in two considerations: leg reach (the maximum forward distance reachable by the leg from a seated position), and clearance for the outstretched leg (the distance between the seat back, or objects located behind the buttocks, and objects in front of the feet).

For clothing, add 1.0 in. for men's shoes, 1.3 in. for military boots, 0.3 in. for heavy clothing behind the buttocks, and 2.5 in. for partial-pressure suits.

11.4.26 Buttock-Knee Length

This is the horizontal distance from the most posterior point on the buttocks to the most anterior point on the knee. The subject sits erect with his knees at right angles. (See Figure 11–27.) The data are given in Tables 11–52 and 11–53.

For U.S. women, subtract 1.0 in from the above values. Add 0.2 in. for light clothing, 0.7 in. or more for heavy clothing, and 2.9 in. for partial-pressure suits.

11.4.27 Buttock-Popliteal Length

This is the horizontal distance from the plane of the most posterior point on the buttocks to the back of the lower leg at the knee. The subject sits erect with his knees at right angles. (See Figure 11–28.) The data are given in Table 11–54.

These values represent the clothed 1st and 5th percentiles minus an arbitrary 0.2 in. For women, subtract 0.7 in. from these values.

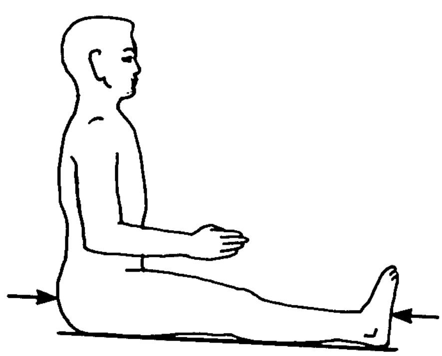


FIGURE 11–26. Buttock-leg length.

TABLE 11–51. BUTTOCK-LEG LENGTH OF U.S. MALE MILITARY PERSONNEL

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	38.2	39.4	42.7	46.1	47.7	2.04
Navy personnel:						
General [2]	36.5	38.0	41.5	44.9	46.4	2.07
Pilots, aircraft [3]	36.8	38.3	42.3	46.3	48.8	----

[1] Hertzberg et al. (1954).
[2] King et al. (1947).
[3] USN (1959).

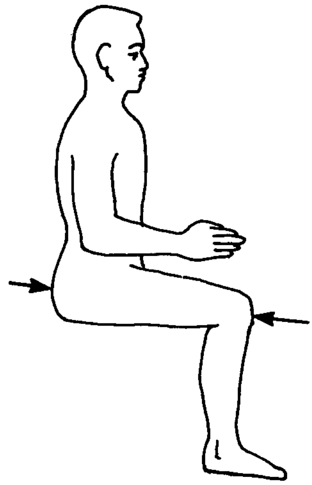


FIGURE 11–27. Buttock-knee length.

TABLE 11–52. BUTTOCK-KNEE LENGTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	21.2	21.9	23.6	25.4	26.2	1.06
Cadets [2]	21.2	22.0	23.6	25.6	26.2	----
Gunners [2]	20.5	21.1	23.1	24.7	25.6	----
Army personnel:						
Aviators [3]	21.4	22.1	23.8	25.8	26.7	1.08
Separatees, white [4]	20.7	21.5	23.4	25.2	26.0	1.12
Separatees, Negro [5]	21.1	21.9	23.8	25.8	26.6	1.17
Navy personnel [6]	20.6	21.4	23.4	25.0	25.8	1.18
Enlisted men [7]	21.7	22.5	24.5	26.5	27.3	1.23
Cadets, aviation [7]	21.8	22.6	24.3	26.2	26.9	----
Truck and bus drivers [8]	21.3	22.1	23.8	25.8	26.5	----
Naval aviators [9]	21.8	22.5	24.1	25.8	26.5	1.00
Civilian men [10]	20.3	21.3	23.3	25.2	26.3	----

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] White (1961).
[4] Newman and White (1951).
[5] USA (1946).
[6] King et al. (1947).
[7] USN (1955).
[8] McFarland et al. (1958).
[9] Gifford et al. (1965).
[10] Stoudt et al. (1965).

TABLE 11–53. BUTTOCK-KNEE LENGTH OF FEMALE U.S. AIR FORCE PERSONNEL

Population	Percentiles (in.)				
	1st	5th	50th	95th	99th
Pilots	20.4	21.1	22.6	24.2	25.0
Flight nurses	20.2	20.9	22.4	24.0	24.8

Randall et al (1946).

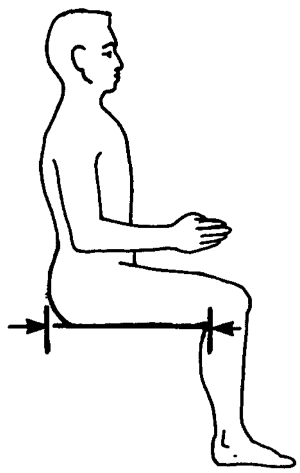


FIGURE 11–28. Buttock-popliteal length.

TABLE 11–54. BUTTOCK-POPLITEAL LENGTH OF U.S. CIVILIAN SAMPLES

Sex	Percentiles (in.)				
	1st	5th	50th	95th	99th
Male [1]	16.6	17.4	18.9	20.8	21.5
Female [1]	16.0	16.8	18.2	20.0	20.6
Male [2]	16.5	17.3	19.5	21.6	22.7
Female [2]	16.1	17.0	18.9	21.0	22.0

[1] Hooton (1945).
[2] Stoudt et al. (1965).

11.4.28 Buttock Depth

This is the horizontal distance between the buttocks and the abdomen at the level of the maximum protrusion of the buttocks. The subject stands erect. (See Figure 11–29.) The data are given in Table 11–55.

11.4.29 Thigh Clearance Height, Sitting

This is the vertical distance from the sitting surface to the top of the thigh at its intersection with the abdomen. The subject sits erect with his knees at right angles. (See Figure 11–30.) The data are given in Table 11–55.

For women, use the same values given for males. Add 0.1 to 0.2 in. for light clothing and 1.4 in. or more for heavy clothing.

11.4.30 Knuckle Height

This is the vertical distance from the floor to the point where the middle finger of the right hand meets the palm. The subject stands erect with his arm and hand extended straight, and his palm flat against the thigh. (See Figure 11–31.) The data are given in Table 11–55.

For clothing, add 1.0 in. for shoes and 1.3 in. for military boots.

11.4.31 Hand Thickness

This dimension is the maximum distance between the dorsal and palmar surfaces of the knuckle of the middle finger where it joins the

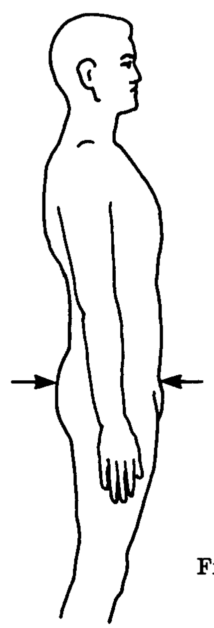


FIGURE 11–29. Buttock depth.

TABLE 11–55. BUTTOCK DEPTH, THIGH CLEARANCE HEIGHT, SITTING, AND KNUCKLE HEIGHT OF MALE USAF PERSONNEL

Dimension	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Buttock depth	7.2	7.6	8.8	10.2	10.9	0.82
Thigh clearance height, sitting	4.5	4.8	5.6	6.5	6.8	.52
Knuckle height	26.7	27.7	30.0	32.4	33.5	1.45

Hertzberg et al. (1954).

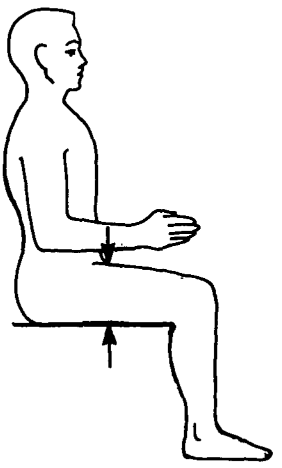


FIGURE 11–30. Thigh clearance height, sitting.

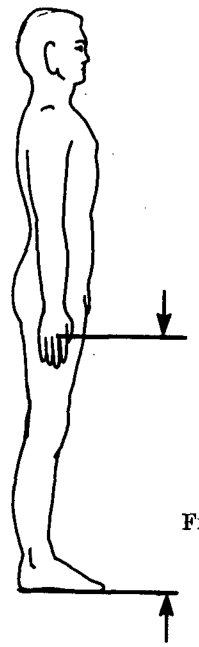


FIGURE 11–31. Knuckle height.

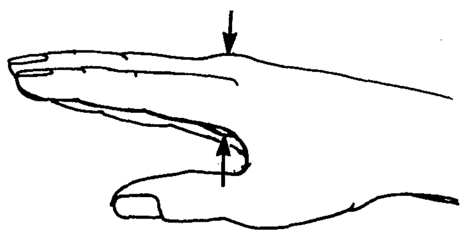


FIGURE 11–32. Hand thickness.

TABLE 11–56. HAND THICKNESS OF MALE AND FEMALE USAF AND NAVY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Male flight personnel [1]	1.0	1.1	1.2	1.3	1.4	0.07
Male basic trainees [2]	1.0	1.1	1.2	1.4	1.4	.09
Female basic trainees [3]	0.8	0.8	1.0	1.1	1.2	.09
Naval aviators [4]	1.0	1.1	1.2	1.4	1.5	.08

[1] Hertzberg et al. (1954).
[2] Daniels et al. (1953b).
[3] Daniels et al. (1953a).
[4] Gifford et al. (1965).

palm of the right hand when the fingers are extended. (See Figure 11–32.) The data are given in Table 11–56.

For women, subtract 0.2 in. from the male values. Add 0.2 in. for wool or leather gloves and about 1.5 in. for arctic mittens.

11.4.32 Hand Length

This is the distance from the base of the thumb to the middle fingertip of the right hand extended straight on the arm. (See Figure 11–33.) The data are given in Tables 11–57 and 11–58.

For women, subtract 0.7 in. from the male values.

11.4.33 Hand Breadth at Thumb

For this dimension, measure the maximum breadth across the palm (at right angles to the long axis of the hand) at the knuckle of the thumb of the right hand with the fingers

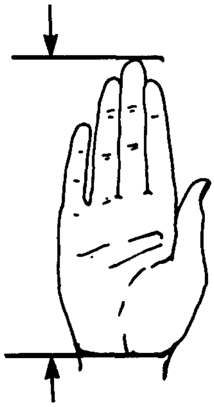


FIGURE 11–33. Hand length.

TABLE 11–57. HAND LENGTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	6.7	6.9	7.5	8.0	8.3	0.34
Cadets [2]	6.8	7.1	7.6	8.2	8.4	----
Gunners [2]	6.6	6.9	7.5	8.1	8.4	----
Basic trainees [3]	6.7	6.9	7.5	8.2	8.5	.38
Army personnel:						
Aviators [4]	6.7	6.9	7.5	8.1	8.3	.34
Separatees, white [5]	6.7	7.0	7.6	8.2	8.5	.36
Separatees, Negro [6]	7.0	7.3	8.0	8.7	9.0	.42
Truck and bus drivers [7]	6.9	7.1	7.6	8.1	8.3	----
Naval aviators [8]	6.8	7.0	7.5	8.1	8.3	.34

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] White (1961).
[5] Newman and White (1951).
[6] USA (1946).
[7] McFarland et al. (1958).
[8] Gifford et al. (1965).

TABLE 11–58. HAND LENGTH OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel:						
Pilots [1]	6.2	6.4	6.9	7.5	7.7	----
Flight nurses [1]	6.3	6.5	6.9	7.4	7.6	----
Basic trainees [2]	6.0	6.2	6.8	7.3	7.6	0.34
Army personnel [3]	6.1	6.4	6.9	7.4	7.7	.33
College students [4]	6.0	6.2	6.7	7.2	7.4	.31

[1] Randall et al. (1946).
[2] Daniels et al. (1953a).
[3] Randall and Munro (1949).
[4] Carter (1932).

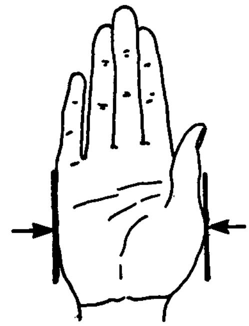


FIGURE 11–34. Hand breadth at thumb.

TABLE 11–59. HAND BREADTH AT THUMB OF MALE AND FEMALE U.S. AIR FORCE AND NAVY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Male flight personnel [1]	3.6	3.7	4.1	4.4	4.6	0.21
Male basic trainees [2]	3.5	3.7	4.1	4.5	4.7	.25
Female basic trainees [3]	3.1	3.2	3.6	4.0	4.1	.23
Naval aviators [4]	3.7	3.9	4.2	4.5	4.6	.19

[1] Hertzberg et al. (1954).
[2] Daniels et al. (1953b).
[3] Daniels et al. (1953a).
[4] Gifford et al. (1965).

extended and the thumb lying alongside and in the plane of the hand. (See Figure 11–34.) The data are given in Table 11–59.

For women, subtract 0.5 in. from the male values. Add 0.3 in. for wool or leather gloves and about 1.0 in. for arctic mittens.

11.4.34 Hand Breadth at Metacarpal

This is the maximum breadth across the ends of the metacarpal bones (where the fingers join the palm) of the index and little fingers of the right hand extended straight and stiff with the fingers together. (See Figure 11–35.) (The term, metacarpal, is the name of an anatomical point located at the distal end of the middle metacarpal bone.) The data are given in Tables 11–60 and 11–61.

For U.S. women, subtract 0.3 in. from the male values. Add 0.3 in. for woolen or leather gloves and about 1.0 in. for arctic mittens.

11.4.35 Knee Height, Sitting

This is the vertical distance from the floor to the uppermost point on the knee. The subject sits erect with his knees at right angles. (See Figure 11–36.) The data are given in Tables 11–62 and 11–63.

For U.S. women, subtract 2.0 in. from the above values. Add 1.0 in. for men's shoes and light clothing, 1.5 in. or more for military boots and heavy clothing, and up to 3.0 in. for women's shoes and light clothing.

11.4.36 Knee-to-Knee Breadth

This dimension is the maximum horizontal distance across the lateral surfaces of the knees. The subject sits erect with his knees at right angles and touching lightly. (See Figure 11–37.) The data are given in Tables 11–64 and 11–65.

For U.S. women, the same values hold. Add 0.5 in. for light clothing, 2.0 in. for heavy clothing, 2.0 in. for heavy clothing, and 9.5 in. for partial-pressure suits.

11.4.37 Popliteal Height, Sitting

This is the vertical distance from the floor to the underside of the thigh immediately behind the knee. The subject sits erect with his knees at right angles and the bottom of his thighs and the back of his knees barely touching the sitting surface. (See Figure 11–38.) The data are given in Table 11–66.

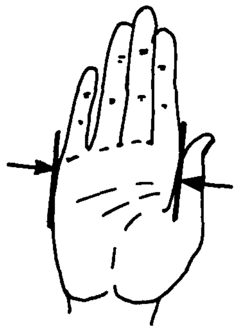


FIGURE 11-35. Hand breadth at metacarpal.

TABLE 11-60. HAND BREADTH AT METACARPAL OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	3.1	3.2	3.5	3.7	3.9	0.16
Cadets [2]	3.0	3.1	3.4	3.7	3.8	----
Gunners [2]	3.0	3.1	3.4	3.6	3.7	----
Basic trainees [3]	3.0	3.2	3.5	3.7	3.9	.18
Army personnel:						
Separatees, white [4]	2.9	3.1	3.4	3.8	3.9	.19
Separatees, Negro [5]	3.0	3.2	3.5	3.8	4.0	.20
Truck and bus drivers [6]	3.1	3.2	3.5	3.8	4.0	----
Army aviators [7]	3.1	3.2	3.5	3.8	3.9	.16
Naval aviators [8]	3.2	3.3	3.5	3.8	3.9	.17

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Newman and White (1951).
[5] USA (1946).
[6] McFarland et al. (1958).
[7] White (1961).
[8] Gifford et al. (1965).

TABLE 11-61. HAND BREADTH AT METACARPAL OF U.S. FEMALE MILITARY PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel:						
Pilots [1]	2.8	2.8	3.0	3.3	3.4	----
Flight nurses [1]	2.7	2.8	3.0	3.2	3.3	----
Basic trainees [2]	2.6	2.7	3.0	3.4	3.6	0.19
Army personnel [3]	2.6	2.7	3.0	3.4	3.6	.20

[1] Randall et al. (1946).
[2] Daniels et al. (1953a).
[3] Randall and Munro (1949).

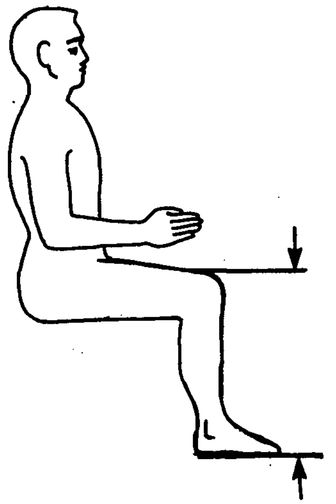


FIGURE 11–36. Knee height, sitting.

TABLE 11–62. KNEE HEIGHT, SITTING, OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	19.5	20.1	21.7	23.3	24.0	0.99
Cadets [2]	19.7	20.4	22.0	23.6	24.3	----
Gunners [2]	19.2	19.8	21.5	23.0	2.37	----
Army personnel:						
Separatees, white [3]	19.0	19.8	21.6	23.5	24.3	1.09
Separatees, Negro [4]	19.6	20.3	22.2	24.0	24.7	1.14
Naval aviators [5]	19.7	20.3	21.8	23.5	24.2	.98
Truck and bus drivers [6]	19.3	20.1	21.7	23.5	24.2	----
Civilian men [7]	18.3	19.3	21.4	23.4	24.1	----

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Newman and White (1951).
[4] USA (1946).
[5] Gifford et al. (1965).
[6] McFarland et al. (1958).
[7] Stoudt et al. (1965).

TABLE 11–63. KNEE HEIGHT, SITTING, OF U.S. FEMALE MILITARY PERSONNEL AND CIVILIANS

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]						
Pilots	18.3	18.7	20.1	21.5	22.2	----
Flight nurses	17.7	18.1	19.5	20.8	21.5	----
Army personnel [2]	16.6	17.2	18.8	20.3	21.1	0.95
Civilian women [3]	17.1	17.9	19.6	21.5	22.4	----

[1] Randall et al. (1946).
[2] Randall and Munro (1949).
[3] Stoudt et al. (1965).

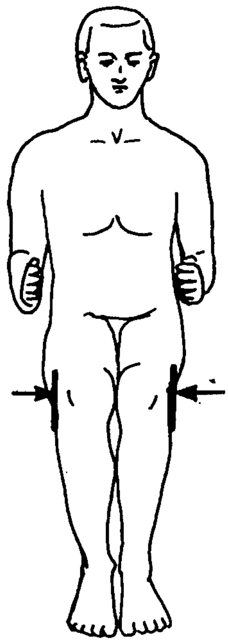


FIGURE 11–37. Knee-to-knee breadth.

TABLE 11–64. KNEE-TO-KNEE BREADTH OF U.S. MALE MILITARY AND CIVILIAN POPULATIONS

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	7.0	7.2	7.9	8.8	9.4	0.52
Cadets [2]	6.8	7.1	7.7	8.4	8.7	----
Gunners [2]	6.7	6.9	7.6	8.2	8.5	----
Truck and bus drivers [3]	6.8	7.3	8.1	9.2	9.5	----

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] McFarland et al. (1958).

TABLE 11–65. KNEE-TO-KNEE BREADTH OF FEMALE U.S. AIR FORCE PERSONNEL

Population	Percentiles (in.)				
	1st	5th	50th	95th	99th
Pilots	6.5	6.7	7.6	8.6	9.6
Flight nurses	6.6	6.8	7.5	8.4	9.6

Randall et al. (1946).

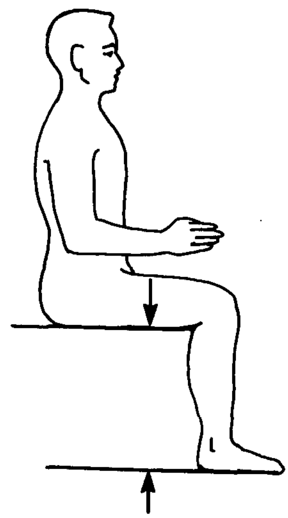


FIGURE 11–38. Popliteal height, sitting.

TABLE 11–66. POPLITEAL HEIGHT, SITTING, OF U.S. MALE AND FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Male Air Force personnel [1]	15.3	15.7	17.0	18.2	18.8	0.77
Male railroad travelers [2]	*16.9	*17.6	*19.0	*20.6	*21.1	----
Female railroad travelers [2]	*16.2	*16.7	*18.1	*19.5	*20.1	----
Naval aviators [3]	15.4	15.9	17.3	18.8	19.3	.86

*Including shoes and light clothing (subtract 2 in. for nude dimension).
[1] Hertzberg et al. (1954).
[2] Hooton (1945).
[3] Gifford et al. (1965).

For U.S. women, subtract 2.0 in. from the above values. For clothing, add 1.0 in. for men's shoes, 1.3 in. for military boots, and up to 3.0 in. for women's shoes.

11.4.38 Foot Length

This is the horizontal distance from the back of the heel to the tip of the longest toe. (See Figure 11–39.) The subject stands with his weight equally distributed on both feet. The data are given in Tables 11–67 and 11–68.

For U.S. women, subtract 1.0 in. from the male values. For clothing, add 1.2 in. for men's shoes, 1.6 in. for military boots, and 2.7 in for heavy flying boots.

11.4.39 Foot Breadth

This dimension is the maximum horizontal distance across the foot (wherever it is found) at right angles to the long axis. (See Figure 11–40.) The subject stands with his weight equally distributed on both feet. The data are given in Tables 11–69 and 11–70.

For U.S. women, subtract 0.4 in. from the male values. For clothing, add 0.3 in. for men's shoes and military boots and 1.2 in. or more for heavy flying boots.

11.5 Functional Body Dimensions

Functional body dimensions are taken from body positions that result from motion. In the past, a distinction has been made between two types of anthropometry, structural (static) and functional (dynamic). Fundamentally, there is only one anthropometry, for the body in motion is only a special case of the body at

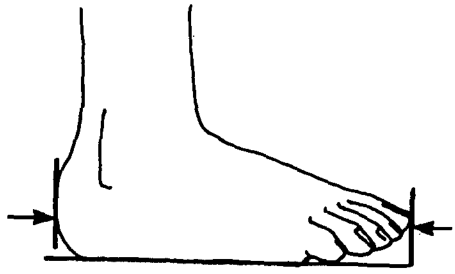


FIGURE 11–39. Foot length.

TABLE 11–67. FOOT LENGTH OF U.S. MALE MILITARY AND CIVILIAN PERSONNEL

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel [1]	9.5	9.8	10.5	11.3	11.6	0.45
Cadets [2]	9.5	9.8	10.5	11.3	11.6	----
Gunners [2]	9.3	9.6	10.4	11.1	11.4	----
Basic trainees [3]	9.2	9.5	10.3	11.2	11.5	.50
Army personnel [4]	9.3	9.6	10.4	11.1	11.5	.47
Separatees, white [5]	9.3	9.7	10.4	11.2	11.5	.48
Separatees, Negro [6]	9.6	9.9	10.8	11.6	12.0	.50
Aviators [7]	9.5	9.9	10.6	11.5	11.9	.49
Truck and bus drivers [8]	9.2	9.6	10.4	11.3	11.6	----
College students [9]	9.2	9.4	10.3	11.1	11.4	.48
Naval aviators [10]	9.5	9.7	10.5	11.3	11.6	.47

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Randall et al. (1951).
[5] Newman and White (1951).
[6] USA (1946).
[7] White (1961).
[8] McFarland et al. (1958).
[9] Bowles (1932).
[10] Gifford et al. (1965).

TABLE 11–68. FOOT LENGTH OF U.S. FEMALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.)					S.D.
	1st	5th	50th	95th	99th	
Air Force personnel:						
Pilots [1]	8.6	8.9	9.6	10.2	10.5	----
Flight nurses [1]	8.7	8.9	9.6	10.3	10.5	----
Basic trainees [2]	8.4	8.7	9.4	10.2	10.5	0.46
Army personnel [3]	8.4	8.7	9.4	10.2	10.5	.44
College students [4]	8.3	8.7	9.5	10.3	10.7	.45

[1] Randall et al. (1946).
[2] Daniels et al. (1953a).
[3] Randall and Munro (1949).
[4] Carter (1932).

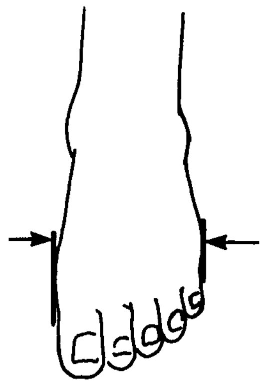


FIGURE 11–40. Foot breadth.

TABLE 11–69. FOOT BREADTH OF U.S. MALE MILITARY AND CIVILIAN SAMPLES

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel [1]	3.4	3.5	3.8	4.1	4.4	0.19
Cadets [2]	3.5	3.6	3.9	4.2	4.3	----
Gunners [2]	3.3	3.5	3.8	4.2	4.3	----
Basic trainees [3]	3.5	3.6	4.0	4.4	4.7	.25
Army personnel [4]	3.4	3.5	3.9	4.2	4.3	.20
Separatees, white [5]	3.3	3.5	3.9	4.3	4.4	.25
Separatees, Negro [6]	3.4	3.6	4.0	4.4	4.6	.25
Aviators [7]	3.5	3.6	4.0	4.4	4.5	.21
Truck and bus drivers [8]	3.6	3.7	4.0	4.3	4.4	----
Naval aviators [9]	3.5	3.6	4.0	4.6	4.9	.30

[1] Hertzberg et al. (1954).
[2] Randall et al. (1946).
[3] Daniels et al. (1953b).
[4] Randall et al. (1951).
[5] Newman and White (1951).
[6] USA (1946).
[7] White (1961).
[8] McFarland et al. (1958).
[9] Gifford et al. (1965).

TABLE 11–70. FOOT BREADTH OF U.S. FEMALE MILITARY PERSONNEL

Population	Percentiles (in.) 1st	5th	50th	95th	99th	S.D.
Air Force personnel:						
Pilots [1]	3.2	3.3	3.6	3.9	4.1	----
Flight nurses [1]	3.2	3.3	3.6	3.9	4.1	----
Basic trainees [2]	3.1	3.2	3.6	3.9	4.0	0.20
Army personnel [3]	3.1	3.2	3.6	4.0	4.1	.22

[1] Randall et al. (1946).
[2] Daniels et al. (1953a).
[3] Randall and Munro (1949).

rest. Obviously, more space is needed for motion than for rest. The origin of the distinction lies in the ease of measurement: the first is relatively simple, hence well developed; the second, because it is three-dimensional, has been cumbersome, time-consuming, and difficult, hence relatively undeveloped. Nevertheless, attempts have been made, as shown below, to extract the essentials from the three-dimensional working situation, and to measure in a simple, one-dimensional fashion. In general, functional dimensions in equipment design relate more to human bodily performance than to human "fit."

11.5.1 Dimensions for Working Positions

The following dimensions are for design of spatially restricted areas where workers, such as mechanics, repairmen for heavy equipment, plumbers, or pipe-fitters, often perform their jobs. Most of these data are from small samples selected to approximate the Air Force population.

Prone

Two dimensions are taken while the subject lies prone with his feet together and comfortably extended, his arms extended forward as far as possible without strain, and his fists clenched. (See Figure 11–41.)

Prone Length (A) is measured horizontally from the tip of the toes to the most forward point on the fists.

Prone Height (B) is measured vertically from the floor to the highest point on the head when the head is raised as high as possible while the chest is on the floor. (See Table 11–71 for data.)

Crawling

Two dimensions are taken while the subject rests on his knees and flattened palms with his arms and thighs perpendicular to the floor and his feet comfortably extended and spaced. His torso is straight with his head in line with the long axis of his body. (See Figure 11–42.)

Crawling Length (A) is measured from the most rearward point on the foot to the most forward point on the head.

Crawling Height (B) is measured from the floor to the highest point on the head. (See Table 11–71 for data.)

Kneeling

Two dimensions are taken while the subject kneels with his knees and feet together and his fists clenched on the floor in front of his knees. His arms are roughly vertical and his head is in line with the long axis of his body. (See Figure 11–43.)

Kneeling Length (A) is measured from the most rearward point on the foot to the most forward point of the head.

Kneeling Height, Crouched (B) is measured vertically from the floor to the highest point of the head. The data are given in Table 11–71. (This title is changed here from Kneeling Height in the first edition of this book, and from the original report—(Hertzberg et al., 1956). See also Kneeling Height, Upright.)

Kneeling Height, Upright, is the height to the top of the subject's head from the floor upon which he kneels, holding his body upright and his head in the Frankfort plane (the top of the ear orifice and the bottom of the edge of the bony eye socket are in a horizontal plane). (See Figure 11–44 and Table 11–71.) (This title is changed here from Kneeling Height in the original report—Alexander and Clauser, 1965—to avoid confusion resulting from a conflict in terminology.)

Bent Torso Height

This is the height of the top of the subject's head from the floor as he stands, feet 18 in. apart, bending forward with hands on knees, but with his head looking straight ahead. (See Figure 11–45.) The data are given in Table 11–72.

Bent Torso Breadth

This is the breadth across the subject's shoulders as he stands, feet 18 in. apart, bending forward with hands on knees, but with head looking straight ahead. (See Figure 11–46.) The data are given in Table 11–72.

Squatting Height

This is the height of the top of the subject's head as he balances on his toes, his body erect and his head in the Frankfort plane. (See Figure 11–47.) The data are given in Table 11–72.

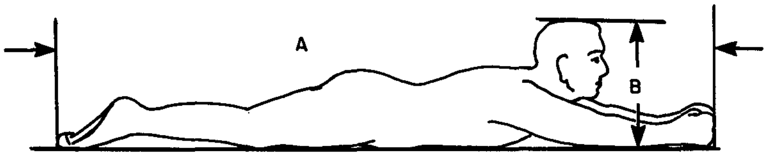


FIGURE 11–41. Prone length (A) and prone height (B).

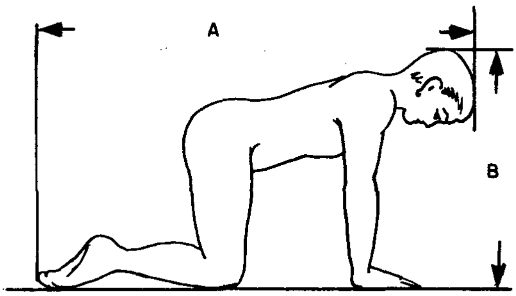


FIGURE 11–42. Crawling length (A) and crawling height (B).

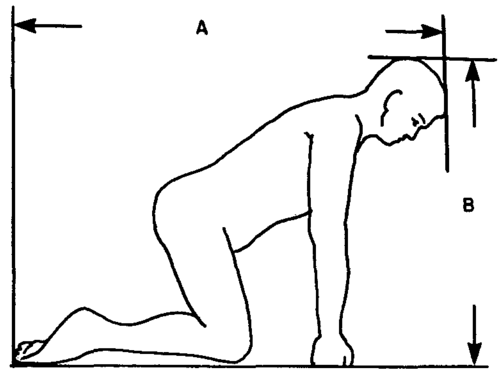


FIGURE 11–43. Kneeling length (A) and kneeling height, crouched (B).

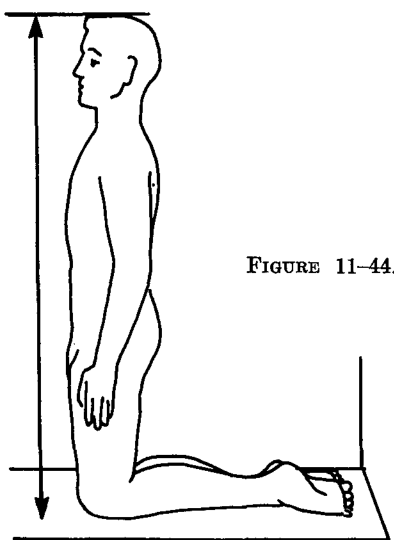


FIGURE 11–44. Kneeling height, upright.

TABLE 11–71. WORKING-POSITION DIMENSIONS OF MALE U.S. AIR FORCE PERSONNEL

Dimensions	Percentiles (in.) 5th	50th	95th	S.D.
Kneeling:				
Height, crouched	29.7	32.0	34.5	1.57
Length	37.6	43.0	48.1	3.26
Crawling:				
Height	26.2	28.4	30.5	1.30
Length	49.3	53.2	58.2	2.61
Prone:				
Height	12.3	14.5	16.4	1.28
Length	84.7	90.1	95.8	3.41

Hertzberg et al. (1956).

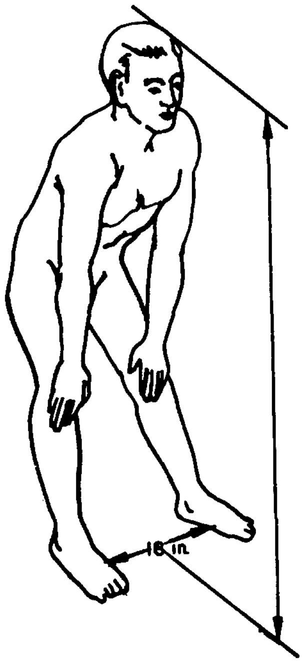


FIGURE 11–45. Bent torso height.

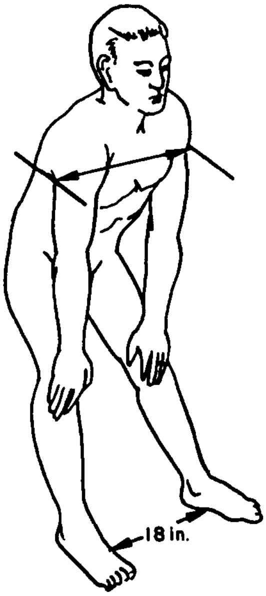


FIGURE 11–46. Bent torso breadth.

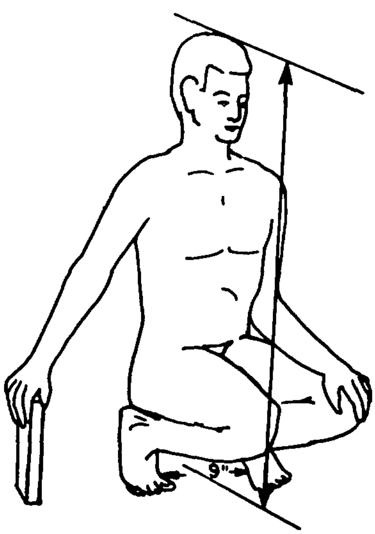


FIGURE 11–47. Squatting height.

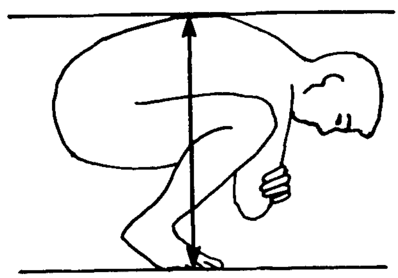


FIGURE 11–48. Minimum squatting height.

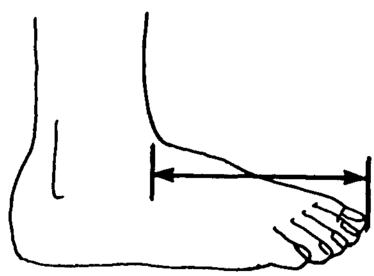


FIGURE 11–49. Functional foot length.

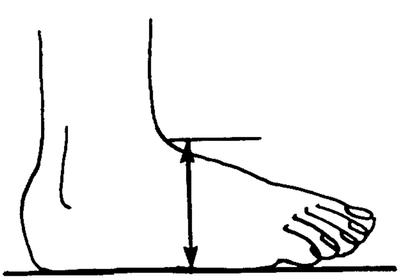


FIGURE 11–50. Functional foot height.

TABLE 11–72. DIMENSIONS FOR WORKING POSITIONS

Dimension*	Percentiles (in.)			S.D.
	5th	50th	95th	
Bent torso height [1]	46.3	52.0	55.9	2.76
Bent torso breadth [1]	16.3	17.5	19.1	.88
Kneeling height, upright [1]	48.2	51.0	54.4	1.77
Squatting height [1]	40.8	43.6	47.0	1.44
Functional foot length [1]	5.48	6.11	6.72	.37
Functional foot height [1]	2.72	2.97	3.33	.21
Vertical reach height, sitting [2]	51.6	55.0	59.0	2.24
Minimum squatting height [2]	21.5	24.8	28.0	2.04

* Additional dimensions can be found in both references cited. Some of the dimensional titles may be confusing; hence the reader is warned to check the original technique before using the data.

[1] Alexander and Clauser (1965).
[2] Snow and Snyder (1965).

Minimum Squatting Height

This is the height from the floor to the top of the squatting subject's horizontally-tilted back. (See Figure 11–48.) The data are given in Table 11–72.

Functional Foot Length

This is the length of the anterior portion of the foot. (See Figure 11–49.) The data are given in Table 11–72.

Functional Foot Height

This is the height of the anterior portion of the foot (See Figure 11–50.) The data are given in Table 11–72.

11.5.2 Reaches for Workspace Layout

Arm reach data are essential for placement of workspace control. Arm length measurements, originally taken on the standing subject as a simple maximum length to quantify anatomical differences, have now proliferated into numerous dimensions, both standing and sitting, involving various combinations of positions of the hand, arm, and shoulder, because different controls demand different degrees of precision of movement and force output. As the reach values typical of the several positions are so different, anthropometrists and designers must distinguish between them. Most arm reach data are taken during anthropometric surveys by means of anthropometers (a calibrated rod with a fixed arm and a sliding arm acting as a caliper, and illustrated in most USAF anthropometric survey reports since 1950), and such data are applicable to cockpit or other workplace design. Data so taken are only for the arm extended straight forward, whereas, for greater workplace refinement, other directions upward and downward also add useful data. Survey data from large samples can thus be augmented by special mockup studies on much smaller but still representative samples. A recent study of this type (Kennedy, 1964; see below) utilizes a sample of 20 men carefully selected as a cross section of the USAF popution.

Overhead Reach

This dimension is measured from the floor to the top of a bar grasped in the subject's right hand and raised as high as he can conveniently reach while standing. (See Figure 11–51.) See Table 11–73 for data.

Overhead Reach Breadth

This is the maximum horizontal breadth of a standing subject while he reaches overhead with both fists touching and arms extended. The subject's heels are 12 in. apart. (See Figure 11–52.) See Table 11–73 for data.

Thump-tip Reach

This is the length of the subject's horizontally extended arm from his back to the tip of his thumb as his thumb and fingertips are pressed together. It was formerly referred to as "functional reach." (See Hertzberg, 1968a.) The subject stands erect against a wall with his

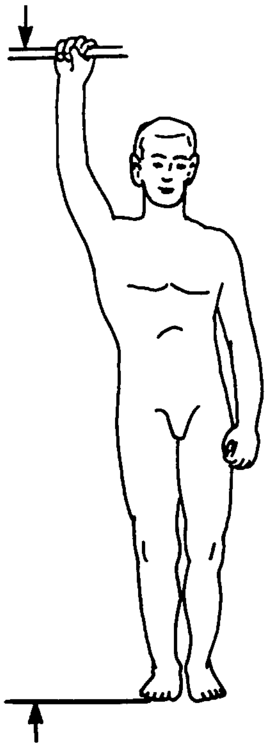


FIGURE 11–51. Overhead reach.

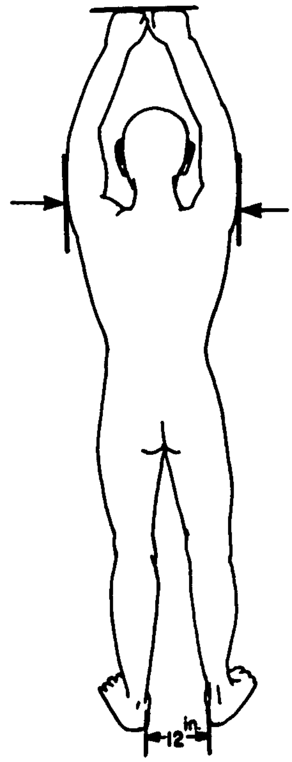


FIGURE 11–52. Overhead reach breadth.

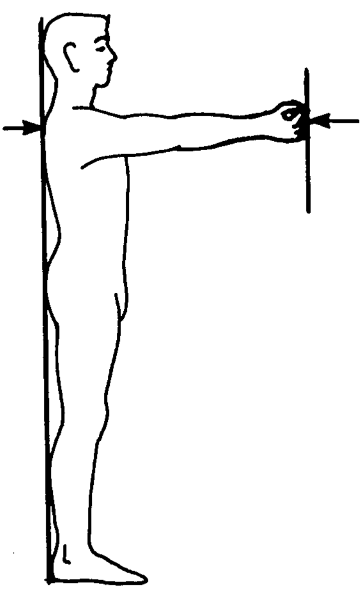


FIGURE 11–53. Thumb-tip reach.

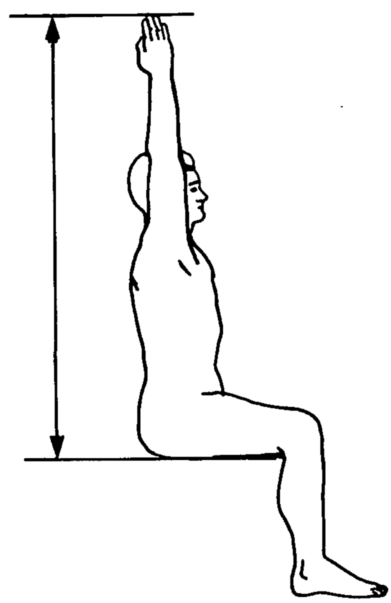


FIGURE 11–54. Vertical reach height, sitting.

TABLE 11–73. VARIOUS ARM REACHES, AND REACH BREADTH OF U.S. MALE MILITARY PERSONNEL

Dimension	Percentiles (in.)					
	1st	5th	50th	95th	99th	S.D.
Thumb-tip reach:*						
USAF [1]	28.8	29.7	32.3	35.0	36.4	1.60
Naval aviators [2]	28.6	29.3	31.4	34.0	35.1	1.42
Overhead reach [3] (1-arm)	----	76.8	82.5	88.5	----	3.35
Overhead reach breadth [4] (2-arm)	----	13.6	14.9	15.9	----	.68

* Formerly Functional Reach; see Hertzberg (1968a).
[1] Hertzberg et al. (1954).
[2] Gifford et al. (1965).
[3] Hertzberg et al. (1956).
[4] Clauser and Alexander (1965).

heels, buttocks and shoulders touching it, and his arm extended forward. (See Figure 11–53.) The data are given in Table 11–73.

Vertical Reach Height, Sitting

This is the height from the seat on which the subject sits erect, to the tip of his vertically extended arm and fingers. (See Figure 11–54.) The data are given in Table 11–72.

Grasping Reach

The equipment for this study used a wooden plywood chair with dimensions and angles (back angle, 103°; pan angle, 6°) that supported the occupant in the same position as that of the pilot in USAF standard fighter cockpits. Encircling the seat in a vertical plane is an arc-like structure supporting 24 measuring staves, held in place by light friction, which radiate from a hypothetical center at the subject's shoulder. Each staff has a knob at its near end for the subject to grasp with thumb and fingers. When the knob has been grasped and pushed away to the subject's reach limit, this distance can be read on a scale on each staff. The seat can be rotated 360°, so that an entire hemisphere or envelope can be measured above Seat Reference Point (SRP) and a good portion of the region below it. (Seat Reference Point is commonly used as a standard starting point of reach dimensions of seated operators, and is defined as the mid-point of the intersection of the plane of the seat surface with the plane of the backrest surface of the seat.)

This study measured the boundaries of "Grasping Reach" (equivalent to Thumb-tip Reach, formerly called "Functional Reach"). The present study treats the shirt-sleeved condition only. Tables 11–73 through 11–79 present the data for each level in turn, as depicted in Figures 11–55 through 11–60. Only the right-arm reach was measured, on the assumption that the body is sufficiently symmetrical. While the right arm is normally a little longer than the left, this difference can be ignored in practice.

Note that the tables do not present the reach dimensions as measured. These were plotted and converted into horizontal contours that center about the vertical line (SRV) erected from the mid-point of SRP. Hence the values for grasping reach and thumb-tip reach cannot be compared directly. Further, the data apply only to populations that are similar in size and proportions to the U.S. Air Force; they cannot be used indiscriminately for all populations. Also, the data are quite specific for the conditions noted: if the seat size or angles change, or if heavy, restrictive clothing must be worn, the reach requirements will change correspondingly.

The proper balance in workplace size is achieved when at least 95% of the population can reach and operate the controls, without causing undue restriction in the large man's performance, or causing the small man undue effort.

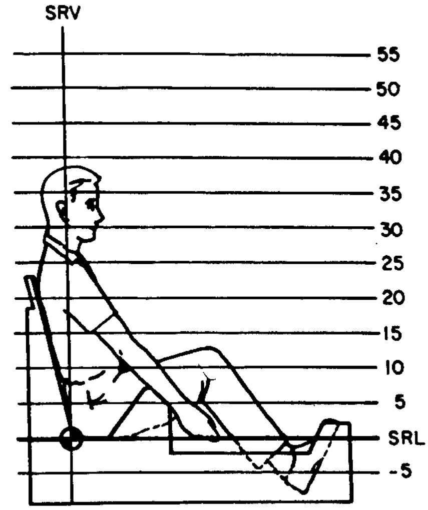
11.5.3 Aperture Sizes and Depths of Reach for One- and Two-Handed Tasks

In modern aircraft and missile technology, mechanics often make adjustments on equipment inside the skin. Dimensions such as aperture breadth and height, location above the floor, depth of reach, mechanic's body position for work, and whether or not he must be

TABLE 11–74. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, SRP LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165					
L150					
L135					
L120					
L105					
L 90					
L 75					
L 60					
L 45					
L 30					
L 15					
0					
R 15					
R 30	19		17.50	20.75	25.00
R 45	20	16.25	19.50	21.75	26.00
R 60	20	17.50	20.50	22.25	26.25
R 75	20	17.25	20.00	22.25	26.00
R 90	20	17.00	19.50	22.25	25.50
R105	20	16.25	18.75	22.00	25.25
R120	20	15.00	18.25	20.75	24.50
R135	20	13.00	16.50	19.00	23.50
R150	19		14.00	16.50	20.25
R165	13			13.00	17.00
180					

Kennedy (1964).



Angular Reach from SRV at the SRP Level

Minimum	R37°* to R140°
5th %ile	R30°* to R151°
50th %ile	R26°* to R166°
95th %ile	R21°* to R175°

* Right side of Leg Support

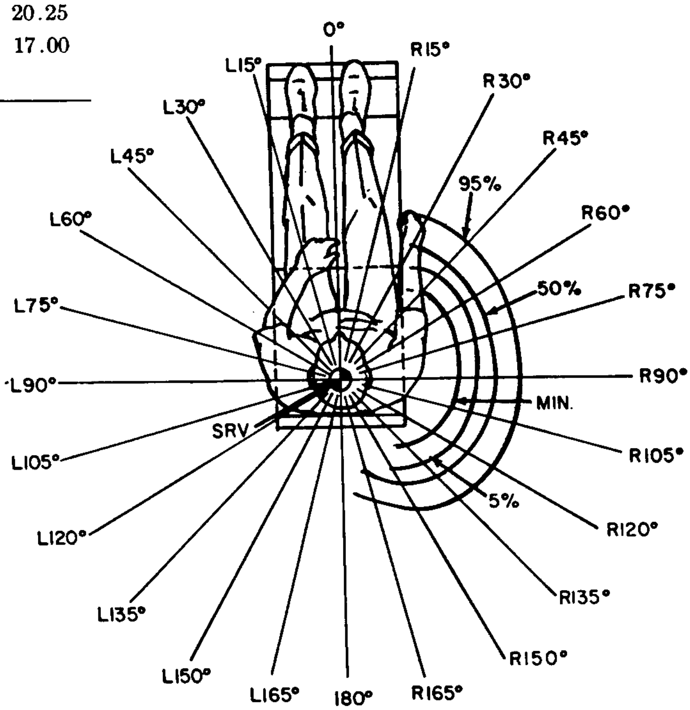
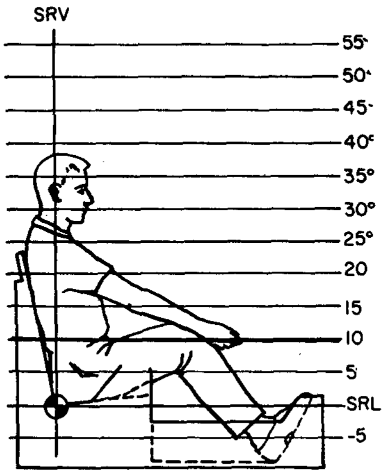


FIGURE 11–55. Shirt-sleeved grasping reach: horizontal boundaries, SRP level (Kennedy, 1964).



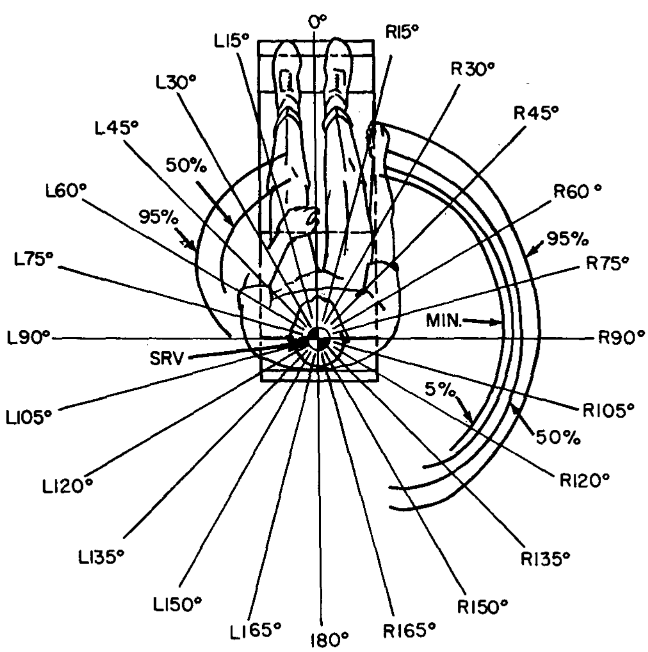


FIGURE 11–56. Shirt-sleeved grasping reach: horizontal boundaries, 10-in. level (Kennedy, 1964).

TABLE 11–75. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, 10-IN. LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165					
L150					
L135					
L120					
L105					
L 90	4				13.50
L 75	4				17.25
L 60	14			16.50	21.00
L 45	15			19.50	23.25
L 30	15			21.00	24.75
L 15	10			22.00	26.25
0					
R 15	20				
R 30	20	26.25	27.00	29.25	33.00
R 45	20	27.25	28.25	30.50	33.75
R 60	20	28.00	29.00	30.75	33.50
R 75	20	28.25	29.25	30.75	33.50
R 90	20	28.25	29.25	31.00	33.50
R105	20	27.75	28.75	30.50	32.75
R120	20	26.75	27.75	29.75	31.50
R135	19		26.25	28.25	30.75
R150	14			25.25	28.75
R165	1				
180					

Kennedy (1964).

Angular Reach from SRV at the 10-inch Level

Minimum	----	R21°** to R130°
5th %ile	----	R20°** to R141°
50th %ile	L65° to L 7°* and	R18°** to R155°
95th %ile	L90° to L11°* and	R16°** to R157°

* Left side of knees
** Right side of knees

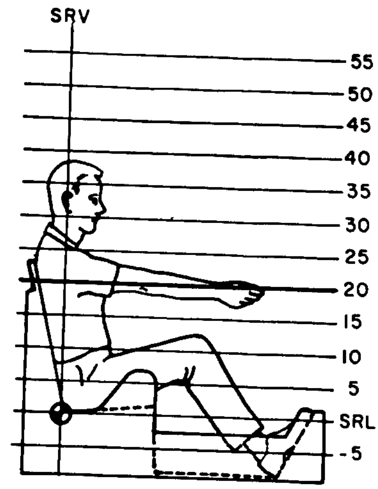
TABLE 11–76. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, 20-IN. LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165					
L150					
L135					
L120					
L105					
L 90	11			14.00	18.75
L 75	16			18.00	21.50
L 60	20	17.00	17.50	20.50	24.50
L 45	20	18.25	19.50	22.75	26.75
L 30	20	20.25	21.50	24.75	28.25
L 15	20	22.50	23.50	26.75	29.75
0	20	25.00	25.50	28.75	31.75
R 15	20	27.25	28.00	30.50	34.00
R 30	20	29.00	30.00	32.00	35.75
R 45	20	30.50	31.00	33.50	36.25
R 60	20	31.50	32.00	33.75	36.25
R 75	20	31.50	32.25	34.00	36.50
R 90	20	31.75	32.25	34.00	36.00
R105	20	31.50	31.75	33.50	35.75
R120	19		30.50	33.00	35.50
R135	9				34.50
R150					
R165					
180					

Kennedy (1964).

Angular Reach from SRV at the 20-inch Level

Minimum	L65° to R110°
5th %ile	L66° to R122°
50th %ile	L90° to R134°
95th %ile	L90° to R146°



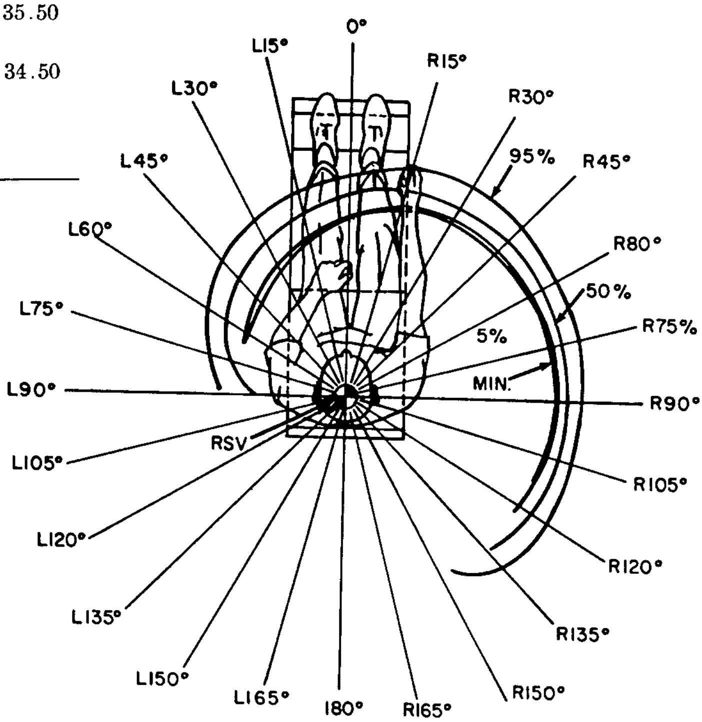
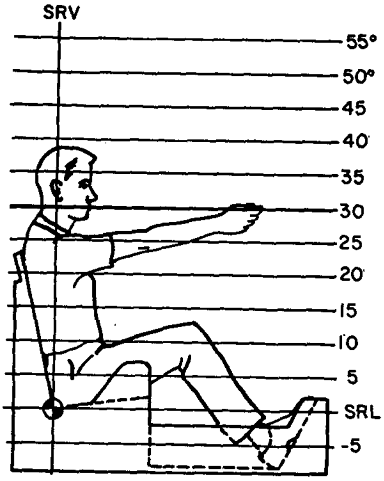


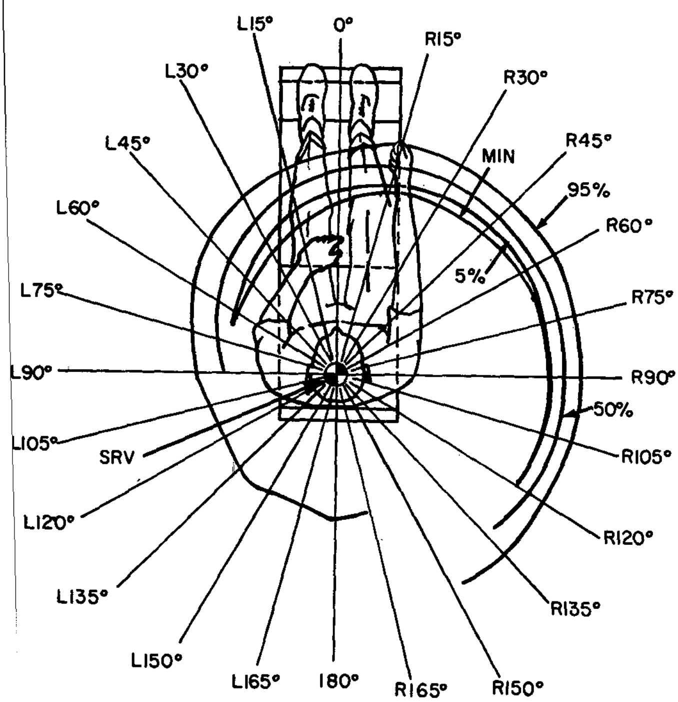
FIGURE 11–57. Shirt-sleeved grasping reach: horizontal boundaries, 20-in. level (Kennedy, 1964).

TABLE 11-77. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, 30-IN. LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165	4				18.75
L150	4				19.25
L135	6				20.00
L120	7				18.75
L105	9				19.00
L 90	16			16.75	20.75
L 75	18			18.75	22.50
L 60	20	17.00	17.25	20.75	24.50
L 45	20	18.25	19.00	22.50	26.50
L 30	20	19.75	21.50	24.50	28.25
L 15	20	22.00	23.75	26.75	29.50
0	20	23.75	25.50	28.50	31.00
R 15	20	26.00	27.25	29.75	33.00
R 30	20	27.75	29.00	31.50	34.25
R 45	20	28.75	30.25	32.25	34.75
R 60	20	30.00	31.00	32.75	35.75
R 75	20	30.75	31.25	33.00	35.50
R 90	20	31.00	31.25	33.25	35.75
R105	20	30.75	31.00	33.00	35.25
R120	19		30.25	32.50	34.75
R135	9				34.50
R150	1				
R165	2				19.50
180	2				20.25

Kennedy (1964).





Angular Reach from SRV at the 30-inch Level

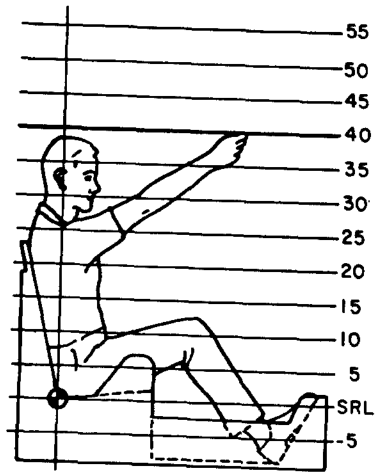
Minimum	L 67° to R111°
5th %ile	L 67° to R122°
50th %ile	L 90° to R134°
95th %ile	R165° to R149°

FIGURE 11-58. Shirt-sleeved grasping reach: horizontal boundaries, 30-in. level (Kennedy, 1964).

TABLE 11–78. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, 40-IN. LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165	14			15.50	21.50
L150	13			14.75	20.00
L135	16			14.00	19.25
L120	19		11.25	13.25	18.50
L105	19		11.75	13.25	18.25
L 90	20	12.00	12.25	12.75	18.25
L 75	20	12.25	12.50	15.00	18.75
L 60	20	12.50	13.25	16.25	20.00
L 45	20	13.00	14.00	17.75	21.50
L 30	20	13.75	15.50	19.50	23.50
L 15	20	15.25	17.00	21.25	24.50
0	20	17.00	19.00	23.00	25.75
R 15	20	18.75	21.00	24.50	28.50
R 30	20	21.00	22.75	22.75	30.50
R 45	20	23.25	24.75	27.75	31.50
R 60	20	24.25	25.50	28.00	31.25
R 75	20	25.00	26.00	28.00	31.50
R 90	20	25.00	26.25	28.25	31.50
R105	20	25.75	26.75	28.50	31.75
R120	19		26.25	28.75	31.50
R135	16			27.00	31.00
R150	8				29.25
R165	10			16.75	23.75
180	10			17.75	23.50

Kennedy (1964).



Angular Reach from SRV at the 40-inch Level

Minimum	L 90° to R119°
5th %ile	L120° to R120°
50th %ile	R156° to R143°
95th %ile	360°

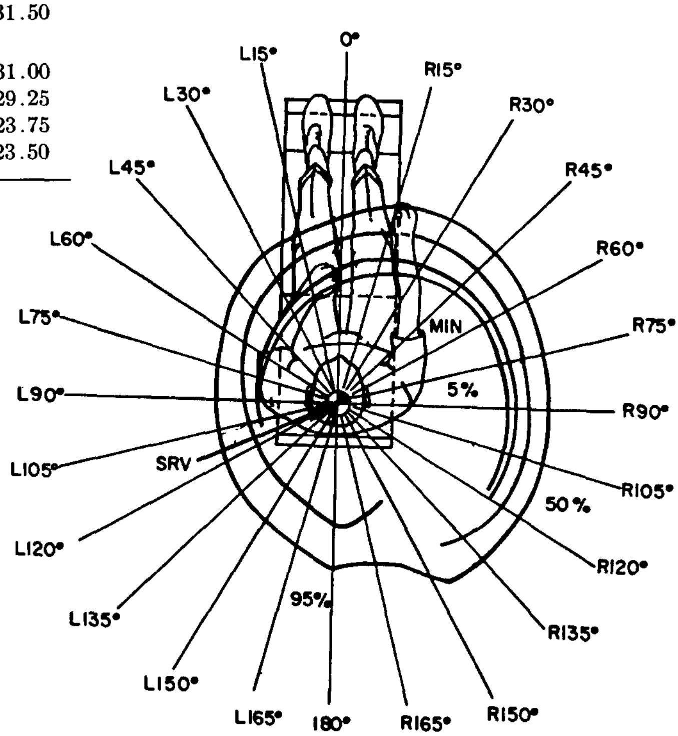


FIGURE 11–59. Shirt-sleeved grasping reach: horizontal boundaries, 40-in. level (Kennedy, 1964).

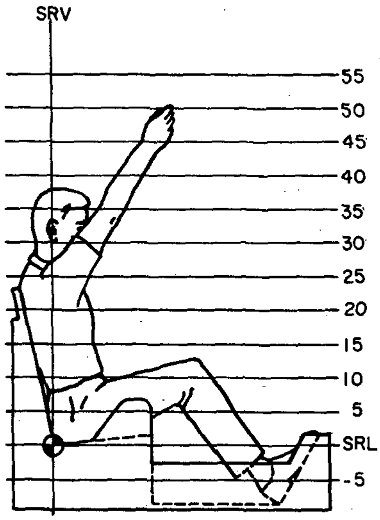
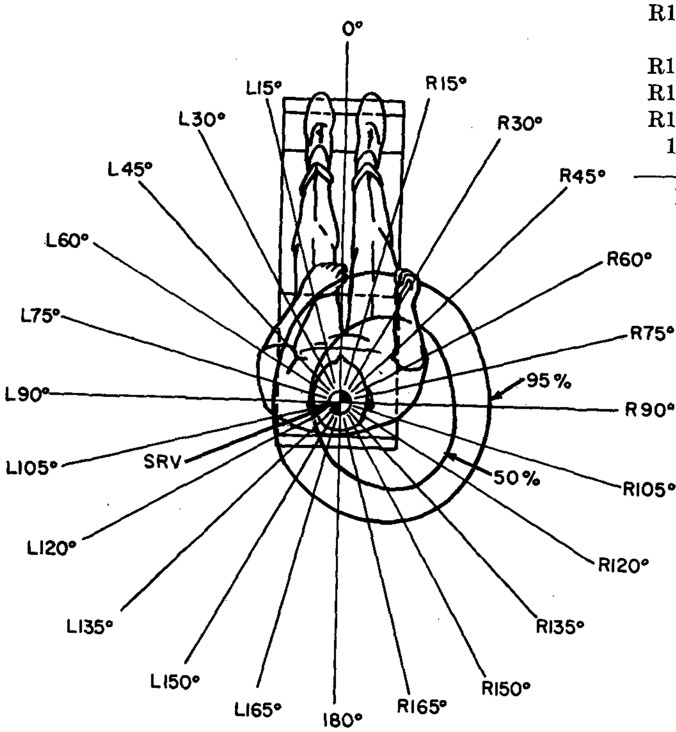


TABLE 11–79. SHIRT-SLEEVED GRASPING REACH: HORIZONTAL BOUNDARIES, 50-IN. LEVEL

Angle (deg)	N	Min	Percentiles (in.) 5th	50th	95th
L165	17			7.50	13.75
L150	17			6.00	13.00
L135	17			5.00	12.00
L120	17			4.50	10.75
L105	17			4.25	9.75
L 90	17			4.25	9.50
L 75	17			4.25	9.75
L 60	17			4.74	10.25
L 45	17			5.25	11.50
L 30	17			6.50	13.25
L 15	17			7.75	15.00
0	17			9.50	17.25
R 15	17			11.75	18.75
R 30	17			14.00	20.00
R 45	17			15.75	21.25
R 60	17			16.75	21.75
R 75	17			16.75	21.75
R 90	17			17.25	22.25
R105	17			17.50	22.25
R120	17			17.50	22.00
R135	17			16.50	20.75
R150	17			14.25	19.00
R165	17			11.50	17.25
180	17			8.75	15.50

Kennedy (1964).



Angular Reach from SRV at the 50-inch Level

Minimum	------------
5th %ile	------------
50th %ile	360°
95th %ile	360°

FIGURE 11–60. Shirt-sleeved grasping reach: horizontal boundaries, 50-in. level (Kennedy, 1964).

able to see, become important. As the complex of dimensions required will vary with the worker's body position, the designer must decide this factor in advance, and choose the dimensions permitting the greatest efficiency and ease of work performance. The weight of clothing that the man must wear will also affect the designer's choices.

In this study (Kennedy and Filler, 1966), the subject was required to grasp a "target tube" 1.5 in. in diameter, and thrust his arm through a mockup surface ("skin") as far as he could conveniently reach. The lower edge of the aperture was then adjusted so as to touch but not impede the underside of his arm, and the upper edge was lowered until he could just see the target while standing and holding his head naturally. The heights of the two edges and the depth of reach were measured by tapes on the mockup, and the breadth required by the arm was taken by anthropometer. The entire procedure was accomplished for one- and two-handed reaches, for standing and seated subjects, and for shirt-sleeved and pressure-suited conditions (A/P 22S-2). Of the sample, 16 men were tested in this suit.

In the ensuing tables, the individual dimensions are presented throughout as follows:

1. Depth of reach. The distance inside the skin at which the man can grasp a 1.5-in. tube, whose center-line closely approximates the center-line of the empty clenched hand.
2. Breadth of aperture. The breadth required for either one arm or two when thrust into the aperture.
3. Floor to top of aperture. The vertical height above the floor for the subject to see to the target tube.
4. Floor to bottom of aperture. The vertical height above the floor to the bottom of the subject's outstreched arms.
5. Vertical dimension of aperture. These values for the several conditions are presented in a separate listing. (See Tables 11–80 and 11–82.)

See the footnotes to Tables 11–81, 11–82, and 11–83 for the detailed measuring technique relevant to each table. (See Tables 11-84 through 11-89 and Figures 11-61 through 11-65.)

11.6 Range of Movement of Body Members

In order to plan for placement and excursion of controls, a designer needs information about the range of movement of the torso, arms, and legs. Required body movements should be kept well within the comfortable limits for maximum efficiency.

11.6.1 Joints and "Links"

The movable joints of the body, articulated by means of ligaments (tough, fibrous bands), are of several types; the three most important are: hinge joints (finger), pivot joints (elbow), and ball and socket joints (shoulder and hip). The range of motion is determined by the joint's body configuration; by the attached muscles, tendons, and ligaments; and by the amount of surrounding fatty tissue, all of which vary to some extent from person to person and in the same person from time to time, i.e., as the person grows older.

Bones articulated at movable joints are the rigid levers of the body's mechanical system. Each of these levers or movable body segments may be likened to the links of mechanical engineering, i.e., an intermediate rod or piece that transmits force or motion. Body links, however, are not identical to mechanical links; and the joints are not "pin-centered" in the engineering sense. The links are functional rather than structural, and may be defined as the distances between the zones of rotation at adjacent joints. Thus, the human body is basically an open-chain system of "links" rotating around joints. The end members of these open-chain links, the hands and feet, can occupy a limitless number of positions in space as a result of the cumulative ranges of these joints (Dempster, 1955a).

11.6.2 Factors Affecting the Range of Joint Movement

The range of joint motion varies from person to person for structural reasons described above and for numerous other reasons. Joint mobility decreases only slightly in healthy people between 20 and 60 years of age. The incidence of arthritis increases so markedly beyond age 45, however, that any older population will have a considerably decreased average joint mobility (Smyth et al., 1959). Women exceed men in the range of movement at all joints but the knee. (See Tables 11–90, 11–91, 11–92 and 11–93.)

TABLE 11–80. STANDING, FORWARD REACH (BOTH ARMS)

	Percentiles (in.)*				
	5th	25th	50th	75th	95th
A. Depth of reach Range: 17.50 to 25.25 SD: 1.50	19.25	21.00	22.25	22.75	24.50
B. Breadth of aperture Range: 15.00 to 20.25 Mean: 17.69 SD: 1.19	15.50	17.00	17.75	18.50	19.50
C. Floor to top of aperture Range: 58.75 to 70.50 SD: 2.34	61.00	63.50	65.25	66.50	69.00
D. Floor to bottom of aperture Range: 51.25 to 61.75 Mean: 56.09 SD: 2.05	52.25	54.75	56.00	57.25	59.00
E. Vertical dimension of aperture	(1)	(1)	(1)	(1)	(1)

Kennedy and Filler (1966).

* The ranges and all percentiles have been rounded off to the nearest 0.25 in.

[1] See Fig. 66 and Tables 86 through 90.

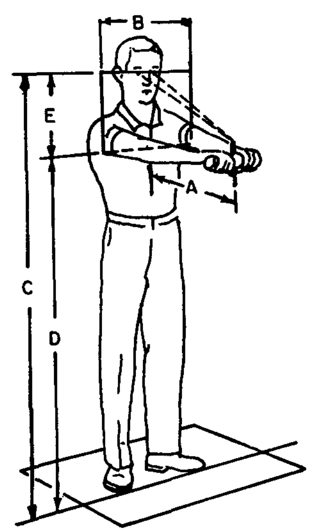


FIGURE 11–61. Standing forward reach, both arms (Kennedy and Filler, 1966).

TABLE 11-81. STANDING, FORWARD REACH (PREFERRED ARM)

	Percentiles (in.)*				
	5th	25th	50th	75th	95th
A. Depth of reach Range: 19.50 to 27.50 Mean: 23.61 SD: 1.82	20.25	22.25	23.75	25.00	26.75
B. Breadth of aperture 12.00					
C. Floor to top of aperture Range: 58.25 to 70.50 Mean: 64.88 SD: 2.36	61.00	63.25	65.00	66.25	69.00
D. Floor to bottom of aperture Range: 51.25 Mean: 56.09 SD: 2.05	52.25	54.75	56.00	57.25	59.00
E. Vertical dimension of aperture	(1)	(1)	(1)	(1)	(1)

Kennedy and Filler (1966).

* The ranges and all percentiles have been rounded off to the nearest 0.25 in.

Note: The measuring technique was as follows: The fixed blade of the beam caliper was placed at the outside of the subject's preferred arm. The subject then closed the corresponding eye and sighted the target-grip cylinder with his other eye while holding his head as straight as possible. The sliding blade of the beam caliper was then moved in until it began to cut off the subject's view of the free end of the target grip. The value was then recorded. For this group of dimensions, N = 30. A breadth of 12.00 in. will accommodate approximately 95 percent of the Air Force population.

[1] See Fig. 66 and tables 85 through 89.

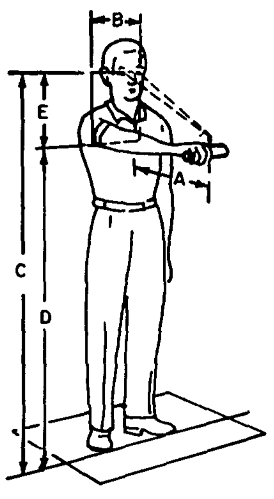


FIGURE 11-62. Standing forward reach, preferred arm (Kennedy and Filler, 1966).

TABLE 11–82. STANDING, LATERAL REACH (PREFERRED ARM)

	Percentiles (in.)*				
	5th	25th	50th	75th	95th
A. Depth of reach Range: 21.75 to 28.63 Mean: 24.65 SD: 1.51	22.00	23.50	24.75	25.75	26.75
B. Breadth of aperture 10.00					
C. Floor to top of aperture Range: 58.25 to 70.00 Mean: 64.70 SD: 2.32	60.75	63.25	64.25	66.00	68.75
D. Floor to bottom of aperture Range: 51.25 to 61.75 Mean: 56.09 SD: 2.05	52.25	54.75	56.00	57.25	59.00
E. Vertical dimension of aperture	(1)	(1)	(1)	(1)	(1)

Kennedy and Filler (1966).

* The ranges and all percentiles have been rounded off to the nearest 0.25 in.

Note: The measuring technique for aperture breadth for one-handed lateral reach is similar to that used to measure aperture breadth for one-handed forward reach. A breadth of 10.00 in. will accommodate approximately 95 percent of the Air Force population.

[1] See Fig. 66 and Tables 86 through 90.

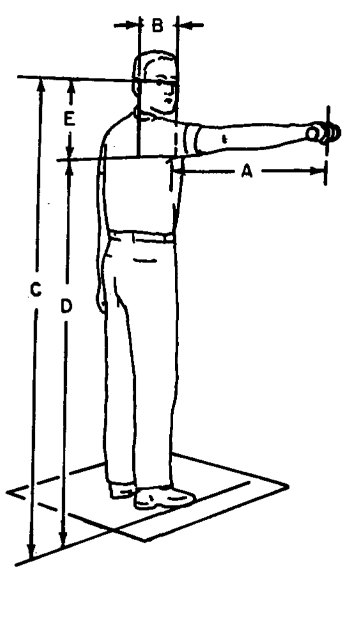


FIGURE 11–63. Standing lateral reach, preferred arm (Kennedy and Filler, 1966).

TABLE 11–83. SEATED, FORWARD REACH (BOTH ARMS)

	Percentiles (in.)*				
	5th	25th	50th	75th	95th
A. Depth of reach Range: 14.00 to 23.50 Mean: 18.26 SD: 2.15	15.00	16.50	17.75	19.50	22.25
B. Breadth of aperture Range: 13.50 to 18.75 Mean: 16.12 SD: 1.25	13.75	15.25	16.00	17.00	18.25
C. Floor to top of aperture † Range: 39.25 to 51.00 Mean: 43.25 SD: 2.05	19.75	41.75	43.00	44.25	46.50
D. Floor to bottom of aperture † Range: 32.50 to 41.75 Mean: 36.59 SD: 1.59	34.25	35.50	36.50	37.50	39.00
E. Vertical dimension of aperture	([1])	([1])	([1])	([1])	([1])

Kennedy and Filler (1966).

* The ranges and all percentiles have been rounded off to the nearest 0.25 in.

† A platform was used to elevate the subject and chair to a level which was compatible with the measuring device. The height of the platform (13.00 in.) was subtracted so that the values given are those that would be found if the chair were on the floor.

[1] See Fig. 66 and Tables 86 through 90.

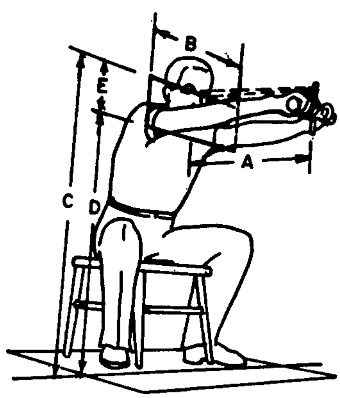


FIGURE 11–64. Seated forward reach, both arms (Kennedy & Filler, 1966).

TABLE 11–84. CROSS-LEGGED SEATED, FORWARD REACH (BOTH ARMS)

	Percentiles (in.)*				
	5th	25th	50th	75th	95th
A. Depth of reach Range: 13.50 to 22.25 Mean: 17.08 SD: 1.91	13.75	15.75	16.75	18.25	20.00
B. Breadth of aperture Range: 13.50 to 18.50 Mean: 15.89 SD: 1.54	13.75	14.75	16.00	16.75	17.75
C. Floor to top of aperture Range: 22.25 to 30.50 Mean: 25.30 SD: 1.54	22.75	24.25	25.25	26.25	28.00
D. Floor to bottom of aperture Range: 17.00 to 23.25 Mean: 19.23 SD: 1.19	17.00	18.50	19.25	20.00	21.25
E. Vertical dimension of aperture	(1)	(1)	(1)	(1)	(1)

Kennedy and Filler (1966).
* The ranges and all percentiles have been rounded off to the nearest 0.25 in.
[1] See Fig. 66 and Tables 86 through 90.

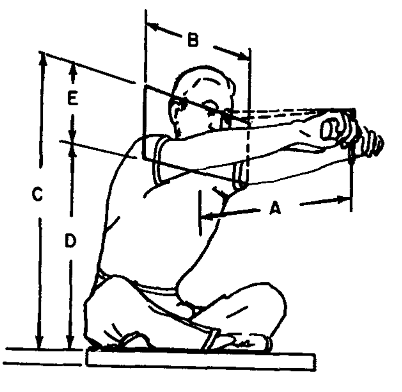


FIGURE 11–65. Cross-legged, seated, forward reach, both arms (Kennedy & Filler, 1966).

TABLE 11–85. RECOMMENDED APERTURE SIZES AND DEPTH OF REACH FOR SHIRT-SLEEVED TECHNICIANS

	Standing positions *			Seated positions *	
	Forward reach, both arms	Forward reach, preferred arm	Lateral reach, preferred arm	Normal, both arms	Cross-legged, both arms
A. Depth of reach, 5th percentile	19.25	20.25	22.00	15.00	13.75
B. Breadth of aperture, 95th percentile	19.50	12.00	10.00	18.25	17.75
C. Floor to top of aperture, 95th percentile	69.00	69.00	68.75	46.50	28.00
D. Floor to bottom of aperture, 5th percentile	52.25	52.25	52.25	34.25	17.00
E. Vertical dimension of aperture (C minus D; see Fig. 11–66)	16.75	16.75	16.50	12.25	11.00

Kennedy and Filler (1966).
* The ranges and all percentiles have been rounded off to the nearest 0.25 in.

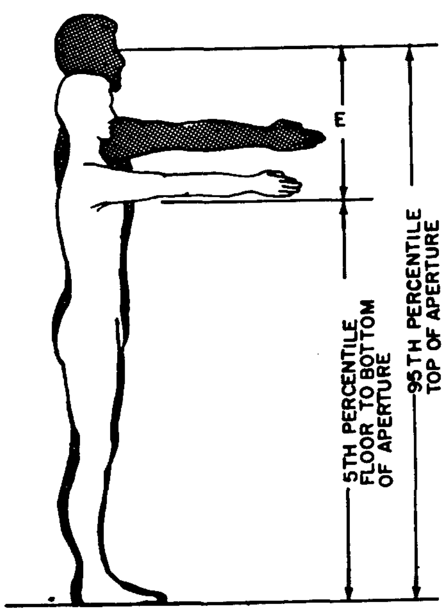


FIGURE 11–66. Reach through vertical aperture (Kennedy & Filler, 1966).

TABLE 11–86. APERTURE SIZES AND DEPTHS OF REACH FOR TECHNICIANS WEARING THE A/P 22S-2 FULL-PRESSURE SUIT

	Standing, forward reach (both arms)				
	5th*	95th*	Mean	S.D.	Range
A. Depth of reach:					
Vented	16.00	-----	19.26	1.57	16.00 to 21.75
1 psi	14.25	-----	17.01	1.62	14.00 to 19.50
3½ psi	7.50	-----	11.30	2.60	7.50 to 15.75
B. Breadth of aperture:					
Vented	-----	23.00	19.24	1.34	17.00 to 23.25
1 psi	-----	25.00	21.18	1.81	18.50 to 25.25
3½ psi	-----	26.00	23.60	1.84	19.25 to 26.00
C. Floor to top of aperture:					
Vented	-----	68.00	63.86	2.88	59.75 to 68.00
1 psi	-----	65.75	62.61	2.01	59.75 to 66.00
3½ psi	-----	66.25	60.38	2.58	56.25 to 66.50
D. Floor to bottom of aperture:					
Vented	50.75	-----	53.49	2.58	50.25 to 57.50
1 psi	50.00	-----	52.38	1.78	49.75 to 55.50
3½ psi	46.50	-----	50.80	2.33	46.25 to 55.50
E. Vertical dimension of aperture (C minus D)†					
Vented	17.25				
1 psi	15.75				
3½ psi	19.75				

Kennedy and Filler (1966).
*Ranges and percentiles have been rounded to nearest 0.25 in.

†Note that these values are not percentiles.

TABLE 11–87. APERTURE SIZES AND DEPTHS OF REACH FOR TECHNICIANS WEARING THE A/P 22S-2 FULL-PRESSURE SUIT

	Standing, forward reach (preferred arm)				
	5th*	95th*	Mean	S.D.	Range
A. Depth of reach:					
Vented	18.00	-----	20.91	1.78	18.00 to 24.00
1 psi	15.25	-----	18.25	1.77	15.00 to 21.00
3½ psi	11.00	-----	14.46	2.25	10.75 to 20.00
B. Breadth of aperture:					
Vented	-----	11.50	10.17	.79	9.25 to 11.50
1 psi	-----	12.50	11.40	.85	9.75 to 12.50
3½ psi	-----	14.00	13.09	.91	11.00 to 14.00
C. Floor to top of aperture:					
Vented	-----	67.75	63.71	2.77	60.25 to 68.00
1 psi	-----	65.75	62.19	2.09	58.75 to 66.00
3½ psi	-----	64.25	59.53	2.31	55.75 to 64.50
D. Floor to bottom of aperture:					
Vented	50.50	-----	53.49	2.58	50.25 to 57.50
1 psi	50.00	-----	52.38	1.78	49.75 to 55.50
3½ psi	46.50	-----	50.80	2.33	46.25 to 55.50
E. Vertical dimension of aperture (C minus D):†					
Vented	17.25				
1 psi	15.75				
3½ psi	17.75				

Kennedy and Filler (1966).
*Percentiles have been rounded to nearest 0.25 in.

†Note that these values are not percentiles.

TABLE 11-88. APERTURE SIZES AND DEPTHS OF REACH FOR TECHNICIANS WEARING THE A/P 22S-2 FULL-PRESSURE SUIT

	Standing, lateral reach (preferred arm)				
	5th*	95th*	Mean	S.D.	Range
A. Depth of reach:					
Vented	20.00	-----	22.52	1.17	20.00 to 24.75
1 psi	19.25	-----	21.70	1.45	19.00 to 25.00
3½ psi	15.25	-----	19.56	1.88	15.00 to 22.00
B. Breadth of aperture:					
Vented	-----	10.25	9.84	.37	9.00 to 10.25
1 psi	-----	11.50	10.73	.71	9.75 to 11.50
3½ psi	-----	14.00	12.76	.89	11.25 to 14.00
C. Floor to top of aperture:					
Vented	-----	66.75	62.70	2.79	58.75 to 67.00
1 psi	-----	65.75	61.54	2.15	57.75 to 66.00
3½ psi	-----	62.75	58.78	2.07	55.75 to 63.00
D. Floor to bottom of aperture:					
Vented	50.50	-----	53.49	2.58	50.25 to 57.50
1 psi	50.00	-----	52.38	1.78	49.75 to 55.50
3½ psi	46.50	-----	50.80	2.33	46.25 to 55.50
E. Vertical dimension of aperture (C minus D):†					
Vented	16.25				
1 psi	15.75				
3½ psi	16.25				

Kennedy and Filler (1966).
*Ranges and percentiles have been rounded to nearest 0.25 in.
†Note that these values are not percentiles.

TABLE 11-89. APERTURE SIZES AND DEPTHS OF REACH FOR TECHNICIANS WEARING THE A/P 22S-2 FULL-PRESSURE SUIT

	Seated, forward reach (both arms)				
	5th*	95th*	Mean	S.D.	Range
A. Depth of reach:					
Vented	13.00	-----	15.82	1.44	13.00 to 18.00
1 psi	11.50	-----	13.96	1.39	11.50 to 15.50
3½ psi	5.00	-----	8.05	2.02	5.00 to 12.00
B. Breadth of aperture:					
Vented	-----	22.50	19.03	1.68	16.25 to 22.75
1 psi	-----	25.00	21.08	1.94	17.50 to 25.00
3½ psi	-----	22.50	19.57	2.22	14.50 to 22.50
C. Floor to top of aperture:					
Vented	-----	45.75	43.62	1.70	40.25 to 45.75
1 psi	-----	46.50	43.46	1.41	40.50 to 46.50
3½ psi	-----	46.50	43.31	1.79	39.25 to 46.50
D. Floor to bottom of aperture:					
Vented	32.25	-----	35.59	1.50	32.25 to 38.00
1 psi	32.25	-----	35.49	1.36	32.25 to 38.00
3½ psi	33.50	-----	35.46	1.24	32.25 to 38.00
E. Vertical dimension of aperture (C minus D):†					
Vented	13.25				
1 psi	14.00				
3½ psi	12.75				

Kennedy and Filler (1966).
* Ranges and percentiles have been rounded to nearest 0.25 in.
†Note that these values are not percentiles.

Slender men and women have the widest range of joint movement. Average and muscular body builds, in descending order, have intermediate ranges (Barter et al., 1957). These differences are often significant, especially those between the thin and the fat groups, where variations of more than 10° in a given movement are not uncommon (Sinelnikoff and Grigorowitsch, 1931).

Physical exercise can increase the range of motion of a joint, but excessive exercise can result in the so-called "muscle-bound" condition in which the range of motion is actually reduced. Some specialized tasks involve the repetition of certain body movements, and, as a result, the range of movement at the affected joints increases.

11.6.3 Posture and Body Position

The range of movement of one part of the body is affected by the position or movement of neighboring parts, e.g., hand rotation can be considerably increased if shoulder movements are added to those at the elbow, and wrist flexion is greater with the hand pronated than it is with the hand supinated. In addition, prone-position movements are not necessarily exactly the same as those made from the standing position.

11.6.4 Static Range-of-Movement Measurement

Joint movement is measured at the angle formed by the long axes of two adjoining body segments (link lines) or, in some cases, at the angle formed by one body segment and a vertical or horizontal plane. The total range of movement is measured between the two extreme positions of the joint. The ranges, in angular degrees, of each of the types of voluntary movement possible at the joints of the body are presented in Tables 11–90 through 11–93. The types of movement measured are as follows (See Figures 11–67, 11–68 and 11–69):

1. Flexion. Bending, or decreasing the angle between the parts of the body.
2. Extension. Straightening, or increasing the angle between the parts of the body.
3. Adduction. Moving toward the midline of the body.
4. Abduction. Moving away from the midline of the body.
5. Medial rotation. Turning toward the midplane of the body.
6. Lateral rotation. Turning away from the midplane of the body.
7. Pronation. Rotating the palm of the hand *downward.*
8. Supination. Rotating the palm of the hand *upward.*

11.6.5 Dynamic Range-of-Movement Measurement

An example of dynamic anthropometry involving body movements is a study of head and eye movements undertaken in connection with gunsight and sighting-panel design (Brues, 1946). Figure 11–70 shows the location of the head and eyes at various angles of gaze, and is based on data taken from measuring the movements of 21 Air Force men. As can be seen, eye movement is a nearly circular curve from 90° above to 45° the horizontal. The ear, on the other hand, does not move in a circular arc, as was commonly supposed, but follows a paraboloid curve. Finally, the pivot point or center of the arc described by the eye is not at the ear but roughly 2 in. below and ½ in. behind it.

11.7 Muscle Strength and Muscle Power

The data appearing in this section are intended only as guidelines, indicating trends and orders of magnitude of human force output. Like body-size data, they must not be construed as valid for all groups. The data have been extracted mostly from American and some European male populations, who are large men by world standards. As effective muscle-strength capability varies with sex, age, body size, physical conditions, motivation, and many other factors, the general data presented herein cannot be construed as universally valid. They are approximations only for their own groups. Furthermore, muscle strength is situation-specific; it varies within the same person according to body position, and up to now there is no single strength test in one position by which strength in other positions can be safely predicted.

Table 11–90. Average increase in range of joint movement of women over men

Movement	Difference (deg)
Wrist flexion and extension	14
Wrist adduction and abduction	11
Elbow flexion and extension	8
Shoulder abduction (rearward)	2
Ankle flexion and extension	4
Knee flexion and extension	0
Hip flexion	3

Sinelnikoff and Grigorowitsch (1931).

Table 11–91. Range of movement at the joint of the neck of male civilians

Movement*	Range (deg)	
	Avg.	S.D.
Ventral flexion	60	12
Dorsal flexion	61	27
Right or left flexion	41	7
Right or left rotation	79	14

Glanville and Kreezer (1937).
* See Fig. 11–67.

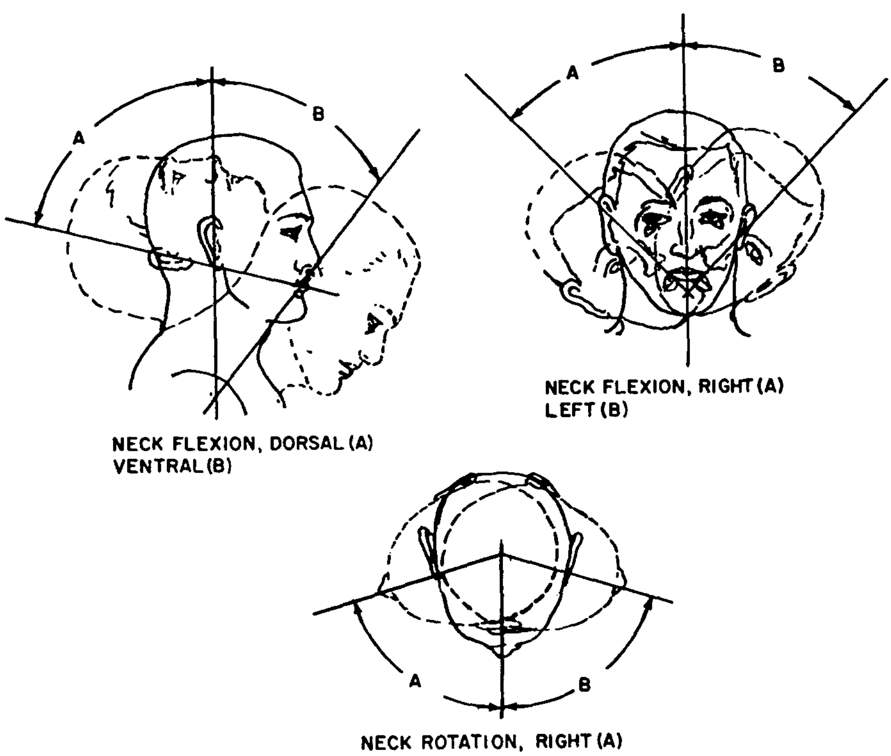


Figure 11–67. Neck movement and rotation.

TABLE 11–92. RANGE OF MOVEMENT AT THE JOINTS OF THE HAND AND ARM OF MALE AIR FORCE PERSONNEL

Movement*	Range (deg) Avg.	S.D.
Wrist flexion	90	12
Wrist extension	99	13
Wrist adduction	27	9
Wrist abduction	47	7
Forearm supination	113	22
Forearm pronation	77	24
Elbow flexion	142	10
Shoulder flexion	188	12
Shoulder extension	61	14
Shoulder adduction	48	9
Shoulder abduction	134	17
Shoulder rotation:		
Medial	97	22
Lateral	34	13

Barter et al. (1957).
* See Fig. 11–68.

TABLE 11–93. RANGE OF MOVEMENT AT THE JOINTS OF THE FOOT AND LEG OF MALE AIR FORCE PERSONNEL

Movement *	Range (deg) Avg.	S.D.
Ankle flexion	35	7
Ankle extension	38	12
Ankle adduction	24	9
Ankle abduction	23	7
Knee flexion:		
Standing	113	13
Kneeling	159	9
Prone	125	10
Knee rotation:		
Medial	35	12
Lateral	43	12
Hip flexion	113	13
Hip adduction	31	12
Hip abduction	53	12
Hip rotation (sitting):		
Medial	31	9
Lateral	30	9
Hip rotation (prone):		
Medial	39	10
Lateral	34	10

Barter et al. (1957).
* See Fig. 11–69.

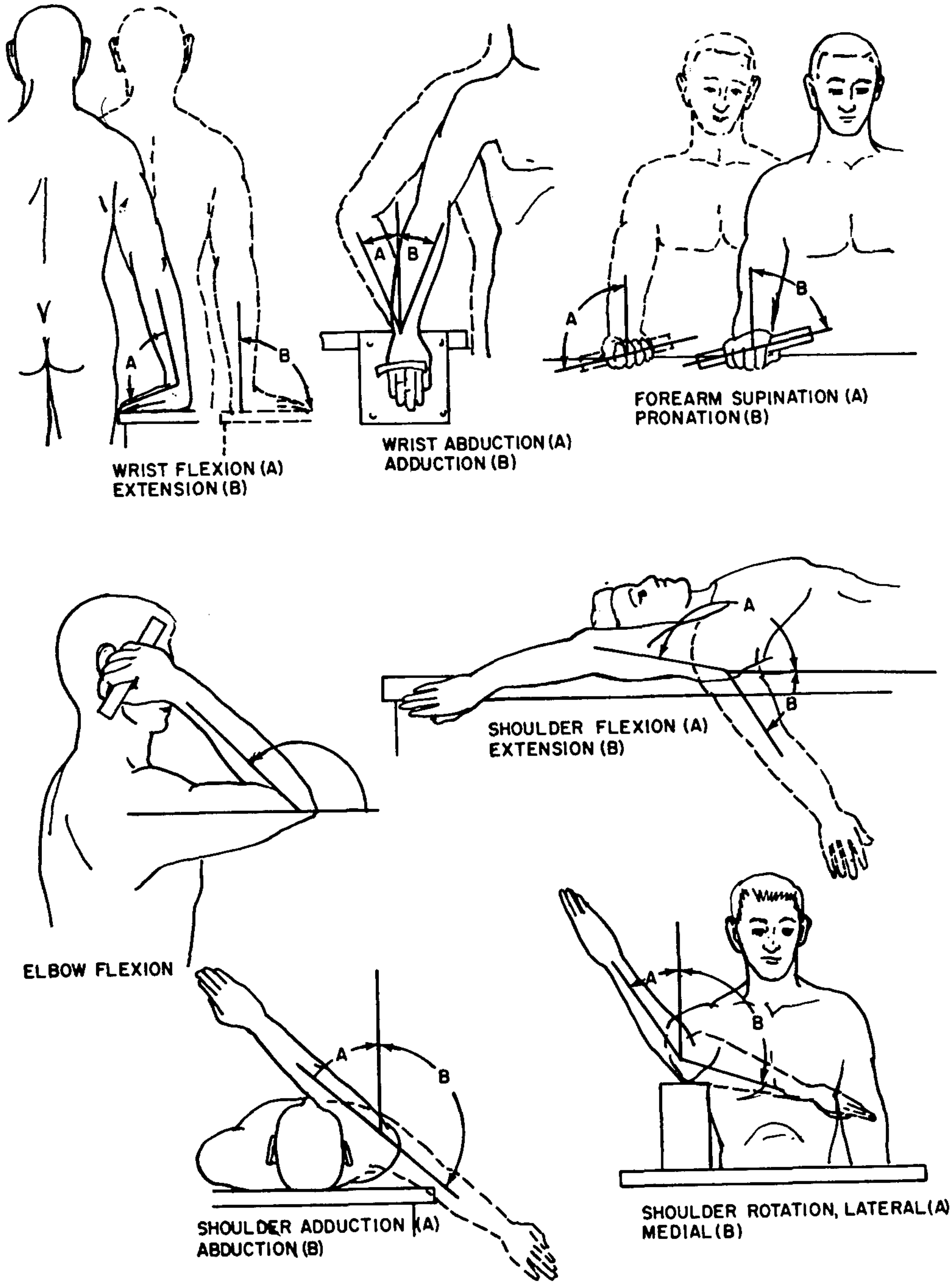


FIGURE 11–68. Wrist, shoulder, and elbow movements.

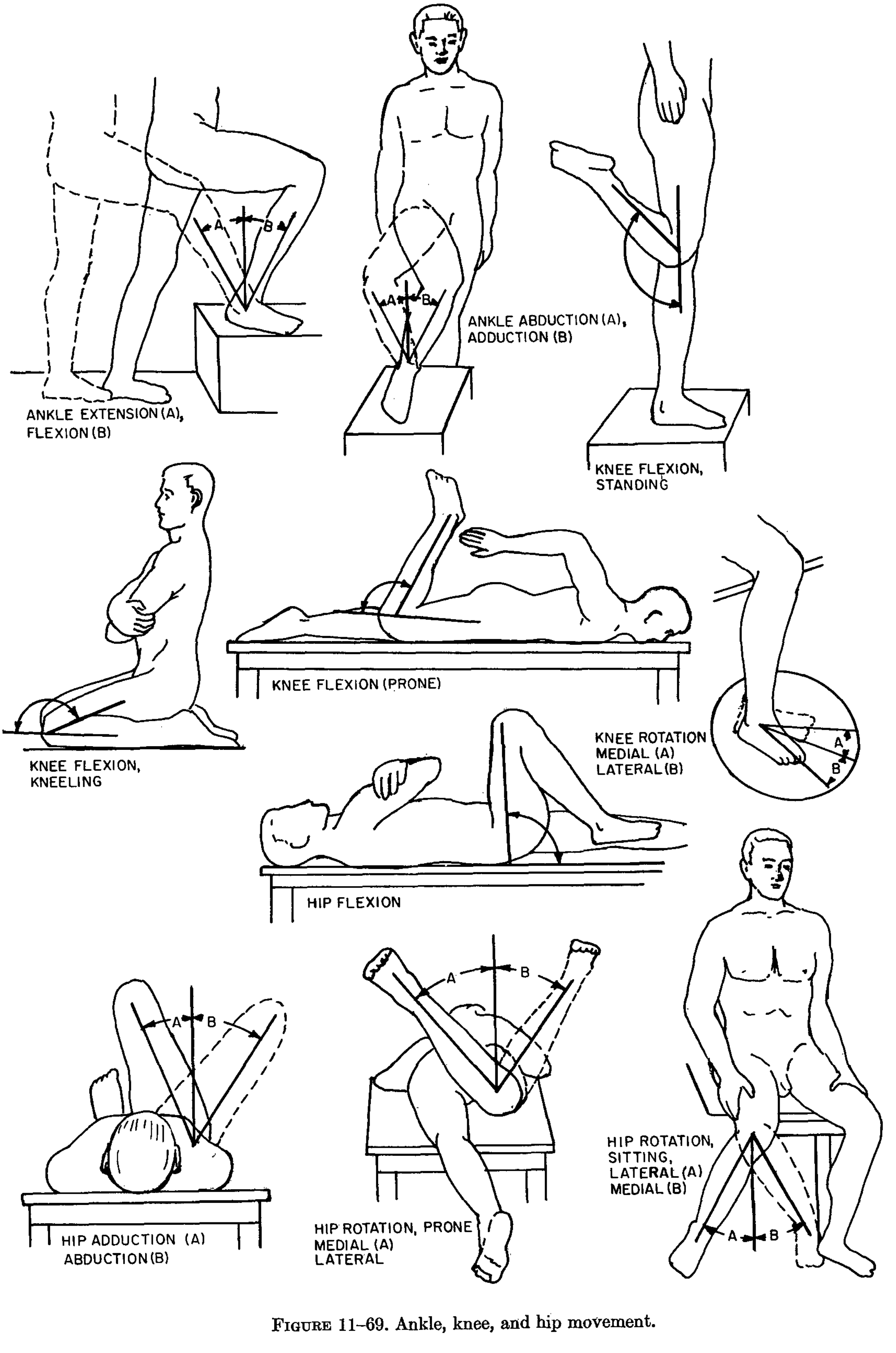


FIGURE 11–69. Ankle, knee, and hip movement.

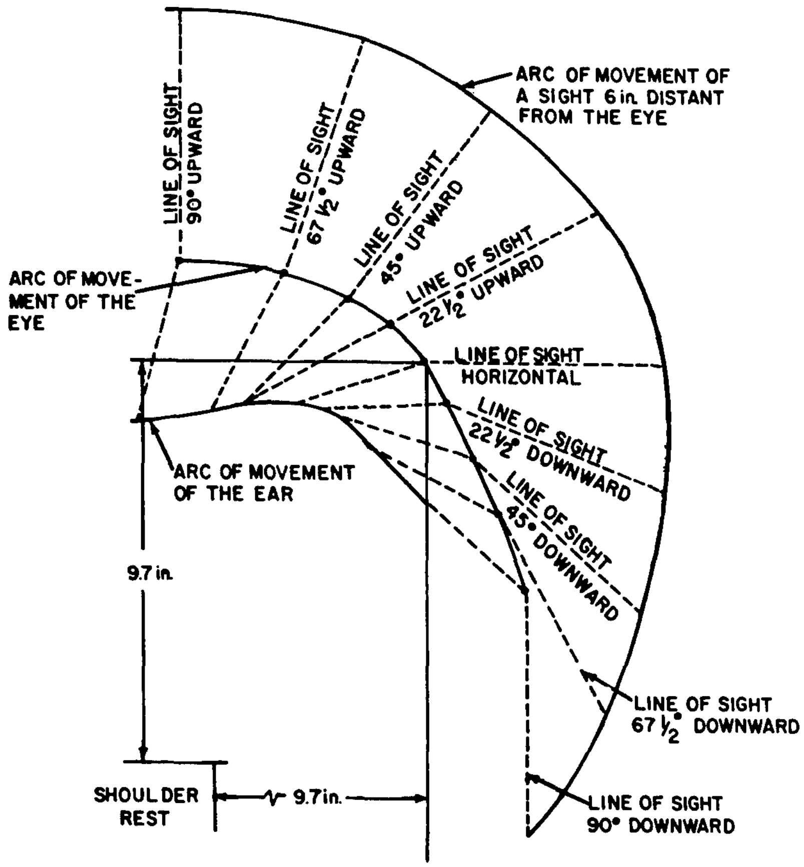


FIGURE 11–70. Angles of sight at various head positions (after Brues, 1946).

The voluntary muscles of the body stabilize the joints, maintain desired body positions, rotate the body segments around the joints, and transmit energy to objects outside the body such as controls or tools. Data on voluntary muscle strength can be used by the design engineer to determine maximum and optimum control resistances, forces required in other manual tasks, and the arrangement of weight for safe, efficient lifting or carrying.

The maximum resistance of a control should be low enough to be overcome by the weakest potential operator. This value should not be exceeded. "Optimal" or "operational" resistance levels, however, should not require the application of maximum power by any operator. Operational resistance levels affect comfort and efficiency and, therefore, should be low enough to prevent fatigue or discomfort, but high enough to prevent inadvertent operation of the control and to provide kinesthetic cues to control movement.

As with control resistance, schedules for other physical tasks or for lifting or carrying must be set with regard to the capability of the weakest operator. Performance at or near the limits of physical abilities should not be required because fatigue will develop rapidly, causing the performance of the operator to deteriorate rapidly. In extreme cases, actual injury (muscle strain or rupture, strained or torn ligaments) might occur.

11.7.1 How the Muscles Work

The amount of force developed by the muscle depends primarily on four factors:

1. The cross-sectional thickness of the muscle: for each square centimeter of muscle tissue, forces of 3 to 4 kilopond can be exerted (Hettinger, 1961; Lehmann, 1961; Scherrer, 1967). ("Pond" is the term standardized in recent years to distinguish force from mass, the latter being still expressed by "gram." It avoids the

ambiguity resulting from the use of "gram" as a unit of force as well as a unit of mass. A mass of 1 gram exerts a force of 1 pond at standard gravity. Thus, one kilopond (kp.) equals 2.2 pounds.)

2. Adequate supply of oxygen to the muscle and removal of carbon dioxide from it.

3. Muscle tension, which "reaches a maximum when the length is greatest, and there is, momentarily, no change in length" (Elftman, 1941). The tension, or contractile force, decreases as the muscle shortens and as its rate of shortening increases (Arkin, 1941; Darcus, 1951; and Elftman, 1941).

4. The mechanical advantage of the body's lever system. The long bones are the lever arms, the joints the fulcra. Force is applied at the points of muscle attachment.

All but the second factor vary with changes at the joints. At the elbow, for example, the force of extension is greatest when the elbow is flexed. This position permits the triceps to attain its greatest length but does not provide the optimum mechanical advantage. In elbow flexion, however, the force is greatest toward the midpoint of the full range of movement (Darcus, 1954). Here the increasing mechanical advantage of the forearm relative to the upper arm more than compensates for the reduction in strength caused by the shortening of the biceps (Haggard, 1946). Although human muscles in maximum contraction can exert large forces—exceeding 1000 lb. in the line of the tendon—the forces applicable to a control handle or pedal are much lower. This is because the mechanical linkages of segments at the joints have evolved to facilitate fast movement rather than exertion of large forces.

11.7.2 Muscle Strength and Working Capacity: Definitions

Many factors—biological, environmental, and occupational—affect muscle strength and working capacity and should be considered in design.

Strength is the maximal force exertable by a muscle, theoretically for an instant of time, in practice for relatively few seconds. Strength is most frequently measured isometrically—i.e., as the maximal force the muscle can exert while its length remains constant—and hence is often called "static" force. Ounces and pounds are the units of isometric force in the English system, and ponds and kiloponds in the metric system.

Working capacity is the amount of external physical work that can be performed. Working capacity is expressed in units of energy (such as ft.-lb., kp.-m., joules). Power is the ratio of the amount of work done to the time needed for it (expressed in horsepower or watts). The working capacity of an operator is normally limited not by his strength but by his endurance (the efficiency of his circulatory system). During physical work, the tension of a muscle changes as well as its length; i.e., it contracts and relaxes alternately. This is often described as "dynamic" work.

11.7.3 Factors Affecting Muscle Strength and Working Capacity

Biological Factors

Age and sex affect human strength. Summarizing various studies (Asmussen and Heeboll-Nielsen, 1962; Hettinger, 1960, Hunsicker, 1955, 1957), the following rough estimates can be made: At about ten years, boys and girls are equally strong. Strength increases rapidly in the teens, but faster among boys than girls. Among men, strength reaches a maximum in the middle to late twenties. It remains on this level for five to ten years and then drops slowly but increasingly. Approximately 95% of the earlier strength is still available at about forty, and 80% at fifty to sixty. Women reach their maximal strength in the early twenties, and remain on this level for about ten years. The final decline in strength is steeper among women than among men. Comparing men and women at about age 30, women can exert approximately 2/3 of men's forces, whereas in their fifties they exert only about half.

Muscle-strength decrement, however, does not proceed at the same rate in all parts of the body. Force capabilities of hand and arm are less affected by age than those of trunk and leg (Asmussen and Heeboll-Nielsen, 1962; Damon et al., 1966).

Working capacity depends on age and sex in about the same manner as strength. While ability for the short-term burst of energy distinctly declines after the twenties, ability for prolonged exertion of energy (i.e., for developing an oxygen deficit) is largely maintained into the forties (Consolazio et al., 1963; Karpovich, 1965; Knuttgen, 1967; Lehmann, 1962; Wilmore and Sigerseth, 1967). In general, the working capacity of women is ⅔ to ⅘ of that of men.

Body build is related to strength and working capacity. Significant but often low correlations have been found between strength and body weight, circumferences, and lengths. Among schemes of body-build classifications (somatotypes), only mesomorphy (athletic build) showed significant positive correlations with strength. For endurance and sustained dynamic work, however, normal or even slender build seems to be advantageous (Damon et al., 1962, 1966; Caldwell, 1963; Laubach and McConville, 1966; Roberts et al., 1959; Tornvall, 1963).

Handedness has relatively small effect on strength and working capacity. In right-handed persons (approximately 90% of all), the right arm is slightly stronger than the left, though usually less than 10% (Damon, et al., 1966; Hunsicker, 1955). Strength differences between the two legs are even less obvious (Elbel, 1949; Hugh-Jones, 1947; Mueller, 1934). For practical purposes, the slight differences in strength of the two sides of the body can be neglected (Hunsicker and Greey, 1957).

Fatigue greatly reduces strength and work capacity. There is a nonlinear, inverse relationship between the fraction of the strength which must be exerted and the time during which it can be exerted—either continuously or sporadically. Maximal force (100% of strength) can be exerted for only a few seconds, and only a fraction—15% to 20%—of maximal strength can be maintained over many hours without fatigue (Caldwell, 1961, 1964; Caldwell and Smith, 1966; Mueller, 1961; Rohmert, 1960a, 1961; Scherrer, 1967).

A similar relationship exists between the amount of dynamic work required and the time over which it can be performed (Clarke, 1962; Hueting and Sarphati, 1966). An increasing pulse rate during work indicates that the operator's body—i.e., his cardiac capacity— is being overstrained and that increasing fatigue will force him to terminate the work (Fascenalli and Lamb, 1966). If the pulse rate levels off during work, the operator is not suffering from fatigue and can go on for many hours. Normally no fatigue occurs if the pulse rate is not more than 40 beats per minute above resting levels (basal pulse rate). This indicates, very roughly, an energy expenditure of ⅓ of the operator's maximal capacity for dynamic work (Rohmert, 1960b; Mueller, 1961).

Fatigue can be postponed or avoided by lowering the required amount of force or energy exertion, which may be done either by reducing its intensity or interrupting the work frequently for short rest pauses. The operator should also be enabled to change his body posture at will and to select for himself the most suitable method to perform his tasks.

Exercise (training) can improve an operator's force and working capabilities significantly within his innate physical potential. Such improvement may be accomplished by simple isometric exercise (Hettinger, 1961; Mueller, 1962), although previous exaggerated hopes for them have not been fulfilled (Royce, 1964). To increase the operator's capacity for dynamic work, exercises should be performed which strengthen both his circulatory system and his muscles (Mueller, 1962). There is only a slight relationship between improvement in dynamic and static capability (Bender and Kaplan, 1966; Berger, 1962; Berger and Henderson, 1966; Kogi et al., 1965).

Environmental Factors

Altitude affects physical capabilities, especially those for sustained energy expenditure. Effects of reduced oxygen pressure of the ambient air depend on altitude, acclimatization, individual variability, and on the nature of the operator's task. Ability for static muscle efforts and energy bursts is little affected by altitude if the duration is less than one minute. Thus, sea-level grip strength could be maintained up to about 23,000 ft., but declined sharply at higher altitudes (Bruener, 1961). Endurance time for moderate muscular work may decrease at altitudes as low as 6,500 ft. (Balke et al., 1965). At 10 to 12,000 ft. the amount of maximal work exertable is usually about 10 to 15% less than at sea level. Men capable of acclimatization to high altitude can restore much of their sea-level capability, but may require weeks or months for it. At about 20,000 ft., even fully acclimatized men have only about half as much muscular endurance as at sea level (Dill et al., 1967; Hansen et al., 1967; Pugh, 1964; Pugh et al., 1964; Vogel et al., 1967). Performance of physical work is deemed impossible above approximately 29,000 ft. (Bruener, 1961); yet, mountain climbing was performed without supplementary oxygen supply up to about 28,000 ft. (Pugh and Ward, 1953). "The exact reason for cessation of work is as obscure at altitude as at sea level" (Consolazio et al., 1966).

Temperature

High ambient temperatures, over approximately 85°F, especially when combined with high humidity, markedly reduce endurance for muscular work, but do not greatly affect short outbursts of energy (Wenzel, 1964; Wyndham, 1962). Unacclimatized men tolerate heat better than unacclimatized women (Morimoto et al., 1967; Weinman et al., 1967).

Low ambient temperatures per se do not greatly reduce the capability to perform physical work, as protective clothing can prevent heat losses of the operator; bulkiness of the clothing, however, can interfere severely with the task. Wind and humidity in addition to cold temperatures can make it very difficult to maintain adequate skin temperatures, especially of the head, hands, and feet (Lehmann, 1962). Dexterity, tactile sensitivity, and overall manual performance can be severely reduced by cold temperatures. Adaptation and acclimatization help to overcome the adverse effects of low ambient temperatures (Fisher, 1957; Fox, 1967).

Occupational Factors

The body position imposed on an operator by his working conditions, by machine design, often determines his force output and working efficiency. For example, a coal miner, standing in a high tunnel, can swing his pick with ease, because his body and load are dynamically balanced; the same man, lying on his side in a low tunnel, can swing it only with great difficulty; the drastic change in body support and balance of load have severely reduced his efficiency. Similarly, the push forces which a seated, braced man can exert are greater than those he can exert when standing upright and unbraced, or even when he is leaning against the object. Hence, the workplace designer must give some thought to the working conditions he creates.

Gravity

The muscle forces that can be exerted on earth at 1G have to be distinguished from those at less than 1 G as in many underwater operations or on the moon, or at 0G as in space. Readers who require information on human work output at less than 1G may find it in Gauer and Haber (1950); Dzendolet and Rievley (1959); Hertzberg (1960); Dzendolet (1960); Streimer, Springer and Tardiff (1964).

At 1G there is a definite difference between an operator's static strength and his ability to perform prolonged dynamic work; data on static strength cannot be applied if the operator has to work dynamically, and vice versa (Kroemer, 1967a). For this reason the data have been divided into three parts:

1. Maximal static forces.
2. Weight-lifting capability.
3. Dynamic working capacity.

Only a few real-world working situations are represented here, and the following data should not be taken as exact figures for all groups and situations, but only as approximations.

11.7.4 Maximal Static Forces

Static (isometric) force as applied to a control means the exertion of muscle force such that the muscles tighten but do not change their length during this tension. The control does not move, or moves negligibly, in relation to the operator's body.

With only few exceptions, the tables present data on maximal efforts. The forces called "optimal," "operational," or "comfortable" are well below maximal strength.

Maximal Static Hand Forces Exerted in Various Standing Positions

Table 11–94 presents forces which were measured on subjects standing erect with their feet 12 in. apart and parallel, without any structures to brace themselves against. In such

TABLE 11–94. MAXIMAL RIGHT-HANDED STATIC FORCES AND TORQUES EXERTED ON A VERTICAL HAND-GRIP BY STANDING SUBJECTS (FEET PARALLEL, 12 IN. APART; N=5)

	Angle (deg)	Location of the handgrip		
		At percentages of maximal grip distance		
		50%	75%	100%
		Force in lb[1]		
Push, horizontal	30	16	24	32
	0	30	35	40
	−30	28	30	32
	−60	28	32	36
Pull, horizontal	30	19	22	26
	0	23	26	29
	−30	28	30	31
	−60	23	28	34
To the left, horizontal	30	35	30	24
	0	42	33	24
	−30	42	34	26
	−60	33	30	26
To the right, horizontal	30	24	22	21
	0	30	25	20
	−30	33	27	22
	−60	25	23	21
Up, vertical	30	28	24	19
	0	34	26	18
	−30	50	40	28
	−60	63	51	41
Down, vertical	30	76	58	41
	0	56	40	33
	−30	35	33	30
	−60	39	36	32
		Torque in ft.-lb.[2]		
Clockwise (supination)	30	12.9	12.2	5.8
	0	13.5	11.7	4.3
	−30	11.5	9.8	3.6
	−60	8.5	6.7	4.9
Counterclockwise (pronation)	30	12.4	11.0	8.0
	0	13.2	11.4	8.0
	−30	15.3	13.4	8.7
	−60	16.8	14.8	10.1

Rohmert (1966) (5 male subjects, selected to match another group of 60).

[1] See Fig. 11–71.
[2] About a horizontal axis at the center of the handgrip.

a case, the forces exerted do not reflect the actual strength of the operators, but only the amount of reaction force available to them. The requirement to stay erect limits the horizontal forces exertable; in this body position, hand forces exerted downward are limited by the operator's weight. (See Figure 11–71.)

If an operator is allowed to bring his body into the most effective position, and if there are rigid surfaces available to which he can anchor his body, very high push forces may be exerted, as Table 11–95 indicates.

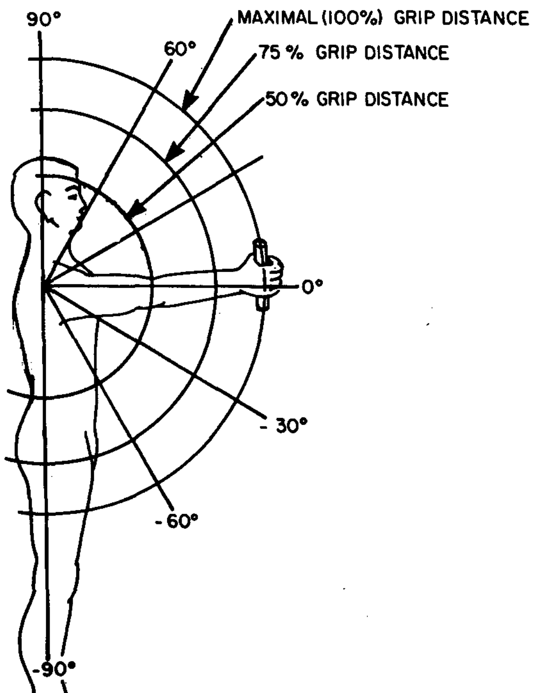


FIGURE 11–71. Hand forces in standing positions (Rohmert, 1966).

Maximal Static Hand Forces Exerted in the Sitting Position

As there are numerous studies of the static forces which sitting subjects can exert (Tables 11–96 through 11–101 and Figures 11–72 through 11–75), a comparison of the data permits the following conclusions:

1. Different samples exert different forces.
2. The force output depends on (a) the position of the arm; (b) the direction of the thrust; (c) the type and location of the body bracing; (d) the instructions of the investigator.
3. Using both arms instead of one usually results in less than double the force output.
4. Strength differences between the two sides of the body are negligible.

Maximal Static Hand Forces Exertable in the Prone Position

The force data presented so far have been taken with the body standing or seated—i.e., with the pull of gravity approximately parallel to either the longitudinal axis of the body or the torso axis alone. Is intrinsic force output reduced when the body axis is perpendicular to gravity, as in the prone position? In a practical situation, as is discussed below, this may be the case, but at present that question cannot be positively answered for lack of comparable data (See Figure 11–76.)

Unfortunately, the data in Tables 11–96 (push and pull forces for the seated position) and 11–102 (push and pull forces for the same men when prone) cannot be used to settle the question, because the two sets of conditions were simply not comparable, for two major reasons.

The first reason is that the conditions of body bracing were totally different. Although the subjects wore no restraint harnesses either seated or prone, the seated subjects nevertheless were braced by the structure—they pulled against foot rests and pushed against the seat back. Their forces were thus considerably augmented by the "toggle" effect at the joints and the use of the large muscles in the legs and torso. In the prone position, the subject was not braced at the shoulders, back or feet, which left body friction on the supporting platform as the limiting factor for push or pull. The prone forces exerted in this test were therefore low.

The second reason is that the directions of arm thrust with respect to the longitudinal axis of the trunk were very different in the two body positions. The seated men thrust in a line straight ahead—an effective direction. The prone subjects thrust in a line over their heads, a direction permitting only relatively low forces. Even adequate bracing could not have remedied that condition.

TABLE 11–95. MAXIMAL STATIC PUSH FORCES EXERTED HORIZONTALLY IN VARIOUS STANDING POSITIONS ON A VERTICAL SURFACE BY MALE ADULTS

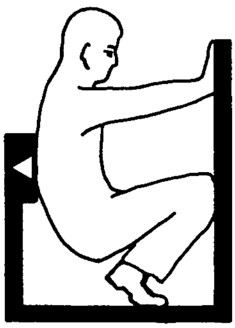
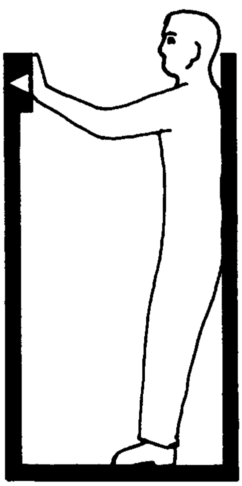
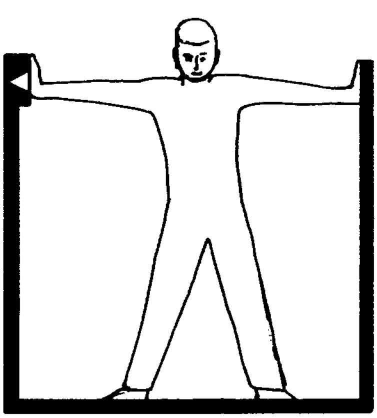
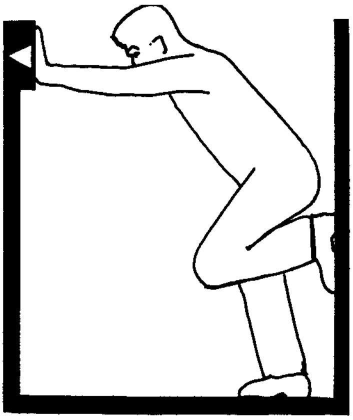
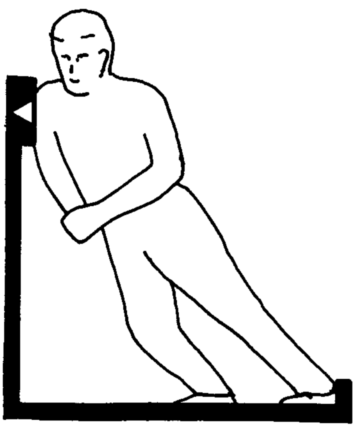
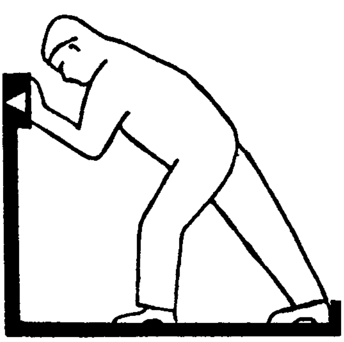
	Force-plate height*	Distances†	Number of subjects	Force (lb)	
				Means	S.D.
	40 percent of acromial height‡	80	43	428	166
		90	43	364	121
		100	43	372	113
		110	43	434	127
		120	43	441	134
		130 percent of thumb-tip reach§	43	397	116
				Both hands	
	100 percent of acromial height§	50	39	131	32
		60	40	150	36
		70	39	221	61
		80	40	289	90
		90	40	220	68
		100	35	145	57
				Preferred hand	
		50	39	59	15
		60	40	67	16
		70	39	81	22
		80	40	117	32
		90	40	111	38
		100 percent of thumb-tip reach§	35	96	39
	100 percent of acromial height§	50	30	83	31
		60	41	78	28
		70	41	117	37
		80	42	159	43
		90 percent of span¶	37	73	30

TABLE 11–95 (continued)

	Force-plate height*	Distances†	Number of subjects	Force (lb) Means	Force (lb) S.D.
	50	80	41	149	40
	50	100	42	174	48
	50	120	37	175	37
	70	80	41	161	36
	70	100	42	164	52
	70	120	37	184	31
	90	80	42	141	33
	90	100	42	152	44
	90	120	37	194	32
	percent of acromial height‡				
	60	70	43	171	38
	60	80	43	192	40
	60	90	43	178	32
	70	60	43	130	25
	70	70	43	157	28
	70	80	43	163	32
	80	60	43	117	29
	80	70	43	139	29
	80	80	43	143	30
	percent of acromial height‡				
	70	70	41	140	33
	70	80	41	155	35
	70	90	41	132	30
	80	70	41	123	28
	80	80	41	122	28
	80	90	41	120	18
	90	70	41	97	21
	90	80	41	101	21
	90	90	41	109	18
	percent of acromial height‡				

Kroemer (1968).

*Height of the center of the force plate (8 in. high by 10 in. long) upon which force is applied.

†Horizontal distance between the vertical surface of the force plate and the opposing vertical surface (wall or footrest, respectively) against which the subjects braced themselves.

‡ See shoulder height.

§See thumb-tip reach.

¶The maximal distance between a person's fingertips as he extends his arms and hands to each side.

TABLE 11–96. MAXIMAL STATIC HAND FORCES AT VARIOUS ELBOW ANGLES EXERTED ON A VERTICAL HANDGRIP BY SEATED MALES (N = 55)

A. Left hand						B. Right hand					
Direction of force	Elbow angle* (deg.)	Means	Percentiles (lb) 5th	95th	S.D.	Direction of force	Elbow angle* (deg.)	Means	Percentiles (lb) 5th	95th	S.D.
Push, horizontal	60	80	22	164	31	Push, horizontal	60	92	34	150	38
	90	83	22	172	35		90	87	36	154	33
	120	99	26	180	42		120		36	172	43
	150	111	30	192	48		150	123	42	194	45
	180	126	42	196	47		180	138	50	210	49
Pull, horizontal	60	64	26	110	23	Pull, horizontal	60	63	24	74	22
	90	80	32	122	28		90	88	37	135	30
	120	94	34	152	34		120	104	42	154	31
	150	112	42	168	37		150	122	56	189	36
	180	117	50	172	37		180	121	52	171	37
To the left, horizontal	60	32	12	62	17	To the left, horizontal	60	52	20	87	19
	90	33	10	72	19		90	50	18	97	23
	120	30	10	68	18		120	53	22	100	26
	150	29	8	66	20		150	54	20	104	25
	180	30	8	64	20		180	50	20	104	26
To the right, horizontal	60	50	17	83	21	To the right, horizontal	60	42	17	82	20
	90	48	16	87	22		90	37	16	68	18
	120	45	20	89	21		120	31	15	62	17
	150	47	15	113	27		150	33	15	64	18
	180	43	13	92	22		180	35	14	62	24
Up, vertical	60	44	15	82	18	Up, vertical	60	49	20	82	18
	90	52	17	100	22		90	56	20	106	22
	120	54	17	102	25		120	60	24	124	24
	150	52	15	110	27		150	66	18	118	28
	180	41	9	83	23		180	43	14	88	22
Down, vertical	60	46	18	76	18	Down, vertical	60	51	20	89	21
	90	49	21	92	20		90	54	26	88	20
	120	51	21	102	23		120	58	26	98	23
	150	41	18	74	16		150	47	20	80	18
	180	35	13	72	15		180	41	17	82	18

Hunsicker (1955).
Note: Subjects wore no restraint harness.

*See Fig. 11–72.

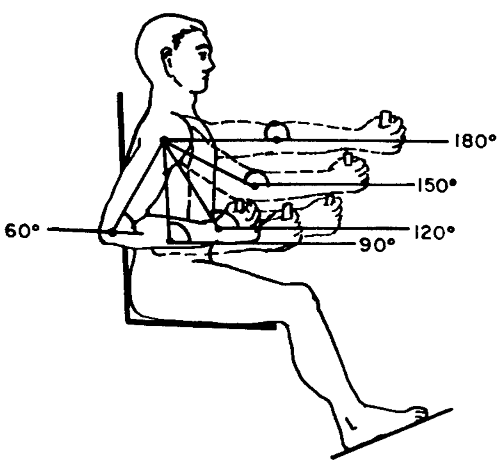


FIGURE 11–72. Maximal static hand forces at various elbow angles exerted on a vertical handgrip by seated males.

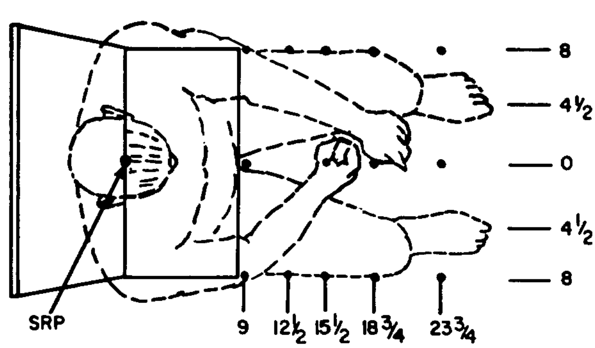


FIGURE 11–73. Maximal one -and two-handed static forces exerted on an aircraft control stick by seated males.

TABLE 11–97. MAXIMAL ONE- AND TWO-HANDED STATIC FORCES EXERTED ON AN AIRCRAFT CONTROL STICK BY SEATED MALES (ONE-HANDED, N = 48; TWO-HANDED, N = 15)

Handle distance from SRP (in.)*	Handle distance from midsagittal plane of body (in.)*	A. Push (Percentiles (lb)) Right hand† 5th	Right hand† 50th	Right hand† 95th	Both hands† 50th	B. Pull (Percentiles (lb)) Right hand† 5th	Right hand† 50th	Right hand† 95th	Both hands† 50th
9	0	26	46	67	99	34	57	86	106
	4½ (left)	18	33	54	88	28	45	66	106
	8 (left)	12	29	44	77	26	40	67	93
	4½ (right)	34	58	82	99	39	62	88	106
	8 (right)	37	65	95	99	39	58	86	106
12½	8 (left)	18	36	68	110	33	53	77	120
	8 (right)	43	74	102	110	49	80	108	120
15½	0	43	86	160	165	54	83	113	146
	8 (left)	23	60	118	121	39	64	98	133
	8 (right)	53	100	164	143	55	89	119	146
18¾	0	64	124	177	154	56	86	127	160
	8 (left)	36	72	114	121	45	74	108	146
	8 (right)	70	125	198	154	58	99	126	160
23¾	0	54	106	141	110	62	102	138	173
	8 (left)	29	64	104	88	51	90	129	173
	8 (right)	56	100	147	99	58	103	133	173

Control distance from SRP (in.)*	Control distance from midsagittal plane of body (in.)*	C. To the left (Percentiles (lb)) Right hand† 5th	Right hand† 50th	Right hand† 95th	Both hands† 50th	D. To the right (Percentiles (lb)) Right hand† 5th	Right hand† 50th	Right hand† 95th	Both hands† 50th
9	0	30	47	66		23	38	49	
	4½ (left)	31	49	67		31	48	64	
	8 (left)	24	44	65		34	55	74	
	4½ (right)	26	46	78		15	27	51	
	8 (right)	26	44	72		12	22	43	
12½	8 (left)	23	44	70		31	48	70	
	8 (right)	22	39	59		16	24	46	
15½	0	24	38	52		20	28	39	
	8 (left)	20	35	58		25	43	63	
	8 (right)	24	40	70		13	22	49	
18¾	0	8	32	53		15	25	35	
	8 (left)	16	30	56		22	36	61	
	8 (right)	22	39	70		14	24	50	
23¾	0	14	29	46		13	20	30	
	8 (left)	11	21	49		19	31	48	
	8 (right)	20	37	76		12	22	51	

Note: In the neutral position, the stick is grasped 13½ inches above SRP.
* See Fig. 11–73.

Unpublished data, Anthropology Branch, Aerospace Medical Research Laboratories.
† All subjects used an aircraft type of seat with standard lap belt and shoulder harness.

TABLE 11–98. MAXIMAL STATIC ONE- AND TWO-HANDED FORCES EXERTED ON AN AIRCRAFT CONTROL STICK BY SEATED MALES (N = 20)

	Preferred hand		Both hands	
Direction of force	Means lb	S.D. lb	Means lb	S.D. lb
Forward push	167	33.4	199	28.3
Backward pull	134	12.7	204	14.3
Lateral push*	88	25.0	101	22.4
Lateral pull†	63	20.1	88	19.9

Watt (1963).
Note: The stick is located 20 in. forward of Seat Reference Point (SRP); see Fig. 11–73.

*Lateral push applied by the palm of either hand against the control handle.
†Lateral pull applied by the fingers of either hand against the control handle.

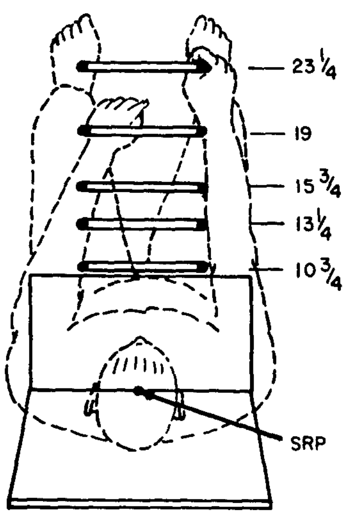


FIGURE 11–74. Maximal one- and two-handed static forces exerted on an aircraft control wheel by seated males.

TABLE 11–99. MAXIMAL ONE- AND TWO-HANDED STATIC FORCES EXERTED ON AN AIRCRAFT CONTROL WHEEL BY SEATED MALES (ONE-HANDED, N = 48; TWO-HANDED, N = 15)

Control distance from SRP (in.)*	Wheel rotation (deg.)	A. Push (Percentiles)				B. Pull (Percentiles)			
		Right hand†			Both hands†	Right hand†			Both hands†
		5th	50th	95th	50th	5th	50th	95th	50th
10¾	0	52	86	135	147	44	66	102	126
	45 (left)	48	84	149	147	40	67	111	126
	90 (left)	32	67	125	103	23	55	109	98
	45 (right)	40	67	128	147	39	67	97	126
	90 (right)	19	52	112	88	18	43	87	98
13¼	90 (left)	32	54	93	88	33	67	120	112
	90 (right)	25	51	83	88	31	60	102	112
15¾	0	61	90	155	177	66	94	145	154
	90 (left)	32	59	139	118	42	71	144	140
	90 (right)	32	53	102	132	49	80	130	140
19	0	64	121	235	265	73	106	169	196
	90 (left)	37	88	171	162	60	88	127	154
	90 (right)	33	67	140	162	61	94	149	168
23¼	0	105	171	242	265	77	125	182	234
	90 (left)	82	131	211	177	73	117	162	182
	90 (right)	49	117	197	191	74	110	186	196

Control distance from SRP (in.)*	Control rotation (deg.)	C. To the left (Percentiles)				D. To the right (Percentiles)			
		Right hand†			Both hands†	Right hand†			Both hands†
		5th	50th	95th	50th	5th	50th	95th	50th
10¾	0	26	46	88	92	20	48	96	91
	45 (left)	21	54	123	102	24	69	121	132
	90 (left)	23	47	91	102	27	59	101	101
	45 (right)	31	54	120	133	24	51	118	111
	90 (right)	21	42	104	122	15	54	112	121
13¼	90 (left)	26	44	86	102	21	52	98	111
	90 (right)	25	45	99	122	19	51	111	101
15¾	0	27	46	112	102	27	59	97	101
	90 (left)	27	43	82	82	19	53	96	101
	90 (right)	29	50	86	112	20	46	91	91
19	0	25	44	95	102	30	63	104	101
	90 (left)	22	43	76	82	27	46	94	101
	90 (right)	33	52	104	122	22	41	87	81
23¼	0	20	39	86	92	35	60	98	101
	90 (left)	21	38	73	71	26	42	82	91
	90 (right)	26	55	109	102	22	40	68	71

Note: Wheel grips are 15 in. apart. In the neutral position, they are 18 inches above SRP.

* See Fig. 11–74.

Unpublished data, Anthropology Branch, Aerospace Medical Research Laboratories.

† All subjects used on aircraft type of seat, with standard lap belt and shoulder harness.

TABLE 11–100. MAXIMAL STATIC HAND FORCES, AT VARIOUS ELBOW ANGLES, EXERTED ON A HORIZONTAL HANDGRIP (HAND PRONATATED) BY SEATED MALES (N = 30)

A. Left hand*

Direction of force	Elbow angle† (deg.)	Means	Percentiles (lb) 5th	Percentiles (lb) 95th	S.D.
Push, horizontal	60	86	33	138	35
	90	60	27	93	28
	120	43	17	71	17
	150	37	15	69	18
	180	32	12	59	13
Pull, horizontal	60	39	20	64	18
	90	37	17	65	18
	120	30	12	56	14
	150	32	15	52	13
	180	34	16	61	15
To the right, horizontal	60	42	20	66	15
	90	38	17	60	12
	120	34	17	53	8
	150	31	17	54	11
	180	28	15	41	8
To the left, horizontal	60	36	18	51	15
	90	27	11	54	11
	120	22	10	39	10
	150	23	9	53	16
	180	20	10	49	13
Up, vertical	60	57	22	100	22
	90	77	37	123	24
	120	91	45	145	30
	150	100	58	159	32
	180	101	47	171	11
Down, vertical	60	74	18	139	35
	90	75	23	136	34
	120	75	29	148	40
	150	79	39	136	29
	180	76	34	138	31

B. Right hand*

Direction of force	Elbow angle† (deg.)	Means	Percentiles (lb) 5th	Percentiles (lb) 95th	S.D.
Push, horizontal	60	94	40	156	36
	90	65	25	100	24
	120	46	23	70	15
	150	40	18	66	18
	180	32	17	59	12
Pull, horizontal	60	37	13	50	16
	90	32	14	54	13
	120	26	13	43	10
	150	29	12	48	10
	180	28	11	48	12
To the right, horizontal	60	41	19	72	19
	90	31	12	64	15
	120	26	9	53	13
	150	21	9	39	11
	180	19	10	34	7
To the left, horizontal	60	48	16	73	18
	90	39	16	59	15
	120	34	15	47	11
	150	32	18	45	7
	180	31	16	57	13
Up, vertical	60	49	23	79	20
	90	69	28	112	29
	120	91	41	138	30
	150	99	43	165	38
	180	95	35	156	35
Down, vertical	60	81	23	158	35
	90	83	22	142	35
	120	92	37	161	35
	150	90	40	154	34
	180	87	41	143	31

Hunsicker (1957).
* See Fig. 11–68.

† See Fig. 11–75. Handgrip was oriented in a vertical lateral plane passing through the subject's external canthus (outside corner of the eye). Subjects wore no restraint harness.

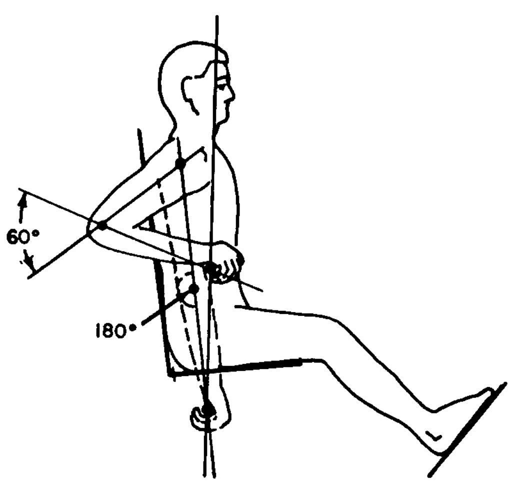


FIGURE 11–75. Hand forces exerted at various elbow angles.

TABLE 11–101. MAXIMAL STATIC HAND FORCES, AT VARIOUS ELBOW ANGLES, EXERTED ON A HORIZONTAL HANDGRIP (HAND SUPINATED) BY SEATED MALES (N = 30)

A. Left hand*

Direction of force	Elbow angle† (deg.)	Means	Percentiles (lb) 5th	Percentiles (lb) 95th	S.D.
Push, horizontal	60	89	35	176	42
	90	59	25	104	27
	120	40	15	80	18
	150	38	13	69	30
	180	30	14	47	10
Pull, horizontal	60	54	23	87	23
	90	42	13	68	21
	120	40	14	66	18
	150	40	16	62	15
	180	40	17	70	18
To the right, horizontal	60	38	16	64	12
	90	32	12	46	12
	120	31	14	55	13
	150	32	12	62	15
	180	29	12	43	9
To the left, horizontal	60	42	17	81	20
	90	33	16	52	12
	120	28	14	45	8
	150	26	12	43	10
	180	27	8	44	10
Up, vertical	60	49	20	89	22
	90	75	24	131	29
	120	94	38	152	33
	150	104	44	164	36
	180	111	45	173	40
Down, vertical	60	58	20	138	41
	90	80	23	160	43
	120	84	35	136	33
	150	84	43	136	29
	180	78	36	124	28

B. Right hand*

Direction of force	Elbow angle† (deg.)	Means	Percentiles (lb) 5th	Percentiles (lb) 95th	S.D.
Push, horizontal	60	96	34	172	39
	90	65	25	117	24
	120	43	20	71	17
	150	36	17	59	14
	180	32	12	58	15
Pull, horizontal	60	51	16	93	25
	90	43	13	74	19
	120	40	11	63	17
	150	37	11	66	17
	180	39	15	73	19
To the right, horizontal	60	44	18	73	19
	90	39	18	72	24
	120	34	17	64	15
	150	32	15	60	14
	180	29	14	48	12
To the left, horizontal	60	36	13	70	17
	90	31	13	48	12
	120	30	12	46	11
	150	31	12	52	14
	180	28	10	44	10
Up, vertical	60	45	17	78	22
	90	63	21	107	27
	120	88	41	143	33
	150	103	37	161	40
	180	113	51	165	34
Down, vertical	60	59	20	132	35
	90	80	17	143	37
	120	92	29	148	13
	150	93	37	150	35
	180	87	44	135	32

Hunsicker (1957).
* See Fig. 11–68.

† See Fig. 11–75. Handgrip was oriented in a vertical lateral plane passing through the subject's external canthus (outside corner of the eye). Subjects wore no restraint harness.

TABLE 11–102. MAXIMAL STATIC HAND FORCES, AT VARIOUS ELBOW ANGLES, EXERTED ON A VERTICAL HANDGRIP BY PRONE MALES

A. Left hand

Direction of force	Elbow angle* (deg.)	Means	Percentiles (lb) 5th	95th	S.D.
Push, horizontal	60	52	17	87	21
	90	54	18	91	22
	120	63	21	108	27
	150	65	24	111	26
	180	67	26	116	28
Pull, horizontal	60	57	17	97	24
	90	66	23	118	26
	120	75	22	126	30
	150	70	21	122	28
	180	61	18	111	26
To the left, horizontal	60	24	8	49	12
	90	22	6	45	10
	120	20	6	38	9
	150	20	5	56	15
	180	22	4	57	19
To the right horizontal	60	44	11	99	24
	90	40	13	92	22
	120	38	9	91	23
	150	34	8	79	23
	180	31	10	67	19
Up, vertical	60	35	13	71	17
	90	40	15	78	18
	120	40	11	81	21
	150	31	7	62	17
	180	18	5	44	12
Down, vertical	60	30	10	51	12
	90	31	12	57	12
	120	31	11	57	14
	150	28	10	48	11
	180	25	7	41	10

B. Right hand

Direction of force	Elbow angle* (deg.)	Means	Percentiles (lb) 5th	95th	S.D.
Push, horizontal	60	66	24	119	26
	90	64	26	103	23
	120	73	29	128	28
	150	74	29	127	30
	180	69	31	123	26
Pull, horizontal	60	61	21	113	26
	90	73	24	121	30
	120	86	31	147	34
	150	81	29	133	33
	180	69	31	118	26
To the left, horizontal	60	49	16	91	22
	90	46	16	87	21
	120	48	15	97	25
	150	45	15	93	26
	180	37	12	71	17
To the right, horizontal	60	30	12	57	12
	90	28	13	51	11
	120	28	11	58	12
	150	28	12	60	14
	180	25	9	61	14
Up, vertical	60	44	13	85	21
	90	52	15	94	22
	120	50	13	91	21
	150	41	13	83	23
	180	23	8	47	12
Down, vertical	60	34	13	61	13
	90	36	16	60	13
	120	35	15	61	15
	150	34	15	60	13
	180	29	13	47	10

Hunsicker (1957).

* See Fig. 11–76. Subjects not braced or restrained.

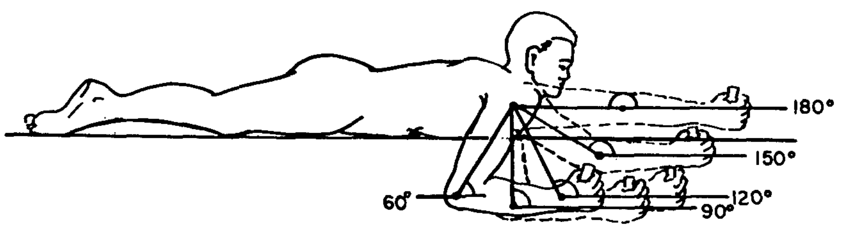


FIGURE 11–76. Hand forces exerted at various elbow angles by prone males.

In ordinary work situations, lower forces can normally be expected in the prone position, because many sets of muscles which are naturally available in the standing or seated positions become inactive. For example, the seated or standing man can often apply not only his arm force to a control, but also those of his torso and legs, whereas the prone man can usually exert only his arm muscles. This inhibitory effect could perhaps be eliminated by designing a prone-position support providing bracing for all thrust directions, and for which all control levers would be located in the proper positions for maximal thrust output. Up to now no test data have appeared to permit a true comparison.

Regarding intrinsic body force output as a theoretical consideration, if comparable body bracings were provided for all body positions, and if the same sets of muscles were tested, it is hard to see why force output should change much merely because body axes become reoriented with respect to gravity.

Maximal static grip forces. The hand squeeze is the exertion of force by movable fingers against the relatively immovable "heel" (thenar and hypothenar eminences) of the hand. The forces are usually measured by means of a dynamometer gripped in the hand, whose fixed portion (which also may be called the "heel") fits against the heel of the hand and whose movable portion (called the "trigger") is actuated by the fingers. There are at least two types of commercial dynamometers, the elliptically-shaped, non-adjustable Collin, and the rectangularly-shaped, adjustable Smedley. (Readers interested in descriptions of these instruments may consult Hunsicker and Donnelly, 1955; and Pangle and Garrett, 1966). As the positions of the gripping fingers are different on the two instruments, the total forces exerted could be expected to vary also.

In the data on grip strength, presented in Tables 11–103 and 11–104, it should be remembered that both types of dynamometers were used in measuring the various samples.

TABLE 11–103. MAXIMAL STATIC GRIP FORCE EXERTED BY MALES

Population	Percentiles (lb) 5th	50th or Mean	95th	S.D.
Air Force personnel, general:[1]				
Right hand	*(59)	104	*(148)	27.3
Left hand	*(56)	94	*(134)	23.7
Air Force personnel, aircrewmen:[2]				
Right hand	105	134	164	18.0
Left hand	96	124	154	16.0
Air Force rated officers:[3]				
Preferred hand	98	124	154	16.8
Army personnel:[4]				
Right hand	106	137	172	----
Left hand	99	132	168	----
Navy personnel:[5]				
Mean of both hands	95	119	143	14.4
Industrial workers:[5]				
Preferred hand	92	117	143	15.4
Truck and bus drivers:[6]				
Right hand	91	121	151	18.1
Left hand	86	113	140	16.4
Rubber industry workers:[7]				
Right hand	*(89)	124	*(159)	21.2
Left hand	*(86)	122	*(159)	22.2
University men:[8]				
Right hand	*(74)	108	*(142)	21.0
Left hand	*(65)	95	*(124)	18.0
Same subjects, force exerted over one minute:[9]				
Right hand	*(42)	62	* (82)	12
Left hand	*(39)	55	*(71)	10

* Percentiles in parentheses were computed from the 50th percentile using the S.D.
[1] Barter et al. (1956). N = 99.
[2] Clarke (1945). N = 914.
[3] Clauser et al. (1967). N = 2420.
[4] Damon et al. (1962). N = 431.
[5] Fisher and Birren (1946). Navy, N = 169; Workers, N = 552.
[6] Damon and McFarland (1955). N = 268.
[7] Damon and Stoudt, unpublished. Quoted in Damon et al. (1966). N = 162.
[8] Tuttle et al. (1950). N = 200.

TABLE 11–104. MAXIMUM STATIC GRIP FORCE EXERTED BY FEMALES

Population	Percentiles (lb)			S.D.
	5th	50th	95th	
Navy personnel *				
Mean of both hands	58	73	87	8.8
Industrial workers:†				
Preferred hand	57	74	91	10.3

Fisher and Birren (1946).
* N = 161.
† N = 96.

How much difference exists between them is unknown, but there is some evidence (Bowers, 1961) that the non-adjustable type tends to give higher readings.

If the fingers are like other muscle systems, there is an angle of the fingers, as they close on the palm, where the force they exert is at a maximum. Hence the distance between the heel of the hand and the inside of the closing fingers assumes some importance in measuring maximal gripping strength. Apparently only one engineering study has been made to find the best distance. In this study, a test using four separation distances—1½, 2½, 4, and 5 in., respectively, on a Smedley dynamometer—each one of 44 subjects (all pilots or flying cadets) pressed more at 2½ in. than at any other distance (Hertzberg, 1955). The comparative results are presented in Figure 11–77. Hence, devices to be triggered by handgrip, as those actuating certain types of ejection seats, should incorporate this spacing, with pressure surfaces approximately parallel. The trigger pull required of the operator should be no greater than the 5th percentile of hand force, and preferably no greater than the 1st.

In that same series of tests, it was also found that, when wearing standard heavy flying gloves (wool liner inside a leather shell), all subjects experienced a considerable reduction of their maximal barehanded gripping force, averaging about 20%, apparently due to slippage between the layers in that hand position. Gloves need not invariably produce a decrement, however; possibly some kinds of gloves in other hand positions may increase the force that can be exerted.

Maximal static finger forces. Tables 11–104 and 11–105 present data, unfortunately rare, on the forces which can be exerted with several fingers combined, or with each finger separately. These forces were measured on different groups of subjects using different hand and finger positions. This may partly explain the fact that forces exerted with the thumb alone (Table 11–105) appear to be smaller than those measured in palmar, tip, and lateral prehension (Table 11–106). These last actions require that the thumb successfully withstand the forces applied by the other fingers. (See Figure 11–78.)

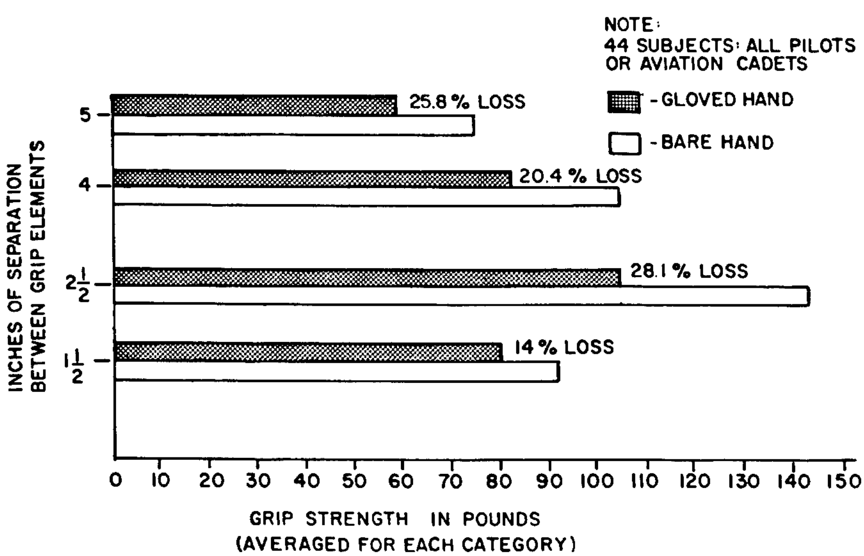


FIGURE 11–77. Grip strength of pilots.

TABLE 11–105. MAXIMAL STATIC FORCE EXERTED BY 100 MALES IN ATTEMPTED FLEXION OF THE EXTENDED FINGERS OF RIGHT HAND

Finger	Mean	S.D.
Thumb	16	3.8
Index	13	2.8
Middle	14	4.3
Ring	11	3.8
Little	7	2.5

Barter et al. (1957).
Note: For measurements on the thumb, the ulnar side of the hand was placed on a flat surface; for measurements on the other fingers, the back of the hand touched the surface. The experimenter held down the subject's wrist. Force was measured perpendicular to the extended finger.

TABLE 11–106. MAXIMAL STATIC FORCE EXERTED BETWEEN THUMB AND FINGERS BY MALE CIVILIANS (N = 15)

	Force (lb)	
Type of prehension *	Avg.	S.D.
Palmar	21.5	5.4
Tip	21.0	4.8
Lateral	23.2	4.8

Taylor (1954).
* See Fig. 11–78.

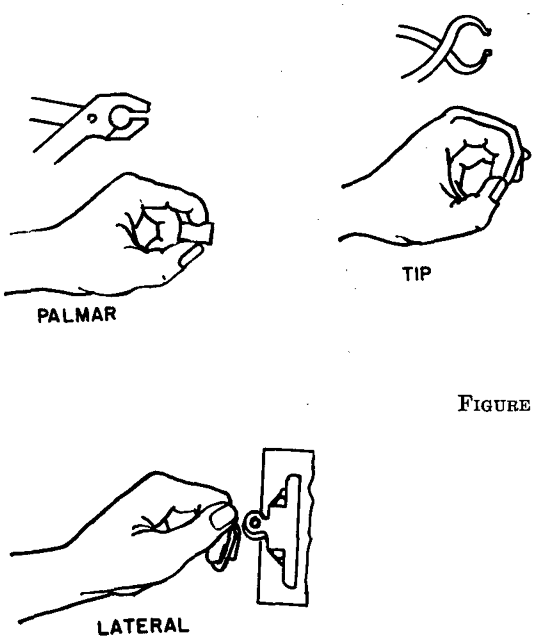


FIGURE 11–78. Maximal static forces exerted between thumb and fingers by male civilians.

Maximal static forces exerted by one foot of the seated operator. Tables 11–107, 11–108, 11–109, 11–110 present the published maximal static forces that have been measured on different kinds of pedals. Tables 11–107 through 11–109 indicate the forces that various samples exerted by thrust of the total leg, i.e., by attempted straightening of the leg at the knee. Table 11–110, however, presents forces exerted by attempted rotation of the foot alone. In these experiments each test pedal was immovable, or virtually so, during force application. The differences in the forces recorded indicate that several factors affect the amount of force exertable through the foot. Among these factors (see Figure 11–79) the most important are:

1. The kind of pedal (which determines the muscle masses used for force production—whether thrusting with the full leg, or rotating the foot about the ankle joint);

2. The kind of seat (a lumbar support for the back is essential if large forces must be exerted);

3. The distance and direction of the pedal with respect to the operator (the operator's largest forces are exerted in or very near the line connecting his hip-joint with the center of the pedal directly in front of him; as the direction of force moves laterally, or downward below the level of his hip, the amount of force exertable becomes smaller—Lehmann, 1958).

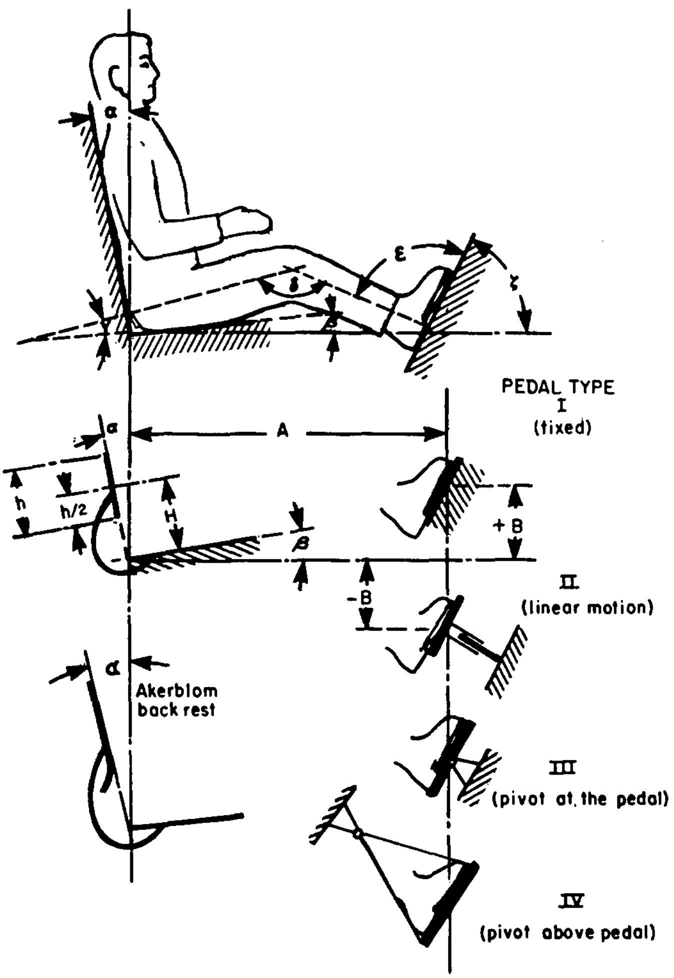


FIGURE 11–79. Dimensional and angular factors affecting pedal force applicability.

TABLE 11–107. MAXIMAL STATIC LEG THRUST EXERTED ON A FIXED PEDAL BY SEATED MALES

Test conditions*									Force			
Type of pedal	Type of seat	α (deg).	β (deg).	γ (deg).	δ (deg.)	ε (deg.)	A % of total leg reach	B (in.)	Means (lb)	S.D. (lb)	Direction	Subjects N
	H = 12 to 14 in. h = 5 in.	(†)	(†)	15	160			5	383	‡68	horizontal.	20 men§
				0	160			−4½	319	‡76		
Type I; fixed pedal pushed with ball of foot.	Åkerblom backrest 20 in. high; most antirior part of the lumbar pad 7 in. above the seat pan.	20, above the lumbar pad.	3			90	95 85	12	305 231	70 65	Approx. in the line from hip joint to the ankle.	60 men.¶
							95 85	4	304 270	65 67		
							95 85	−2	315 283	69 65		
							95 85	−8	298 254	63 70		
							95 85	−14	235 235	52 67		

* See Fig. 11–79.
† Probably 0°, but not explicitly stated in the original publication.
‡ Calculated from $SD = \sqrt{N \cdot SE}$.
§Rees and Graham, 1952.
¶Rohmert, 1966.

TABLE 11–108. MAXIMAL STATIC LEG THRUST, EXERTED BY SEATED SUBJECTS, ON A PEDAL PIVOTED NEAR THE INSTEP

Test conditions							Force		
Type of pedal	Type of seat	α (deg.)	β (deg.)	γ (deg.)	δ (deg.)	ε (deg.)	Means (lb)	Direction	Subjects N
	H = 8 in. h = 5 in.	(†)	(†)	10–20 ----	130–150 ----	90	appr. 440 appr. 340	5° to 15° below horizontal.	1 man.‡ 2 women.‡
Type III; pivoted near the instep; large enough to accommodate the entire foot.	Back rest supports pelvis and back.		(†)	−6 −15 −10 −10 −9	94 149 162 165 167	90	73 227 385 346 250	Approx. in the line from the hip joint to the center of the pedal.	6 "powerfully built men".§
				8 10 5	93 136 164	90	87 270 559		
				19 16 15 15 15 15	67 129 160 160 160 169	90	89 319 845 ¶(691) **(689) 530		
				36 33 34	88 106 125	90	135 184 443		
				48 49	72 81	90	133 130		

*See Fig. 11–79.
† Probably 0°, but not explicitly stated in the original publication.
‡ Mueller (1936).
§ Hugh-Jones (1947).
¶ 32 drivers of the Royal Armoured Corps.
**12 school boys, aged 14–18.

TABLE 11–109. MAXIMAL STATIC LEG THRUST EXERTED ON AIRCRAFT-TYPE PEDALS BY SEATED MALES

Test conditions*						Force			Subjects N
Type of pedal	Type of seat	δ (deg.)	ϵ (deg.)	A (in.)	B (in.)	Means (lb.)	S.D. (lb.)	Remarks	
Probably Type IV; Junkers Ju 35 cockpit						480	66	Subjects fresh.	11 (pilots and engineers).‡
						147	24	Subjects fatigued.	
Type II; B-24 aircraft cockpit (pedals accommodating the entire foot).		111 ± 5	120 ± 5			565	Approx. 100.		515 (student pilots).§
Type IV; pivoted near the shin (4).	Not described.		90	33.2 ± 1.8	2.4 ± 5.8	724 (median).		Direction of force parallel to the lower leg.†	166 (tank personnel).¶

* See Fig. 11–79.
† Force application pushed the pedal assembly into the direction of thrust.
‡ Hertel (1930).
§ Elbel (1949).
¶ Martin and Johnson (1952).

TABLE 11–110. MAXIMAL STATIC FOOT-ROTATIONAL FORCES ABOUT THE ANKLE, EXERTED ON AN AIRCRAFT BRAKE PEDAL BY SEATED PILOTS

Test conditions							Force (†)			
Type of pedal	Type of seat	α (deg.)	β (deg).	ζ (deg.)	A (in.)	B (in.)	Means (lb)	S.D. (lb)	Direction (deg. below horizontal; perpendicular to pedal).	Subjects N
Type IV; standard F–80 aircraft pedal. (Ball of foot placed on a low ridge fixed on the pedal pedal parallel to the axis, and 6.15 in. from the pedal axis lying under the the subject's heel.)	Hard surfaced plywood mock-up of a standard aircraft seat.	13	6.5	85 80 75	36.58 ± 1.18	−3.5	99 112 124	42 58 61	5 10 15	100 USAF pilots. (†)
				70 65 60	"	"	132 137 137	65 68 67	20 25 30	
				55 50 45	"	"	127 115 107	60 54 50	35 40 45	
				40 35	"	"	91 78	42 37	50 55	

* See Fig. 11–79.
† Convertible into torque around pedal axis by multiplying by 6.15 in. (length of lever arm).
‡Hertzberg and Burke, 1971.

Extremely large forces may be exerted by an operator if pedals are arranged at about seat (hip) height, and if his knee angle is about 170°–i.e., the hip-to-pedal distance is such as to allow a toggle action of the knee during the thrust. This fact, however, should not lead to the conclusion that all pedals which must be operated with sizable force should be located so high. To facilitate foot movements between two pedals, and to avoid the muscle strain just to hold the foot on a high surface, pedals should be arranged rather low. A comfortable knee angle (for application of moderate forces) is near 110° when the thigh is horizontal or only slightly inclined.

Tables 11–107 through 11–110 illustrate facts and problems frequently encountered with muscle strength data. Some of these are:

1. Although great strength differences exist between different individuals in the population, subjects in experimental test samples are usually highly selected and often do not adequately represent the range of prospective operators.
2. Almost no data have been published on the strength of women.
3. Although the pedal type and the manner of force application affect the amount of force exertable, not all pedal types or methods of operation have been investigated.
4. Likewise, although the position of leg and pedal, and the type of pedal and seat greatly affect the amount of force exertable, these experimental parameters are seldom clearly enough reported in the original research descriptions for engineering use.
5. Seemingly small changes in the experimental conditions may bring about substantive differences in the strength scores. For example, changing the knee angle (δ, see Table 11–108) can cause the force exerted to drop or increase by as much as several hundred pounds.
6. Finally, users of muscle-strength data must realize that such tabulations cannot provide absolute data, but only indicate trends and orders of magnitude.

11.7.5 Weight-Lifting Capabilities

Static lift strength. Tables 11–111 through 11–113 show the maximal static forces subjects exerted in pulling upward on a low horizontal bar, while facing it, thus indicating the relative efficacy of a pull by straightening the bent legs as compared with straightening the bent back. The data clearly show the superiority of the legs. This body position not only produces the larger forces but also reduces the danger of back injury.

Such tests, although still static, can be thought of as the beginning phase of the transition from the static effort to the dynamic action of lifting.

Loads Men Lift

Weights are lifted by a complete but momentary body action, whose beginning is analagous to static force application. The occasional lifting of "black boxes" can thus be considered as a class of muscle-strength performance that is transitional between strictly static effort and long-continued dynamic work.

One of the important factors is the solidity of the worker's footing. Ideal conditions for exerting the necessary force to do a job do not always exist. Slippery, uneven, or gravel-covered ground may severely reduce the man's capability; and if he is a soldier, he may have to contend with one or more of these impediments while under fire from an enemy.

In addition to solid footing (and other factors like age, health, condition, and motivation,), the angular relationship between body segments is important. In the one-man lifting of heavy containers—a special form of force exertion—the weight that the man can lift depends in part on the size of the package; i.e., on the horizontal distance between the centers of

TABLE 11–111. MAXIMAL STATIC UPWARD PULL FORCES EXERTED BY MALE AIR FORCE PERSONNEL: BACKLIFT VERSUS LEGLIFT (N = 914)

	Percentiles (lb)			
Type of lift	5th	50th	95th	S.D.
Backlift *	375	520	665	90
Leglift †	1010	1480	1950	290

Clarke (1945).

Note: The pull was two-handed, vertical, on a horizontal bar adjusted to the arms extended downwards in front of the legs.

* Legs are straight and the back is slightly bent and then straightened for the lift.

† Back is straight and the legs are slightly bent and then straightened for the lift.

TABLE 11–112. MAXIMAL STATIC UPWARD PULL FORCES EXERTED WITH TWO HANDS ON A HORIZONTAL BAR

Population	N	Percentiles (lb) 5th	50th or Mean	95th	S.D.
British women:*					
College students	†460	‡(160)	216	‡(272)	34.4
Factory workers:					
Employed	†3076	‡(119)	183	‡(247)	38.8
Unemployed	†413	‡(101)	165	‡(229)	39.2
British men:§					
College students	†1704	‡(271)	367	‡(463)	58.9
Factory workers:					
Employed	†10344	‡(251)	363	‡(474)	67.7
Unemployed	†1250	‡(214)	315	‡(415)	60.8
U.S. Air Force:					
Cadets (men)**	†914	375	520	665	90
"Leglift"**	**914	1010	1480	1950	290

Note: The bar in front of the subjects was adjusted so that each subject could just grasp it with his arms extended downward.

* Cathcart et al. (1927).

† Subjects tested with legs straight and back slightly bent (sometimes called backlift).

‡ Percentiles in parentheses were computed from the mean using the S.D.

§ Cathcart et al. (1935).

¶ Clarke (1945).

** Subjects underwent an additional trial (designated as leglift) under opposite conditions: They stood facing the bar with their backs straight and their legs slightly bent.

TABLE 11–113. MAXIMAL WEIGHTS LIFTED TO VARIOUS HEIGHTS BY MALE AIR FORCE PERSONNEL (N = 19)

Height lifted (ft)*	Percentiles (lb) 5th	50th	95th	S.D.
1	142	231	301	47
2	139	193	259	40
3	77	119	172	31
4	55	81	112	19
5	36	58	83	16

Emanuel et al. (1956).

* Subjects lifted a maximally weighted ammunition case (25½ × 10¾ × 6 in.) from the floor and placed it on platforms of various heights.

TABLE 11–114. "REASONABLE" WEIGHTS LIFTED WITHOUT STRAIN TO VARIOUS HEIGHTS BY AMERICAN MALES (N = 75)

Stature ranges of the subjects	Height lifted* (ft)	Average weight (lb)	S.D. (lb)
63.3–66.6	1½	124	20.8
	3½	73	11.6
	5¼	53	8.6
68.9–69.6	1½	138	22.6
	3½	92	12.5
	5¼	65	10.8
71.6–74.9	1½	146	30.9
	3½	96	14.2
	5¼	67	8.5

Switzer (1962).

Subjects lifted a box (12 × 12 × 6 in.), equipped with two handles, from the floor and placed it on platforms of different heights. The box was weighted so that the subjects felt this was a reasonable weight which they could lift repeatedly without strain.

gravity of the man and the package. (See Figure 11–80.) Thus, the efficient design of equipment for human use today increasingly requires biomechanical data as well as engineering data.

It is worth noting here that different experimenters specify varying criteria in their lifting tests; for example, Table 11–113 shows the "maximal" weight that one sample was able to lift to various heights with both hands, and Table 11–114 lists the "reasonable" weight that another sample chose to lift similarly. This again shows that "maximal," "optimal," and "reasonable" constitute quite different criteria for design.

Switzer (1962) measured the two-handed lifting abilities of 75 college students in selected stature groupings. His summarized results, categorized into three percentile groups and three heights of lift, are presented in Table 11–114. Switzer concluded that designers should confine the weight of their "black boxes" to the load that the 1st-percentile man can conveniently lift, using two hands, to the highest test level of 62.5 in. In his sample it was about 27 lb.

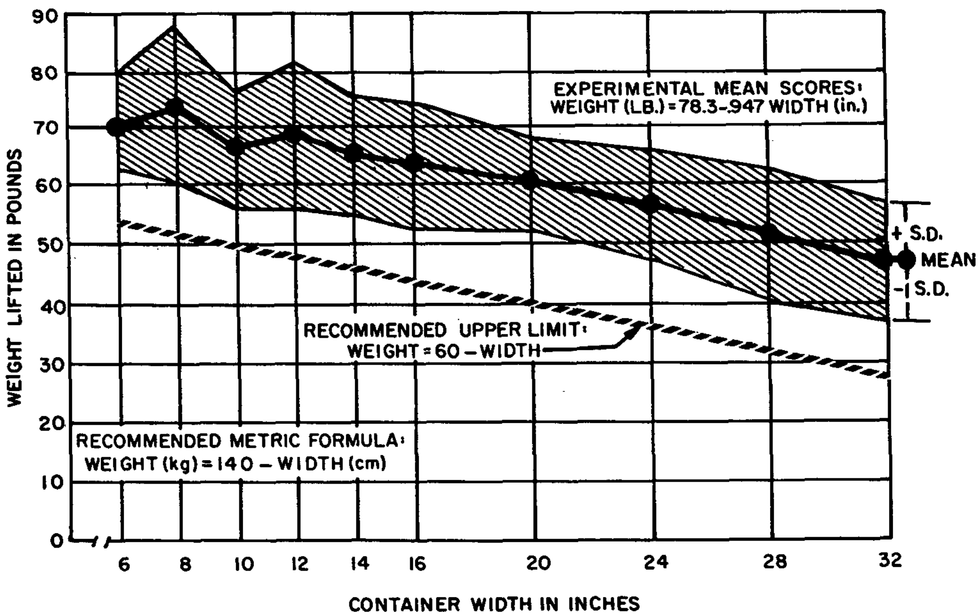


FIGURE 11–80. Container weight lifted as a function of container width.

Snook and Irvine (1966), using nine subjects, conducted a weight-lifting study in which they attempted to hold reasonably constant certain variables not previously controlled. Comparing their results with other recommendations in the literature, they conclude that ". . .50 lbs. is the maximum permissible weight of a compact object that should be lifted by unselected, adult male workers." For more highly selected personnel the ". . .maximum permissible weight of lift may approach 75 to 80 lbs., for lifts from floor level to knuckle height, and 65 to 70 lbs., for lifts above knuckle height."

In a study of one-handed lifting, McConville and Hertzberg (1966) investigated the relationship between weight and increasing package width. The boxes, all 10 in. high with a handle centered on top, varied in width from 6 to 32 in. (See Table 11–115). They were lifted from the floor onto a 30-in.-high table by a sample of 30 college men in two groups of 15 each, who had been cautioned not to overstrain. The graph of the averaged weight-lifting scores, and the original regression formula, are shown in Figure 11–80.

That original formula, of course, should not be used directly for package design. If a package to be lifted with one hand by 95% of the population, the mean weight in that formula must be reduced by 2 S.D. to the 5th percentile. The equation for those 5th-percentile values has been simplified, with negligible error, to an easily remembered formula, also graphed as a dashed line in Figure 11–81, which is recommended for design of one-handed loads.

$$\text{Weight (lb.)} = 60 - \text{Width (in.)}$$

TABLE 11–115. MAXIMAL WEIGHTS IN BOXES OF VARIOUS WIDTHS LIFTED WITH THE RIGHT HAND (N = 30)*

Container width (in.)	Weight lifted (lb)† (means)	S.D.
6	70	8.6
8	73	14.4
10	66	9.7
12	69	12.7
14	65	9.6
16	64	11.1
20	60	7.0
24	56	9.2
28	50	11.1
32	48	9.8

McConville and Hertzberg (1966).
* Two groups of 15 male subjects each, cautioned not to overstrain.
† All boxes 10 inches high, with a handle centered on top. All boxes lifted from floor to table top, 30 in. high.

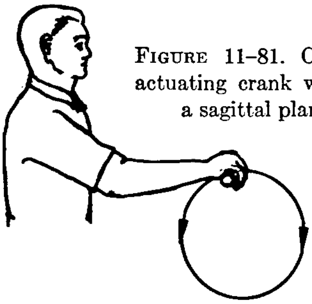


FIGURE 11–81. Operator actuating crank wheel in a sagittal plane.

or, in metric terms,

$$\text{Weight (kg.)} = \frac{140 - \text{Width (cm.)}}{5}$$

11.7.6 Data on Dynamic Working Capacity

When performing dynamic work, an operator moves his limbs by changing the length of his muscles through alternate tension and relaxation. The amount of energy or power (power = energy divided by time) which the operator can expend depends on the individual, the nature of the task, and the external conditions.

The usable external power-output of the human body is limited to the following ranges (quoted from Wilkie, 1960):

1. In single movements (of duration less than 1 sec.) to less than 6 hp.; by the intrinsic power production of muscle, and by the difficulty of coupling a large mass of muscle to a suitably matched load.
2. In brief bouts of exercise (0.1 to 5 min.) to 2 to 0.5 hp.; by the availability in the muscles of stores of chemical substances that can yield energy by hydrolysis.
3. In steady-state work (5 min. to 150 min. or more) to 0.5 to 0.4 hp.; by the ability of the body to absorb and transport oxygen.
4. In long-term work, lasting all day, to perhaps 0.2 hp.; by wear and tear of muscles, the need to eat, etc.

These amounts of power were generated under ideal conditions by champion athletes, while bicycling, or simultaneously bicycling and cranking with the hands—these being the most effective methods of using the human body as motor. Hence these figures represent about the maximum that can be achieved by superbly conditioned and motivated men. For ordinary healthy men, 20 to 30% should be deducted from the figures quoted, for women about twice these percentages. Less favorable environmental conditions, or less effective equipment, further reduce the power exertable by human operators.

If a machine operator is to exert muscular energy repeatedly or over a long period, rotating a crank by legs or arms is the most effective and least fatiguing way. A less desirable alternative is the reciprocal motion of a lever. The following rules will help to maximize the worker's physical capabilities while reducing unnecessary strain.

Rotary Motion

Rotary pedal cranks, as on bicycles, are to be preferred if high power is required. With rotary pedals, healthy men can easily achieve approximately 60 watts over many hours, or 100 watts for about half of an hour (Mueller, 1939).

1. Normally, the pedal crank arm should have a radius of 7.9 in., and be operated at 40 to 90 r. p. m. depending on the power output desired, and the body dimensions of the operator.
2. The line from the pedal crank axis to the saddle should be inclined 20° to 30° behind vertical.
3. The distance between saddle and foot pedals should be such that the operator, keeping his trunk immobile, must completely extend his legs in order to place his heels on the pedals in their furthest positions.
4. The inertia of the rotating masses should be large enough that the pedals can maintain their angular velocity for one or two revolutions if the operator removes his feet.

A hand crank, rotated with both arms by a standing operator, permits continuous exertion of as much as 100 watts for several minutes (Lehmann, 1962).

1. The crank arm should have a radius of about 12 in. and be operated at approximately 30 r.p.m.
2. The crank axis should be at about waist height and parallel to the operator's frontal plane, so that his hands move in his midsagittal plane. (See Figure 11–82.)
3. The crank handle should have a sleeve that rotates smoothly and freely about the handle axis, so as to eliminate the need for wrist flexion while cranking.
4. Two-handed actuation requires a handle about 15 in. long, so that the operator's inside arm will not bump against the crank arm.

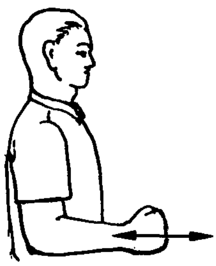


FIGURE 11–82. Operator using fore-and-aft reciprocal motion.

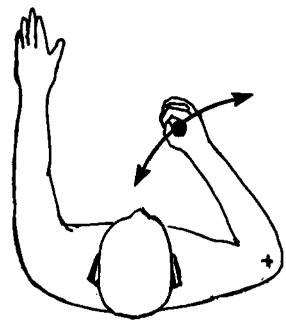


FIGURE 11–83. Lateral movement for small forces only.

Reciprocal Motion

Energy is transmissible by fore-and-aft actuation of a lever, if rotary operation is inadvisable for some mechanical reason.

1. No wrist action, or at most very little, should occur during control operation. This means that if a rigid lever is contemplated, the handle excursion must be small; otherwise a crank is preferable, with a rotatable sleeve on the handle.
2. The handle travel should be no more than 20 in. and preferably between 8 and 12 in.
3. At the beginning and at the end of the hand travel, the resistance should be low. Large resistances are best overcome by pull.
4. Hand travel should occur in a line that is at about waist height, roughly straight, horizontal, and in front of the operator; or in his midsagittal plane, if the control is operated with both hands. It should occur in the sagittal plane through the shoulder if only one arm is used. (See Figure 11–82; Kroemer 1966, 1967b.)
5. Only for small resistances may lateral travel of the handle occur in front of the operator (Stier, 1959). In such a case his hand should swing horizontally about the elbow, the upper arm merely rotating about its longitudinal axis. (See Figure 11–83.)

References

Aird, W. M., Bond, F. R., & Carrington, J. M. Anthropometric survey of male members of the Australian Army. Part I: Clothing survey. Report No. 3/58, June 1958, Australian Army Operations Research Group, Australian Military Forces, Canberra, Australia.

Alexander, M., & Clauser, C. E. Anthropometry of common working positions. Report No. AMRL–TR–65–73, 1965, Aerospace Medical Research Labs. Air Force Systems Command, WPAFB, Ohio.

Alexander, M., Garrett, J. W., & Flannery, M. P. Anthropometric and workspace dimensions for Air Force pressure-suited flight personnel. AMRL–TR–69–6, 1969. Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base, Ohio.

Arkin, A. M. Absolute muscle power. *Archives of Surgery*, 1941, **42,** 395.

Asmussen, E. & Heeboll-Nielsen, K. Isometric muscle strength in relation to age in men and women. *Ergonomics*, 1962, **5,** 167.

Balke, B., Nagle, F. J., & Daniels, J. Altitude and maximum performance in work and sports activity. *Journal of the American Medical Association*, 1965, **194:6,** 176.

Barter, J. T., et al. A statistical evaluation of joint range data. Report No. WADC–TN–57–311, 1957, Wright Air Development Center, WPAFB, Ohio.

Baxter, J. H. Statistics, medical and anthropological, of the Provost-Marshal-General's Bureau. Washington, D.C.: U.S. Government Printing Office, 1875.

Bayer, L. M., & Gray, H. Anthropometric standards for working women. *Human Biology*, 1934, **6,** 472.

Behnke, A. R., & Siri, W. E. The estimation of lean body weight from anthropometric and X-ray measurements. Research and Development Technical Report USNRDL–TR–203, 1957. U.S. Naval Radiological Defense Laboratory, San Francisco, California.

Bender, J. A., & Kaplan, H. M. Determination of success or failure in dynamic (isotonic) movements by isometric methods. *Research Quarterly*, 1966, **37**, 3.

Berger, R, A. Comparison of static and dynamic strength increases. *Research Quarterly*, 1962, **33**, 329.

Berger, R. A., & Henderson, J. M. Relationship of power to static and dynamic strength. *Research Quarterly*, 1966, **37**, 9.

Bowers, L. E. Investigation of the relationship of hand sizes and lower arm girths to hand grip strength as measured by selected hand dynamometers. *Research Quarterly*, 1961, **32**, 308–14.

Bowles, G. T. *New types of old Americans at Harvard and at eastern women's colleges.* Cambridge: Harvard University Press, 1932.

Brozek, J., & Behnke, A. R. (Co-chairman), Conference on Body Composition, 1963. The New York Academy of Sciences, New York, N. Y.

Bruener, H. Luftdruck and Luftzusammensetzung. In G. Lehmann (Ed.), *Handbuch der gesamten Arbeitsmedizin, Vol I. Arbeitsphysiologie*, Berlin, Germany: Urban and Schwarzenberg, 1961.

Brues, A., Movement of the head and eye in sighting. In: Randall et al., Human body size in military aircraft and personal equipment. Report No. AAFTR-5501, 1946, Aero Medical Lab., WPAFB Ohio.

Bulgarian Academy of Sciences. Physical development and fitness of the Bulgarian people from the birth up to the age of twenty-six. Volume 1, 1965, Bulgarian Academy of Sciences, Ministry of Education, Ministry of Public Health and Social Welfare. Sofia, Bulgaria: Bulgarian Academy of Sciences Press.

Caldwell, L. S. The relationship between the maximum force exertable by hand in a horizontal pull and the endurance of a sub-maximal response. Report No. 470, 1961, U.S. Army Medical Research Lab., Fort Knox, Ky.

Caldwell, L. S. Relative muscle loading and endurance. *Journal of Engineering Psychology*, 1963, **2**, 155.

Caldwell, L. S. The load-endurance relationship for a static manual response. *Human Factors*, 1964, **6**, 71.

Caldwell, L. S., & Smith, R. P. Pain and endurance of isometric muscle contraction. *Journal of Engineering Psychology*, 1966, **5**, 25.

Carter, I. G. Physical measurements of "old American" college women. *American Journal Of Physical Anthropology*, 1932, **16**, 479.

Cathcart, E. P., The physique of women employed in industry. Proceedings of the Royal Institute of Great Britain (1927), Vol. **25**, 185–199.

Cathcart, E. P., Hughes, D. E. R.., & Chalmers, J. G. The physique of man in industry. Report No. 71, 1935, Industrial Health Research Board, Medical Research Council, London, England.

Chen, K. P., Damon, A., & Elliot, O. Body form, composition and some physiological functions of Chinese on Taiwan. *Annals of the New York Academy of Sciences*, 1963, **110**, 760–777.

Clarke, D. H. Strength recovery from static and dynamic fatigue. *Research Quarterly*, 1962, **33**, 349.

Clarke, H. Analysis of physical fitness index test scores of air crew students at the close of a physical conditioning program. *Research Quarterly*, 1945, **16**, 192.

Clauser, C. E., McConville, J. T., & Young, J. W. Weight, volume and center of mass of segments of the human body. Report No. AMRL–TR–69–70, 1969. Aerospace Medical Research Laboratory, WPAFB, Ohio.

Coakley, J. D., Fucigna, J. T., & Barmack, J. E. A functional application of anthropometric data to the design of the workspace of PPI scope operators. Report No. WADC–TR–53–3, 1953, Wright Air Development Center, WPAFB Ohio.

Comas, J. La Antropologia Fisica en Mexico y Centro-America. Instituto Panamericano de Geographia e Historia. Mexico, 1943.

Consolazio, C. F., Johnson, P. E. & Pecora, L. J., *Physiological measurements of metabolic functions in man.* New York: McGraw-Hill, 1963.

Consolazio, C. F. Nelson, R. A., Matoush, L. O., & Hansen, J. E. Energy metabolism at high altitude. *Journal of Applied Physiology,* 1966, **21,** 1732.

Crowley, L. V., Ryner, R. R., III, & Pollack, H. Studies on nutrition in the far East. Relations between body weight, height, and skinfold thickness measurements in Chinese Nationalist troops. *Metabolism,* 1956, **5,** 272–275.

Dahlberg, G. *Race, reason, and rubbish.* London: Allen and Unwin, 1942.

Damon, A. Physique and success in military flying. *American Journal of Physical Anthropology,* 1955, **13,** 217.

Damon, A. Constitutional factors in acne vulgaris: Prevalence in white soldiers and lack of association with ancestry, complexion, body hair, blood group, handedness, plasma pepsinogen, and physique. *American Medical Association Archives of Dermatology,* 1957, **76,** 172.

Damon, A. Discrepancies between findings of longitudinal and cross-sectional studies in adult life: physique and physiology. *Human Development,* 1965, **8,** 16–22.

Damon, A., & Critchton, J. M. Body disproportions and occupational success in truck drivers. *American Journal of Physical Anthropology,* 1965, **23,** 63–68.

Damon, A., & McFarland, R. A. The physique of bus and truck drivers: With a review of occupational anthropology. *American Journal of Physical Anthropology,* 1955, **13,** 711.

Damon, A., Stoudt, H. W., & McFarland, R. The human body in equipment design. Cambridge, Mass.: Harvard University Press, 1966.

Damon, A., Bleibtreu, H., Elliot, O., Giles, E., & Stoudt, H. W. Predicting somatotype from body measurements. *American Journal of Physical Anthropology,* 1962, **20,** 461.

Daniels, G. S. The "average man"? Report No. WCRD–TN–53–7, 1952, Wright Air Development Center, WPAFB, Ohio.

Daniels, G. S., Meyers, H. C., & Worrall, S. H. Anthropometry of WAF basic trainees. Report No. WADC–TR–53–12, 1953. Aerospace Medical Research Laboratory, WPAFB Ohio.

Daniels, G. S., Meyers, H. C., & Churchill, E. Anthropometry of male basic trainees. Report No. WADC–TR–53–49, 1953. Aerospace Medical Research Laboratory, WPAFB Ohio.

Darcus, H. D. The range and strength of joint movement. In W. F. Floyd & A. T. Welford (Eds.), *Symposium on human factors in equipment design.* London: H. K. Lewis & Co., 1954.

Darcus, H. D. The maximum torques developed in pronation and supination of the right hand. *Journal of Anatomy,* 1951, **85,** 55.

Davenport, C. B., & Love, A. G. Army anthropology. In Vol. 15 of Medical Department of the U.S. Army in the World War, 1921, Office of the Surgeon General, Department of the Army, Washington, D.C.

Dempster, W. T. The anthropometry of body action. *Annals of the New York Academy of Science,* 1955, **63,** 559.

Diehl, H. S. Heights and weight of college men. *Human Biology,* 1933, **5,** 445.

Diehl, H. S. Heights and weights of college women. *Human Biology,* 1933, **5,** 600.

Dill, D. B., Myhre L., G., Brown, D. K., Burrus, K., & Gehlsen, G. Work capacity in chronic exposure to altitude. *Journal of Applied Physiology,* 1967, **23,** 555.

Donelson, E. G., Ohlson, M. A., Kunerth, B., Patton, M. B., & Kinsman, G. M. Anthropometric data on college women of the Middle States. *American Journal of Physical Anthropology,* 1940, **27,** 319.

Ducros, E. Statistiques de Biometrie Medicale Elementaire Relatives au Personnel Navigants de l'Armee de 1'Air Francaise. In AGARDograph 5, *Anthropometry and Human Engineering*. London, England: Butterworths Scientific Publications, 1955.

Dzendolet, E. Manual applications of impulses while tractionless. Report No. WADC-TR-60-129, 1960, Wright Air Development Center, WPAFB, Ohio.

Dzendolet, E., & Rievley, J. F. Man's ability to apply certain torques while weightless. Report No. WADC-TR-59-94, 1959, Wright Air Development Center, WPAFB, Ohio.

Elbel, E. R. Relationship between leg strength, leg endurance, and other body measurements. *Journal of Applied Physiology*, 1949, **2**, 197.

Elftman, H. The action of muscles in the body. *Biological Symposia*, 1941, **3**, 191.

Emanuel, J., Chaffee, J. W., & Wing, J. A study of human weightlifting capabilities for loading ammunition into F-86H aircraft. Report No. WADC-TR-56-367, 1956, Wright Air Development Center, WPAFB, Ohio.

Evrard, E. Statistiques Biometriques relatives au Personnel Navigant de la force Aerienne Belge, May 1954, Ministère de la Defense Nationale Force Aerienne, Direction du Services de Santé.

Fascenelli, F. & Lamb, L. E. Biomedical monitoring during dynamic stress testing: III. Maximum exercise tolerance, ergometer. *Aerospace Medicine*, 1966, **37** (9), 928-935.

Fisher, F. R. (Ed.) *Protection and functions of the hands in cold climates.* Washington, D.C.: National Academy of Sciences, National Research Council, 1957.

Fisher, M. B., & Birren, J. E. Standardization of a test of hand strength. *Journal of Applied Psychology*, 1946, **30**, 380.

Fisher, M. B., & Birren, J. E. Age and hand strength. *Journal of Applied Psychology*, 1947, 31, 490.

Fox, W. F. Human performance in the cold. *Human Factors*, 1967, **9**, 203.

Gauer, O., & Haber, H. Man in gravity-free conditions. In Vol. 1 of Aviation Medicine in World War II, 1950, Department of the Air Force, Washington, D.C.

Gibbons, T. B., Phillips, I. A., Budensiek, R. K., & Gilbertson, J. R., Age, height and weight of 2173 men entering recruit training during 1953. Naval Medical Research Unit No. 4, 1953, U. S. Naval Training Center, Great Lakes, Ill.

Gifford, E. G., Provost, J. R., & Lazo, J. Anthropometry of naval aviators-1964. Report No. NAEC-ACEL-533, October 1965, U.S. Naval Air Engineering Center, Philadelphia.

Glanville, A. D., & Kreezer, G. The maximum amplitude and velocity of joint movements in normal male human adults. *Human Biology*, 1937, **9**, 197.

Gould, B. A. *Investigations in the military and anthropological statistics of American soldiers.* Cambridge: Hurd and Houghton, 1869.

Guha, B. S. *Census of India*, **1931**, *Vol. I, Part III: Ethnographical. A. Racial Affinities of the Peoples of India.* Simla: Government of India Press, 1935.

Haggard, H. W. The mechanics of human muscles. *Mechanical Engineering*, 1946, **68**, 321.

Hansen, J. E. Vogel, J. A., Stelter, G. P., & Consolazio, C. F. Oxygen uptake in man during exhaustive work at sea level and high altitude. *Journal of Applied Physiology*, 1967, **23**, 511.

Hart, G. L., Rowland, G. E., & Malina, R. Anthropometric survey of the armed forces of the Republic of Korea. Technical Report No. 68-8-PR, 1966, U.S. Army Natick Labs., Natick, Mass.

Hathaway, M. L., & Foard, E. D. Heights and weights of adults in the United States. Report No. 10, 1960, Human Nutrition Research Division, Agricultural Research Service, U.S. Department of Agriculture, Washington, D.C.

Heath, C. W. *What people are.* Cambridge: Harvard University Press, 1945.

Hertel, H. Determination of the maximum control forces and attainable quickness in the operation of airplane controls. Technical Memo No. 583, 1930, National Advisory Committee on Aeronautics, Washington, D.C.

Hertzberg, H. T. E. Some contributions of applied physical anthropology to human engineering. Symposium on "Dynamic Anthropometry", *Annals of the New York Academy of Sciences*, 1955, **63**, 616–629.

Hertzberg, H. T. E. Dynamic anthropometry of working positions. *Human Factors*, 1960, **2**, 3.

Hertzberg, H. T. E. The conference on standardization of anthropometric techniques and terminology. *American Journal of Physical Anthropology*, 1968a **28**, 1–16.

Hertzberg, H. T. E. World diversity in human body size and its meaning in American aid programs. OAR Research Review, 1968b, **7**, No. 12, 14–17, Office of Aerospace Research, Arlington, Virginia.

Hertzberg, H. T. E., Daniels, G. S., & Churchill, E. Anthropometry of flying personnel-1950. Report No. WADC–TR–52–321, December 1954, Wright Air Development Center, WPAFB, Ohio.

Hertzberg, H. T. E., Emanuel, I., & Alexander, M., The anthropometry of working positions: I. A preliminary study. Report No. WADC–TR–54–520, 1956, Wright Air Development Center, WPAFB, Ohio.

Hertzberg, H. T. E., Churchill, E., Dupertuis, C. W., White, R. M., & Damon, A., Anthropometric survey of Turkey, Greece, and Italy. AGARDograph 73, New York: Pergamon Press, 1963.

Hertzberg, H. T. E., & Burke, F. E. Foot forces exerted various aircraft brake-pedal angles. *Human Factors*. 1971, **13**, 5.

Hettinger, T. Die Muskelkraft bei Frauen und Maennern. *Zbl.Arbeitswissenschaft*, 1960, **5**, 79.

Hettinger, T. *Physiology of strength*. Springfield, Ill.: Charles C. Thomas, 1961.

Hooton, E. A. *A survey of seating*. Gardner, Mass.: Heywood-Wakefield Co., 1945.

Hueting, J. E., & Sarphati, H. R. Measuring fatigue. *Journal of Applied Psychology*, 1966, **50**, 535.

Hugh-Jones, P. The effect of limb position in seated subjects on their ability to utilize the maximum contractile force of the limb muscles. *Journal of Physiology*, 1947, **105**, 332.

Hulse, F. S. Exogamie et Heterosis. *Archives Suisses d'Anthropologie General*, 1957, **22**, 104–125.

Hunsicker, P. A. Arm strength at selected degrees of elbow flexion. Report No. WADC–TR–54–548, 1955, Wright Air Development Center, WPAFB, Ohio.

Hunsicker, P. A. A study of muscle force and fatigue. Report No. WADC–TR–57–586, 1957, Wright Air Development Center, WPAFB, Ohio.

Hunsicker, P. A., & Donnelly, R. J. Instruments to measure strength. *Research Quarterly*, 1955, **26**, 408–420.

Hunsicker, P. A., & Greey, G. Studies in human strength. *Research Quarterly*, 1957, **28**, 109.

Institut der Wehrmedicinalstatistic. Tables of weight, statures, etc., of recruits. 1964.

Ishii, Y. Studies on the physique and physical strength of workers. *Journal of Scientific Labour* (Japan), 1957, **33**, 259.

Karpinos, B. D. Height and weight of selective service registrants processed for military service during World War II. *Human Biology*, 1958, **30**, 292.

Karpinos, B. D. Current height and weight of youths of military age. *Human Biology*, 1961, **33**, 335–354.

Karpovich, P. V. *Physiology of muscular activity*, Philadelphia: W. B. Saunders Co., 1965.

Kay, W. C. Anthropometry of the ROKAF pilots, 1961. Headquarters, ROK Air Force, Soeul, South Korea.

Kennedy, K. W. Reach capability of the USAF population. Report No. AMRL–TDR–64–

59, 1964, Aerospace Medical Research Labs., WPAFB, Ohio.

Kennedy, K. W., & Filler, B. E. Aperture sizes and depths of reach for one- and two-handed tasks. Report No. AMRL-TR-66-27, 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

King, B. G., Morrow, D. J., & Vollmer, E. P. Cockpit studies—The boundaries of the maximum area for the operation of manual controls. Report No. 3, Project X-651, 1947, U.S. Naval Medical Research Institute. Bethesda, Md.

Knuttgen, H. G. Aerobic capacity of adolescents. *Journal of Applied Physiology*, 1967, **22**, 655.

Kogi, K., Mueller, E. A., & Rohmert, W. Die relative Wirkurg isometrischen und statischen Trainings auf die Ausdauer bei dynamischer Arbeit. *Int. Z. ang ew. Phsyiol.*, 1965, **20**, 465.

Kroemer, K. H. E. Heute zutreffende Körpermasze. *Arbeitswissenschaft*, 1964, **3**, 42–45.

Kroemer, K. H. E. Ueber die ergonomische Bedeutung der raeumlichen Lage kreisbog enfoermiger Beweggungsbahnen von Betaetigungsteilen. Forschungsbericht Nr. 1687 des Landes Nordrhein-Westfalen, 1996, Westdeutscher Verlag, Koeln, Germany.

Kroemer, K. H. E. Maximal static force versus stress measurements as criteria for establishing optimal conditions. Preprints of the Annual Scientific Meeting of the Aerospace Medical Association, April 1967, 166.

Kroemer, K. H. E. *Was man von Schaltern, Kurbeln und Pedalen wissen muss.* Beuth Vertrieb, Berlin, Germany, 1967.

Kroemer, K. H. E. Push forces exerted in sixty-five common working positions. AMRL-TR-68-143, 1968, Aerospace Medical Research Lab., WPAFB, Ohio.

Laubach, L. L. & McConville, J. T. Muscle strength, flexibility, and body size of adult males. *Research Quarterly*, 1966, **3**, 384.

Lehmann, G. *Praktische Arbeitsphysiologie* (2nd ed.). Stuttgart, Germany: G. Thieme, 1962.

Lehmann, G. (Ed.), Arbeitsphysiologie. Vol I of the *Handbuch der gesamten Arbeitsmedizin.* Berlin, Germany: Urban and Schwarzenberg, 1961.

Lippert, S. Estimated body dimensions of the adult male Jewish population of Israel. TAE Report No. 40, June 1954, Technion Israel Institute of Technology, Department of Aeronautical Engineering, Haifa, Israel.

Martin, L. Etude biometrique de grandeurs somatiques recueilles sur les conscrits et recrues belges et de leur evolution. Institute National de Statistique, Brussels, Belgium, 1958.

Martin, W. B., & Johnson, E. E. An optimum range of seat positions as determined by exertion of pressure upon a footpedal. Report No. 86, 1952, U.S. Army Medical Research Lab., Fort Knox, Ky.

McConville, J. T., & Hertzberg, H. T. E. A study of one-handed lifting. Report No. AMRL-TR-66-17, 1966, Aerospace Medical Research Lab., WPAFB, Ohio.

McCormick, E. J. *Human Factors Engineering.* 2nd Ed., 1964. New York: McGraw-Hill.

McCormick, M. Y. Physical characteristics of the 1946 airline transport pilot population, 1947, Civil Aeronautics Administration, Washinton, D. C.

McFarland, R. A., Damon, A., & Stoudt, H. W. Anthropometry in the design of the driver's workspaces. *American Journal of Physical Anthropology*, 1958, **16**, 1.

Monod, H., & Pineau, H. Nouvelles donnees anthropometriques concernant les jeunes adultes francais. *Biotypologie*, 1958, **19**, 24–31.

Morant, G. M. Surveys of the heights and weights of Royal Air Force personnel. V. Heights and weights of flying personnel of the Royal Air Force and Commonwealth Air Forces. Report No. FRPC 711(d), January 1951, RAF Institute of Aviation Medicine, Flying Personnel Research Committee, Farnborough, Hants, England.

Morimoto, T., Slabochova, Z., Naman, R., & Sargent, F. Sex differences in physiological reactions to thermal stress. *Journal of Applied Physiology*, 1967, **22**, 526.

Mueller, E. A. Die guenstigste Anordnung im Sitzen betaetigter Fusshebel. *Arbeitsphysiologie*, 1936, **9**, 125.

Mueller, E. A. Der Einfluss der Traegheit auf das Arbeitsmaximum beim Radfahren. *Arbeitsphysiologie*, 1939, **10** (4), 436–439.

Mueller, E. A. Die Muskelfunktion, und die physische Ermuedung. In G. Lehmann (Ed.) *Handbuch der gesamten Arbeitsmedizin*, Vol 1, Arbeitsphysiologie. Berlin, Germany: Urban and Schwarzenberg, 1961.

Mueller, E. A. Physiology of muscle training. *Canadian Review of Biology*, 1962, **21**, 303.

Newman, R. W. The body sizes of young men. In Bennett et al. (Eds.), *Human factors in technology*. New York: McGraw-Hill, 1963.

Newman, R. W., & White, R. M. Reference anthropometry of Army men. Report No. 180, 1951, Environmental Protection Branch, U.S. Army Quartermaster Research and Development Center, Natick, Mass.

O'Brien, R., & Shelton, W. C. Women's measurements for garment and pattern construction. Pub. No. 454, 1941, U.S. Department of Agriculture, Washington, D.C.

Olivier, G., & Coblentz, A. Etude Anthropometrique du Personnel Navigant Francais l'Aeronautique Civile et Militaire. Association d'Anthropologie Appliquée, Septembre 1965, Doc. AAA/08/65.

Oshima, M., et al., & Alexander, M. Anthropometry of Japanese pilot. Reports of the Aero-Medical Lab., Vol. 2, No. 2, 1962, Tokyo, Japan. (Also published as AMRL-TR-65-74, same title, 1965, Aerospace Medical Res. Lab., WPAFB, Ohio.)

Pangle, R., & Garrett, L. Origin of the spring-scale dynamometer. *Research Quarterly*, 1966, **37**, 155–156.

Pett, L. B., & Ogilvie, G. F. The Canadian weight-height survey. *Human Biology*, 1956, **28**, 177–188.

Pett, L. B., & Oglivie, G. F. The report on Canadian average weights, heights and skinfolds. *Canadian Bulletin on Nutrition*, 1957, **5**, 1.

Pugh, L. G. C. E. Cardiac output in muscular exercise at 5,800 m (19,000 ft). *Journal of Applied Physiology*, 1964, **19**, 441.

Pugh, L. G. C. E., & Ward, M. Physiology and medicine. In *The Ascent of Everest*. London: J. Hunt, Hodder & Stoughton, 1953.

Pugh, L. G. C. E., Gill, M. B., Lahiri, S., Milledge, J. S., Ward, M. P., & West, J. B. Muscular exercise at great altitudes. *Journal of Applied Physiology*, 1964, **19**, 431.

Randall, F. E. Survey of body size of Army personnel, male and female: IV. Body dimension of Army females—methodology and general considerations. Report No. 123, 1947, Environmental Protection Branch, U.S. Army Quartermaster Research and Development Center, Natick, Mass.

Randall, F. E., & Munro, E. H. Reference anthropometry of Army women. Report No. 159, 1949, Environmental Protection Branch, U.S. Army Quartermaster Research and Development Center, Natick, Mass.

Randall, F. E., Damon, A., Benton, R. S., & Patt, D. I. Human body size in military aircraft and personal equipment. Technical Report No. 5501, 1946, Air Materiel Command, WPAFB, Ohio.

Rees, J. E., & Graham, N. E. The effect of backrest position on the push which can be exerted on an isometric foot-pedal. *Journal of Anatomy*, 1952, **86**, 310.

Ries, W. Das Koerpergewicht unter Beruecksichtigung von Alter und Geschlecht. *Zeitschrift fur Alternsforschung*, 1963, **17**.

Roberts, D. F., & Bainbridge, D. R. Nilotic physique. *American Journal of Physical Anthropology*, 1963, **21**, 341–371.

Roberts, D. F., Provins, K. A., & Morton, R. J. Arm strength and body dimensions. *Human Biology*, 1959, **31**, 334.

Roberts, L. B. Size increase of men wearing various clothing combinations. Project No.

9, File No. 741–3, 1945, U.S. Army Medical Research Lab., Fort Knox, Ky.

Roebuck, J. A. Anthropometry in aircraft engineering design. *Journal of Aviation Medicine*, 1957, **28**, 41.

Rohmert, W. Ermittlung von Erholungspausen fuer statische Arbeit des Menschen. *Internationale Zeitschrift fuer angewandte Physiologie einschliesslich Arbeitsphysiologie*, **1960**, **18**, 123–164.

Rohmert, W. Die Armkraefte des Menschen im Stehen bei verschiedenen Koerperstellungen. *Internationale Zeitschrift fuer angewandte Physiologie*, 1960 b, **18**, 175.

Rohmert, W. Untersuchung statischer Haltearbeiten in achtstuendigen Arbeitsversuchen. *Internationale Zeitschrift fuer angewandte Physiologie*, 1961, **19**, 35.

Rohmert, W. *Maximalkraefte von Maennern im Bewegungsraum der Arme und Beine*. Koeln, Germany: Westdeutscher Verlag, 1966.

Royal Canadian Air Force. RCAF anthropometrical survey of pilots and navigators. Report No. 322, 1963, RCAF Institute of Aviation Medicine, Toronto, Canada.

Royce, J. Re-evaluation of isometric methods and results, a must. *Research Quarterly*, 1964, **35**, 215.

Samuel, G. D., & Smith, E. M. B. A comparison of seven anthropometric variables of American, British and Canadian pilots. Report No. IAMR 322, April 1965, RAF Institute of Aviation Medicine, Farnborough, Hants, England.

Scherrer, J. *Physiologie de Travail*. Paris, France: Masson et Cie, 1967.

Sen, R. N. Some anthropometric studies on Indians in a tropical climate. *Proceedings of the Lucknow Symposium* (UNESCO/ India *Symposium*) *on Environmental Physiology and Psychology in Arid Conditions*. Arid Zone Research, 1964, **24**, UNESCO, Paris.

Sharp, E. D. A comparison of three full-pressure suits in terms of control activation time. Report No AMRL–TR–64–126, 1964, Aerospace Medical Research Labs., WPAFB. Ohio.

Sheldon, W. H., Stevens, S. S., & Tucker, W. B. *The varieties of human physique*. New York: Harper, 1940.

Sinelnikoff, E. & Grigorowitsch, M. Die Beweglikeit der Gelenke als sekundares geschlechtliches und konstitutionelles Merkmal. *Z. Konstitutionslehre*, 1931, **15**, 679.

Smyth, C. J., et al. Rheumatism and arthritis. *Annals of Internal Medicine*, 1959, **50**, 366.

Snook, S. H., & Irvine, C. H. Maximum acceptable weight of lift. *American Industrial Hygiene Association Journal*, 1967, **28**, 322–329.

Snow, C. C., & Snyder, R. G. Anthropometry of air traffic control trainees. Report No. AM 65–26, September 1965, Federal Aviation Agency, Oklahoma City.

Stier, F. Die Geschwindigkeit von Armbewegungen. *Internationale Zeitschrift fuer angewandte Physiologie*, 1959, **18**, 82.

Stoudt, H. W., & Marett, J. R. *The physical anthropology of Ceylon*. Ceylon National Museum's Ethnographic Series, Publication No. 2, September 20, 1961. (Edited by P. E. P. Deraniyagala, Dir. of National Museum, The Colombo Museum, Colombo, Ceylon.)

Stoudt, H. W., Damon, A., & McFarland, R. A. Heights and weights of white Americans. *Human Biology*, 1960, **32**, 331.

Stoudt, H. W., Damon, A., McFarland, R. A., & Roberts, J., Weight, height, and selected body dimensions of adults, United States, 1960–1962. Report No. 8, Series 11, June 1965, National Center for Health Statistics, U.S. Dept. of Health, Education and Welfare, Washington, D. C.

Streimer, I., Springer, W. E., & Tardiff, C. C. Human output characteristics during specific task performance in reduced traction environments. *Human Factors*, 1964, **6**, 121–126.

Switzer, S. K. Weight-lifting capabilities of a selected sample of human males. Report No. MRL–TDR–62–57, 1962, Behavioral Sciences Lab., WPAFB, Ohio.

Taylor, C. L. The biomechanics of the normal and of the amputated upper extremity. In

P. E. Klopsteg & P. D. Wilson (Eds.), *Human limbs and their substitute.* New York: McGraw-Hill, 1954.

Tornvall, G. Assessment of physical capabilities. *Acta Physiologica Scandinavica,* 1963, **53,** Suppl. 201.

Tuttle, W. W. Janney, C. D., & Thompson, C. W. Relation of maximum grip strength to grip strength endurance. *Journal of Applied Physiology,* 1950, **2,** 663.

Tyroler, H. A. Preliminary cross-sectional documentation of physical measurements of male population at risk by geographic location and age, 1958. Health Research Foundation, Asheville, N. C.

Udjus, L. G. Anthropometrical changes in Norwegian men in the twentieth century. Anthropological Department, University of Oslo, Norway, 1964.

United States Air Force. B-52 airplane mock-up human factors study. Report No. WCRD–TN–53–130, 1953b, Aero Medical Lab., Wright Air Development Center, WPAFB, Ohio.

United States Army. Unpublished data, 1946, Environmental Protection Branch, U.S. Army Quartermaster Research and Development Center, Natick, Mass.

United States Marine Corps. Unpublished data, 1949, U.S. Marine Corps Research Lab., Marine Clothing Depot, Philadelphia.

United States Navy. Weight distribution of recruits. *Statistics of Navy Medicine,* 1949a, **5,** (6), 2.

United States Navy. Height distribution of recruits. *Statistics of Navy Medicine,* 1949b, **5,** (7), 2.

United States Navy. Unpublished data, 1957, U.S. Naval Medical Research Lab., U.S. Navy Submarine Base, New London, Conn.

United States Navy. Anthropometric data, *U.S. Navy Medical News Letter,* 1959, **33,** (4), 33.

Vogel, J. A., Hansen, J. E., & Harris, C. W. Cardiovascular responses in man during exhaustive work at sea level and high altitude. *Journal of Applied Physiology,* 1967, **23,** 531.

Walk, D. Finger dexterity of the pressure-suited subject. Report No. AMRL–TDR–64–41, May 1964, Aerospace Medical Research Labs., WPAFB, Ohio.

Watt, D. N. Maximum forces exerted on a stick type control. Report No. 63–RD–1, 1963, RCAF Institute of Aviation Medicine, Toronto, Canada.

Weinman, K. P. Slabochova, Z., Bernauer, E. M., Morimoto, T., & Sargent II, F. Reactions of men and women to repeated exposure to humid heat. *Journal of Applied Physiology,* 1967, **22,** 533.

Wenzel, H. G. Die Wirkung des Klimas auf den arbeitenden Menschen. In G. Lehmann (Ed.), *Handbuch der gesamten Arbeitsmedizin,* Vol. 1, Arbeitsphysiologie, Berlin. Germany: Urban & Schwarzenberg, 1961.

Wenzel, H. G. Moeglichkeiten und Probleme der Beurteilung von Hitzebelastungen des Menschen. *Arbeitswissenschaft,* 1964, **3,** 73.

White, R. M. Anthropometry of Army Aviators. Report No. EP-150, June 1961, Headquarters, Quartermaster Research and Development Command, U.S. Army, Natick, Mass.

White, R. M. Anthropometric survey of the Royal Thai Armed Forces. (Sponsored by the Advanced Research Projects Agency, Washington, D. C.), 1964, U.S. Army Natick Labs., Natick, Mass.

White, R. M. Anthropometric Survey of the Armed Forces of the Republic of Vietnam. (Sponsored by the Advanced Research Projects Agency, Washington, D. C.), 1964, U.S. Army Natick Labs., Natick, Mass.

White, R. M., Kobrick, J. L., & Zimmer, T. R. Reference anthropometry of the Arctic-equipped soldier. Technical Report No. EPT-2, 1964, U.S. Army Natick Labs., Natick, Mass.

Wilkie, D. R., Man as a source of mechanical power, *Ergonomics,* 1960, **3,** 1.

Wilmore, J. H., & Sigerseth, P. O. Physical work capacity of young girls, 7–13 years of age. *Journal of Applied Physiology*, 1967, **22,** 923.

Wyndham, C. H. Tolerable limits of air conditions for men at work in hot mines. *Ergonomics*, 1962, **5,** 115.

User expectations, capabilities, intentions
Interfaces with other systems and the environment
Required
Potential

12.2.3 Definition of Maintenance Levels and Areas

Maintenance areas are defined by one or more of the following:

1. Nature of work done (checkout, repair, replacement, inspection, etc.)
2. Extent of maintenance performed (man-hours, number of parts replaced, etc.)
3. Location of work (operational site, limited repair shop, overhaul shop, etc.)
4. Subsystems worked on (propulsion, guidance, flight control, etc.).

Levels of maintenance tend to be differentiated by the extent of maintenance performed and remoteness from the site of system operation. A given level of maintenance may include one or more separate areas.

System users and maintenance-logistics organizations will have existing definitions and expectations for maintenance areas and levels. The designer may be concerned with modifying existing organizations and facilities. At least he must establish early a set of assumptions concerning areas and levels to provide a context for functional analysis.

12.2.4 Identification of Maintenance Functions and Work Flow

Maintenance functions analysis is a key step in the development of a maintainability program. It involves defining what maintenance will be required and what method will be used. This step allows definition of technical performance requirements and the nature of support equipment that will be required. The functions analysis should:

1. Identify functions down to the level required to make decisions about man–machine tradeoffs, and
2. Organize and classify these functions for efficient and effective decision making.

In performing a function analysis, the following functions should be considered.

Adjusting and servicing. Periodic adjustments and required servicing should be identified. These should include only preventive maintenance actions requiring minimum down time. Removal of any major component should be included under recycle and overhaul analysis. Identification of required preventive maintenance actions can be accomplished by reviewing reliability figures for each component. Where excessive preventive maintenance needs are identified, redesign should be considered to eliminate the needs.

Verification. Identification of functions verifying that the system is operational requires a detailed functional breakdown of the system. Analysis should begin at the maintenance area that is most closely in support of mission performance or operational readiness and proceed to functions at rearward areas.

Troubleshooting. A troubleshooting function is involved when alternative sources of a malfunction produce the same symptom during verification. The analysis should include identification of symptoms that indicate which alternative is malfunctioning. The level of detail required for this analysis depends on the level of maintenance that will be performed to return the equipment to an operational state. For example, a "no-go" for a given verification may result from a malfunction in any one of dozens of parts. If these parts are all located in a replaceable module, further troubleshooting may not be required. The malfunctioning module is removed and replaced by a functioning module. If the parts are not located in a replaceable module, troubleshooting may have to be performed to the piece part level. If this involves excessive troubleshooting, redesign might be required.

Fault correction. For each cause of a malfunction indication, the appropriate corrective function should be identified. These functions will indicate the action required to insure that a malfunction has been eliminated. Corrective actions involve adjustments, removal and replacement of the malfunctioning unit, cable repair, alignment, etc. These actions should be specified to the level required to determine skill, space, access, and support requirements.

Recycle and overhaul. Recycle and overhaul maintenance are needed because regular preventive and corrective maintenance are usually insufficient for full operational capability. In some systems, such as aircraft engines, it is necessary to remove components for overhaul in order to restore acceptable reliability.

Analysis of recycle maintenance functions begins by identifying what units, if any, must be recycled or have major overhaul on site. Available reliability information and logistics considerations are factors in making this decision. Recycle also includes original installation of equipment.

12.2.5 Allocation of Maintenance Functions

The functions identified in previous analyses can be organized by type, area of performance, time of performance, subsystem or component involved, and sequence of the function in relation to other functions. This organization of functions will provide a structure for decisions concerning man–machine tradeoffs.

Tradeoff information about each function should describe the nature and extent of performance to be provided by the technician and by specified types of equipment. Cross-reference to similar tradeoffs for other functions should be made to determine the number of functions that can be performed by a single element, man, or machine.

Some of the tradeoff variables in deciding between manual or machine performance of functions are:

Cost. Machine functions tend to require higher initial investment, but may offset this with more rapid performance during use. In general, automatic equipment must have a high use factor to be justified from a cost standpoint.

Flexibility. Men are commonly most flexible, although automatic test equipment having a variety of programs and special adaptors may be quite flexible.

Reliability. Quality equipment with self-checking features probably compares favorably with human performance.

Speed. Automatic equipment is usually much faster than manual performance, particularly for high-speed checking. Automatic checkout is usually indicated where operational requirements demand speed.

Available manpower. Where the number of highly skilled technicians is adequate, the desirability of automatic support equipment is lowered. The reverse is true, but personnel requirements will actually be reduced only if planning for automation is carefully integrated into the total design of the maintenance system.

Type of function. Automation lends itself best to verification, except for inspecting. It may be useful for troubleshooting; at present, it has limited capability for adjusting, servicing, replacing, and repairing.

The above are general guidelines at best, and should be modified for any specific application.

12.2.6 Establishment of Maintainability Program Plans

Of particular concern in establishing maintainability program plans are interfaces between maintainability and other aspects of human engineering design, effective flow of information between maintainability design and personnel programming, and assurance that designers making decisions affecting technician performance have ready access to relevant human engineering information. Once these factors have been considered, the designer must: (a) establish a maintainability concept, and (b) describe a development plan.

The maintainability concept document. The maintainability concept document is a comprehensive description of maintenance design as it exists at some point in time or as it is projected into the future. During later stages of development, it is sometimes known as the maintenance plan.

Preparation of the maintainability concept as a formal document has four purposes:

1. It will permit widespread review and opportunity for suggested improvements.
2. It will provide maintainability designers with a convenient record of a milestone in its continuing development.
3. It will provide a convenient source document for design groups that must make assumptions about the maintenance subsystem in the course of their own development work.
4. It, by its very nature, will dictate certain

design decisions which in turn, will serve as constraints on the subsequent detailed design processes.

The maintainability concept document may include a series of revisions required to keep up-to-date design information available for all interested groups. Consideration should be given to the inclusion of the following areas of content:

Relationship to operational objectives—relating the maintenance system to mission success, turnaround, down time, reliability, etc.

Cost analysis—identifying the principal cost factors that have affected or are expected to affect design of the maintenance system.

Maintenance organization—identifying the maintenance levels or echelons, indicating the extent of maintenance and organization of each.

Personnel requirements—describing the probable needs for field engineering and military technicial personnel. This information is developed under the Personnel Subsystem as the qualitative and quantitative personnel requirements information program.

Logistics and supply—describing the relationship of the maintenance system to the supply system.

Facilities—indicating the probable existing and new facilities that will be required for operation of the maintenance system.

Prime equipment—describing characteristics of the prime equipment that will influence maintainability.

Support equipment—identifying government furnished equipment requirements and describing special support equipment under development.

Auxiliary job aids—describing the manuals, checklist, and other job aids planned for each maintenance position.

Like all other major elements of the system, maintenance design requires critical time phasing in its development and long lead-time planning if it is to fit effectively into the total system phasing. Even more than most aspects of systems, almost everyone on the design team seems to be responsible for a piece of the maintainability design effort. Unless explicit responsibilities are identified, considerable overlap in some areas may occur and others may be neglected.

12.3 Maintainability Design Assessment

Maintainability design assessment is a continuing process. It begins with maintainability predictions which are frequently updated and ends with a demonstration that maintainability requirements have been met. Maintainability predictions during design serve several purposes: (a) to alert the designer to specific design deficiencies, (b) to indicate the extent to which a particular design will meet maintainability requirements, and (c) to establish manpower requirements. Maintainability demonstrations included in military hardware acceptance tests are to verify the predictions and assure that maintainability requirements have been met.

12.3.1 Maintainability Prediction

There have been several efforts to develop effective maintainability prediction techniques. Some of the better-known prediction methods are described in Military Handbook, MIL STD 472, "Maintainability Prediction" (1966). However, none of the methods developed so far has been sufficiently impressive to gain general acceptance. While no attempt will be made to review all related research here, a few representative efforts will be summarized.

A Checklist for Evaluating Signal Corps Equipment

One of the first maintainability prediction efforts was sponsored by the Army Signal Corps (Munger and Willis, 1959). The result was a first approximation to an objective method for quantitative evaluation of maintainability during the development cycle, although the predictive validity of the technique was not established in terms of operational criteria. The method involved the use of a checklist of design features to obtain comparative scores for different electronic equipments.

Aircraft Maintainability Prediction Via Time Synthesis

Another maintainability prediction effort was undertaken by Convair (Schafer et al., 1961) under contract with the Aerospace Medical Research Laboratories. Under this charter, standard times for basic task elements were established and used to predict times for unscheduled maintenance actions. This technique is sometimes called "task time synthesis." It is based on the assumption that times for basic elements of tasks are linearly additive.

Major steps in the Convair prediction procedure include:

1. Analysis of subsystems to determine expected malfunction symptom patterns.
2. Selection of a sample of symptom patterns.
3. Identification of suspected components.
4. Determination of replacement times for suspected components.
5. Determination of time to prepare, test, align, and clean each component.
6. Calculation of mean maintenance time for each subsystem based on estimates of relative frequency that each symptom pattern occurs.
7. Calculation of maintenance load for the whole system based on estimated failure rates for subsystems.
8. Calculation of mean maintenance time for a whole system.

Standard times used for replacement and repair task elements were obtained by timing mechanics as they performed representative tasks several times and taking the average time for each task element. Replacement tasks consisted largely of fastener (access screws, bolts and nuts, electrical disconnects, safety wires) removal and replacement. Elements of preparation, test, alignment, and cleanup tasks included switch and handle operations, meter reading, valve operation, changes of position, inspection, etc.

Partial validation of the Convair prediction scheme was accomplished by (a) selecting a sample of 58 components from an operational weapon system; (b) predicting replacement times for each component on the basis of standard times for fastener operations, parts handling, and torque readings involved; and (c) determining the relationship between the resultant predictions and independent estimates made by mechanics on the basis of their experience with the equipment and maintenance tasks. Regression analysis showed that 83% of the variance in estimated replacement times could be accounted for in terms of the predicted times.

Schafer and his associates were of the opinion that failure to obtain a higher coefficient of regression was partially due to inadequate consideration for accessibility problems involved in replacement tasks. Time did not permit them to systematically explore effects of difficult work positions and restricted visual or physical access; hence, their standard times for task elements were derived largely from operations in which the mechanic had nothing in his way, could see what he was doing, and was in a comfortable position. Although the Convair prediction effort was not designed to establish quantitative relationships between specific design features and maintainability indices, it did suggest the importance of fastener considerations and accessibility.

Design Checklist Predictors for Electronic Equipment

A different approach to maintainability prediction was taken by RCA in another Air Force sponsored effort (Maintainability Techniques Study, and Maintainability Engineering, RCA, 1963). RCA's principal objective was to establish relationships which could be used to assess maintenance requirements early in the equipment design stage. Multiple correlation techniques and regression analysis were the principle tools used in developing the RCA prediction technique. Maintenance time was selected as the criterion of maintainability.

First, the RCA study team developed a set of descriptive items which could be used to score maintenance tasks. Items were chosen to represent three major maintenance parameters: personnel, support, and design. Personnel considerations included skill levels, attitudes, experience, and technical proficiency. Support items pertained to supply problems and the availability of tools, test equipment, manuals, and technical orders associated with the equipment. Design items covered equipment characteristics related to demands placed upon maintenance personnel, tools, and test equipment. Preliminary results were consistent with conclusions drawn from earlier research in that the design items showed the highest correlations with maintenance time. Personnel and support items showed low, non-significant correlations and were eliminated from further consideration.

The significant design items comprised three independent checklists. Checklist A was named "Physical Design Factors" and consisted of items primarily pertaining to access, fasteners, packaging, units, displays, fault indicators, test points, adjustments, and safety. Checklist B, "Design Dictates Facilities," included items related to provisions for test equipment utilization such as connectors, jigs, and fixtures, and requirements for interactions among technicians, operators, and supervisors or contractor personnel. Checklist C, "Design Dictates Maintenance Skills," provided for an evaluation of requirements related to human capabilities such as physical strength, manual dexterity, visual acuity, memory, patience, alertness, initiative, planning, and analysis.

Application of the RCA technique requires that a sample of representative features and associated maintenance tasks be selected for the system in question. System information required to apply the checklist scoring technique can be obtained from schematic diagrams, physical layouts, number of parts by class, predicted failure rates, descriptions of tools and test equipment, a knowledge of the operational environment and equipment functions, and a maintenance analysis of diagnostic procedures.

The following example, one of fifteen items from Checklist A, illustrates the guidelines provided for scoring equipment by the RCA method:

Access (External):

1. Access adequate both for visual and manipulative tasks (electrical and mechanical)	4
2. Access adequate for visual, but not manipulative, tasks	2
3. Access adequate for manipulative, but not visual tasks	2
4. Access not adeuate for visual or manipulative tasks	0

Item No. 1 *External Access*

Determines if the external access to systems, subsystems, equipments, or components is adequate for visual inspection and manipulative actions. Scoring applies to access problems for actions performed on the exterior of the external units and considers the extent of disassembly required to gain access to the interior of the components. Items to be considered include: access plates, panels, dust covers, cables, and other obstructions which interfere with maintenance.

Scoring Criteria

1. To be scored when the external access, while visual and manipulative actions are being performed on the exterior of the component, does not cause delay because of obstructions (cables, panels, supports, etc.).

2. To be scored when the external access is adequate (no delay) for visual inspection, but not for manipulative actions. External screws, covers, panels, etc., can be located visually; however, external packaging or obstructions hinder manipulative actions (removal, tightening, replacement, etc.).

3. To be scored when the external access is adequate (no delay) for manipulative actions, but not for visual inspections. This applies to the removal of external covers, panels, screws, cables, etc., which present no difficulties; however, their location does not easily permit visual inspection.

4. To be scored when the external access is not adequate (delay) for both visual and manipulative tasks. External covers, panels, screws, cables, etc., cannot be easily removed nor visually inspected because of external packaging or location.

Retterer (1965) and his associates at RCA found that scores on their three checklists, or predictors, accounted for more than 50% of the variance (a multiple correlation of 0.74) in a sample of maintenance times for electronic equipment.

Validation of the RCA technique as a predictor of electronic equipment maintainability showed a tendency for underestimation and rather large standard errors. Nevertheless, this effort tends to confirm the importance of design considerations in achieving maintainability. Physical design factors, as represented by Checklist A, consistently proved to be the best single predictor investigated by RCA.

Prediction of Shipboard Equipment Maintainability Via Time Synthesis

The United States Navy also has sponsored research on maintainability prediction. A prediction technique developed by the Federal Electric Corporation is included in the Bureau of Ships electronic equipment design guide ("Maintainability Design Criteria Handbook for Designers of Shipboard Electronic Equipment," Federal Elec. Corp., 1965). The Federal Electric Corporation method is similar to that developed by Convair (Schafer et al., 1961) in that it utilizes tables of standard times empirically determined for what are regarded as basic elements of maintenance tasks.

The Federal Electric researchers prepared nine timetables, one for each functional level at which maintenance may be performed. Functional levels were identified as follows: part, stage, subassembly, unit, group, equipment (set), subsystem, and system. The task elements consisted of the technician's activities in achieving subgoals of the maintenance process in general, i.e., fault localization, isolation, disassembly, interchange, reassembly, alignment, and checkout.

To apply the Federal Electric technique, the analyst must have sufficient information to establish the nature of malfunctions representative of the system in question. Guidance is provided to permit application of the prediction technique at various stages of system development. Having selected a sample of malfunctions, the analyst determines the functional level at which corrective maintenance will be performed. He then refers to the tables for appropriate task element times. The total repair time for each representative malfunction is multiplied by the failure rate for the type of component involved. The total of the resultant products provides a maintainability index for the equipment item involved.

Federal Electric Corporation used two Navy electronic equipments, a transmitter and a transceiver, to validate their prediction technique. Criterion data were derived from failure-report data on the two equipments. Predicted mean maintenance times were regarded as valid if they were within a 99% confidence interval of the geometric mean of reported times. Results were generally positive.

The general utility of prediction techniques which depend on empirically derived task times has been questioned because past experience indicates that accurate field data on maintenance times are difficult to obtain and often applicable only to the specific type of equipment and technician population sampled. Generalization to other situations can be expected to have validity only at a very gross level (Rigney et. al., 1966). Federal Electric's times were obtained from and validated against the performance of Navy technicians maintaining shipboard electronic equipment. Convair derived all of its data from the performance of company mechanics maintaining a specific type of aircraft. Therefore, large prediction errors might be expected from attempts to apply these techniques to systems involving task elements significantly different from those used for standardization.

This does not mean that other prediction techniques are not equipment- or situation-specific. Neither Retterer nor Munger and Willis attempted to establish the validity of their techniques for other than restricted classes of equipment.

Human Engineering Design Dimensions As General Predictors

A maintainability prediction effort conducted by Topmiller (1964) was motivated partially by the fact that no attempt had been made to test the generality of prediction techniques across heterogeneous items of equipment. Topmiller was also interested in identifying design dimensions which could be related quantitatively to maintainability criteria obtainable from a standard maintenance data reporting system, i.e., that provided for by U.S. Air Force Manual 66–1 (1963).

Topmiller chose to investigate the predictive validity of human engineering design recommendations included in maintainability design guides, e.g., Folley and Altman (1956). Pertinent design recommendations were extracted from the literature and converted into 117 questionnaire items, each representing a design feature. The questionnaires were used to obtain scores for electrical, hydraulic, and mechanical subassemblies for three different Air Force weapon systems: a tanker, a bomber, and a missile. The questionnaires were applied by mechanics or technicians familiar with the equipment in question. One of five alternatives had to be checked for each questionnaire item:

1. The feature is clearly a design characteristic of the equipment.
2. The feature is reflected in the design of the equipment to a great extent.
3. The feature is not applicable to the equipment.
4. The feature is reflected in the design of the equipment to a small extent.
5. The feature is not possessed by the equipment.

Criterion data for evaluating the predictive validity of questionnaire scores consisted of mean maintenance times derived from standard Air Force maintenance data reports required for the equipment in question by U.S. Air Force Manual 66–1.

Factor analysis of the resultant data disclosed seven factors, or subtests, which were significantly correlated with the maintainability criterion. An eighth factor (Subtest III below) was identified, but it did not contain enough items to have significant predictive validity. The sets of correlated items comprising each subtest were as follows:

Subtest I. Maintenance Safety

1. Are there any cables attached to the front of cabinets where they could be bumped by passing equipment or personnel?
2. Do units in excess of 45 lbs. but less than 90 lbs. have provision for two-man lift where lifting height is not in excess of 5 ft?
3. Are there any equipment units presumably designed for one-man manipulation that are too large or too heavy to be conveniently lifted, carried, pulled, pushed, or manually turned?
4. Are components so arranged in a given compartment that one type of specialist must wait for another type of specialist to remove his part before the first technician can get to his part?
5. Do access doors to high-voltage areas have provisions for automatically shutting off power (safety interlocks on door, etc.)?
6. Are there any lines routed such that the failure to securely clamp them in place would allow the lines to drop and be struck by moving parts?
7. Are all access covers, that are not completely removable, self-supporting in the open position?
8. Are cables or lines attached to units that can be partially removed (chassis on slide racks) attached so the unit can be replaced conveniently without damaging the cable or interfering with the securing of the unit?
9. Are controls and displays that are used for maintenance or calibration only, and that are not necessary to actual operation, kept distinct from operation controls?
10. Are cables routed so they cannot be pinched by doors, lids, etc. and so they will not be stepped on or used as hand holds?
11. Do labels identifying control handles appear on the handle as well as the panel, since the panel may be removed during maintenance?
12. Are access openings free of sharp edges of projections that constitute a safety hazard?
13. Where screwdriver adjustments must be made blind, are either mechanical guides provided or the screws mounted so that the screwdriver will not fall out of line?

Subtest II. Maintenance Information

1. Do display labels provide full identifying information?
2. Do display labels appear on every item that you must recognize, read, or manipulate?
3. Do display and control clearly indicate their functional relationship? Displays should be labeled by functional quantity rather than operational characteristics (i.e., gal., p.s.i., ohms, etc.)
4. Are access points individually labeled?
5. Are the functions of each control clearly labeled?
6. Are display labels imprinted, embossed, or attached so that they will not be lost, mutilated, or become otherwise unreadable?
7. Are displays and controls labeled to correspond to notations found in system diagrams, in T.O's., etc?
8. Does (do) access(es) have labels indicating what auxiliary equipment is needed for service, checking, etc., at this point?
9. Does (do) the access(es) have labels that tell what can be reached through this point?

10. Are appropriate labels used for each test point?

11. Are parts that require access from two or more openings properly marked to indicate this requirement?

12. Do access covers have permanent part numbers marked on the cover?

13. On component covers, are there display labels that provide relevant information concerning electrical, pneumatic, or hydraulic characteristics of the part?

14. Are all potted parts labeled with current, voltage impedance, terminal information, etc.?

15. Are displays and their associated controls clearly labeled to show their relationship?

16. Are color codes or other symbolic coding schemes used for identifying test points or tracing wire or lines easily identifiable under all illumination conditions?

17. Are electrical terminals clearly marked "+" or "−"?

18. Are U-lugs rather than O-lugs used for clamping?

19. Do display labels for each termination have the same code symbol as the wire or line attached to it?

Subtest III. Handling and Removal for Replacement

1. Are handles and grasp areas so located that at least 2 inches of clearance from obstructions is provided?

2. Are units designed to be removed and replaced provided with handles or other suitable provisions for grasping, handling, and carrying?

3. Are cables routed so they need not be bent or unbent sharply when being connected or disconnected?

4. In your judgment was the equipment designed to consider all environmental conditions (cold weather, darkness, etc.) under which it was to be maintained?

5. For removable units which have irregular, fragile, or awkward extensions, such as cables, waveguides, hoses, etc., are these extensions easily removable before the unit is handled?

Subtest IV. Fasteners and Tools

1. Are fasteners standardized to minimize the number of tools required?

2. Are access cover fasteners of the captive type?

3. Are mounting bolts semipermanently captive (preferably with snap-on collars)?

4. Are finger-operated fasteners used whereever feasible?

5. Are identical screw and bolt heads used to enable various panels and components to be removed and installed with one type of tool?

6. Are tool-operated screws of the type that can be operated by several tools (screwdriver, wrench, or pliers)?

7. Do hinged doors or covers have captive, quick-opening fasteners wherever possible?

8. Are there a greater number of fasteners (screws, bolts, clamps, etc.) on access panels than are required to maintain structural integrity of the unit?

9. Are field units and assemblies replaceable with nothing more than common handtools?

10. Are functionally similar units interchangeable between systems or subsystems?

11. Are combination-head mounting bolts with a deep internal slot (either single blade or phillips) arrangement and hex head used?

12. Are all cables, liquid or pressure lines. color coded and both ends tagged?

Subtest V. Alignment and Keying

1. Are assembly parts designed with orienting seats, pins, etc., to save time getting parts in proper position for fastening?

2. Is the proper orientation of units within their respective case made obvious, either through design of the cases or by means of appropriate labels?

3. Are components or access covers designed so they can be easily oriented for fastening by providing alignment pins or grooves?

4. Are all interchangeable units physically coded (keyed) so that it is impossible to insert a wrong unit?

5. Are units coded (labels, colors, etc.) to indicate the correct unit and its orientation for replacement?

6. Are guide pins or other means used on units and assemblies for alignment during mounting?

7. Is there some means of physical design to prevent mismating or interchanging con-

nections (unique fastener or socket, routing of cables or lines so lengths vary, spreader blocks, etc.)?

8. Are connector plugs designed so that pins cannot be damaged (aligning pins extend beyond electrical pins)?

9. Is the design such that it is physically impossible to reverse connections or terminals in same, or adjacent, circuits?

10. Are plugs, connectors, and receptacles clearly marked to show proper position of keys for aligning pins for proper insertion position?

11. Are keyways for tubes and tube sockets suitably marked so you do not have to rely on "feel" to find the proper position?

12. Are adjacent soldered connections located far enough apart so work on one connection does not compromise the integrity of adjacent connections?

13. Are control linkage attachments designed so that reversed assembly is not physically possible?

Subtest VI. Manual Control Layout

1. Do controls that are intended to have a limited degree of motion have adequate mechanical stops to prevent damage?

2. Are control knobs that only require occasional resetting provided with a cover seal or otherwise guarded against inadvertent actuation?

3. Are pointer knobs designed so there is no possibility of mistaking which end of the knob is the pointing end?

4. Are cranks designed and positioned with respect to the speed or load which they administer; that is, small cranks at elbow height for fast wrist action, light loads; large cranks oriented for full arm motion?

5. Do the controls that are multi-position selectors have detent positions to prevent leaving the switch between detent positions?

6. Do controls that are meant to have continuous movement (not detented) actually possess a small amount of smooth, even resistance?

7. Are all lubrication points accessible?

8. Are all controls located where they can be seen and operated from the normal working position—without disassembly or removal of any part of the installation?

9. Generally, do controls appear on panels in the sequential order in which they are to be used?

10. Are controls marked to indicate which direction to operate the control?

11. When more than one control is used for adjusting a single (or interacting) function, are the controls placed on the same panel?

12. Do units over 90 lbs. have provisions for mechanical or power lift?

13. Are externally mounted controls that must be operated without visual reference located in front rather than to the side or behind the operator?

14. Are internally mounted controls located away from dangerous voltages?

15. For fixed-procedure operation is the sequence of use of controls indicated by number?

16. Are exterior access doors on flight line equipment so large or heavy that they require more than one man to open and remove them?

17. Are all units weighing 45 lbs. or more prominently labeled with their weight?

18. Are sensitive adjustments located or guarded to prevent inadvertently bumping them out of adjustment?

19. Are access openings generally located at convenient working positions?

Subtest VII. Workspace Configuration

1. Is the equipment located so that awkward working positions are unnecessary?

2. Is there sufficient space to use test equipment and other tools required during checkout?

3. Can controls (switches, knobs, etc.) be easily reached for operation from a convenient working position?

4. Are components located so that physical interference among technicians working on the same or adjacent areas is minimized?

5. Is there sufficient space between connectors so they can be grasped firmly for connecting and disconnecting?

6. Does (do) the access(es) have labels which specify the frequency for maintenance either by calendar or operating time?

7. Are displays located so they can be observed without removal of other equipment or disassembly of any portion of the installation?

8. When maintenance activities demand the collaboration of a team effort, are traffic flow and communication adequate?

9. When group activity demands the use of a central visual display, are the lines of sight to the display blocked by poor arrangement of people or equipments?

10. Can displays (dials, gages, GO-NO-GO indications, etc.) be easily observed and/or read from a convenient working position?

11. Is printed matter displayed upright from your normal viewing position?

12. Are cables or lines which must be routed through walls or bulkheads designed for easy installation and removal without necessity for cutting or splicing of lines?

Subtest VIII. Accessibility

1. Are access openings designed so the technician can see what he is doing? (Clearance for hand only obscures the thing the technician is working on.)

2. Do access openings used for frequent visual inspection have transparent or "quick-opening" type covers?

3. Are units placed so they are not in recesses, behind or under stress members, floorboards, seats, hoses, pipes, or other items which are difficult to move?

4. Are high failure rate components easily accessible for replacement?

5. Are bulkheads, brackets, other units, etc. designed not to interfere with the removal or opening of covers of units within which work must be done?

6. Are access covers designed so that it is obvious how they are to be opened or closed?

7. Are access covers that remain attached to the basic equipment designed so they do not have to be held open or "dangle" in the way?

8. Are quick-disconnect devices (fractional turn, quick-snap action, press-fit, etc.) used wherever possible?

9. Are all the throw-away assemblies or parts located to be accessible without removal of other components?

10. Are maintenance displays located on the side normal for maintenance tasks of inspection, checkout, troubleshooting, removal, and replacement?

11. When selector switches have to be used with a cover panel off, are there duplicate switch-position labels on the internal unit so you don't have to refer to the label on the case or cover panel?

12. Are units that are frequently pulled out of their installed position for checking mounted on roll-out racks, slides, or hinges?

13. Is adequate clearance provided between switches and knobs to prevent inadvertent actuation, while allowing sufficient finger room for manipulation?

14. Are self-locking safety catches provided on connector plugs rather than safety wire?

15. Are cases designed as lift-off units rather than having the units lifted out of the cases?

16. Do covers or shields through which mounting screws must pass for attachment to the basic chassis of the unit have large enough holes for passage of the screw or bolt without perfect alignment?

17. Do miniaturized components as compared with nonminiaturized components tend to be mounted too close together for easy removal and replacement?

18. Are limit stops provided on rollout racks and drawers?

19. If components require "stacked" mounting to conserve space, are the components with the highest failure rate located on top?

Table 12–2 shows predictive validities for the seven significant subtests. Regression analyses showed that total scores on the seven-factor subtests accounted for 27% (multiple correlation coefficient = 0.523) of the criterion variance across all items of equipment. Division of the equipment into more homogeneous classes, i.e., electronic and nonelectronic, increased the predictive efficiency as follows: 37% of the criterion variance was accounted for in the case of electronic equipment; 47% in the case of nonelectronic equipment. All three percentage values may seem rather small unless one considers the several sources of variance in the underlying model. For example, a portion of the criterion variance is attributable to differences in the performance ability of the various technicians who accomplished the maintenance tasks. Additional criterion variance is attributable to errors in reporting, recording, and transmitting maintenance task times. And, of

TABLE 12–2. CORRELATIONS BETWEEN HUMAN ENGINEERING DESIGN SUBTESTS (FACTORS) AND REPORTED MAINTENANCE TIMES

	Factors	All equipment N = 45	Electronic N = 25	Nonelectronic N = 20
I	Maintenance safety	−0.185	−0.249	−0.408
II	Maintenance information	−*0.357	−0.298	−0.401
IV	Fasteners and tools	−*0.310	−0.353	−0.324
V	Alignment and keying	−*0.356	−0.134	−**0.536
VI	Manual control layout	−**0.464	−*0.455	−*0.481
VII	Workspace configuration	−0.190	−0.080	−0.289
VIII	Accessibility	−0.190	−0.186	−0.061
	R	*0.523	*0.609	*0.682
	R^2	0.274	0.371	0.465

Topmiller (1964).
* = 0.05 and ** = 0.01, significance values based on Wallace and Snedecor Tables (1931). (R = multiple correlation; R^2 = percent of total criterion variance accounted for by R).

course, accuracy in prediction was lost because of low interrater reliability in applying the questionnaire. (Topmiller obtained a median between-rater reliability of 0.62.)

Nevertheless, both Retterer (1965) and Topmiller (1964), in independent maintainability prediction efforts, showed that equipment design factors account for a significant proportion of variance in maintenance times. These findings, combined with those of Shapero et al., (1960) which indicate that 20 to 50% of all missile equipment malfunctions result from human errors attributable to poor human engineering, are convincing evidence of the need for continuing emphasis on maintainability design considerations.

Note that each of Topmiller's subtests exhibits some heterogeneity with respect to the nature of design features represented among the items. For example, three items which are normally regarded as accessibility considerations fall under "Manual Control Layout." The "Alignment and Keying" subtest also includes some items pertaining to connector design. This sort of complication is common to factor analysis applications and makes factor *naming* difficult. Consequently, a more traditional equipment-design taxonomy is used in subsequent discusssion.

12.3.2 Maintainability Demonstrations

Partial test plans for use in evaluating achieved maintainability are included in Military Standard (MIL STD) 471, "Maintainability Demonstration" (1966). All maintainability measures provided for by Military Standard 471 involve maintenance time, either preventive or corrective or both. Military Standard 471 specifies sampling procedures which are to be used to assure that maintenance tasks selected for demonstration purposes are representative. Comparable instructions are not provided for sampling the performance of maintenance personnel. The standard merely states that personnel used in the demonstration "shall be of the type, number and skill level representative of the personnel who will perform maintenance during the operational phase." Nevertheless, care should be exercised to assure that personnel performance is appropriately sampled. Maintenance performance can be expected to vary between individuals even when their designated skill levels are the same (Topmiller, 1964), and the same individual will not perform the simplest task with the same accuracy and speed every time.

Although it is difficult to generalize beyond the specific context in which maintainability demonstration data are obtained (because there are so many uncontrolled variables and the number of observations is so small), maintainability assurance personnel typically rely upon insights gained from demonstration data as a basis for possible design improvements. A summary of maintenance data obtained from 12 maintainability demonstrations for satellite control stations is shown in Table 12–3 (Rigby, 1965). This information suggests the importance of display design and accessibility in satellite control stations.

TABLE 12-3. SUMMARY OF MAINTENANCE ACTIVITY DATA FROM SATELLITE CONTROL STATION MAINTAINABILITY DEMONSTRATIONS

Maintenance activity	Percent time
Watching panel indicators	19
Gaining and closing accesses	13
Consulting with supervisor and each other	12
Finding and using test points	9
Getting and setting test equipment	9
Removing and replacing LRU/spares	9
Removing and replacing connectors	6
Finding and using documentation	5
Checking and aligning (adjustments)	3
Other	15
Total	100

Rigby (1965).

12.4 Maintainability Design Features

The following sections of this chapter include discussions and examples of design features which have been found useful in enhancing system maintainability. These features, although generally applicable, are not to be applied blindly. Blanket adherence to any set of design features is no substitute for the planning and analytical steps outlined in the introductory sections of this chapter. A thorough understanding of underlying principles and overall system objectives is of prime importance. Tradeoffs on a given design problem will, at times, require that certain design features be modified or adapted. In the long run compromises which work to the disadvantage of those responsible for operation or maintenance of the equipment may create personnel selection and training problems and prove generally detrimental to system operation.

12.4.1 Accessibility

General Considerations

It is fundamental that a maintenance technician be able to gain access to the equipment he is to maintain. Inaccessibility can render the simplest task impossible. There must be sufficient clearance to use the tools necessary to accomplish the job. There must be adequate space to permit convenient removal and replacement of components, and there should be adequate visual access to the task area. If adequate access is not provided, attempts to accomplish maintenance often result in equipment damage, skinned knuckles, and lost time. Moreover, simple tasks, such as routine inspections or other preventive maintenance, become tedious and time-consuming and may be neglected or done poorly. The system designer should give special attention to the location of the work place in relation to the task, postures dictated by the task, anthropometric dimensions and physical capabilities of the technician, and environmental and clothing factors. In a few instances, relatively complete and specific data are available in the literature for use in determining minimum access size requirements. When this is not the case, the designer may have to estimate space requirements on the basis of task analyses, dimensions and functional characteristics of tools and equipment components, and anthropometric data such as are presented in Chapter 11 of this *Guide.*

Space requirements for tool use. In planning for accessibility it is important to consider the specific nature of tasks to be performed through or within each access area in order to determine the tools necessary for the task and space required for their use. Kama (1963, 1965) has collected data on the effect that various degrees of access restriction can have on performance times for simple manual tasks. For example, Figure 12-2 shows the average removal and replacement time for a component as a function of the size of the access aperture and the depth at which the component is located within the access area. The component was secured by means of four screws which were removed and replaced by means of an ordinary screwdriver. Note the marked effect of depth at the smallest aperture size.

The importance of equipment orientation is brought out by the contrast between Figure 12-3 and 12-4. Conditions under which the data represented by these two figures were collected were the same except for the following differences. For Figure 12-3, the base of the component was mounted to the left side of the aperture so that the technician approached a side of the component. For Figure 12-4, the component was mounted at the back of the aperture so that its top faced the technician. (See Figure 12-5.) The change in component

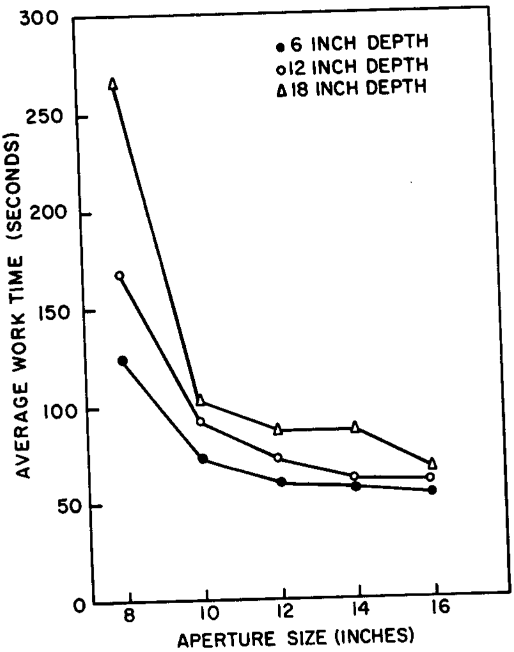


FIGURE 12–2. Average work time for removing and replacing a component as a function of depth and aperture size (Kama, 1963).

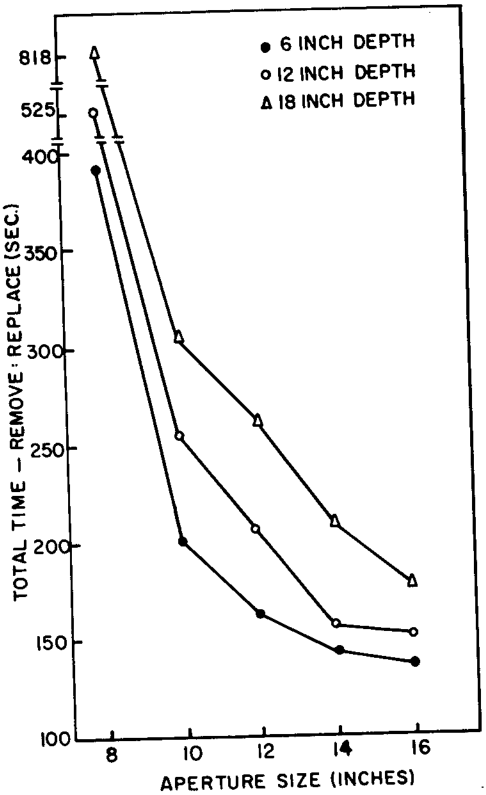


FIGURE 12–3. Average total work time for removing and installing a side-mounted component as a function of depth and aperture size (Kama, 1963).

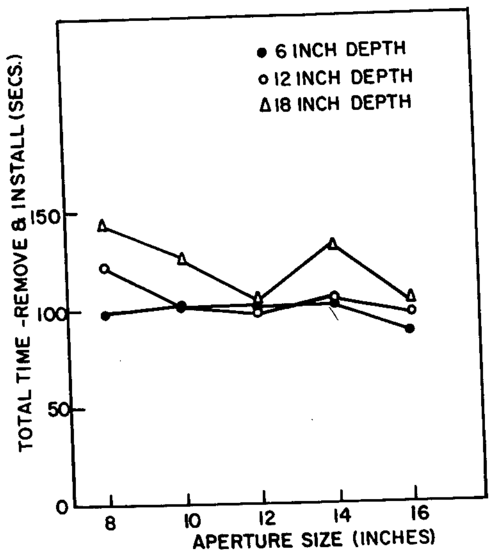


FIGURE 12–4. Total work time for removing and installing a backmounted component (Kama, 1963).

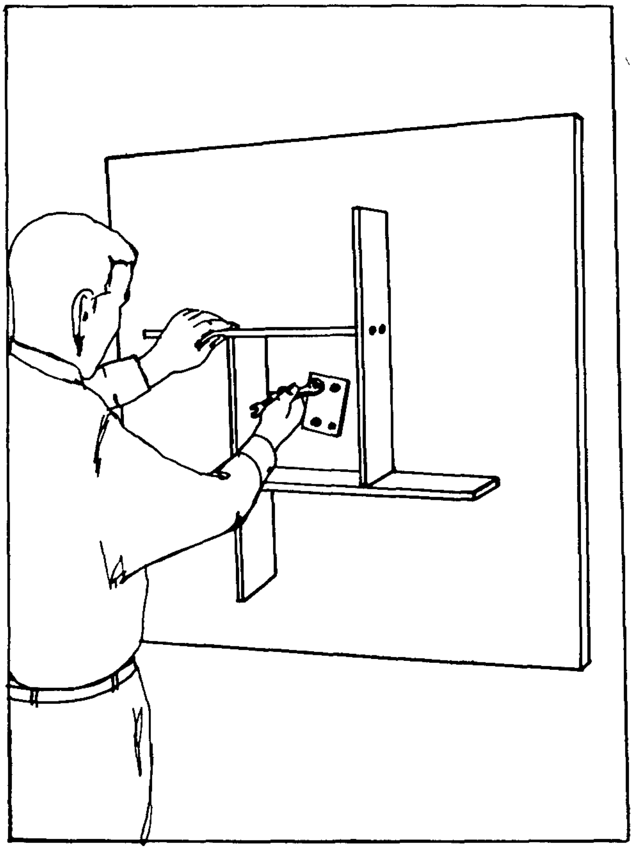


FIGURE 12–5. Apparatus used in study of component location and aperture size effects ("back-mounted" condition) (Kama, 1963, 1965).

orientation mounting position from side to back tends to nullify effects of both depth and aperture size. Relatively simple but important interactions such as this may be overlooked unless careful attention is given to maintenance requirements during design.

Some empirical data on the volume swept in using various hand tools have been tabled and illustrated by Altman et al., (1961). Although the data are based on a single subject, they provide useful guidelines.

Work position. The position which the man must assume to do his job must be carefully considered. What a man can and will do varies as a function of the place and position which must be assumed. A relatively simple and conveniently accessible task may, in fact, be very difficult or impossible if the technician must assume an awkward working position, use a platform, or stand on an insecure or slippery surface. Additional related information on human body dimensions, muscle strength, and reach capabilities is included in Chapter 11.

Protective clothing effects. The size of access openings and spacing between components must be large enough to accomodate effects of any protective clothing which maintenance personnel may be required to wear. The necessity for wearing heavy gloves, pressure suits, and similar garments makes intricate tasks difficult to perform even under optimal access conditions. Experiments have shown performance times for a simple maintenance task to be 2.5 times more for pressure-suited conditions (5 p.s.i.) than for shirt-sleeve conditions (Seeman et. al., 1966).

Gillespie (1965) has tabled data which show how much pressure suits add to specific body dimensions.

Visual access. Minimum access sizes often reflect only the amount of space required for physical access or manual manipulation involved in tasks. Provision also should be made for visual access so that the technician will not have to work "blind." Special lighting should be considered for critical tasks likely to be obscured from standard facility lighting. The need for ready access for inspection purposes should also be considered. Transparent viewing ports may meet such needs; however, sometimes it is also necessary to touch or handle components in order to detect temperature discrepancies, breaks, corrosion, wear, etc.

Types of accesses. Access to equipment is usually afforded in one of the following ways, listed in order of preference from the accessibility viewpoint:

1. Exposed equipment; when compatible with structure, environmental, operational and safety requirements, equipment should be left exposed for maintenance. In a building or large vehicle, test and service points, maintenance controls and displays, and rack-mounted modules ("black boxes") are usually amenable to exposure except as precluded by safety requirements.
2. Exposure of equipment via quick-opening devices; for example, equipment mounted in drawers or in racks with full-length doors, quick-opening hoods, or dust covers.
3. Limited access openings without covers.
4. Limited access openings with covers.

Design Recommendations for Accessibility

Size and shape of accesses. Accesses need not be of regular geometric shape. Whatever configuration best satisfies both structural and accessibility requirements should be used. Dimensions of accesses should be determined by factors such as the size of tools and replaceable components, critical anthropometric dimensions, the extent and direction of forces that must be applied during removal and replacement actions, etc. One large access is generally better than two smaller ones, but when structural or other constraints dictate, separate openings may be provided for visual and physical access.

Location of accesses. Access openings should be located so other equipment will not have to be removed from the installed position in order to perform the required maintenance action. The final arrangement of equipment in the system must be considered to assure that accesses are located on exposed surfaces. It is also desirable that maintenance accesses be located on the same face of the equipment as are related displays, controls, test points, etc.

The technician's work place, i.e., where he must stand or sit, should be considered in locating accesses. The height of an access should conform to anthropometric capabilities as described in Chapter 11. When accesses must be located at heights which exceed the effective reach capabilities of personnel, provisions must be made for appropriate support equipments such as work stands, steps, carts, or ladders. Accesses for heavy units (over 100 to 150 lbs.) should be located so powered assists, e.g., fork lifts, cranes, etc., may be used for removal and replacement.

Access covers. Hinged doors, provided with mechanical devices for holding them open to positions allowing maximum access, are usually preferred to cover plates or panels. Lift-off covers are acceptable if they do not fit so tightly as to require painstaking care in replacement. Covers should be large enough to prevent damage to wires and fragile components during removal and replacement. Hinged covers require space equal to their size in order to open. When space is limited a tongue-and-slot design may be used in lieu of hinges.

Access cover fasteners. Fasteners for access covers and panels should be kept to a minimum, hence, the desirability of hinges or tongue-and-slot design. Quick-opening catches or fasteners which require only part of a turn are often adequate. If possible, all fasteners should be captive and hand operable. Unconventional hand tools should not be required. Ordinarily, the same type of fasteners should be used for all covers and cases unless a different fastener serves a special purpose. Then the special fastener should be used only where it is required; in fact, it probably should be incompatible with other applications so that fasteners will not be interchanged by mistake. When screws or bolts are used, heads which will accommodate two types of tools, e.g., wrench and screw driver, are preferred. Fasteners should be located so that they are accessible without removal of other units of equipment.

Labeling accesses. Labels convey the following information about accesses:

1. If the manner in which a cover is opened is not obvious from its structure, it should be labeled with appropriate instructions.
2. When a tube or plug has to be put through a small hole, a conspicuous label can be used to show how the pins on the tube or plug will line up with the holes in the socket.
3. An access may be labeled with a number, letter, or other symbol that identifies it in the maintenance instructions.
4. Labels may be used to indicate items that can be reached through an access and the service equipment, if any, to be used there.

Drawers, racks, and hinged units. Making the chassis an integral part of the equipment structure, e. g., pull-out shelves or drawers, is a technique frequently employed in the design of large electronic equipments. Integral-chassis designs usually provide significant space savings over removable panels or hinged access doors. Pull-out drawers make equipment units accessible from several sides without removal of the console from its normally installed position. However, maintenance personnel still must have work space around the drawer. For example, a minimum of 18 in. clearance on each side of an extended drawer is recommended. If access is not required on all three sides and the unit can be removed and carried by one man, the clearance may be reduced to as little as 4 in. on one side (MIL–STD–803A–1).

Doors, hatches, and passageways. Traffic areas, including hatch openings, doorways, and stairs leading to aisles, corridors, and other passages, concern the maintainability engineer for two reasons. First, these areas must accommodate items of prime and support equipment which will be carried through them during the course of maintenance. Second, although not generally regarded as work area, traffic space frequently affords space essential to operations involved in the removal, replacement, or checkout of subsystems and components. Consequently, the maintainability engineer must guard against traffic space tradeoffs in favor of increased equipment volume or operator conveniences that jeopardize system maintainability. Recommendations for the design of traffic areas in general are included in Chapter 10.

Final installation. Retention of accessibility should be a prime consideration in planning for final installation of equipment. An item of equipment which is easily maintained in isolation may become a maintenance burden within the system context. In fact, the entire accessibility design effort can be nullified by improper installation. Since available space is almost always limited, efficient use of space is essential to achievement of accessibility objectives. Some useful principles for planning the layout of the final installation are listed below:

1. Arrange equipment items so that maintenance by one specialist will not require removal or handling of equipment by another specialty type.
2. Arrange equipment components according to the sequence of operations involved in basic maintenance procedures.
3. Give prime locations to equipment requiring more frequent maintenance.
4. Do not stack or otherwise arrange equipment so that good units must be removed to gain access to malfunctioning units.
5. Design for removal of units through the front rather than the back of consoles.
6. Do not hide test points, adjustment controls, etc., behind units of equipment.

Safety features. The following suggestions call attention to safety considerations sometimes neglected in the design of accesses:

1. Provide internal fillets or other protection on the edges of accesses that might otherwise injure the technician's hand or arms.
2. Access covers, cases and handles should have rounded corners and edges to minimize the possibility of injuries and equipment damage.
3. Accesses that lead to equipment with high voltage should be equipped with safety interlocks so that electrical circuits will be opened upon removal of the access cover. If the technician should need to work on the equipment with the power on, provide a "cheater" switch that automatically resets to its safety-protection position upon replacement of the access cover.

4. Provide conspicuous warning labels on all accesses leading to high voltages, rotating machinery, or other hazards.

5. Design for positive indications to show that access covers are unsecured even though in place.

12.4.2 Unitization and Packaging

Systems are almost always composed of several smaller units of equipment which perform some logical system function. The smaller units, in turn, usually consist of an assembly of still smaller components. Packaging equipment into separate units, if done properly, can significantly enhance maintainability. For example, units of equipment are often designed so that their functional characteristics are precisely defined and easily measured; this facilitates troubleshooting and fault isolation procedures. (See "Maintenance Procedures," Section 12.4.11.)

Unitization can also reduce handling problems, e.g., small, fragile components are easier to handle when packaged as a group and larger units can be broken down into subassemblies which are lighter and smaller. Similarly, less reliable units can be designed and mounted for quick replacement without removal of larger, expensive assemblies of equipment. In general, if system designers, maintenance engineers, and user organizations plan together, packaging arrangements and unit design will contribute to efficient division of maintenance responsibilities within the total maintenance structure.

Packaging Techniques

It is standard practice to arrange and package components primarily on the basis of requirements related to heat dissipation, available space, size and weight of components, and similar variables. However, since resultant layouts of components and circuitry vary significantly from one type of equipment to another, design engineers would do well to consider other approaches which not only conform to standard engineering requirements but also produce arrangements which are logical and consistent. Three alternative packaging principles for electronic equipment have been investigated experimentally. Each method facilitated maintenance more than the standard practice described above. They are listed and described below in order of demonstrated merit.

Logical-flow packaging. Logical-flow packaging facilitates systematic troubleshooting and reduces the need for complicated schematics and supporting diagrams. It is amenable to automatic or semiautomatic diagnostic techniques and appears to be compatible with current trends in integrated circuitry and microminiaturization. Advantages of logical-flow design derive primarily from the following features:

1. Circuitry and component arrangement clearly indicate the direction of current flow.
2. Modules and subassemblies parallel stages of the block diagram.
3. Modules perform specific functions and are readily replaced.
4. A fault can be isolated to a specific unit by a simple input-output check.

Circuit packaging. Adherence to this principle requires that each circuit be packaged with its associated parts as a replaceable module. All identical or similar circuits usually are grouped together. The resultant modules are usually either identical or very similar to those used in the logical-flow method. Such modules represent meaningful functions which are easily checked. Circuit packaging is also adaptable to automatic checkout equipment.

Component grouping. Consistent use of this method requires that all components of identical or similar nature be grouped together. Inexpensive components, e.g., resistors and condensers, may be mounted on plug-in boards. A goal of this design approach is to facilitate mass replacement of components for which troubleshooting is a tedious and time-consuming task.

Location of Components

There are a number of maintainability design features which apply to unit design provided that they are properly integrated with the packaging techniques. For example, delicate components should be placed so that they are not likely to be damaged during the course of repair to other components, but this does not mean that they should be removed from their proper position within a logical-flow layout. Similarly, components should not be stacked so that replaceable components are not accessible. The orderly mounting of components on a flat surface is generally preferred, e.g., parts on one side of a board or chassis and wiring circuits on the other. Units and components also should be located and arranged so that the number of inputs, outputs, and crisscrossing of conductors is kept to a practical minimum. Troubleshooting problems are sufficiently difficult even with a minimum of circuit tracing and connector repair.

Mounting and Assembly

Units of equipment which will require frequent inspection, repair, or replacement should be designed as plug-in modules or mounted in rollout drawers, if feasible. Drawers typically require less access space. (See "Drawers, Racks, and Hinged Units," in Section 12.4.1.) Smoothly operating bearings facilitate operation and removal of racks or drawers, but they should be equipped with locking mechanisms to hold them in place during maintenance and between maintenance actions. Drawers and racks should be designed with limit stops, but the stops should be conveniently releasable to permit drawer removal. Drawer and rack console design should preclude tipping of the entire assembly when the drawers are extended.

Mounting fasteners should be minimal in number and as easy to remove as stress requirements will permit.

Alignment, Support, and Handling Aids

Maintenance problems can be reduced by using supports, guides, keys, and key ways to assist in handling, aligning, and positioning units. Such provisions are especially valuable when visual and physical access to equipment is restricted. Bottom-mounted alignment pins may be used for units which are easily lifted by one man, i.e., those weighing twenty pounds or less. Heavier units should have side-mounted brackets or alignment devices so they can be slid, rather than lifted, into place.

Units weighing more than 10 lbs. should have handles, located above the center of gravity, for use in removing, replacing, and carrying the equipment. Handle dimensions should accomodate hand sizes as specified in Chapter 11. Units which weigh more than 25 to 40 lbs., depending on the bulkiness of the unit, should be designed for two-man handling. Special lifting eyes should be considered for units heavier than 100 to 150 lbs.

Sometimes it is desirable to incorporate rests, stands, or other supports into a unit design to hold the equipment in place and protect fragile components during maintenance as illustrated in Figure 12-6.

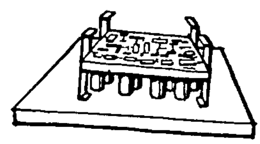


FIGURE 12-6. Illustration of one method providing stands which protect components during maintenance.

Identification of Units

Each unit of equipment should be clearly identified with respect to functional characteristics and other information pertaining to its proper use and replacement. Permanently etched or embossed lettering is preferred to paint or ink. If surface labels must be used, decals or stamped labels are usually easier to read than stenciled labels. Explicit, numerical specification of information such as electrical characteristics of resistors, condensers, etc., is more desirable than color-coding. If color-coding is used, the colors should be of a permanent nature and consistently used throughout the system. Code meaning should be unambiguous and clearly explained on a visible equipment panel, if possible, and in technical manuals for the equipment. Additional information on labeling and coding is found in Section 12.4.6 of this chapter and in Chapter 3.

Interchangeability

When designing units and selecting equipment components, the advantages of standard, interchangeable, regularly stocked components should be considered. The availability of replacement parts is critical to efficient maintenance. Similar units which are *not* interchangeable should be clearly distinguished to prevent mismating; design for physical incompatibility is preferred to labeling alone.

Adjustment, Inspection, and Lubrication Provisions

Functional modules should be designed so that each unit can be checked and adjusted independently with minimal requirements for additional adjustment when they are assembled to form the system or subsystem. Units which include lubricated mechanical assemblies should be equipped with external fittings so that disassembly will not be necessary when lubrication is required. Similarly, units which require regular inspection for signs of excessive wear sometimes can be designed with cover plates or openings which make complete disassembly unnecessary, e.g., brake linings.

12.4.3 Connectors

Cable and line connectors should be designed for maintainability as well as for reliability and manufacturing economy. For example, soldering wires directly to terminals may be the most efficient method for achieving reliable electrical connections provided no maintenance is required, but, when heat is applied to connections during removal and replacement of associated components, adjacent connections and components may be damaged. If a problem of this type is anticipated during the preliminary design stage, the solution can be relatively simple, e.g., provision for adequate spacing between soldered connections and adjacent components or, perhaps, the use of some other type of connector. Similar maintenance problems related to the design of connectors and possible solutions are suggested in the following sections.

Connector Types

Ordinarily, connections should not be made permanent or semi-permanent if removal and replacement of the component is expected. Hand-operated plug-in contacts, which are easy to connect or disconnect, are especially desirable for on-line maintenance. When a positive, stress-resistant connection is desired, preference should be given to hand-operated, quick-disconnect devices which require less than one turn for release. For ease of maintenance, screw terminals and lugs are better for electrical connections than solder connections. U-lugs are preferred to O-lugs because the latter cannot be disconnected without complete removal of the screw. Bolt or screw assemblies and threaded connectors require more time for operation and should be used only when necessary. Connector operation should require only common hand hand tools. (For more detailed information on connector types, see Aeronautical Systems Division Technical Report 61–424, Rigby et al., 1961.)

Location and Spacing

When possible, the designer should consider the final installed position for a component or unit of equipment and locate connectors so that they will be accessible both visually and physically. Insufficient clearance is likely to result in damage to equipment, injuries to personnel, and lost time.

Reference should be made to anthropometric dimensions in determining whether connectors are within the functional arm reach of maintenance personnel. Similarly, connections must be spaced so that sufficient clearance is allowed for operation whether by bare hand, gloved hand, or with tools. For example, terminals for soldered connections should be spaced far enough apart to permit proper clearance for a soldering iron. The end of the wire also should be left exposed to facilitate removal, (See Figure 12–7.) (For detailed information on space required for hand operations, see Aerospace Medical Research Laboratories Technical Documentary Report 61–424, Rigby et al, 1961; MIL STD 803–A–1, Aeronautical Systems Division Technical Report 61–381, Altman, et al, 1961; and Chapter 11 of this text.)

Keying and Alignment

A single connector or plug-in unit is often designed to include multiple pins or contact points. This eliminates the need for making each connection independently. However, misalignment with receptacles can result in dam-

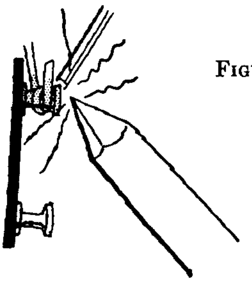


FIGURE 12-7. Clearance between terminals should be left for a soldering iron, and the end of the wire should be exposed to allow for removal.

age to the pins, bad connections and, consequently, inoperable equipment. To facilitate proper alignment of such connectors, key ways and keys or alignment pins, which extend beyond the contact pins, should be provided. Stripes or arrows on the sides of matching plugs and receptacles are also valuable aids to proper alignment and orientation. Symmetrical arrangements of alignment pins which are conducive to 180° misalignment errors should be avoided.

Identification and Coding

Mismating of receptacles and connectors may result in equipment damage and personal injuries. Design for physical incompatibility is the best safeguard against such errors. Physical incompatibility can be achieved through the use of different sized connectors, variations in alignment pin and key way arrangements, etc.

Visual cues may also be used to alleviate connector problems. Permanently installed receptacles should always be labeled to indicate the nature of the output and the appropriate connector. Adjacent connectors can be coded numerically or by color to facilitate differentiations among similar connectors. Test points designed into multiple-contact electrical connectors should be clearly coded or labeled to indicate which conductor is tapped at each point.

Coding schemes should be consistent throughout the system. If standardized codes are available, they should be used. Principles for visual displays presented in Chapter 3 should be observed where appropriate.

Safety Features

Connectors should be designed, located, and protected to minimize dangers to personnel and equipment from pressures, voltages, or other hazards associated with the release or handling of connectors. Most of the preceding connector design recommendations contribute to safer working conditions as well as improved maintainability, e.g., appropriate labeling as to outputs, safeguards against mismating of connectors, adequate access and spacing provisions, etc. Interlocks between connectors and access panels or covers so that the power source is disconnected automatically upon removal or opening of the equipment are sometimes desirable for safety reasons. Recessed or shielded receptacles provide added protection against high voltages. Receptacles should always be "hot" and plugs, "cold." Electrical plugs with self-locking safety catches are preferable to plugs which must be safety wired by the technician.

12.4.4 Conductors

Accessibility

Although conductors vary in design and function, they have maintainability features which apply to conductors in general. For example, conductors which must be disconnected for removal and replacement of components should be located for accessibility and equipped with convenient connector types in accordance with guidance provided in the preceding section of this chapter. Conductors should be located for convenient removal and replacement when required, but it should not be necessary to remove or disconnect them to gain access to unrelated items of equipment. If it is expected that maintenance actions will require some movement of the equipment, e.g., rolling out drawers, with the conductors connected, sufficient conductor length and flexibility should be incorporated.

Routing

Conductors should be sufficiently rugged or be routed and protected to avoid damage to them during operation and maintenance. Conductors should not be routed so that they will serve as handholds if such use is apt to damage the conductor or injure personnel. Conveniently located handholds also should be provided. Conductors should not be routed against objects with sharp edges. (See Figure 12-8.)

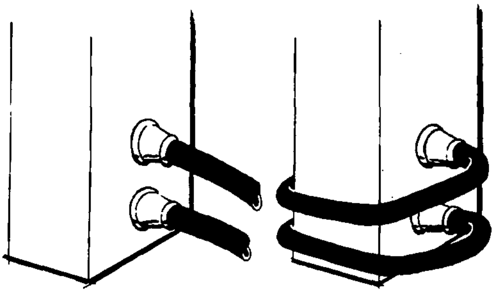


FIGURE 12-8. Routing cables againt objects with sharp edges is not good design practice.

Identification and Coding

When multiple conductors are used in a system, color-coding can be used for differentiating them and for tracking specific conductors. A standard color-coding technique should be adopted and used consistently for each category of conductors, i.e., electrical, hydraulic, pneumatic, etc. (See U.S. Air Force Exhibit 58–20, and U.S. Air Research and Development Command Manual 80–5, for additional information.)

Specific Types of Conductors

Cables. Electrical cables and wires, although generally flexible and somewhat elastic, are subject to damage with attendant hazardous conditions if high voltages are involved. They require special protection against sharp bending, pinches, and cuts from sharp metal edges. When the full length of long cables is not in use at all times provision should be made for automatic rewinding or recoiling. Long internal cables should be secured to the chassis by clamps appropriately spaced to keep the cable out of the way and clear of damage. When multiple conductors cannot be combined into a single cable, it may be advantageous to form the conductors into harnesses.

Appropriate materials should be used for insulation to protect wiring from anticipated hazards such as high temperatures or corrosive substances. Electrical cables should be routed to avoid proximity to high-temperature sources or fluid conductors. Use of clear insulation facilitates the detection of breaks in wiring.

Tubing. Tubing should be fastened in place to prevent damage from vibration or bending. Tubing is especially subject to damage from use as handholds or steps and should be appropriately protected. Labels should be provided to indicate fluid type, operating pressure, direction of flow, and operating temperature. Standard tubing sizes and fittings should be used. Actually, when pressure and environmental conditions permit, flexible hose is preferred to tubing because it is easier to install or relocate. High-pressure lines should also be fastened securely to prevent injuries from whipping.

Piping. Pipe is more rigid than tubing or cables and is difficult to remove; hence, space allowed for tool operation during the connecting and disconnecting of unions is often critical to efficient maintenance. Pipes, like tubing, should be labeled to indicate type of fluid carried, direction of flow, pressure, and temperature. Piping also is subject to use as steps or handholds so extra routing precautions may be necessary to avoid damage to equipment and injury to personnel.

12.4.5 Fastener Design

Fastener design principles are relatively simple; however, neglect of certain considerations can lead to problems. In fact, Schafer et al. (1961), found the number of fasteners involved in a maintenance action to be one of the best predictors of maintenance time. Difficulties in fastener removal are always frustrating and time-consuming. Equipment can fail because personnel neglect to replace cover plates having fasteners that are difficult to remove and replace. Dropped fasteners also may cause extensive damage to mechanical equipment. Such problems can be reduced significantly with a minimum of effort and cost if the causes are anticipated during the planning and design stages.

Ease of Operation

The number of turns required to secure or loosen a fastener should be kept to the minimum consistent with actual sealing, stress, or safety requirements. Similarly, bolts should not be too long and screw threads should not be finer

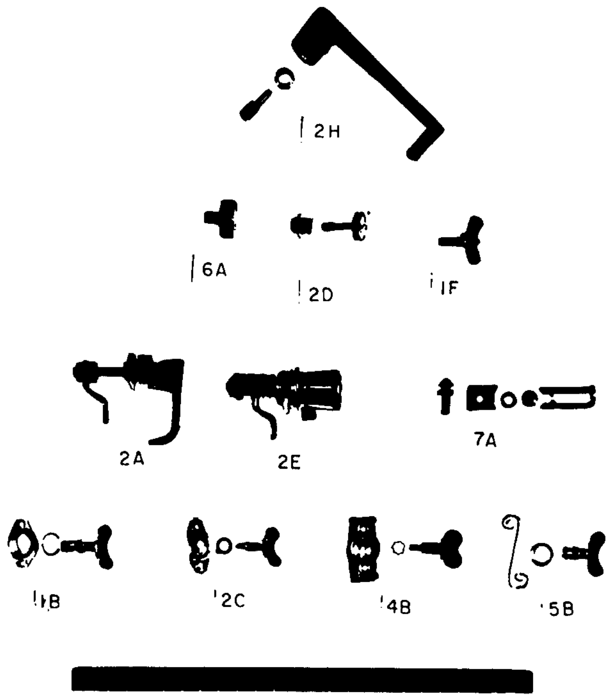


FIGURE 12–9. Some alternate designs for quick-release handoperated fasteners (Martin et al., 1968).

than necessary. Quick-opening devices which snap on and off or require less than one turn to operate are preferred. Tool requirements should be kept to a minimum. Hand-operated fasteners are preferred when practical.

Some proposed designs for quick-release hand-operated fasteners are shown in Figure 12–9. These fastener designs have been evaluated experimentally under normal gravity (static) and simulated zero-gravity conditions (Martin et al., 1968). Principal criteria were operation time, force, and torque requirements required for removal and installation of an access panel held by two fasteners. Mean operation times for 10 subjects (3 trials each per fastener type) for the two gravity conditions are shown in Figure 12–10. Mean resultant momentum values (average peak force multiplied by average operation time) are shown in Figure 12–11.

Permanent fasteners, such as rivets, should not be used to mount components which will require removal or replacement. Self-locking devices are preferred to those which require safety wires or pins. Cotter pins are easier to remove with common tools if they have large heads and do not fit too tightly. It is also important that proper clearance be allowed for the use of tools. Fasteners located too close to internal corners can be especially troublesome.

Durability

Although the use of corrosion and rust-resistant materials is important, fastener durability is also a function of design. For example, shallow-slotted screw heads are easily damaged. Hexagonal heads are generally better, in this respect, than slotted, square, or knurled heads. Bolt and screw heads which provide for internal and external wrenching are better; when one surface is damaged, the fastener may still be removed with an alternate conventional tool.

Captive fasteners are desirable especially when access to one end of the fastener is restricted. Dropped nuts are often lost and may result in equipment damage. However, captive fasteners should be designed for ease of replacement when they become damaged or worn. (See Figure 12–12.)

Fasteners used for similar purposes should

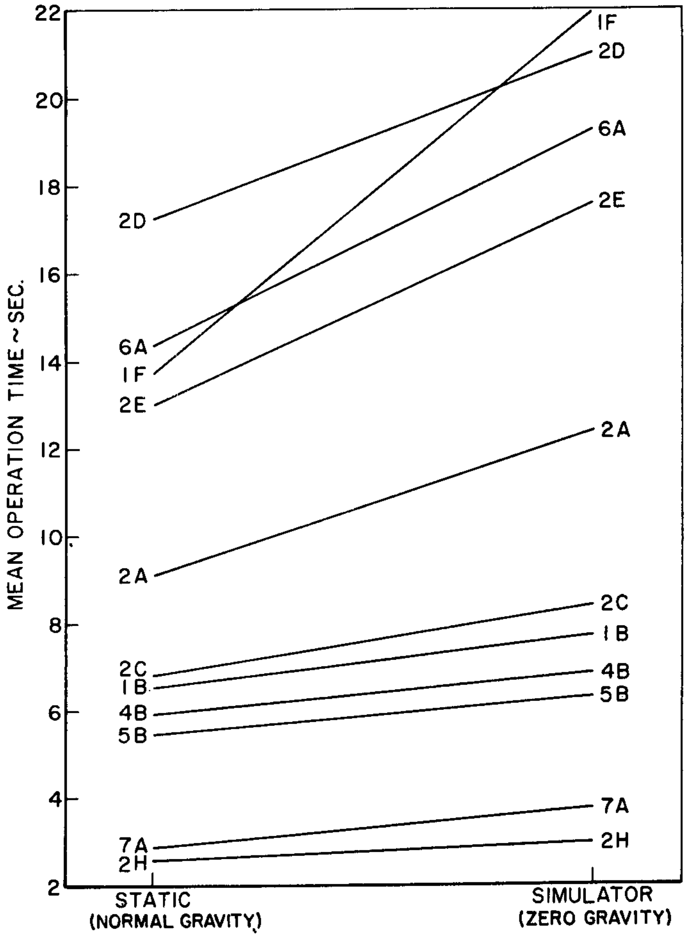


FIGURE 12–10. Mean operation time for fasteners shown in Figure 12–9 (Martin et al., 1968).

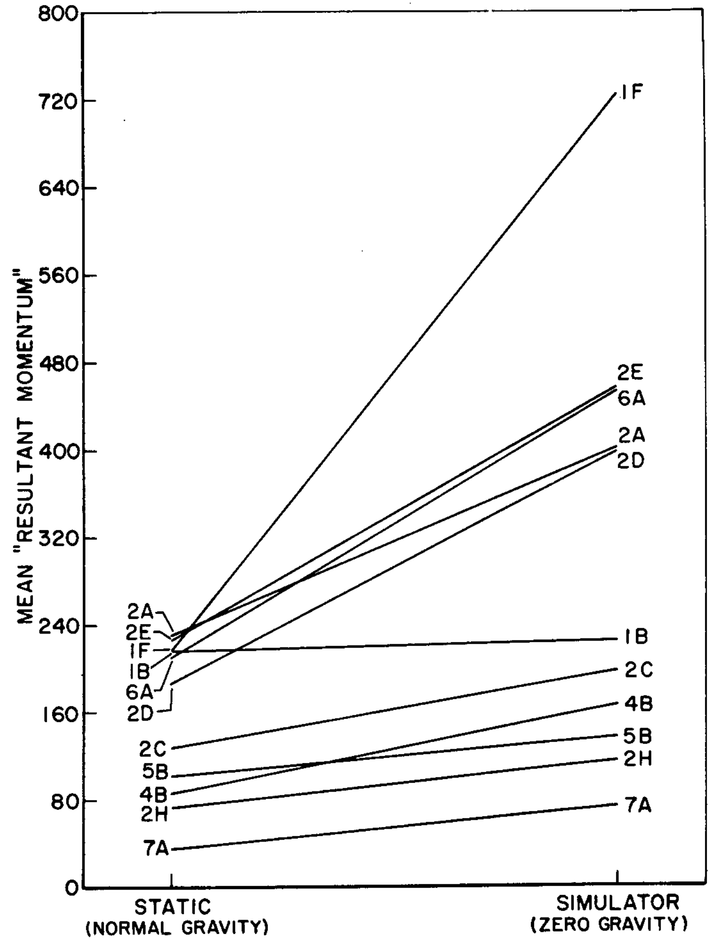


FIGURE 12–11. Mean resultant momentum associated with operation of the eleven fasteners shown in Figure 12–9 (Martin et al., 1968).

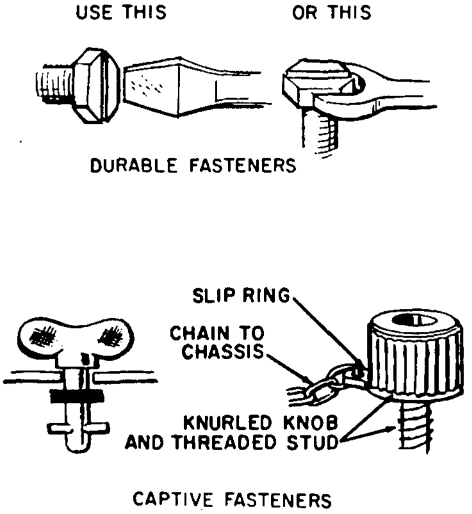


FIGURE 12–12. Illustration of one way to provide durable fasteners and two types of captive fasteners.

be of the same type and size to reduce the number of different tools required and decrease the probability that the wrong tool size or type will be used. When a special-purpose fastener is used, it should be distinguishable from other fasteners. The letter "M" is sometimes embossed on the head of mounting bolts for this purpose. When bolts with left-hand threads are required, they should be labeled similarly. When torque is critical, it also should be specified clearly on or near the fastener.

More specific information on fastener design considerations can be found in sources such as the "Guide to Integrated System Design for Maintainability" (Rigby et al., 1961).

12.4.6 Labeling

The discussion which follows is limited to labeling considerations which are specific to maintenance-related labels. General principles applicable to labeling methods are presented in Chapter 3.

Purpose and Content

Maintenance labels generally serve one or more functions: (a) identification of specific units, parts, controls, displays, test points, etc., with respect to purpose or functions; (b) presentation of critical information with respect to proper maintenance procedures for an equipment item; and (c) presentation of safety information, i.e., warnings which lead to prevention or avoidance of conditions hazardous to maintenance personnel or the equipment itself.

Word choice and sentence structure should take into consideration the ability of the lowest-level technician who will read and comprehend the label. (See Table 12–4 for "Expectations from the novice technician".) Unusual technical terms should be used only when required to impart exact information. Although labels should be brief, abbreviations and abstract symbols should be avoided unless known to be familiar to maintenance personnel in general.

Legibility

Certain conditions common to maintenance environments can have adverse effects on labels. For example, grease or dirt can damage decals or stenciled labels; e.g., ink is often smeared. Consequently, etched or embossed labels are generally preferred. However, the contrast and resolution afforded by engraving may prove marginal or inadequate under poor lighting. The use of larger letters and numerals may partially compensate for poor contrast and resolution, but high-contrast paints, e.g., black letters on white background, in combination with engraving should be considered for greater assurance of legibility over extended periods of time.

12.4.7 Displays for Maintenance

General principles for the presentation of visual and auditory information are found in Chapters 3 and 4. The following discussion pertains to display considerations which seem especially important for effective maintenance.

Shared Displays

Proper diagnosis of system characteristics for maintenance may require different or more detailed information than is needed for operation or control. This may present a problem if the same displays are used for both operation and maintenance because it is ordinarily undesirable to include more information than the display user actually needs. If the additional information required by one user is actually detrimental to the performance of the other,

TABLE 12–4. EXPECTATIONS FROM THE NOVICE TECHNICIAN

1. Can read and write at the high school level.
2. Will know names of and be able to identify common colors like red, orange, blue, etc. Will probably not be able to match terms such as crimson, aqua, coral, cerise, fuschia, turquoise, etc., with colors without a color chart. Will probably only be able to make about seven reliable absolute discriminations across the color spectrum.
3. Can distinguish between different geometric forms such as circles and squares, but cannot make reliable judgments between differences within the same class such as ellipses having different ratios of major to minor axes.
4. Can do simple arithmetic computations, but will frequently make mistakes. Is likely to be confused by algebraic formulae.
5. Must use watch or other timing mechanism to make reliable estimates of time intervals.
6. Is unlikely to know all the common electrical, pneumatic, hydraulic, or mechanical symbols by memory.
7. Is not aware of all shock or other potential equipment hazards.
8. Will be able to interpret oscilloscope indications correctly only with guidance of job aids.
9. Will vary widely from other technicians in his interpretation of imprecise quantitative words such as "approximately," "about," "close to."
10. Will know or be able to tell from the context of written maintenance instructions roughly the size of an equipment item he is trying to locate.
11. Will tend to overlubricate.
12. Can use conventional household tools including soldering tools. Will not know all of the simple cautions such as matching screwdriver blade and screw slots or avoiding overheating of parts near solder joints. Will have to be told how to use tools such as plumb bob and ruler, contact benders, spring hooks, feeler gauges, calipers, etc.
13. Can manipulate fasteners but may not know the correct tool to use.
14. Knows, or can easily figure out, how to remove or open dust covers, doors, latches.
15. Knows, or can easily figure out, how to manipulate tool-operated controls, toggle switches, rotary controls, cranks, and other common controls. Will have to be told the correct settings and will need help in correctly using verniers.
16. Will make numerous assembly and disassembly errors without extensive specific experience, instructions, or illustrations.
17. Does not know how to position components and parts for alignment unless alignment instructions or aids are provided.
18. Knows or can easily figure out how to attach test leads to test points and to test equipment, but first must be told what equipment items are involved and their locations.
19. Will follow a relatively inefficient troubleshooting strategy unless guidance on diagnosis and systematic checking is provided.
20. Should be able to point to the following if given the appropriate name: conventional household tools; chassis; cabinets; racks; pull-out drawers; dust covers and cases; jacks; transistors; tubes, sockets, and holders; screw, bolt, and clamp fasteners; conventional controls; conventional displays; plugs and cables in general (but not specific kinds); terminals; motors and common motor parts; washers, gears, shafts, bases, containers, joints, housings, straps, pulleys, wheels, resistors, transformers, capacitors, crystals, relays, antennas, microswitches, insulators in general (but not necessarily all specific types).
21. Probably would not be able to point out the following with any degree of reliability if given the appropriate name: tools other than conventional household tools; unlabeled cans; different motors as a function of their shape; controls not having conventional knobs and levers; most mechanical parts; waveguides and other antenna plumbing; difference between chokes, r-f coils, and certain wire resistors; oil-filled, bathtub, button, and tubelike condensers; magnets; parts dangerous to life when touched.

Adapted from Swain (1957).

separate displays should be provided for maintenance and operation purposes.

Similar problems may arise with location of displays intended for use by both technicians and operators. Maintenance often becomes a two-man operation simply because displays cannot be seen from the position where repairs or adjustments are being made. In such instances, separate displays for operation and maintenance may reduce manpower requirements and communications problems.

If maintenance is anticipated during system operation, pertinent displays should be located so the maintenance activity will not interfere with operator functions. Displays should be located near (preferably above) associated controls which the technician must use. If it is desirable to hide maintenance displays (and controls) from the operator's view lest they confuse him, they may be placed behind appropriately labeled, quick-opening access panels.

Scaling

Multi-scale displays can be easily misread and should be avoided. Only one scale should be visible at a time. At the very least, each scale should be clearly labeled to indicate its function or purpose. Color-coding each scale to match a corresponding control position is also a good practice when appropriate.

Scale markings should be consistent with the degree of reading accuracy required. Unnecessary scale markings tend to increase the frequency of reading errors. The requirement for interpolation or transformation also increases error frequency. If transformations cannot be avoided, a conversion table should be included in the display panel. Decimal transformations are preferred.

Identification

Displays should be labeled to indicate clearly the function which they measure, e.g., "TURRET DRIVE VOLTAGE." When a display has an associated control, labels for the two should be similar, if not identical.

Use of Secondary Senses

Although visual displays constitute the primary source of maintenance-related information, invaluable cues are often obtained via other sensory modes. However, except for auditory warning devices which may be built into equipment, such cues are a chance product of equipment design. The following examples serve to illustrate: Overheated equipment "feels" hot. Some overheated components, such as selenium rectifiers, produce characteristic odors. Leaking substances, such as coolants, e.g., ammonia, and antifreezes, can also be detected by smell. Worn bearings or gears usually produce detectable sounds and vibrations. Components, or tools, of known configurations can be identified by touch when they cannot be seen. Occasionally, substances which look and smell the same can be differentiated on the basis of taste. Experienced maintenance personnel could increase the list of examples manyfold. It is also very probable that design engineers and system planners could extend the usefulness of such fortuitous cues through formal design provisions and procedural aids. Potential maintenance applications for the olfactory and cutaneous senses are emphasized in two recent publications of the Aerospace Medical Research Laboratories (Goldbeck et. al., 1966; Crawford and Barnes, 1966).

12.4.8 Maintenance Controls

In most instances, criteria for the selection and design of operator controls apply to maintenance controls as well. Hence, the reader is referred to Chapter 8 for general recommendations and principles related to control design. Since a control often has one or more associated displays, some of the preceding discussion on displays also pertains to control design. For example, technicians frequently use operator controls during maintenance; hence, the desirability of provisions whereby operator control functions can be effected from the maintenance site, e.g., beneath the engine cowling.

Location

Equipment should be designed so that maintenance-related controls are located on exposed surfaces which are readily accessible during maintenance. If it is essential that such controls be concealed or protected, they may be placed behind properly labeled quick-opening access doors or panels. Controls which are routinely used in a particular sequence during maintenance or check-out should be arranged in that order unless it conflicts with operator requirements Sometimes, it is acceptable to group together controls with similar functions such as adjustment controls, light switches, etc. Controls should not be located near high voltages, hot components, rotating machinery, etc.

Recessed and Tool-Operated Controls

Controls which are designed and located so that they are not conveniently operated without the benefit of tools may discourage tampering, but they also increase the complexities of the technician's tasks. They should be avoided if frequent adjustments are anticipated. When they are necessary, certain features facilitate their use. For example, a channel or groove from the access opening to the control will guide the tool, usually a screw driver or spin wrench, into position. Control axes which are oriented vertically so that the tool can be inserted from above are usually better than other orientations because the tool position is easier to maintain.

Identification

Controls should be labeled clearly to indicate the function controlled and relate the control to its associated display. Labels for controls which are always used in a fixed sequence should include appropriate sequence numbers. Labels should fully describe the conditions or actions affected by changes in control position.

12.4.9 Maintenance Support Equipment

Equipment items such as vehicles, stands, dollies, etc. are used in support of many maintenance operations. Support equipment design can have substantial impact on the efficiency and safety of maintenance operations. Support equipment, which is difficult to use, which poses personnel hazards, or which fails to provide features necessary to fulfill its intended purpose, will severely detract from the maintainability of the prime equipment. For example, if a maintenance stand proves to be difficult to position, provides no space for auxiliary tools or equipment, or poses other problems to the technician, probably, it will not be used. In its place, technicians will often substitute a jury-rigged device which, though possibly of more immediate utility, may not maximally contribute to safe and efficient maintenance.

Failure to insure that maintenance support equipment is itself reliable and conveniently maintainable will add to the overall system maintenance burden. In designing maintenance-support equipment, the design steps and suggestions discussed in preceding sections of this chapter are no less important than in the case of prime equipment. The designer should be particularly alert to the design implications of the following questions:

1. What tasks must the technician perform in using the device? Can the required tasks be easily and safely accomplished by the technician?
2. Does the item of support equipment under consideration adequately provide for safe and efficient task support in terms of both the equipment-equipment interface and the man-equipment interface?
3. What effect will the expected physical and climatic environments have on the utility of the device? What design characteristics must be employed to insure that the utility of the equipment item is not adversely affected by environmental conditions?

Considerations related to the design of some types of support equipment are discussed in the subsequent sections.

Maintenance Stands

Properly designed maintenance stands are essential to technicians in most maintenance applications. Adherence to the following design criteria will contribute to ease of operation and general utility of maintenance stands.

1. Stands should be designed to correspond in height (or be adjustable in height) to the level of pullout racks which they are designed to accommodate.
2. If space permits, provide built-in shelves and stands on prime equipment. These are particularly useful if frequent maintenance entailing component removal or the use of test equipment must be accomplished on the prime equipment.
3. Consider designing hinged access covers so that they may be used in their open position as maintenance stands. This possible application should not detract from the convenience of opening or closing the access cover.
4. Tables and stands should be designed so that they can be used without tipping on a 15° incline even if the weight is applied to the lower edge of the stand.
5. "Drop leaf" tables or stands should be designed to insure stability when loads are applied to the outer edge of the extended leaf.
6. Capacity of the stand or table, in pounds, should be clearly stated in a prominent label.

Work Platforms and Walkways

In many applications, it is not possible to design prime equipment so that the maintenance technician has ready access from a ground-level position to all items of equipment he may be called upon to replace or repair. In these instances, an adequately designed work platform is indispensable. The following design recommendations have been found useful in enhancing the safety and utility of work platforms and associated facility items.

1. Provide sufficient space on the platform to accommodate both the technician(s) in the posture(s) which will have to be assumed to do the job and any associated tools or test equip-

ment. As a minimum, six square feet of floor space is required per technician.

2. The platform must permit the technician to have both hands free for work. It must also protect him from inadvertent falls and should not itself interfere with access.

3. If the platform is on wheels, wheel locks or brakes must be provided.

4. Design platforms with suitable places for securely resting test equipment at a convenient operating level.

5. If the technician must carry equipment onto an elevated platform, access stairs with an angle of climb no greater than 35° should be provided.

6. Design the work stand so that it can be easily positioned. A stand that is difficult to position invites substitution of jury-rigged and often dangerous means of task access.

7. Provide handrails or handgrips on platforms, stairs and around floor openings or wherever personnel may fall from an elevation. The dimensions shown in Table 12–5 are recommended for design of handrails, handholds, or guardrails.

TABLE 12–5. RECOMMENDED DIMENSIONS FOR WORK PLATFORM FEATURES

Handrails	Height—32 in. above floor or step Diameter—1¾ to 2 in.
Steps	Riser—7¼ in. Tread—10½ in.
Handholds	48 to 56 in. above step
Guardrails on platforms.	Dual rail; lower rail 22 in. above platform floor; upper rail 45 in. above platform floor.

Altman et al. (1961).

8. Walkways should be a minimum of 12 in. in width and should be guarded by rails or other means. If work is accomplished from the walkway, the width should be increased materially (i.e., to 2 ft.).

9. Provide nonskid surfaces on all areas of prime or support equipment on which people are likely to walk or stand.

10. Be sure that platforms and walkways are designed to sustain the weight of the number of people and associated equipment required for the maintenance task in that area. A 95th-percentile man weighs about 200 lbs. (Hertzberg et al., 1954).

Tools

Tools tend to be a source of maintenance problems for a number of reasons. For example, instead of carrying heavy tool boxes containing numerous tools which are seldom used, technicians often carry only a few general purpose tools. Special or rarely used tools are often lost, misused, misplaced, or locked up for safekeeping. Hence, maintenance requiring special tools may be done with improper tools, which can result in equipment damage or personal injury, or invaluable time may be lost searching for rarely used tools. Special tools are usually more expensive and difficult to replace, too. Because of such difficulties, system designers and maintainability engineers must carefully consider tool requirements during the system planning phase as well as during the actual development of the system.

Considerations related to space requirements for tool operation are discussed briefly in a preceding section of this chapter under "Accessibility," Section 12.4.1. It is worthwhile to test tools and related equipment on the actual hardware as early as possible during system development. This not only affords an opportunity to assess the adequacy of access provisions and general procedures; it also provides an opportunity to identify and correct deficiencies such as the following: (a) unnecessary design features or restrictions which require extra tools or special devices, (b) support equipment items which are inadequate for conditions under which they must be used, and (c) duplication or omission of essential items of equipment.

General guidelines and design practices which tend to minimize the occurrence of problems or deficiencies related to tool requirements include the following:

1. Minimize requirements for different kinds of tools by designing equipment so that the same tool can be used in as many tasks as possible.

2. Give priority to design features which facilitate maintenance (a) by hand without tools, e.g., hand-operated fasteners; or (b) with no more than standard tools, in that order of preference.

3. Require special tools only because of

reasons peculiar to the equipment or situation, e.g., when it is essential that items of equipment be locked or otherwise controlled to prevent tampering or maintenance by other than selected specialists.

4. Use standard tool kits provided to specialists by user organizations as guides in determing tool requirements. When additional tools are essential, take action to assure that they are provided.

5. Tool handles should provide adequate gripping surfaces. Knurled or grooved surfaces are appropriate if edges are not too sharp. Handle size should be based on encumbering effects of protective clothing as well as anthropometric dimensions of the nude subject. For example, tool handles with diameters of 1.5 in. have been used effectively in conjunction with pressurized gloves (Seeman et al., 1966). Rather extensive tool modification is usually necessary to achieve an effective interface with remote-handling devices such as are used for handling radioactive materials.

6. When tools are being selected, torque requirements should be carefully considered. If factory assembly is done with torque wrenches, these may also be required for maintenance. If precision torquing is required, wrenches should have variable torque settings. If torque requirements are high, power tools or wrenches with lever type handles are preferred. Ratchet screw drivers, which can be operated with one hand, are good if torque requirements are low and space is limited. The need for offset screw drivers should be avoided because it is difficult to control the direction of forces applied by them and screw head damage frequently results.

7. Incorporate special design features such as tool guides or safety guards to facilitate proper tool positioning and mating with controls or fasteners when hazardous or difficult operations may be involved.

12.4.10 Test Equipment and Bench Mockups

Test equipment and bench mockups are, basically, just pieces of equipment, and, like any other equipment, such test units must themselves be checked, calibrated, and maintained. For this reason recommendations about the design of units, covers and cases, cables and connectors, test points, displays, and controls apply just as much to test equipment and bench mockups as to prime equipment. Consult all of the prior articles of this section for recommendations about these aspects of the design of test equipment and bench mockups.

Test Equipment

There are four general types of test equipment used in maintenance work, and these can be listed as follows:

1. Built-in test equipment, which is an integral part of the prime equipment. This type can be a complex automatic checker or a simple voltmeter with external leads.
2. Go-no-go test equipment, which provides only one of two possible answers to the question: Is the given signal in or out of tolerance?
3. Automatic test equipment, which checks two or more signals in sequence without help from the technician. (The test usually stops when the first out-of-tolerance signal is detected.)
4. Collating test equipment, which presents the results of two or more checks as a single display, e.g., a light might come on only if a number of different signals are in tolerance.

These four types of test equipment are not mutually exclusive, however, and a given test unit might have all or some of the features of any of them in any combination. There seems to be no good general rule for specifying how much of each type should be built into a given item of test equipment. The relative advantages and disadvantages of each feature must be judged in terms of the demands that will be placed on the equipment and on maintenance technicians in the field. (See Table 12–6.)

Very elaborate test equipment can simplify the job of the line technician and reduce preparation or turn-around time for systems like inteceptor aircraft and missiles, but this does not necessarily reduce the total maintenance load. In fact complex test equipment sometimes requires so much maintenance that it increases the total amount of maintenance required. Above all, test equipment should be designed so that it is easy, fast, and safe to use.

Another point to keep in mind is that the weight of portable test units interacts with the

TABLE 12–6. ADVANTAGES AND DISADVANTATES OF FOUR TYPES OF TEST EQUIPMENT

Type	Advantages	Disadvantages
Built-in	Cannot be lost or damaged independently of prime equipment. Requires no special storage facilities. Does not need to be transported to prime equipment.	Might add appreciably to size and weight of prime equipment. Will require greater total number of test units because there must be one for each prime equipment. Calibration of each test unit might be difficult or inconvenient. Might increase complexity of, and amount of maintenance needed on, prime equipment.
Go-no-go	Presents information that is clear and easy to read. Simplifies decisions and tasks for maintenance man.	Unique circuitry usually required for each signal value to be tested. The additional number and complexity of circuits often adds to cost of test unit and to time required for its development and, later, maintenance. Usually of little help in checking common voltages or simple waveshapes except in long sequences that must be checked quickly.
Automatic	Can make rapid series of checks with little or no chance of omitting any steps.	Usually large, heavy, and expensive. Usually highly specialized with little versatility. Almost essential that it have self-checking features to detect its own malfunctioning, which adds to cost and difficulty of maintaining it.
Collating	Reduces number of indicators technician must read and so reduces checking time and errors. Simplifies trouble shooting if it provides indication of which signal, if any, is out of tolerance.	Disadvantages are similar to those for go-no-go and automatic test equipments. If it merely indicates that all signals are, or are not, in tolerance, it will not aid in trouble-shooting.

amount of rough treatment they receive. Table 12–7 shows the number of obstacles encountered by Navy electronic technicians in making fifteen regular service trips with instruments of various weights. Note the direct correlation between the weight of the instruments and the number of times they bumped against steel.

Design Recommendations

The following recommendations should be observed when designing test equipment:

1. Minimize the number of controls and displays.

2. Reduce the number and complexity of the steps required to operate the set (e.g., by "ganging" certain controls or by making certain operations automatic).

3. Prepare clear operating and maintenance instructions.

4. Provide "reverse interlocks" that turn off the set when the cover is closed, and/or use both warning lights and written warnings on the test equipment to remind the technician to turn off the equipment when he is through with it.

5. Use selector switches rather than plug-in connections on test equipment. (See Figure 12–13.) Selector switches are quicker to use and they reduce the likelihood of faulty connections.

6. Provide a signal to show when the set is warmed up and ready to use. If such a signal cannot be provided, a label near the warmup switch should state clearly how much warmup time is required.

7. Provide a simple check to show when the set is out of calibration or is not working properly.

8. The outer case and all removable parts should be clearly labeled.

9. A label on the cover or case of the set should state its purpose and the precautions that should be observed in using it.

10. Full instructions for using the set should be stored in it, and/or a checklist for operating

TABLE 12–7. NUMBER OF OBSTACLES ENCOUNTERED BY MAINTENANCE MEN CARRYING TEST EQUIPMENT ABOARD U.S. NAVY VESSELS

Weight of instrument (lb)	Relative size of ship	Number of obstacles encountered					Number of times instrument bumped on steel
		Doors*	Coamings†	68° ladders‡	90° ladders§	Misc.¶	
4	–Small	52	43	21	13	16	2
	–Large	97	84	63	3	91	5
8	–Small	46	40	19	3	19	6
	–Large	92	83	59	3	88	19
16	–Small	38	32	17	0	16	21
	–Large	88	78	43	1	70	30
> 16	–Small	21	27	5	0	9	29
	–Large	33	21	23	0	32	36

* Most doors are 26 in. wide (a man walking comfortably is 21 in. wide).

† Watertight doors have a 7 to 12-in. coaming over which the man must step.

‡ Ladders are usually at a 68° angle with 25-in.-wide steps. Handrails of chain or rope (which give 4 to 8 in.) are used although solid railings are more common.

§ Watertight hatches are generally 18 in. in diameter and always have a vertical (90°) ladder leading to them.

¶ Most passageways are 32-in. wide, allowing a man only about 6 in. on each side of him when not carrying an instrument.

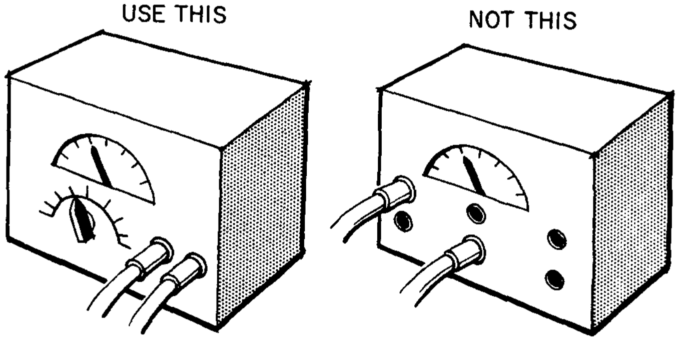


FIGURE 12–13. Comparison of test equipment using a selector switch and using plug-in connections.

the equipment should be printed on a metal plate and attached to it.

11. Label every item that the technician must recognize, read, or use.

12. Provide circuit-protection devices, such as circuit breakers and fuses, to protect the set against damage if the wrong switch or jack position is used.

13. Provide adequate, integral storage space for such removable items as test leads.

14. Provide fasteners or holders so that accessories will be held securely and safely in the storage compartment.

15. Label what goes into the storage compartment and show how it should be stored.

16. Test leads should require no more than a fraction of a turn for attachment to the prime equipment.

17. Portable test equipment should be rectangular in shape for convenient storage.

18. Handles on the cover and/or case should be recessed or hinged for convenient storage.

Bench Mockups

The term "bench mockup" refers to prime-equipment units set up in a maintenance shop or depot to check or locate faults in units brought in from the field. Such mockups may consist of an entire system or only part of a system, and they may be provided with signal generators and dummy loads to simulate inputs and outputs.

A unit of the mockup is replaced with a unit from the field that is suspected of a malfunction. The bench mockup feeds signals into and receives signals from the suspected component. If an out-of-tolerance signal appears when the suspected unit is installed, but not when the regular unit is installed, the technician knows there is something wrong with the suspected unit.

Although bench mockups are usually put together from production units of the prime equipment, they should be designed as items of test equipment because they often require additional maintenance aids such as signal generators, dummy loads, and extra junction boxes, terminal strips, test points, controls, and displays.

Design Recommendations

The following recommendations should be observed when designing bench mockups:

1. Provide extension cables for all units so that they can be removed from the mockup for checking. (See Figure 12–14.)

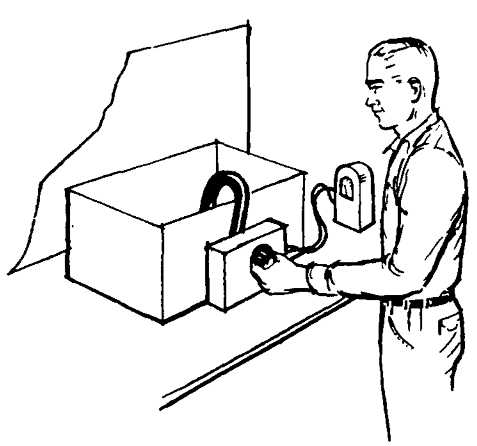


FIGURE 12–14. Illustration of the advantages of providing extension cables so units can be removed for checking.

2. End connectors on all mockup cables should be of the quick-disconnect type that requires only a strong push or pull to connect or disconnect it. (Mockup cables need not withstand strong vibration or shock, but, because of the way mockups are used, they do need to be connected and disconnected frequently.)

3. Provide extra-heavy coverings on mockup cables (for example, vinyl tubing) to protect them from wear resulting from frequent connection and disconnection.

4. Mockup cables, including extension cables for units, should be provided with test points to check the signal flow through each wire. One satisfactory design is to provide test points at the connector. (See Figure 12–15.)

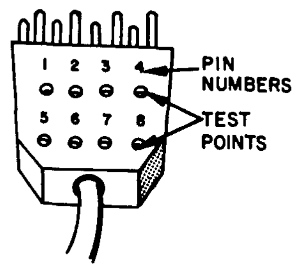


FIGURE 12–15. One method of providing test points on connectors.

5. Operating instructions for the mockup should give correct signal values and tolerances for each test point.

6. Use transparent plastic covers on mockup units that contain parts the operation of which may be checked visually. (Caution: Do not follow this recommendation if a metal cover is needed for electrical shielding.)

7. Mockup units should be installed so that every unit is accessible without removing any other unit.

8. The layout of the mockup should provide enough space so that the technician can get at the units.

9. Provide a pullout shelf or some other method of supporting the test equipment while it is being used (see Figure 12–16).

12.4.11 Maintenance Procedures

Ideally, every maintenance operation should be covered by written procedures prepared and evaluated during the development of the system. Instructions, flow diagrams, schematics, and decision trees should be included. These aids should be compatible with the capabilities and limitations of the personnel who will use them. Anticipated environmental conditions, equipment characteristics, and task requirements must also be considered. General categories of

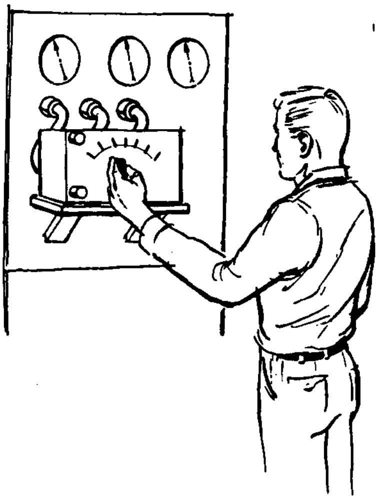


FIGURE 12–16. Illustration of the advantages of providing a shelf for test equipment when it is being used.

maintenance tasks ordinarily anticipated include inspecting, troubleshooting, adjusting, replacing, repairing, and servicing.

General Recommendations

The following general recommendations should be observed when preparing maintenance procedures:

1. Keep maintenance procedures as brief as possible without sacrificing necessary information.
2. Prepare procedures that give unambiguous results.
3. Make tolerances realistic for the level at which maintenance will be performed, i.e., line, field or shop, factory, or depot maintenance.
4. Procedures should not be difficult for technicians to follow. Keep the ideas and words as simple as is practicable. Procedures sometimes can be simplified by designing the test equipment to do some or all of the programming.
5. Keep the number of decisions the technician must make as few as is reasonable.
6. Keep decisions simple by reducing the number of alternatives and prior conditions the technician has to keep in mind.
7. Use exact step-by-step procedures, and use the same ones as often as possible. (This will make it easy for inexperienced technicians to learn the procedure.)
8. Be sure each procedure states how to start up the equipment and how to shut it down.
9. Do not require the technician to work near dangerous voltages or delicate components.
10. Always provide systematic trouble-shooting procedures for the technician to follow. Failure to do this may result in the technician following inefficient or even dangerous procedures.

Troubleshooting

The complex diagnostic problems and decision processes involved in troubleshooting make fault isolation the most demanding of maintenance tasks. Research suggests that effective troubleshooting depends largely upon the extent to which the equipment and procedural aid design facilitates the application of a systematic strategy. To illustrate, McKendry and Stover (1962) identified five sources of fault isolation problems in electronic equipment:

1. The extent to which the field of alternative trouble sources is organized or structured. For example, a technician frequently traces a fault to an area where components and circuitry are a tangled mass; then, he may forget or "lose track" of what he has already done and is almost certain to make redundant time-consuming checks (Rigby et al., 1961).
2. The nature of the organization involved. (See the discussion of "Packaging Techniques" 12.4.2 in this chapter.)
3. The quantity and quality of information available concerning the sources of trouble, e.g. tolerance ranges for "normal" outputs, component reliabilities, symptom descriptions, etc.
4. Size of the field of possible trouble sources.
5. Stability of the "trouble field" over time.

The equipment designer usually has latitude with respect to equipment characteristics related to the above-listed problem areas—especially the first three. Therefore, he is in a good position to alleviate future maintenance problems through planning and coordination with maintainability engineers responsible for preparing procedural aids during system development. Although the optimum troubleshooting procedure for a given situation may depend upon personal preferences and experience, in the military, where taining and experience are often minimal, significant advantages are gained through the standardization of procedures and design characteristics.

Troubleshooting Strategies

If an optimal troubleshooting procedure is to be provided, the designer should be aware of approaches currently in use. For example, troubleshooting procedures for electronic equipments usually include four major steps (Myers et al., 1964):

Step 1. Visual Checks—for clues such as smoke, loose connections, missing or damaged components.

Step 2. Operational Checks—an evaluation of readings obtained from meters, gauges, and other indicators, e.g., "front-panel checks." (See Tables 12–8 and 12–9 for examples of checklists which can be provided to facilitate steps in the troubleshooting process.)

Step 3. Intermediate Checks—to isolate the malfunction to a particular "stage." (An equipment is viewed as consisting of three levels: The first level consists of a major unit such as a video decoder. The second level involves the various "stages," e.g., detector or first i–f amplifier. The third level is comprised of components, i.e., resistors, capacitors, etc.)

Step 4. Systematic Checks— having isolated the malfunction to a particular stage or stages, the technician proceeds to check pertinent circuits in some order.

Step 3 above normally involves one or more of the following strategies:

1. Reliability Strategy. This approach involves the consideration of failure rates. The technician simply checks the least reliable components or stages first. Reliabilities may be: (a) estimated on the basis of the technician's experience or (b) obtained from reliability data supplied with the equipment.

2. Conditional Probability Strategy. The technician first identifies effects of the malfunction on functional characteristics of the equipment, i.e., he determines the "symptoms." Then he checks that stage or component which is the most *probable* cause of the symptoms. As in the case of the reliability strategy, he may draw upon his own past experience or he may refer to manuals in which various symptoms are listed together with components which are most likely to be the source of the problem.

3. Syndrome Analysis. This technique is similar to that of "conditional probability." The principal difference is that syndrome analysis is based on the technician's knowledge of

TABLE 12–8. OUTLINE OF PREFLIGHT CHECK, B–7 FIRE-CONTROL SYSTEM

- 1. Radar-set check
  - 1.1 Power-supply check
    - 1.1.1 +300v d-c regulated
    - 1.1.2 +150v d-c regulated
    - 1.1.3 −150v d-c regulated
    - 1.1.4 Repeller voltage
    - 1.1.5 Power bus (a-c)
    - 1.1.6 Relay supply voltage (d-c)
  - 1.2 Transmitter-output checks
    - 1.2.1 Frequency check
    - 1.2.2 Inspect frequency spectrum
    - 1.2.3 PRF check
    - 1.2.4 Inspect modulation envelope shape
    - 1.2.5 Power-output check
  - 1.3 Receiver test
    - 1.3.1 Sensitivity check
    - 1.3.2 Receiver-output check
    - 1.3.3 AFC check
- 2. Search-mode operational check
  - 2.1 Antenna search-pattern check
    - 2.1.1 Narrow-pattern check (right and left)
    - 2.1.2 Wide-pattern check (center)
    - 2.1.3 Antenna spin-pattern check
  - 2.2 Display calibration
  - 2.3 Artificial-horizon check
- 3. Manual-track operational check
  - 3.1 Antenna-control check
  - 3.2 Manual-track test
  - 3.3 Lock-on check
    - 3.3.1 Check sensitivity of lock on
    - 3.3.2 Check display function
- 4. Auto-track operational check
  - 4.1 Attack-phase operational check
  - 4.2 Rocket-fire check
  - 4.3 Pull-out-warning check
- 5. Miscellaneous checks
  - 5.1 Check antenna-table dither
  - 5.2 Check radar anti-jam control operation
  - 5.3 Check pilot's scope controls

TABLE 12–9. ANTENNA SEARCH-PATTERN CHECK, B–7 FIRE-CONTROL SYSTEM

Step	Component	Action	Indication	Remarks
2.1	Antenna	Mount azimuth and elevation protractors		
2.1.1	Pilot's control box:	Set to:	Antenna travel on protractors should be:	
	Operation Sw.	Auto-Search		
	Antenna-Az. Sw.	Center-Narrow	Az: $\pm 23\frac{1}{2}°$	Adjust pot. P2371
	Antenna-El. Con.	Center	El: $\pm 2\frac{1}{2}°$	Adjust pot. P2346
	Pulse length Sw.	Long		
2.1.2	Pilot's control box:	Set to:	Antenna travel on protractors should be:	Same as for 2.1.1
	Antenna-Az. Sw.	Center-broad	Az: $\pm 67°$	
	Antenna-El. Con.	Full clockwise	El: $+30°$ $-25°$	

circuit function and operational theory rather than on the probability of failure. In using syndrome analysis the technician evaluates symptom patterns derived from observations of the system under operational conditions and/or measures of functional characteristics obtained by test equipment, e.g. voltmeters, ammeters, oscilloscopes, etc. (See Table 12–10 for example of an aid to symptom pattern analysis.)

4. Signal Tracing Strategies. This approach involves inserting a signal into an equipment and tracing it through stages or components until a faulty unit is identified on the basis of an inappropriate deviation in its output. The several variations in this technique are identified on the basis of distinguishing characteristics such as "backtracking," "middle-to-trouble," "front-to-back," and "back-to-front."

5. The Half-Split Technique. This strategy derives from decision-making theory. It was devised to optimize fault localization in the absence of probability information with respect to alternative sources of the malfunction. In applying the technique, the technician attempts to reduce the trouble field by one-half with each check that he makes; i.e., he selects each test point so that approximately one-half of the remaining potential test points lie to either side of it. (See Figure 12–17.)

6. Bracketing. This technique involves the use of schematics or block diagrams in a rather unique way to aid the troubleshooting process. After making the initial inspection of the malfunctioning equipment, the technician identifies by means of brackets drawn on a schematic or block diagram the area or areas of uncertainity. As the troubleshooting process continues, he relocates the brackets to include a progressively narrowing area of uncertainty until the malfunction is isolated to a single component or replaceable unit. Bracketing is frequently used in combination with one or more of the strategies previously discussed.

Myers et al. (1964) found that of the six principal strategies described above, the last three, Signal Tracing, the Half-Split Technique and Bracketing, required the least training and experience, and of these three, the Half-Split and Bracketing techniques tended to be the more efficient.

More recently, research and development efforts have been directed toward the development of computer techniques for further optimizing the trouble-shooting process (Folley and Pieper, 1964; Hannom et al, 1967). Computer techniques can be used to generate troubleshooting decision trees on the basis of specific system data including signal flow, probability of malfunction, and cost of making the check for all possible test points. Previous research has shown that troubleshooting decision trees for particular equipments significantly improve the performance of both novices and experienced technicians. Further development of computer-aided techniques promises to provide an effective general method for producing such trees for both electronic and nonelectronic equipments.

Test pattern * Directions

TABLE 12–10. SYMPTOM PATTERNS OBTAINABLE FROM PREFLIGHT CHECK B–7 FIRE CONTROL SYSTEM

Test pattern *			Directions
1.1	1.2	1.3	
X	O	O	Do not run test 1.2 and 1.3 until power supply is adjusted correctly or replaced
O	X	O	Use data flow diagram 1.a for trouble shooting
O	O	X	Use data flow diagram 1.b for trouble shooting
O	X	X	Use data flow diagram 1.c for trouble shooting
2.1	2.2	2.3	
X	O	O	Use data flow diagram 2.a for trouble shooting
O	X	O	Use data flow diagram 2.b for trouble shooting
O	O	X	Use data flow diagram 2.c for trouble shooting
X	X	O	Replace master controller (LRU 7)
O	X	X	Use data flow diagram 2.d for trouble shooting
X	O	X	Replace master controller (LRU 7)
X	X	X	Replace master controller (LRU 7)
3.1	3.2	3.3	
X	O	O	Use data flow diagram 3.a for trouble shooting
O	O	X	Use data flow diagram 3.b for trouble shooting
X	X	O	Replace master controller (LRU 7)
O	X	X	Use data flow diagram 3.c for trouble shooting
X	O	X	Use data flow diagram 3.d for trouble shooting
X	X	X	Use data flow diagram 3.e for trouble shooting
4.1	4.2	4.3	
X	O	O	Use data flow diagram 4.a for trouble shooting
O	X	O	Use data flow diagram 4.b for trouble shooting
O	O	X	Use data flow diagram 4.c for trouble shooting
X	O	X	Use data flow diagram 4.d for trouble shooting
O	X	X	Replace master computer "B" (LRU 13)
X	X	X	Use data flow diagram 4.e for trouble shooting
5.1	5.2	5.3	
X	O	O	Use data flow diagram 5.a for trouble shooting
O	X	O	Use data flow diagram 5.b for trouble shooting
O	O	X	Use data flow diagram 5.c for trouble shooting
X	X	X	Check for voltage at main power bus. If this checks OK, replace pilot's control box (LRU 3)

Note: Patterns not shown in this table probably never will occur.

* X = malfunction symptom observed, O = intolerance indication obtained.

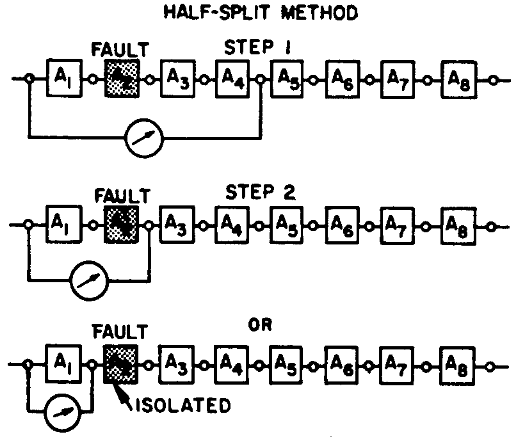


FIGURE 12–17. Illustration of the half-split method.

12.4.12 Maintenance Job Aids

Job aids are information sources such as manuals, handbooks, checklists, schematics, and diagrams. They may include voice recordings, movies, filmstrips, etc. During the past several years, computer technology developments and increased emphasis on total system performance have combined to increase the concern of engineering psychologists for machine-to-man communication links and related information processing problems (Topmiller, 1965). Nevertheless, research on job aids has been limited (Folley and Munger, 1961), and there are no simple formulas for selecting among design alternatives or making tradeoffs between training and job aid complexity. Hence, it is essential that designers be aware of typical problems and current developments in job aid design.

Problems in the Field

Job aids do not always facilitate job performance or reduce training requirements. Poorly designed job aids tend to offset the advantages of human engineering features in prime equipment and increase the chances for human error. In many cases the job aid is not even used in the field (Folley and Altman, 1956). By way of illustration, let us review some deficiencies that have been identified in the Air Force technical order system by maintenance personnel. (Losee et al., 1962).

1. Inappropriate or Incomplete Procedures. Troubleshooting information was often inap-

propriate or out-of-date because of modifications in the original equipment design. Critical information was ambiguously presented, difficult to find, or, in some instances, missing altogether. Of special concern were tolerance specifications for "normal" conditions, part numbers, illustrations, etc. Technicians often expressed a desire for more pictures, step-by-step procedures, and simplified diagrams. Lists of probable malfunctions and associated symptoms were regarded as inadequate or incomplete and, hence, of limited usefulness. The amount of descriptive information was often disproportionate to the complexity of the equipment or task, i.e., simple operations were described in detail and complex ones were over-simplified. Problems in understanding alignment and calibration procedures were common.

2. Deficient Schematics or Diagrams. Schematics and wiring diagrams were noted for errors and lack of clarity. Some diagrams included so many circuits, components, etc., that frequent retracing was required. Circuits were difficult to follow because of poor continuity between related schematics, e.g., between major assembly and module schematics. Color-coded data flow indications were suggested as a partial solution to some of the tracing problems. Technicians also wanted schematics to show critical parameter values, including voltages, waveforms, tolerance ranges, scale factors, etc. Frequently used schematics and diagrams should be printed on durable materials because they were often soiled or torn. Large-scale copies suitable for displaying on the wall as key charts and diagrams were also desired.

3. Late, Inaccurate, or Unrevised Data. Technical orders often contained erroneous information either because revisions called for by equipment modifications were not made or long time lags were involved. Contradictions among different technical orders on the same system were not uncommon; for example, the pilot's handbook might say one thing and the maintenance manual something else. Occasionally, there were also incompatibilities between illustrative materials and associated procedural descriptions.

4. Locating the Necessary Information. The average respondent reported that he spent 30% of his total job time looking for necessary information. The technical order indexing system was confusing to some. Descriptive terminology used in technical orders was sometimes inappropriate or unfamiliar to the technician; hence, equipment items and related information were difficult to identify or locate. The technical order numbering system was regarded as complicated and hard to remember. Material was poorly organized, e.g., illustrations might be several pages removed from pertinent descriptive information. The numerous revisions common to technical orders were difficult to follow and sometimes missing altogether. Parts were difficult to identify on the basis of infortion in parts lists or illustrated parts breakdown, i.e., federal stock number, manufacturer, etc. Many technicians think that maintenance publications should include more and clearer information on theory of operation.

5. Excessive Technical Orders and "Referrals." Technicians complained because too many different technical orders were required to do a single job, e.g., cabling diagram in one book, theory of operation in another, internal schematics in a third, etc. This often entailed the duplication of information. Technicians also resented the large number of referrals within the same manual which consumed time and made procedures difficult to follow.

6. Level of Writing. Technicians were of the opinion that procedural guides provided to them were either written for engineers or were written by someone who was unfamiliar with maintenance procedures. Whatever the reason, the level of writing was frequently beyond the technician's comprehension. As a result, the technicians suggested that someone more representative of the actual user should participate in the publication development and validation process.

7. Excessive Size and Weight of Manuals. Because the orders contained redundant information, technical orders were excessively bulky and heavy. Maintenance manuals were especially troublesome outdoors in windy or rainy weather. Manuals also occupied too much space on work benches. Compact, pocketsized editions containing only essential information were highly desired.

New Concepts for Handling Maintenance Information

Current attempts to alleviate some of the maintenance data problems include an Air Force sponsored program to develop criteria for "Presentation of Information for Maintenance and Operation." Referred to as Project PIMO, this effort has as its ultimate objective the application of advanced computer methods and audio-visual display techniques to develop an effective system for storing, retrieving, and presenting accurate, job-related, maintenance data. The program also seeks to provide for trouble-shooting strategies based on systems analyses rather than personnel experience.

Ordinarily, job aids should complement training by reducing the amount of job-related information which must be learned and remembered (McCormick, 1964). Therefore, anticipated changes in maintenance data systems will interact with maintenance training programs as well as with job aid design requirements. Shriver and Trexler (1966) have investigated several new concepts which could have significant effects on training. Most of these concepts include features which appear consistent with the objectives of Project PIMO. For example, Bell Telephone Laboratories' Maintenance Data System (MDS) concept is a data management system to alleviate the maintenance information storage and retrieval problem through automated storage and display techniques. The MDS concept also calls for the preparation and organization of pertinent information into a troubleshooting structure by means of linear and branching programming techniques. "Logic Sequence Diagrams" are used to present the operations required for a technician to obtain a specific result. References are included in the presentation to permit either: (a) confirmation of programmed checks, or (b) branching to subroutines on the basis of test results.

The technician retrieves information from the MDS central file by punching a code into a readout device. Information can be selected for retrieval on the basis of the technician's past experience and his knowledge of the current equipment status; e.g., experienced technicians may choose to bypass certain steps. As conceptualized, MDS would support two principal classes of maintenance personnel: (a) on-site operator-maintainers, and (b) central repair technicians.

Maintenance instructions prepared for MDS have been classified into two categories:

1. Active diagrams which provide programmed steps for:

a. Operating prime equipment.
b. Setting up system on test equipment.
c. Isolating faults to subsystem level.
d. Isolating faults to chassis or terminal connection.
e. Routine (periodic) maintenance checks.
f. Testing unit chassis.

2. Passive diagrams which provide supporting data (but no instructions for trouble analysis) as follows:

a. Block diagrams—showing chassis functions related to input/output and reference designators.
b. Interface diagrams—showing wiring between chassis.
c. Rack wiring diagrams (standard manufacturer's data).
d. Schematic of logic diagrams.
e. Chassis wiring lists or diagrams (manufacturer's data).
f. Parts-list information.

Development of the MDS is incomplete at this writing. However, the audio-visual system (A-VIS) which it would incorporate has been tested experimentally. The results showed that the presentation of audio-visual information via random-access machine was no more effective than the use of technical manuals. However, the visual information prepared for a A–VIS proved 40% more effective than standard manual preparations whether presented by machine or in manuals. Hence, the mode of presentation is less important than the choice of format, content, etc.

Some of the new maintenance concepts reviewed by Shriver and Trexler would do little more than reorganize the content of current job aids. Others propose significant prime equipment design changes. For example, a concept known as ADMIRE (Automatic Diagnostic Maintenance Information Retrieval) would extend automatic checkout to the "piece-part" level. Fault isolation would be achieved through built-in sensors and a diagnostic computer system (Anderson and Lee, 1963). Thus, the MDS described above may be viewed as something of a compromise. It proposes to alleviate problems related to data storage and retrieval but still provides for the application of technician abilities in making measurements and decisions.

All of the proposed concepts promise some improvement in maintenance performance. Shriver and Trexler were not prepared to compare them on the basis of their relative contributions to the effectiveness of the total man-machine system, but some effectiveness considerations were suggested for use in such evaluations. They are: 1. Manual Savings. Some concepts may reduce the bulk of manuals required but require expensive special art work or color. Others reduce manual costs but increase hardware costs.

2. Training Savings. Some concepts will reduce training costs more than others.

3. Analysis Costs. All the concepts require an analysis of the equipment to optimize the troubleshooting strategy and generate supporting information. Removal of this task from the technician's responsibilities is one of the more significant advantages of advanced concepts. (See the discussion of troubleshooting strategies under "Maintenance Procedures" in Section 12.2.4.) This cost is partially a function of equipment complexity. A small increase in maintenance effectiveness would be expected to more than offset this cost.

4. Performance Effectiveness. The concepts vary with respect to the extent to which they reduce errors and task times.

More detailed specifications for job aids doubtless will accompany the development and adoption of new and improved maintenance data-handling systems. The following paragraphs contain guidelines for the preparation of procedural guides and related maintenance materials. For recommendations concerning the presentation of information, in general, the reader is referred to Chapter 3. A series of studies conducted by the Aerospace Medical Research Laboratories provides useful suggestions concerning the design of check-lists in particular (Rees, 1959; Rees and Copeland, 1959; and Rees and Kama, 1959).

A Procedure for Developing Job Aids (Folley and Munger, 1961)

1. Prepare a complete list of all tasks in the system that are to be performed by the individual on the job for which the aid is being prepared.

2. Obtain, or prepare, to the extent possible, a step-by-step procedure for each of the tasks listed in step 1, above.

3. Examine each step of each task, and the task as a whole, and make the necessary judgments regarding what information could efficiently and effectively be contained in a job aid, rather than be learned by the job incumbent.

4. Determine the manner in which the job incumbent will attempt to obtain and use the information presented in the job aid.

5. Conduct a small-scale tryout of the proposed job aid.

6. Collect the job-aid information into a form most usable in the context of on-the-job performance. Select a suitable format and size for the job aid, and prepare the aid for on-the-job use.

Presentation Techniques for Troubleshooting Aids

Folley and Altman (1956) list five kinds of presentations based on differences in the purpose for which the information is to be used. (See Table 12–11.) Similarly, Rogers and Thorne (1965) describe five presentation techniques for electronic troubleshooting:

1. Schematic Diagrams. These are used to depict intra- and interchassis wiring in detail so that every part is represented. Voltages, resistances, etc., may be included. In fact, it is preferred that the characteristics of the signal, including tolerances, be specified on the schematic at each point.

2. Functional Diagrams. These diagrams indicate the function performed by each major part in a system. They are, in effect, schematic diagrams in which minor elements, such as resistors and capacitors, are not shown.

3. Block Diagrams. Block diagrams differ from functional diagrams primarily in that they are less detailed. They show only units which

TABLE 12-11. TECHNIQUES FOR PRESENTING MAINTENANCE INFORMATION

Purpose	Kind of presentation
Describing procedures in detail	Step-by-step instructions.
Describing procedures for experienced technicians	Checklists.
Presenting physical features	Drawings and photographs.
Presenting large amounts of data	Tables and charts.
Describing processes or interrelationships	Diagrams.

Folley and Altman (1956).

generate and transmit major signals, e.g., tubes and stages within a chassis usually are not shown.

4. Troubleshooting Charts. These indicate or suggest what to check when an abnormal indication is detected in the system or subsystem. In other words, they relate parts to specific symptoms or symptom patterns.

5. Written Instructions. These presentations often tell the technician *how* to do what he has decided to on the basis of information obtained via the previous four types of presentation. They usually consist of prosaic descriptions of procedural steps involved in maintenance actions such as checkout, calibration, removal, and replacement of components, etc.

Selecting the Form of Presentation

No single presentation technique is suitable for presenting all the information needed for troubleshooting; however, some have been found to be more valuable than others (Atchley et al., 1964). For example, whereas schematic diagrams are practically indispensable, block diagrams often contain only information which is learned rather readily during training. Because size limitations make all-inclusive troubleshooting charts impractical, they, too, are of limited value to well-trained or experienced technicians. Written instructions are expecially valuable for the initial checks in troubleshooting, but like troubleshooting charts, cannot contain a complete description of every maintenance task if they are to be of reasonable size.

When one is selecting the form of presentation to be used, the following recommendations should be observed:

1. Be sure that the form of presentation is appropriate for its purpose.

2. Use step-by-step instructions for maintenance tasks. Paragraph formats should be avoided since they obscure the step-by-step nature of procedures and require many more words.

3. To supplement or clarify written instructions and to provide extra information to technicians, use drawings or photographs. Use good quality drawings or photographs in which relevant details have been highlighted and irrelevant details blanked in to illustrate a point.

4. Use tables for presenting large quantities of data, being sure that the instructions state how and when each table should be used.

5. Be sure that tabular data can be used by the technician without having to make conversions or transformations.

6. Use diagrams to describe processes and interrelationships.

7. Show components in data-flow diagrams in the same relative position that they are in the equipment, if practicable.

8. Show only the electrical characteristics of the signal in data-flow diagrams, not the electrical characteristics of the component.

Instructional Content

After determining the appropriate form of presentation, it is necessary to determine that the instructional content of the manuals is also appropriate. The following recommendations should be observed when considering the instructional content of maintenance manuals.

1. Provide the technician only with the information he needs to do his job; instructions should contain only job-relevant information, avoid excessive detail, unnecessary theory, and verbiage.

2. Instructions should be specific to the level of maintenance to be performed, i.e., shop or depot, rather than having a single set of instructions for use at all three levels.

3. Be sure that all symbols and names in the

instructions agree with those actually on the equipment.

4. Give the in-tolerance signal characteristics and the acceptable tolerances for each test point.

5. Job instructions should be fully indexed. The index should contain words the technician is likely to look for in locating a particular item.

6. Include appropriate stock numbers with all parts listed This facilitates ordering of replacement parts.

7. Be sure that the information in the manuals is accurate and up-to-date.

References

Altman, J. W., Marchese, A. C., & Marchiando, B. W. Guide to design of mechanical equipment for maintainability. Technical Report No. ASD TR 61–381, August 1961, Aeronautical Systems Division, WPAFB, Ohio.

Anderson, N. S., & Lee, R. Application of human factors in maintenance of electronic systems. Fourth National Symposium on Human Factors in Electronics: Men, Machines, and Systems. IEEE, Wahsington, D. C., May 1963.

Atchley, W. R., Ver Hulst J. & Lehr, D. J. Specification of optimum fault isolation presentation, AMRL Technical Documentary Report 64–57, Wright-Patterson Air Force Base, Ohio, June 1964.

Barton, J. R., Electronic systems division policy on maintainability engineering. *Proceedings of ESD Maintainability Conference*, Hanscom Field, Mass., March 1963.

Crawford, B. M., & Barnes, N. K. The nature and capabilities of the cutaneous senses. Report No. AMRL–TR–66–108, July 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

Federal Electric Corporation. Maintainability design criteria handbook for designers of shipboard electronic equipment. March 1965, Prepared for Department of the Navy by Federal Electric Corp., Paramus, N.J.

Folley, J. D., Jr., & Altman, J. W. Guide to design of electronic equipment for maintainability. Report No. WADC–TR–56–218, April 1956, Wright Air Development Center, WPAFB, Ohio.

Folley, J. D., Jr., & Munger, S. J. A review of the literature on design of informational job performance aids. Report No. ASD–TR–61–549, October 1961, Aerospace Medical Research Labs., WPAFB, Ohio.

Folley, J. D., Jr., & Pieper, W. J. Development of a method for generating troubleshooting decision trees. Final Report, Contract No. NASw–577, 1964, Office of Manned Space Flight, NASA, Washington, D. C.

Gagne, R. M. (Ed.) *Psychological principles in system development.* New York: Holt, Rinehart, and Winston, 1962.

Geise, J., & Holler, W. W., (Eds.) Maintainaability engineering. Report prepared under Contract No. DA–31–124–ARO–D–100–34, July 1965, for U. S. Army Materiel Command with Martin-Marietta Corp. and Duke University.

Gillespie, K. W. Comparative evaluation of USAF Standard A/P22S–2 and improved A/P22S–2A high altitude, full pressure flying outfits. Technical Report No. 65–9, April 1965, Systems Engineering Group, Research and Technology Div., WPAFB, Ohio.

Goldbeck, R. A., Kaeding, J. H., & Feroglia, W. E. Odor coding for malfunction detection and diagnosis. Report No. AMRL–TR–66–122, August 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

Hannom, T. J. B., Azzari, A. J. Brooks, A. I. Hunter, J. R. Steinberg, L. Chubb, G. P. Development of a computer program for generating trouble-shooting decision trees, AMRL Technical Report 67–83, Wright-Patterson Air Force Base, Ohio, September 1967.

Hertzberg, H. T. E., Daniels, G. S., & Churchill, E. Anthropometry of flying personnel—1950. Report No. WADC–TR–52–321, December 1954, Wright Air Development Center, WPAFB, Ohio.

Kama, W. N. Volumetric workspace study: I. Optimum workpsace configuration for using various screwdrivers. Report No. AMRL-TDR-63-68 (I), August 1963, Aerospace Medical Research Lab., WPAFB, Ohio.

Kama, W. N. Volumetric workspace study: Part II. Optimum workspace configuration for use of wrenches. Report No. AMRL-TDR-63-68(II), December 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Leuba, H. R. Maintainability prediction—the next generation. Proceedings of the 1967 Spring Seminar on Reliability, Maintainability . . . Elements of System Effectiveness, IEEE, Boston: April 1967.

Losee, J. E., Allen, R. H., Stroud, J. W., & Ver Hulst, J. A study of the Air Force maintenance technical data system. Report No. AMRL-TRD-62-85, August 1962, Aerospace Medical Research Labs, WPAFB, Ohio.

Martin, W. L., Crawford, B. M., Kama, W. N. & Herman, J. Quick-release fasteners for space applications, AMRL Technical Report 68-117, Wright-Patterson Air Force Base, Ohio, July 1968.

McCormick, E. J. *Human factors engineering.* New York: McGraw-Hill, 1964.

McKendry, J. M., & Stover, R. E. Equipment design considerations for improving maintenance procedures. *Electro-Technology,* July 1962, 93–95.

MIL STD. Human engineering design criteria for aerospace systems and equipment. Military Standard 803A-1(USAF), January 1964, Department of the Air Force, Washington, D. C.

MIL STD. Maintainability prediction. Military Standard No. 472, 1966, Washington, D. C.

MIL STD. Maintainability demonstration. Military Standard No. 471, 1966, Washington, D. C.

Munger, M. R., & Willis, M. P. Development of an index of electronic maintainability. Report No. AIR-B75-59-FR-207(I), 1959, American Institutes for Research, Pittsburgh.

Myers, L. B., Carter, R. G., & Stover, R. E. The evaluation of trouble shooting strategies. Technical Report No. NAVTRA-DEVCEN 1231-1, January 1964, U. S. Naval Training Device Center, Port Washington, N. Y.

Radio Corporation of America. Maintainability techniques study. (Final Report). RADC Technical Document Report No. 63-85, Vol. 1, February 1963, RCA Service Company, A Division of RCA.

Radio Corporation of America. Maintainability engineering. RADC Technical Documentary Report No. 63-85, Vol. II, February 1963, RCA Service Company, A Division of RCA.

Rees, D. W. Guide to design of Air Force check-list publications. Report No. WADC-TR-59-758, December 1959, Wright Air Development Center, WPAFB Ohio.

Rees, D. W., & Copeland, N. K. The effects of serial position in check-list design. Report No. WADC-TR-59-522, September 1959, Wright Air Development Center, WPAFB, Ohio.

Rees, D. W., & Kama, W. N. Size of tabs: A factor in handling of guides and check-lists. Report No. WADC-TR-59-158, March 1959, Wright Air Development Center, WPAFB, Ohio.

Retterer, B. L., Griswold, G. H., McLaughlin, R. L., & Topmiller, D. A., The validation of a maintainability prediction technique for an airborne electronic system. Report No. AMRL-TR-65-42, May 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

Rigby, L. V. Maintainability demonstration results for satellite control subsystems, SSD Technical Documentary Report 65, Space Systems Division, AFSC, Inglewood, Calif. October 1965.

Rigby, L. V., Cooper, J. I., & Spickard, W. A. Guide to integrated system design for maintainability. ASD Technical Report No. 61-424, October 1961, Aeronautical Systems Div., WPAFB, Ohio.

Rigney, J. W., & Bond, N. A., Jr. Maintainaability prediction: Methods and results. Technical Report No. 40, June 1964, Electronics

Personnel Research Group, University of Southern California, Los Angeles.

Rigney, J. W., Schuster, D. H., Runyan, T. L., & Budoff, I. J., Analysis of fault-location behavior in electronic equipment. *IRE Transactions on Human Factors in Electronics*, 1961, **2**, 84–67.

Rigney, J. W., Cremer, R. H., Towne, D. M., & Mason, A. K., A computer program for generating inherent corrective maintenance workload times: II. The Application of ARMAN to maintainability analysis of two radar systems. Technical Report No. 50, November 1966, Department of Psychology, University of Southern California, Los Angeles.

Rogers, J. P., & Thorne, H. W. The development and evaluation of an improved troubleshooting manual. Technical Report No. 65–1, March 1965, Human Resources Research Office, George Washington University, Alexandria, Va.

Schafer, T., Benson, N., & Clausen, H. Development of criteria and quantitative predictors of maintainability of Air Force equipment. Report No. ASD–TR–61–502, September 1961, Aero Systems Div., WPAFB, Ohio.

Seeman, J. S., Smith, F. H., & Meuller, D. D. A technique to investigate space maintenance tasks. Report No. AMRL–TR–66–32, April 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

Shapero, A., Cooper, J. I., Rappaport, M., Schaeffer, K. H., & Bates, C., Jr., Human engineering testing and malfunction data collection in weapon system test programs, WADD Technical Report 60–36, February 1960.

Shriver, E. L., & Trexler, R. C. A description and analytic discussion of ten new concepts for electronics maintenance. Technical Report No. 66–23, December 1966, Human Resources Research Office, George Washington University, Alexandria, Va.

Swain, A. D. Maintenance diagrams for preventive maintenance of ground electronic equipment: Phase III, final report. Report No. AIR–B20–57–FR–157, October 1957, American Institutes for Research, Pittsburgh.

Topmiller, D. A. A factor analytic approach to human engineering analysis and prediction of system maintainability. Report No. AMRL–TR–64–115, December 1964, Aerospace Medical Research Labs., WPAFB, Ohio.

Topmiller, D. A. Application of behavioral science to performance aid development. Report No. AMRL–TR–65–146, August 1965, Aerospace Medical Research Labs., WPAFB, Ohio.

United States Air Force. Handbook of instructions for ground equipment designers. ARDC Manual No. 80–5, May 1957, Air Research and Development Command, Baltimore, Md.

United States Air Force. Gas and fluid line identification for use in missile and space systems. AFBM Exhibit 58–20, October 1958, Air Force Ballistic Missile Div., Inglewood, Calif.

United States Air Force. Organizational and field maintenance manual. Air Force Manual No. 66–1, September 1963, Department of the Air Force, Washington, D. C.

U.S. Army Material Command. Engineering design handbook; maintainability guide for design. Report No. AMCP 706–134, August 1967, U.S. AMC, Washington, D. C.

Selected Readings

Ankenbrandt, F. L. (Ed.) *Maintainability design*. Elizabeth, N. J.: Engineering Publishers, 1963.

Goldman, A. S. & Slattery, T. B. *Maintainability: a major element of system effectiveness.* New York: John Wiley, 1964.

Rigby, L. V. &. Cooper, J. I. *Problems and procedures in maintainability.* ASD Technical Note 61–126, October 1961, Aerospace Medical Laboratory, Wright-Patterson Air Force Base, Ohio.

Chapter 13

Training System Design

Glenn L. Bryan

Office of Naval Research
Washington, D.C.

James J. Regan

Naval Training Device Center
Orlando, Fla.

This chapter deals primarily with conceptual design. It is included here because there have been many advances in educational technology within recent years which offer considerable promise. This chapter seeks to point out these opportunities. Emphasis is also placed upon the application of system design techniques to the training area. Until recently there were unfortunate tendencies to design and build training equipment in isolation. Training isolated from the job was regarded as an end in itself. It is now widely appreciated that such compartmentalization is unwise. Alternatively, an integrated systems approach seeks to integrate relevant training considerations with software design, hardware design, system evaluation and utilization—all in terms of the success of the training as demonstrated by the effectiveness of the trained man doing his job.

CONTENTS

This chapter was reviewed by Gordon A. Eckstrand and Robert Glaser.

13. Training System Design

13.1 Learning and Training

A designer who seeks to conceptualize a training system must understand the interrelationships between learning and training. Learning is an internal process. Training is a form of process control. Its function is one of manipulation outside of the learner so as to initiate and regulate the basic process of learning. The training situation is analogous to the relationship between *process* and *process control* at an oil refinery where the basic process consists of the rearrangement of organic chemical bonds and is controlled through the management of pressures and temperatures with respect to time. It would be foolish for the designer of a cracking tower to ignore the chemistry of the refining process, although he himself is not expected to be an organic chemist. So it is that the designer of training equipment seeks to incorporate as much as he can regarding the basic learning processes in his designs. Otherwise, how can he hope to control these processes? By the same token, his training arrangements must eventually be judged in terms of changes in performance which these arrangements cause to occur. To paraphrase an old saw, "the proof of the training is in the learning."

13.1.1 Stages and Types of Learning

Fleishman's work (1962) has provided an analytic technique and rationale for a phenomenon which we have all experienced: namely, that learning progresses through stages. The abilities called for during the early stages of learning are often quite different from those required later on. For instance, in learning how to shift the gears of a car the beginner must first deal with the symbolic and verbal material. He learns that each gear shift lever position has a name. He is told the purpose of shifting. The instructor explains the way the clutch works. And so on. During this early stage of learning, those trainees who lead the class are the ones who make high scores on verbal tests. Later on in the training, coordinated skill requirements loom larger. At that point, all of the trainees know what it is that they are supposed to be doing, but they haven't yet become skillful in doing it. How well various students do will depend upon differences in their ability to coordinate their limbs, and to judge the engine speed by ear, and so on.

If a training system is designed to take a learner all of the way through the various stages of learning (or to accommodate a number of different students at different learning stages), it is advisable for the system to be capable of differentially emphasizing those aspects of the learning that are appropriate to each learner's current stage. A number of possibilities suggest themselves. For example, in learning to shift gears on a simulator it might be helpful if the simulated transmission were set to be quite tolerant of synchronization and rate errors during the early phases of training (deemphasizing physical coordination). Once the trainee has demonstrated that he knew the appropriate sequence of gear shifting, the sensitivity of the transmission could be adjusted so that the simulator would stall, jerk, or gallop if the clutch, accelerator, and gear shift lever weren't properly synchronized. Also, early in the training, the system could be set to develop his verbal and conceptual skills to some criterion level before attempting to develop and integrate his psychomotor skills.

Figure 13-1 presents laboratory evidence that different skills come into play at different stages of the learning process. It shows that the different combinations of abilities predominate as practice continues and as proficiency increases. The changes are shown to be progressive and systematic throughout the practice period. This clearly indicates that the combination of abilities contributing to differential success among indi-

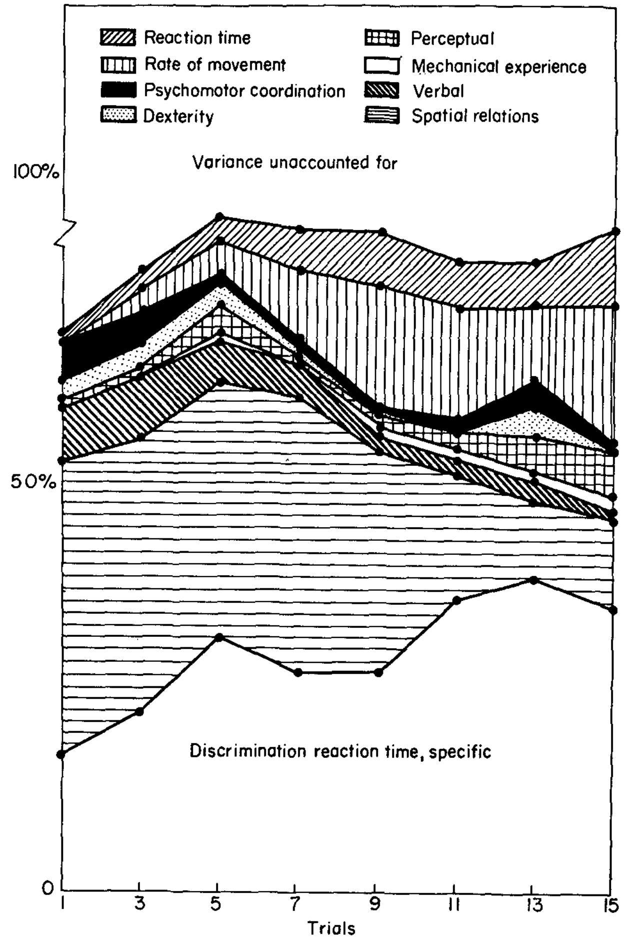


FIGURE 13-1. The changing composition of variance during the course of learning (Fleishman and Hempel, 1955).

viduals at an early stage of training is quite different from the combination which comes into play later.

It has long been recognized that distinctions can be made between different kinds of learning, e.g., skill learning, concept learning, discrimination learning, and rote learning. If there is a requirement to teach the student how to identify common features among a number of different situations, concept learning is involved. It is necessary to present the student with a wide range of positive and negative instances of the concept and provide some way of letting him know when he has correctly identified each. If a student must learn to distinguish between two easily confused situations, discrimination learning is involved. It is advisable to present them simultaneously so that he can make direct comparisons. If the student must have an opportunity to go over the material repeatedly, rote learning is involved.

Transfer of Training

The concept of transfer of training is based on the fact that there is continuity to man's behavior. How we react in any new situation is largely determined by the way we reacted in similar situations in the past—and how well satisfied we were with the outcomes of these past situations. This means that any trainee comes to each new training situation with a backlog of experiences. That backlog influences the way that he performs during his training. Furthermore, everything that is done to him during his training is likely to have some influence upon his subsequent success. When the things that a man learns in one stage of his training carry over to benefit him during later stages of his training, we say that *positive transfer of training* has occurred. A well-designed training sequence seeks to control the training environment so as to maximize positive transfer.

On the other side of the ledger, negative transfer of training and habit interference can also occur. The former term refers to the case where it is more difficult to learn Task B as a result of having learned Task A. The latter case refers to a response (habit) which occurs inappropriately in the new learning situation because it is triggered by an old stimulus. Thoughtful design of training equipment can do much to minimize these two forms of interference. One very useful procedure is to avoid piecemeal design. Perhaps the greatest advantage of the system approach is that it prevents piecemeal efforts and focuses attention upon such matters as the effects of one training experience on another training experience, and concerns itself about the vital relationships between training and performance on the job.

Figure 13–2 shows the relationships among these concepts. On the left-hand side of the graph, a hypothetical learning curve is presented. This indicates that the student acquires proficiency as a result of going through the training trials. At the conclusion of that series of trials he is about 65% proficient. The decelerating curve indicates that further practice, while beneficial, would produce diminishing returns.

On the right-hand side of the graph, three different conditions are depicted. The uppermost curve indicates that positive transfer has occurred because the student "begins where he left off" and continues to improve with additional training trials as he learns Task B. Clearly, he is better off than the student represented by the curve labelled "Negative Transfer." Here the student shows no benefit from the training which he received on Task A and the slower rate of learning progress on Task B strongly suggests that the learning of Task A may actually be interfering with the learning of Task B.

The lowermost curve shows a case of habit interference. The initial plateau (that portion of the curve having zero slope) indicates no improvement with practice. What usually happens at this point is that the student is giving the wrong responses to the stimuli presented. In the case of true habit interference, the responses given could be identified as those appropriate to Task A (but not to Task B).

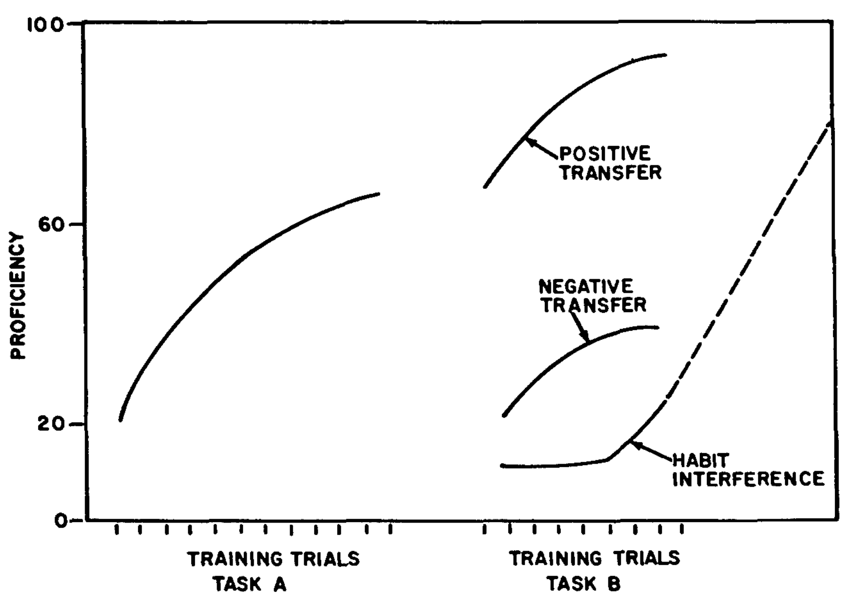


FIGURE 13–2. Hypothetical curves showing the relationship between the learning of one task (Task A) and three different possibilities for transfer to the subsequent learning of a second task (Kinkade, 1969).

Thus far this discussion of transfer has been concerned with the relationships among the various stages of training. The same sort of considerations apply to carry over from the training situation to the job situation. Unless care is exercised, it is very easy to create, inadvertently, training materials and devices which prove to be negatively related to on-the-job performance. The way to maximize positive transfer, from training to the job, is to coordinate job task analysis information with training conditions. One approach to this problem is to design the training equipment to be as "realistic" as possible. Unfortunately, such realism is costly and the increased training benefits may not be justified.

Realism

The term "realism" refers to the extent to which training equipment strives to be identical to the regular on-the-job equipment. Occasionally serious discussions are undertaken to settle "once and for all" the question of just how realistic training equipment ought to be. Such discussions are futile. A piece of training equipment can be very realistic and very good, or it can be very realistic and not very good. There is no direct correlation between degree of realism and the training effectiveness of a training system. Decisions regarding the degree of realism should be based upon the same question as all of the other conceptual design decisions, namely, what contribution will the addition of *this* particular realistic feature make to the effectiveness of *this* particular process-control arrangement? It is not desirable to include expensive realism for its own sake. Training equipment should not be designed to *be* something (e.g., to be realistic) but, to *do* something (control learning). Learning processes place their own special requirements upon the design of training systems. Equipment designed in this spirit will include many features peculiar to training objectives. For the most part, well-designed training equipment intentionally deviates from "reality" in order to promote learning.

Two ways in which a training device might be designed to deviate intentionally from operational (on-the-job) equipment deserve special attention. First, hardware components in training equipments should be selected for their durability because they are used over and over again by unskilled personnel. For example, a simple "On-Off" switch on a piece of training equipment may receive far more use in a single week of "procedures" training than it would ever get in a normal lifetime of routine operation. Similarly, electronic equipment which has to be repeatedly dismantled and reassembled during electronic maintenance laboratory training deteriorates rapidly unless it is especially made to withstand such usage. Suffice it to say, a detailed analysis of how the device is to be used during training provides adequate guidance for the experienced design engineer in this area.

A second reason for the designer to deviate intentionally from realism is to introduce certain "unrealistic" features so as to promote learning. There are many possibilities of this sort. They are described in Section 13.3 of this chapter.

An important thing in training is to provide a training situation that elicits appropriate student behavior. Realism introduced too soon may preclude this, and realism at the end of training may be unnecessary in terms of expense or in terms of the ability of the students to transfer to the real thing.

13.2 Phases of Training System Design

The massive training burden of World War II brought home the great value of thorough analysis to determine *what* should be included in training. Such analyses also proved helpful in suggesting *how* things should be taught.

Reasonably enough, the principal basis for deciding what a man should be taught is the behavior that he is expected to exhibit on the job; more precisely, it is what a man does not already know and what he cannot already do on the job that should determine what is included in his job training.

This is so evident that it is hardly worth mentioning except that it is often "honored in the breach" or done in terms of some vague idea of the job, instead of in terms of the actual job as it is in the real world. More often than not, when the contents of a training course are validated against a job, the training curriculum contains much information which is of little use to the man on the job. Thus, curriculum design seeks to assure the inclusion of essential job-relevant material and the exclusion of non-essential material. A number of effective techniques for collecting accurate, comprehensive, training relevant job information were developed within the past decade. The brief description which follows illustrates such techniques.

13.2.1 Describing the Job

As a first step, each job for which training is to be designed must be adequately described. In most large organizations, a position description already exists for each job. At first glance, these descriptions might appear to be useful design guides. But, they aren't. They are too general. Also, they include much which is irrelevant to training design. Ordinarily, such position descriptions don't actually describe what a man does on his job, how he does it, how well he has to do it, or what he had to learn before he could do it.

Hence, it is almost always necessary to produce a new job description for each job covered by a major new training system development. Several methods are available for acquiring this job-descriptive information. Men already trained can be observed doing the job. Or, incumbents can be interviewed to find out what they really do. Or, they can be asked to fill out questionnaires, or to keep more adequate records of how they spend their time. Another source of job information is the supervisor. Supervisors can be asked to indicate what they expect a newly trained man to know and what they want him to be able to do. The equipment used on the job can be studied to find out precisely what has to be done to it by the trained man in order to do his job. And so on. All of these methods are both used and useful. However, the fact that there are so many different methods betrays the fact that none is entirely adequate.

The most important consideration in selecting a job-descriptive method is to avoid sampling bias. To do this, seek information from a wide variety of sources. When feasible, employ more than one technique. These precautions are necessary since the "same" job is seldom standard across large organizations and among different workers.

One useful way to organize a description of the human performance required in a man-machine system is by means of an Operational Sequence Diagram (Brooks, 1960), an example of which is shown in Figure 13-3.

This figure shows a small portion of one block of activity involved in the "setup" and "checkout" phases of the operation of a modern sonar system. The diagram graphically portrays the sequential flow of activity. It indicates the nature of each man's duties by means of geometric symbols. Specific activities are spelled out. Interrelationships among activities are depicted. Starting at the upper left corner of the diagram, we note that the supervisor is required to give a verbal order for all equipment to be energized. This order is addressed to Scan Console Operator No. 1 who, upon receiving the order, depresses the system power switch. This action causes indicator lights to light up on each console. And so on. This format permits a great deal of information to be packaged in compact format.

The relationship between this particular block of activity and other blocks is shown in Table 13-1. As can be seen, the block used in the previous figure is Block Number 2 in this table. It follows the block named "Compute settings and predict performance." It is followed by the "Search phase."

The block analysis organizes gross activities according to their normal order of occurrence and provides some guidance with respect to the allocation of training emphasis among blocks. For example, inspection of Table 13-1 reveals that Blocks 3 through 7 are essential to the job. Clearly some training should be directed to them. On the other hand, the remarks accompanying Block 8 indicate that the activities labelled, "Conduct special communications," occur under unusual conditions and probably merit relatively low training priorities. It will also be noted that the activities referred to in Block 1 seem to be the sort of thing that may be fundamental to all of the other blocks. If so, these should be taught first or receive other special treatment.

The activities in each block can be further expanded to present a more detailed "Task List," which is the next step in the development of an adequate job description. Examination of Table 13-2 reveals that the descriptive process entails decomposing the job into its major parts, and then the decomposing of the parts. Each successive level gets more detailed. At the level

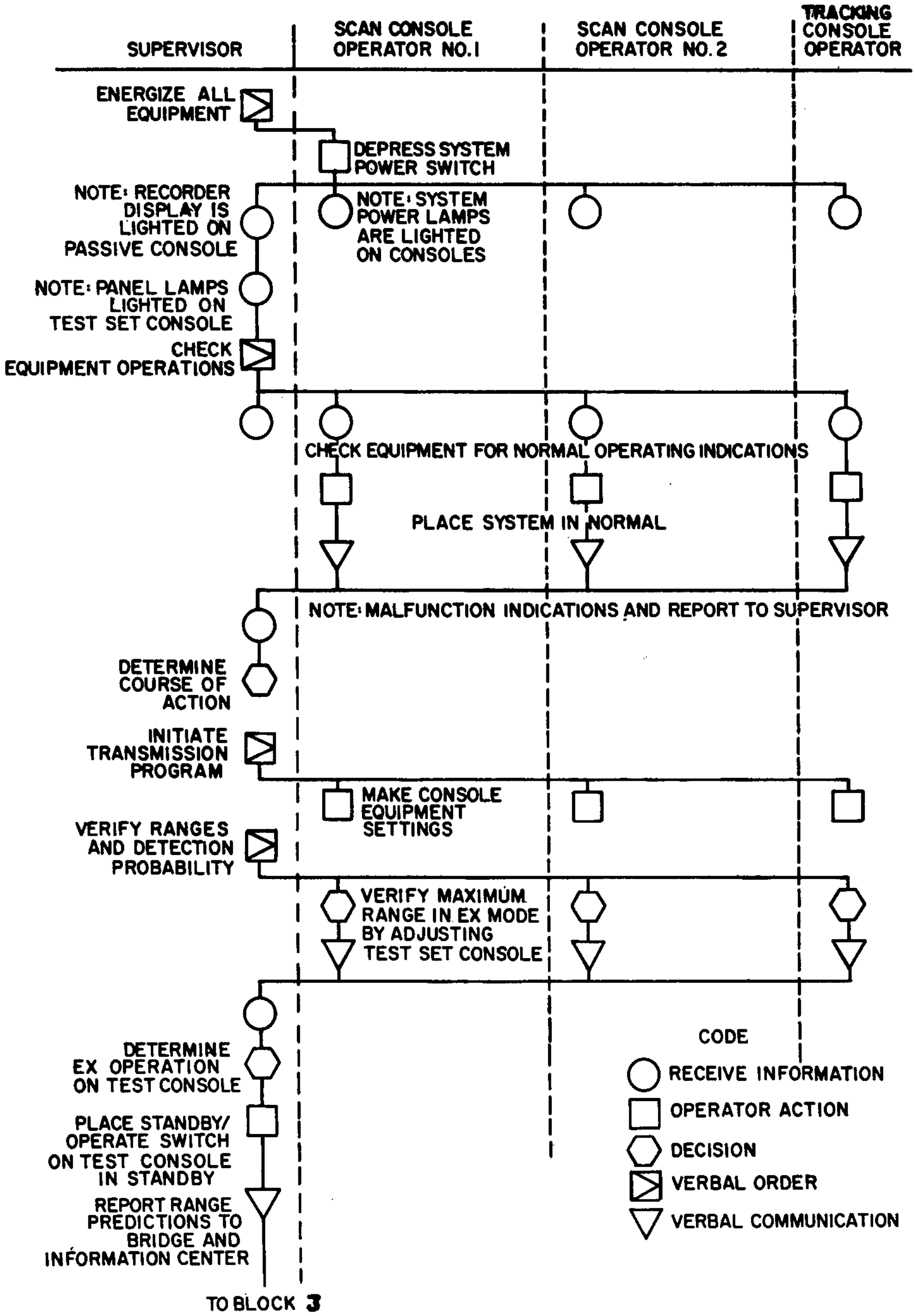


FIGURE 13–3. Operational sequence diagram for setup and checkout of sensor equipment (adapted from Brooks, 1960).

TABLE 13-1. SENSOR SYSTEM BLOCK ANALYSIS

Block No.	Block name	Remarks
1	Compute settings and predict performance.	First action required when watch begins. May be repeated one or more times during watch period.
2	Setup and checkout of equipment.	Performed at beginning of watch, and may be repeated during watch. May be omitted if watch continues previous watch activity.
3	Conduct search (search phase).	Normal predominant activity on watch. Ends when consistent signal is detected.
4	Contact phase	Follows block 3. Begins when consistent signal is observed, or definite audio signal is heard. Ends when ON-target button is pressed and signal is classified as definite non target, and block 3 activity is resumed.
5	Track target	Begins when ON-target button is pressed. Ends when FCS TRACK is accepted at sensor console.
6	Conduct attack (attack phase).	Begins when sensor console has accepted TRACK, ends when target is lost or destroyed.
7	Conduct lost	Begins when three successive signals fail to appear after contact has been established.
8	Conduct special communications.	Special condition of use not related to basic systems operation. Omit from study.

TABLE 13-2. DETAILED TASK LIST FOR SENSOR SYSTEM

Block 3.	Conduct search and surveillance.
3.1	Conduct search in BZ mode.
3.1.0	Check that XMTR on lamp is lighted.
3.1.1	Make specific equipment settings for BZ mode.
3.1.2	Select desired bearing.
3.1.3	Monitor scan display and audio channel for signal.

Note: System will recycle continually every 55 sec. with 5 returns per sec.

3.1.4	Operator detects signal on scan display.
3.2	Conduct search in CC/TD mode.
3.2.0	Check that XMTR on lamp is lighted.
3.2.1	Make specific equipment settings for CC/TD mode.
3.2.2	Select desired bearing.
3.2.3	Monitor scan display and audio channel for signal.
3.2.4	Operator detects signal on scan display.
.	
.	
.	
3.6	Observe indication on passive recorder.
3.6.0	Energize unit and RCDR switch is on.
3.6.1	Periodically monitor recorder display for signal by noting variation in pen traces.

Note: Paper advances at rate of 1 in. per min. Since 5 in. of paper are visible, display should be observed at least once per 5 min.

shown in Table 13-2, the various tasks which each trained operator must be able to perform are spelled out. A numerical coding scheme relates each task to its proper block and to other tasks in the list.

In a similar manner, a detailed time breakdown for the tasks is prepared. This detailed time-by-task information is used to compute system block times. These data are summarized to estimate system block times as shown in Table 13-3. This figure provides detailed estimates of the amount of time required to complete the tasks in each block, assuming that sensing conditions are good and the operators are highly skilled. Such information is important in setting performance standards for men who are to be trained by the new training system.

13.2.2 Task Analysis

In keeping with the general strategy of breaking the job down into smaller and smaller bits, it is now appropriate to decompose the tasks into the discrete activities which make up each task.

A task analysis format devised by Chenzoff and Folley (1965) involves placing the tasks in behavioral classes related to training design decisions. Figure 13-4 shows this format applied to the sonar system upon which the previous tables were based. Each activity is number coded to relate it to the task of which it is a component. These numbers are listed in the first column. The next column indicates who performs the activity. The next column indicates which of five kinds of activity it is (procedural, monitoring, perceptual motor, communication, decision making). This is followed by an indication as to whether the activity is part of a fixed (F) or variable (V) sequence. The next column contains ratings ((0,) not essential; (1,) necessary but not

TABLE 13–3. ESTIMATE OF SYSTEM BLOCK TIMES FOR SENSOR SYSTEM

Block 1	14 min., complete.
Block 2	12 min., complete.
Block 3	Continuous operation throughout watch, interrupted only by going to Block 4.
Blocks 4–7	Depend on specific mode of operation.

	BZ mode	BB mode	TD mode	RT mode	ConTc.
Block 4	192 sec	213 sec	72 sec	81 sec	23 sec.
Block 5	119 sec	189 sec	61 sec	52 sec	
Block 6			209 sec	216 sec	
Block 7	143 sec	162 sec	41 sec	51 sec	

	TD mode	*RT mode*
Minimum attack time	425 sec	444 sec

TASK CODE	OPERATOR	TYPE OF ACTIVITY	SEQUENCE	CRITICALITY	COORDINATION	SPECIALIZED BEHAVIOR	DIFFICULTY	DYNAMIC CONDITION	REMARKS
4.1.1	SCAN CONSOLE	COMMU-NICATION	F	1	3	1	--	--	
4.1.2	SCAN CONSOLE	PERCEPTU-AL MOTOR	V	2	1	2	--	D	
4.1.3	SUPER-VISOR	COMMU-NICATION	V	1	2	1	--	--	
4.1.12	SUPER-VISOR	DECISION	V	2	2	3	1	D	

FIGURE 13–4. Functional analysis/task details (adapted from Schrenk, 1969).

demanding; (2,) critical) of the criticality of each activity. Whether the task can be done by one man acting alone or whether it involves team coordination is shown in the next column ((1) 1 man, (2) 2 men, (3) 3 or more). The amount of training required to convert the typical trainee from his entry-level behavior to his desired exit-level behavior is given under the heading "Specialized Behavior"—((0) not related to previous experience, (1) readily learned, (2) short training period, (3) extensive practice). The last three columns deal with the degree of difficulty in performing the activity once it is learned ((0) not difficult, (1) difficult, (2) very difficult)—whether it occurs when the sonar system is in the static (S) or dynamic (D) condition, and the inevitable "Remarks." Dashes indicate that the item is indeterminate or not applicable.

13.2.3 Functional Training Requirements

Working with all of the material collected thus far, the gross functional requirements of any training system which meet the needs of the operational sensor system can now be spelled out. Review of all of the materials collected in the case of the sonar system (from which the previous figures were drawn) led an experienced analyst to a number of conclusions, some of which follow:

1. The monitoring and decision tasks will be difficult to perform because of the masking effects of noise and reverberations.
2. Coordination and inter-operator communication are important because of a high degree of console interdependencies.
3. Since detection and classification of sonar

targets will be very difficult, training stimulus signals should be carefully chosen, especially to provide a wide range of operator experience.

4. Three levels of training will be needed: individual. subteam, and team.

5. The training situation should encompass all modes (monitoring, operating, communicating, etc.) so that critical interactive skills can be practiced.

6. Task performance times are available to set training standards.

7. Both perceptual-motor and decision skills are required, thus a wide range of simulation and gaming techniques will be involved.

8. The difficult tasks are proportionately small and are identified, thus permitting appropriate use patterns to be designed.

Such a set of conclusions provides a broad context for "roughening out" the requirements which a training system should seek to satisfy if it is to provide adequately trained sonar operators.

13.2.4 Trade-Off Analysis

Training resources are seldom sufficient to permit the unconstrained design of training systems. Consequently, it is usually necessary to review functional training requirements in terms of the projected costs of satisfying them. Studies of the type required at this point resemble those ordinarily made to "sell" the basic concept of the trainer in the first place to those who control the financing of such developments. However, at this point in the design cycle these studies serve quite another purpose. They seek to estimate the costs and relative benefits of various training features in order to determine what will be included and what will be left out of the new training system. If, as is almost certain to be the case, different training features cost different amounts of money and serve different ends, it should be possible to "trade-off" one training feature against another to insure the optimum expenditure of the available dollars.

It is important for the cost estimates used in this connection to cover *all* expenses associated with the *entire* system projected over its *full* period of utilization. To do otherwise is to run the risk of producing a system whose support requirements exceed the operating budget of the training establishment, its ability to provide staff and instructors. and other fixed constraints. Unfortunately, it is not unusual for a training school to discover belatedly that a new training equipment requires programs, film clips, and magnetic tapes that are, themselves, quite expensive. Even more expensive is the art work, the narrators, the actors, the technical writers and all of the other support personnel necessary to provide up-to-date libraries of materials suitable for use with the new training equipment. As a result, elaborate audio-visual aids often are not used to the extent originally anticipated. The burden of providing new material to run through the equipment is too great to depend upon the volunteered efforts of the teachers. If we are to avoid storerooms crowded with unused teaching aids, funds must be provided to cover the cost of procuring the materials necessary to support their continued operation.

A considerable amount of information is also required regarding the plans for utilizing the training system. How many men will have to be trained? How long will each man be available to receive training? How many shifts a day can the training system be allowed to operate? What are the minimum standards of performance that a trainee must surpass before he is considered to be sufficiently well trained? Given this type of information, the design can be tailored to the needs of the prospective user within a set of realistic constraints.

A number of difficulties arise in connection with training tradeoff studies. Two major ones deserve comment. One stems from the tendency of the person designing the study to restrict the range of alternatives too narrowly. For example, a study might compare one costly hardware alternative with another costly hardware alternative—but not with any nonhardware alternatives. Similarly, from a methodological point of view it would be more informative to examine the interactions of several variables and to represent each variable by several points along its extent instead of a single point in any given study. Failure to plan such tradeoff studies carefully frequently leads to sweeping generalizations based upon a simple comparison between specific instances. In the worst cases, one alternative has been intentionally chosen to be a straw man, purposely selected to justify a decision already made. While this might be a good marketing tactic, it is a poor way to produce tradeoff information relevant to training system design. Such rigged studies have made many people quite skeptical of the credibility of studies which compare design options.

A second common difficulty in performing tradeoff analysis is of a much more fundamental type. Briefly, it has to do with the "softness" of the estimated "benefits" which serve as the denominator in costs/benefits ratios. To say the least, it is difficult to derive persuasive and accurate "benefits estimates." Needless to say, accuracy is desirable, but the real concern is that the benefits be expressed in appropriate terms.

Several different ratios are useful expressions of benefit. For example, "cost per adequately trained man" or "cost per hour of effective operation" or "cost per unit learned" are useful. In comparing training alternatives A and B, we might have reason to believe that it would take longer to learn up to a specified standard if Alternative A were chosen. However, Alternative A might be a good bit cheaper to implement. Clearly, what we would need to know under such circumstances is just *how much* slower A is than B, and just *how much* more expensive B is than A in terms of real costs. Only when this type of quantification is available can design decisions be made with full confidence that the resources are being expended as wisely as possible.

Before going on, one more point is in order. In addition to comparing various *training* alternatives with each other, comparative costs and benefits of *non-training* alternatives can also be reckoned. Training is but one way to correct a mismatch between jobs to be done and the available supply of talent. Other ways exist. To name a few—efforts can be increased to recruit or retain, on a selective basis, a greater number of men who already know how to do the job and thus don't have to be trained. Or, improved efforts can be made to secure more effective distribution of men currently being trained so as to maximally utilize trained manpower and minimize the number of men to be trained. Or, jobs can be redesigned so that they can be filled by people with lower qualifications or with less training. And, of course, equipment can be "human engineered" so that relatively untrained people can operate it.

Channels should exist to allow prompt and appropriate consideration of these "non-training" alternative means for coping with unusually tough or expensive training problems exposed by the preliminary analyses. Hopefully, by such means, training device development would go forward only when it was proven to be cost-beneficial and essential. And, under those circumstances, the devices selected for development would receive such funding and design attention as necessary to accomplish the prescribed training.

13.2.5 Task Analysis in Perspective

A traditional task analysis technique has been presented here without comment. However, before leaving this subject, some personal impressions are offered to emphasize that the designer should look for the training-relevant forest through the trees of task analysis detail.

During the past 15 years or so, job and task analyses have received rather heavy emphasis. When the U.S. Air Force introduced the first carefully worked-out set of procedures for obtaining Qualitative and Quantitative Personnel Requirements Information (QQPRI), there was a great deal of rejoicing among behavioral scientists. At long last it seemed that they had a mechanism which would allow them to make substantial contributions to applied manpower and training problems. These scientists foresaw continuous broad-based research support which would allow this modest beginning to grow into a systematic program of carefully documented, research-based, job-validated training techniques. The QPRI and the QQPRI were much-needed first steps. But, they were just crude beginnings. Unfortunately, the support for the backup research has been modest. Job and task analyses have become popular as ends in themselves. Critically needed basic research programs required to expand and validate the behavioral information have not been adequately supported. As a result, job analyses are called upon to do more than they can possibly do. Job-descriptive information alone cannot provide sound bases for translating job requirements into personnel and training requirements. Job and task analyses, although essential to modern system development, are no better than the science on which they are based. The scientific basis is sometimes weak and uncertain. Given this condition, improvements in the scientific bases are more important than improvements in analytic techniques as such. At present a great deal more information is needed regarding the relationships between various kinds of tasks and various kinds of learning. Not enough is known about the way that the training should be adapted to the different backgrounds and capacities of different learners. More will be said about this later. In brief, there is a tendency for job analysis to be oversold. The analyst becomes enthralled with breaking the workaday world into ever-smaller bits in order to produce an involved and cumbersome document. It is this reductionistic approach without sufficient regard to the meaningfulness of the behavioral elements that should be avoided.

13.3 Training Device Design Concepts

As mentioned earlier, modern educational technology has emphasized the importance of a number of functional characteristics which have been shown to be of great value in programmed instruction, computer-assisted instruction, closed-circuit television, and language laboratories. Although none of these features is entirely new, there is little evidence that the designers of large-scale simulators have attempted to incorporate such concepts in new hardware design. The purpose of this section is to describe these features and to suggest some of the possibilities for including them in new simulators. Undoubtedly, the creative designer will think of many other possibilities.

13.3.1 Prompting and Cuing

Prompts are signals which indicate that the time has come for a specific action to occur and direct the student to perform that act. One example of a training device which employs prompts extensively is the Solatron Automatic Keyboard Instruction (SAKI) used in teaching keypunch operators the operation of the keyboard of card-punching equipment. During the early stages of their training, a light comes on behind each key on the keyboard when it is to be pressed (Pask, 1958). As training proceeds, the prompt is momentarily delayed to give the student an opportunity to make the appropriate key selection. If he doesn't make it during that interval, the key is lighted to prompt his response. The same technique might be applied in teaching a pilot trainee how to scan his instrument panel. In that case, he could be prompted to inspect each instrument in a fixed sequence by having each light up according to a programmed sequence.

Cuing is similar to prompting. Although the terms may be used interchangeably, the term "cue" usually refers to a simple signal that indicates it is "time to act." Thus, a cue is much less directive than a prompt. For example, in a cockpit simulator, a light could come on indicating that some (undesignated) action should be taken. This would be a cue. After a scheduled interval had elapsed, some specific control, such as the landing gear retractor handle, might glow to *prompt* the trainee to operate it.

It is considered good practice to employ prompting techniques during early stages of training and then to switch to cuing as training progresses. Whenever either cues or prompts are used, arrangements should be made to withdraw them systematically as the trainee gains confidence and competence. It is a responsibility of the training system designer to wean the trainee away from those artificial aids not present in the "real world." Two ways can accomplish the weaning: by reducing the frequency with which the crutches are used, and by reducing their intensity. In either case, the prompts or cues gradually disappear and the trainee relinquishes his dependence upon them as he learns to identify the stimulus conditions present in the real world which acquire (through association during training) the power to prompt the appropriate responses.

13.3.2 Feedback to the Learner

A common process-control arrangement involves feedback loops. A signal is picked off at some point in an information chain and fed back to an "earlier" stage in the chain in order to regulate an on-going process. Sometimes this regulation is handled automatically, but this is not a necessary condition. For present purposes, let us consider feedback arrangements which can be designed into training systems to provide the learner with information regarding his own learning even as it takes place.

The concept of feedback is surprisingly rich as it is applied to learning. Some things that we learn to do are generous in providing "built-in" feedback. Others aren't. As an example, consider the youngster learning how to drive a nail with a hammer. As he struggles to acquire mastery, the very situation continuously furnishes feedback information to him. He sees the hammer miss the nail and strike the wood beside it. The dent in the wood tells him the direction and extent of his error. Even as he swings, his muscles report the feel of the swing and he experiences the solid satisfaction associated with a good blow delivered to the nail's head. In time he will get so he can interpret the authoritative whack of the hammer when it strikes the nail correctly or the annoying twang of the nail as it vibrates from a glancing blow. With repeated trials, he learns to recognize these indications and to experiment with various grips and stances hoping to match them with better performance. Even the casual observer will see the skill developing. There are fewer mashed fingers, fewer bent-over nails. Each nail is driven into the wood with fewer blows. The swing becomes more rhythmical. The boy seems more relaxed and self-assured as he achieves mastery.

One reason boys enjoy learning to pound nails is because of the task's rich feedback characteristics. Lack of feedback may contribute to their difficulties in training for other tasks and their related dislike of such learning tasks. Many scholars (especially Smith and Smith, 1966) feel that the act of learning consists essentially of adjusting one's activities in terms of the feedback received.

Feedback is scant or ambiguous in many learning tasks. The job of the system designer is to figure out ways to supply feedback in the training situation to expedite learning.

One type of feedback is called *knowledge of results*. In its simplest form it tells the student whether a response is "right" or "wrong." Most proponents of the extensive use of knowledge of results urge that the student be informed regarding the correctness of his action as quickly as possible after he has concluded his response,

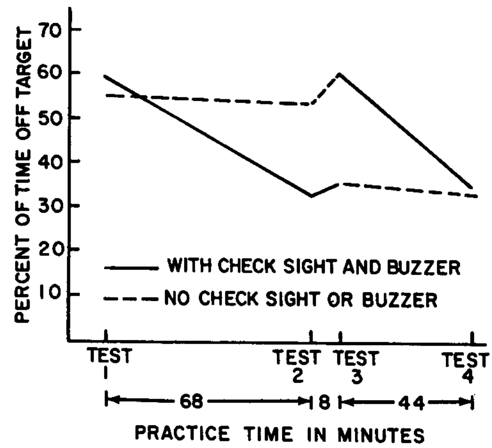


FIGURE 13–5. Effect of immediate knowledge of results on speed of learning to track aerial targets (Wolfle 1945).

hence, the expression, "immediate knowledge of results."

Figure 13–5 shows the effect of knowledge of results. One group was informed immediately when off target and learned faster than a group who trained conventionally. When the latter group was given immediate feedback they quickly caught up. This has both theoretical and common sense appeal. However, Lavery (1962) argues that feedback following *each* trial leads the trainee to try to correct the "variable error" component of his performance, while feedback provided only at the conclusion of a *block* of trials leads the trainee to correct his "constant error" component. In some instances, the objective of training is to minimize the constant error. To be on the safe side, trainers could be designed to allow either trial-by-trial or end-of-block feedback. In addition to designing so that trial-by-trial feedback can be withheld until the end of a block of trials, it is useful to design so that feedback can be withheld altogether under instructor control. In this way, both the instructor and the student find out how well the student can perform without knowledge of results.

It is quite simple to provide the student with a signal to indicate whether or not his last response was correct. However, before such a signal can be presented to the student, it is necessary to evaluate the response to determine whether or not it was, in fact, correct. This is considerably more difficult to implement. Up until quite recently, about the only feasible procedure for doing this "automatically" was to limit the student's response alternatives to some small fixed set of responses (similar to multiple choice questions). When this was done, it was possible to present "canned" evaluations since the student could make no responses except those anticipated by the program writer. Such restrictions are subject to all of the adverse criticisms leveled at true-false tests and multiple-choice tests. They are often misleading, arbitrary, and frustrating. Ideally, one would hope to be able to allow the student to make whatever response in the training situation which he felt was appropriate. That response would then be judged and reported back so that the student would know whether the system regarded his response as "right" or "wrong." The advent of digital computers has demonstrated the potential for quickly evaluating individual student's free responses (responses not selected from a set of fixed alternatives) and automatically supplying feedback in accordance with contingencies written into the computer program. A description of computerized training systems is contained in a later section of this chapter. (See Section 13.3.5.)

It has been shown (Katcher and Hunter, 1957) that there are some benefits to be derived from recording salient aspects of the trainee's performance for the purpose of feeding it back to the student for later review and analysis. This analysis can be made by the trainee himself. Such critical self-analysis can be of direct training benefit. Its advantages include the fact that the student can see *sequences* of behavior as unitary and not just a series of "responses." It also obviates the equipment requirement for built-in response evaluators. From the standpoint of training system design, in order to implement such training, the training systems designers have to decide what data to acquire, how these data should be stored, and how to retrieve the stored data for later exercise recapitulation and critique. Ideally, this information should be made available almost immediately after each session while the trainee can still recall details about the exercise. Such prompt service may require the use of automatic recording devices and mechanical data processors. For certain types of training, video-tape replay offers attractive possibilities. A word of caution may be in order, however. Even with the use of sophisticated recording and processing equipment and a TV camera that "sees all," it is important to specify completely, in advance, exactly what it is that the student is expected to accomplish during his review and self-appraisal. And, detailed arrangements should be made to assess the student's progress toward those particular objectives.

A special problem with regard to feedback comes up in the area of providing training to "teams." This kind of training brings together a group of men who presumably have mastered the individual skills necessary to do their jobs, but now must learn to participate in the integrated activities of the team. A common method of training is to exercise the team under simulated conditions which resemble situations that the team will be expected to cope with when it has completed its training. In many real-life instances, particularly in the military, the teams are organized so as to function effectively under adverse conditions and to adapt "gracefully" their performance in the face of personnel or equipment failures. Such reliability in the face of adversity is accomplished by means of redundancy, parallelism, and the capability of one team member to compensate for another team member's inadequacy. When it comes to training such teams, these conditions which contribute to reliability of team performance confound the training feedback situation. As a result, it is often difficult for an individual trainee to evaluate his own performance as distinct from the team's. There is experimental evidence (Klaus and Glaser, 1968) that under conditions such as these, the trainee may actually strive to stabilize inappropriate and inadequate behaviors in the mistaken belief that these behaviors are actually contributing to the success of his team. From the standpoint of training system design, what is needed in these cases is some direct evaluative feedback provided directly to each separate team member concerning *his own* actions as well as providing feedback information to the team as a whole concerning the team's success. Supplied with this information, the individual trainee can then try to maximize the rate of his own learning and thus, at the same time, provide ever-better inputs to the other trainee members of the team.

13.3.3 Difficulty Level and Error Control

Certain training systems, especially those built around simulators, depend upon exercises or "problems" to provide the controlled experiences which the training system designer expects to result in appropriate learning. Well-designed equipment will accommodate a range of exercises varying widely with respect to their difficulty. A common fault among training devices is that they make inadequate provision for altering the difficulty level. For example, in air defense trainers there is undue reliance upon the idea of changing the number of targets in an incoming raid as the basis for controlling problem difficulty. Strictly from a training standpoint, it would be helpful if other factors (such as target speed) could be varied over a very wide range. It is even desirable to permit the learner to take his first few faltering steps in a slow-paced mode (where the targets would come in so slow as to permit the neophyte time to react). It is not enough to design the simulation so that it faithfully reflects reality. It should also be capable of manipulation to control task difficulty during the various stages of learning by deviating from reality. Another important reason for designing training systems to accommodate a very broad spectrum of problem difficulties is to permit careful control of the number of errors made by the learner. Errors should only be allowed to happen to the extent that they contribute to the learning process. While it is sometimes true that we learn by our mistakes, it is not always the case. Unless capably controlled, student errors can lead to erroneous associations. Frequent errors by the student are danger signals. Over the years it has become increasingly evident that errors can seriously disrupt the orderly progress of learning. Skinner (1954) has made a strong case for arranging training conditions in such a way that very few errors occur. Not all informed investigators agree with his "low error rate" position. However, almost all would agree that the error rate should be under firm control of the training system (whether the rate be controlled at a high or a low level). The regulation of problem difficulty and appropriate sequencing of the material to be learned are two effective ways of controlling the error rate. Unless conscious design effort is directed toward flexibility with respect to both the difficulty level of the exercise material and the sequencing of the material to be learned, the eventual user of the training system is apt to be saddled with naive (or even discredited) training strategies inflexibly built into the system.

From the design point of view, this means that the input parameters should be extended well beyond the demands of the real situation. For example, if a training simulator requires the presentation of information regarding airborne targets, it is good practice to allow for a far larger number of independently maneuverable targets than originally thought necessary. This provides the option of employing unrealistically high (or low) target loads for training purposes. The same thing applies to target speeds. For training purposes, it may be desirable to be able to speed the targets up, perhaps twice as fast as they would "really" go. Or, the instructor may wish to slow the targets down to a crawl. He might even find it useful from a training standpoint to stop them altogether—or to back them up.

13.3.4 Flexible Sequencing

Training materials designed for flexible sequencing can be presented in any order at all. Series of events can be interrupted, reversed, or reordered at will. It should not be necessary for a learner to always "begin at the top." The instructor should be able to move quickly to any point in a scenario that he chooses and begin the action there. He should be able to stop the action when he wants to do so, recover information about the student's performance quickly, and reset to the very same starting point as many times as he chooses. The system should be designed so that such changes are very easy to make and not dependent upon time-consuming software modifications. In this regard, it is important to note that the simple addition of a so-called "general purpose" digital computer does not guarantee this degree of flexibility. Even though such devices are reprogrammable, such reprogramming can be very expensive and time-consuming. A better solution is to provide an adequate systems program in the first place that can switch easily from mode to mode.

Some strings of behavior are called "behavior chains." These chains are made up of a sequence of activities which always occur (in real life) in a fixed order. Thus, if the separate activities which make up the sequence are represented by letters of the alphabet, then A, B, C, D, E would represent a chain consisting of five activities. Since these activities always occur (in real life) in the indicated order, it would seem reasonable to design a "procedures-trainer" to accommodate that particular order and no other (assuming a procedures-trainer was called for). However, such a design decision would eliminate an effective training procedure called *backward chaining.*

In backward chaining, even though A, B. C, D, and E always occur in that order in real life, they are not taught in that order. Rather, activity E (the last) is taught first. Then D is taught and D-then E are practiced. Training continues to "back through" the chain until all activities have been learned.

Although backward chaining may not always prove to be a superior technique for training people to learn to perform a behavioral chain, it should always be considered. The main point in including this discussion here is to emphasize again the desirability of building great flexibility into training equipment and to be prepared to deviate from the "real world" for the purpose of expediting training. Thus, even though a sequence of activities *always* occurs in a given order in real life, it may be desirable to teach them in a kind of reverse order under training circumstances. (See Gilbert 1962). Careless design of training equipment can unintentionally rule out the use of powerful new training techniques.

13.3.5 Adaptive Training

The term adaptive training refers to a range of training situations in which the material presented to the trainee depends on his current state of knowledge, his training goals, and his personal background. These response-contingent situations typically, but not necessarily, involve the use of digital computers. An early form of adaptive instruction exists in the tutorial situation with a ratio of one student for one teacher. Here the student's progress can be continually assessed and the instructional methods and materials adjusted to his progress. In fact, this is precisely what a good tutor does.

In certain of its forms, programmed instruction is an incomplete example of adaptive training. For example, branching programs permit different students to proceed through the material by different routes depending on how each responds to the subject matter provided. A program of instruction with many branching alternatives can deal with a range of trainee differences in ability, experience, and rate of learning. In this manner, weaknesses and confusion can be detected and remedial training can be provided. Slow learners can proceed more slowly, take smaller steps, or start out with simpler problems.

The training equipment designer should be concerned with adaptive instruction because its most promising implementation depends upon sophisticated training hardware. More specifically, the form of adaptive instruction of most interest to the designer is computer-assisted instruction (CAI). A typical CAI system involves a large digital computer, student input stations, and a variety of software components. The student is presented with the instructional material (language, history, basic electronics, statistics) at his station. Such stations incorporate some method of displaying the material and some provision for the student to enter his response to it into the computer system. The computer controls the display, reviews the student's responses, and presents additional material determined by an analysis of information it has about the student and his prior responses. It keeps a record of the instruction and, when properly programmed, it can devise and modify rules for presenting material based on the trainee's activity.

Major development problems remain to be solved before the potential of CAI can be realized. This is particularly true in military settings where much of the subject matter to be taught differs from that in the institutions where most of the CAI has been developed. Languages easily used by instructors to address the computer need to be developed. The interface between the student and the computer needs to be enriched in various ways. For example, we need to know what the characteristics are of a student station for teaching procedural or manipulative skills and decision making. A third important development problem exists in the area of the instructional strategies. What should be the basis for presenting to the student a given block of material, for setting a level of difficulty, a rate of presentation, or a particular method of instruction? The computer specialist, psychologist, and designer need to cooperate in attacking these problems.

The training equipment designer will be faced in the future with the design of CAI systems. A detailed discussion of this emerging training technology is beyond the scope of this chapter. The interested reader is referred to a bibliography (Hickey, 1968) for reference.

The concept of adjustive or adaptive instruction need not be restricted to CAI systems. Such features can be instrumented in important ways in conjunction with conventional training systems. The designer should consider this possibility in every new training system design. Increasing individualized training and the automation of the instructor's role can distinctly increase the quality of training possible. (Automated instructor stations are in exploratory development now.) Even in existing systems, the computer software areas offer an opportunity for incorporating, at least in a limited sense, adaptive features. For example, a command control system weapons control station can provide various configurations of target data and weapons availability in a range of time frames, any combination of which can be presented to the station operator. A given presentation can be keyed to the previous decisions of that operator which had been recorded and analyzed automatically. The progress of the trainee can be measured in terms of the complexity of the stimulus situations which he can handle. A record of his performance can be compared with other student performance and with operational requirements. The student can thus be readily certified to enter the full team as a qualified operator, or if he shows exceptional promise, he can be singled out for special subsequent assignment. In like manner the entire team can be dealt with on an individual basis.

A wide range of learning materials must be available to the designer who is developing a system with adaptive features. The form the student presentation takes varies with the task, of course, but control of the presentation sequence rests with the computer. The rules programmed for computer use will likely be modified with experience. The computer programs should include features which will permit these improvements to take place—histories of students' responses, for example.

In considering the incorporation of adaptive features in present trainers, the designer will usually be faced with digital computer software problems. These problems include not only the programming but the pedagogical rules to be adopted. Good *general* rules do not exist. The programs should be designed with the explicit objective of collecting data during use of the training system which will form the basis for improved presentation rules. The designer may be faced with special student and instructor station design considerations. Since much of the operation is automatic, the instructor position should be keyed to monitoring current student progress and improving future training procedures for use with future students. The student station may require special response features which permit the student to respond (for example, through key sets) directly and unequivocally to the computer. The position may also require special feedback displays since the on-line role of the instructor is very likely reduced in adaptive instructional settings.

Although many of the instructor's functions are automated in adaptive training, the instructor is not displaced entirely, as he has new roles to play. He can act as monitor, troubleshooter, and decision-maker since he will be relieved of many routine tasks. The details of his new role need yet to be identified but his effectiveness will surely increase.

In design situations where full-scale digital equipment is not required or available, various alternate methods for introducing adaptivity should be considered. In a vehicular control trainer, for example, adaptive circuits can be incorporated to vary system dynamics and system inputs as the trainee progresses. In small, essentially manual, training devices the instructor can vary the presentations in some systematic way. The designer should provide equipment (e.g. a random access device), presentation materials, and manuals which will permit and encourage the instructor to function adaptively.

13.3.6 Designing the Training Device

When it comes down to the actual design of the large-scale simulator itself, there are a number of points worth making. While these are in no sense unique to simulator design, they are important.

First, a word about *reliability*. Nobody likes equipment that doesn't work. But, many people approach training equipment with chips on their shoulders—almost as if they hope it won't work. Actually, reliability is harder to deliver in a large-scale simulator for several good reasons. A given simulator is apt to be one-of-a-kind and hence may contain many of the "bugs" that crop up in any complex first attempt. Also, trainers are often subjected to hard usage. And, simulators can be misused (in the sense that they may be asked to do things they weren't designed to do). In any case, the designer should design so as to emphasize reliability. The use of standard solid-state circuitry, modular construction, redundancy, ruggedization, and large safety factors all contribute to equipment reliability.

Building-block design. Some efforts have been made to employ building-block design philosophies in the planning of large-scale simulators. For example, several recently designed aircraft simulators employ a central computer which services two or more different aircraft cockpit trainers at the same time. Plans called for more cockpits to be substituted as they became available. In like manner, the central processor could be replaced by another general-purpose computer when gains in the state-of-the-art warranted such a change. There is great benefit to be derived from planning families of simulators which can be readily modified by changing building blocks or designing new building blocks that are compatible with the old. When a large-scale simulator has not been designed according to building-block principles, it is hard to modify and update. Eventually it becomes hopelessly inadequate, falls into disuse, and then has to be replaced *in toto*. The modern simulator is too expensive to permit such periods of disuse and such large-scale obsolescence. Building-block designs also provide a basis for expanding small-scale simulators, for using common blocks to serve several different simulators, and to reduce the maintenance and support burdens which stem from a multiplicity of unique devices.

Ease of reprogramming. One hedge against change is the design of training simulators around programmable general-purpose digital computers. These remarkable devices can be programmed in a variety of ways to control many aspects of the training simulator. While this is true, it doesn't follow that the initial programming is inexpensive or that a program, once written, is easily modified to incorporate new and unanticipated features. However, a great deal can be done at the time that the program is originally written to provide for its subsequent modification. Adequate documentation and proper program organization go a long way toward making a program easily modified. In any case, the complete training system including all of the hardware and all of the software can be planned against the certainty that reprogramming will eventually become necessary.

Multi-use and multi-user training system. There are definite advantages to designing new training systems so that they can be used for different purposes and for different kinds of users—ideally in such a manner that these manifold users can use the training system simultaneously. Consider, for example, a modern command and control system. The ideal training system should be able to provide useful exercises for an entire interacting team within a simulated context. But, it should also be capable of being subdivided in such a manner that any (or all) subteam(s) could be segregated for the purpose of training with its function automatically supplied to the remnant of the team by a computer program. In the same manner, any single team member should be able to withdraw from his team for private training while the functions of the missing man are automatically supplied to the team and the man's training is automatically administered to him by the computer program.

Many other modern possibilities can be suggested. But, it is appropriate to turn now to the design of the major components of a large-scale simulator. Additional points and examples are included in the discussion of the design of the student stations and the instructor's station.

13.3.7 Design of Instructional Stations

Up to this point a number of characteristics of learning have been presented along with their system design implications. Just exactly how these characteristics are to be taken into account in each particular case must be left to the discretion of the designer since hard and fast general rules are not available. In this brief section, we turn from broad system design considerations to a more customary organization focusing on the design of the student stations and the instructor's station.

Student Stations

The student station is the location of the learner functioning as an element in the instructional system. It could be just the chair that he occupies in the classroom. However, for this discussion, attention will be directed to those instances where the learner physically manipulates training hardware. The flight simulator or the sonar team trainer would be cases in point.

The characteristics of a proper student station stem from the designer's desire to maximize positive transfer of training. Warnings against expensive realism for its own sake are repeated here again since the student station is *the* place where costly and unnecessary realism is apt to show up in a simulator.

One helpful approach to the design of the student station is to analyze the role that it plays in the training system. It is most important, in this regard, to take seriously the fact that the student station stands *between* the student and the material which he is expected to learn. Obviously, anything that stands between the learner and the material which must be learned occupies a critical position in the system. The station can be designed to give the learner clear access to the skills and knowledge that he is seeking. Or, it can be designed so as to becloud, obscure, confuse, distort, and disrupt. Looked at from this point of view, the student station serves as a communication link to the material to be learned. It can be a passive link serving as a channel along which information can flow. It can also be an active link which participates in the training by initiating, guiding, and directing communication. It can call the student's attention to things about the material to be learned and to aspects of the student's own behavior which deserve his attention. Training devices of this sort are not just something for the student to practice on. Properly employed, under expert supervision, they involve the learner actively participating in his own improvement. He forsakes the role of a docile container waiting to be filled with knowledge or the role of a piece of raw material waiting to be turned on the lathe of training. Rather, he is an active element in the training system personally seeking through controlled experience to acquire those capabilities which will enable him to assume new responsibilities beyond the training situation. Every detail of the design of the student station should be reviewed to determine the extent to which it helps the student to learn. If something in the training situations isn't part of the subject matter which is itself to be learned (like the location of the throttle), then it should be there to expedite communication between the learner and his subject matter. Otherwise it should be eliminated.

Instructor Stations

Every training complex requires special provisions so that the instructor may exercise his important system functions. Of particular interest from the standpoint of design, are those functions whereby the instructor seeks to apply his own special experience as an instructor to the benefit of the trainee. Too often, instead, the training situation pits the instructor against his trainee. Lecture topics are (perhaps unconsciously) chosen to show off the instructor's knowledge. Tricky problems are sometimes introduced to fool the students. Homemade shortcuts are sometimes taught to discredit "official" procedures thus shaking the student's confidence in the material he is trying to master. These are very human tendencies. It is especially important that the training systems designer doesn't aggravate them. Rather, the instructor's station must be designed in a way that allows the instructor to be a good instructor. And, what are those ways? For one, the instructor needs to have full knowledge of the range of material that is available for his immediate use in the trainer. The designer must arrange for such material to be available. These materials should be easily loaded into the system. If a training system cannot be quickly and accurately configured by the instructor in accordance with the demonstrated needs of the learner, then that system will, sooner or later, become stereotyped and doctrinaire. Or, maybe it will simply not be used at all. From the designer's point of view, this means that easy foolproof mode changes have to be designed.

If film cassettes, cards, or tapes are required, they should be stored convenient to the instructor so he can make changes easily. The system should be capable of being "made ready" in a minimum amount of time.

The instructor's station must also provide selected information to the instructor about the student's performance. This information allows the instructor to analyze the student's progress and diagnose his difficulties. To provide these features, the designer needs to know exactly what information should be collected, at what rate it should be sampled, and how it should be displayed to the instructor or recorded for his use. Some of this information can be obtained from the preliminary analysis. Some of it can be suggested by instructional personnel. Much can be obtained by developing and testing of complete exercises to accompany the training equipment when it is delivered. Although it is seldom done, in many cases it would be possible to simulate a new training simulator and to try out various exercises and information flow arrangements on the simulated version of the simulator prior to the time that the instructional system is actually built.

The possibility of automatic data acquisition should be considered as a basis for assisting the instructor to interpret the student's performance in the simulator. Preset, adjustable alarms, indicating that the student's performance had gone beyond some predetermined bounds, could be designed into the training equipment. Ordinarily, arrangements should be made to allow the instructor to intervene in the training and alter the dynamics of the training system to help the trainee over troublesome spots and to reroute him around areas that can't be dealt with in the context of the current training if such actions appear to be necessary. Similarly, the instructor should be able to skip ahead if the trainee is capable of moving along faster, or is ready to undertake problems of greater complexity, or is ready to work in dirtier environments. In short, the designer should repeatedly ask himself what the instructor must be able to *do* to assist the student maximally, and then arrange for the instructor to be able to do those things with the machine and not in spite of it.

In addition to features which tell the instructor how well each student is doing, there is a need for features which will let the instructor know how well the instructional system, itself, is doing. What proportion of time is the system being operated? Of that time, how much is distributed among the various modes in which the system operates? What is the production rate of trained students as a function of various types of instructional strategy? What system errors have cropped up? And so on. Just as the aircraft pilot requires instruments to fly under certain conditions, so the controller of a large train'ng device needs to receive information regarding the system's current status relative to its goals.

An Example

Specifically, what might be the design consequences of applying the guidelines discussed in this chapter? How wou'd a designer go about specifying the student and instructor stations of, for example, an aircraft trainer? He might begin with a detailed task list such as the one identified in Table 13–2. With this list, augmented by system block times (Table 13–3) and task characteristics (Fig. 13–4), a tentative specification of the training tasks to be performed can be drawn up. Following are some of the tasks that can be trained in a flight simulator:

Ground Checklist: Pre-start, Pre-takeoff, Instrument, Shutdown.

Ground Communications: Route Clearance, Takeoff Clearance.

Maneuvers: Instrument Takeoff and Checklist, Instrument Departure, Enroute Navigation, Instrument Navigation, Instrument Arrival and Checklist, Instrument Approach and Checklist.

The above maneuvers may be made up of some combination of: Straight-and-level, constant-airspeed flight, Level turns, Straight climbs, Straight descents, Climbing turns, Descending turns, Acceleration and deceleration.

Navigation with reference to instruments: Homing, Station passage, Track interception, Track following, Vectoring, Holding, Unusual attitude recovery.

Malfunctions. Some examples: Engine, Radio, Electrical, Hydraulic.

Airborne Communications: Departure, Enroute, Arrival, Approach.

In addition to these flight tasks, a flight simulator can incorporate provisions for training in weapons checklist and delivery, radar navigation, electronic target acquisition, fire control and, with visual simulation, landing, and takeoff.

Consideration is then given to means for evaluating reliably in the trainer the performance of the student pilot doing each of the identified tasks. Referring to the list of tasks above, the ground and airborne checklists can be recorded (see below) as they are performed. Provisions should be made for freezing the problem when an error has been made and performance scores should be available for callup by the instructor at his request. Ground and airborne communications may be partially "canned" and controlled by the computer. But, they must be monitored manually and taped for subsequent review. Flight maneuvers, including navigation, can be recorded and compared by the computer with prerecorded standards. Instrument navigation and malfunctions can be introduced and scored automatically. Finally, scoring of the non-flight tasks can be placed under computer control.

Effective scoring, be it manual or automatic, is dependent on the performance measurements selected.

In this example the task performance measures might properly include:*

Checklists (ground & airborne).	Number of errors.
	Total time.
	Number of omissions.
	Response latencies (length of time from completion of one item to the initiation of the next).
Communications (ground & airborne).	Response latencies (time from end of instructor-initiated message and initiation of student response).
	Content analysis—e.g., correspondence with required voice procedures and the accuracy and completeness of message.
	Amount of communication number of messages individual message length.
Flight maneuvers.	Integrated error scores—distance from prescribed pattern integrated over time which can be computed in various ways (e.g., integrated error squared and/or visual comparison of the tracks of the generated and prescribed maneuver).
	Frequency of out of tolerance excursions.
Instrument navigation.	Sample of instrument readings compared with expected readings.
	Integrated error.
	Time required and accuracy in reaching a designated position.
	Visual scan patterns of trainee in instrument flight.
Malfunctions.	Time to respond to malfunction indication.
	Accuracy and completeness of corrective sequence.
	Appropriateness (including timing) of action taken (e.g., abandon aircraft).
Non-flight tasks.	Accuracy of procedures.
Weapons delivery.	Miss distance.
	Appropriateness of tactics.
Radar detection and tracking.	Time between entry of target and detection.
	Number of missed targets.
	Number of false targets.
	Time to lock-on.

All of these measurements must be evaluated for their reliability (common scores should be obtained under common conditions), and their sensitivity (changes in performance with practice should be reflected in changing scores). These characteristics can be established by pilot experiment and by subsequent experience with the flight simulator. Establishing the relevance and validity of the measurements depends on

* For a more detailed discussion of performance measures, see Bowen et al, 1966.

prior experience and expert judgment at the design stage. Transfer experiments with the flight simulator when in use are necessary to verify these decisions.

Once a detailed list of training tasks (such as the above) has been assembled, they should be ordered (e.g. in terms of their design implications), and those for which training device provision is to be made should be identified. Detailed learning blocks must then be prepared for each selected task to reveal the interactions among student and instructor personnel. Figure 13–6 (Flexman et al, 1968) is an example of one such instructional block—Radio Voice Procedures.

This block diagram describes a one-hour training session (although time may vary among students). Three levels of instructor activity (operator, line instructor, and monitor) are specified. The chart reflects certain preliminary design decisions (e.g., the use of closed-circuit television to monitor selected student activities) and explicitly identifies periods of peak instructor load. Further, it identifies the kinds of instructor activity (e.g., discuss specific problems) expected at given times. The designer should have available to him charts of this sort as he designs the trainer. The number and nature of these blocks are based on the number and nature of tasks to be taught. In this example, over twenty such training block diagrams are required.

The learning block pictured in Figure 13–6 was selected to demonstrate a "worst case" of instructor loading. It is an example of manual instruction. However, other instructional blocks can be automated. For example: (a) orientation to the trainer and the trainee's learning task can be accomplished with audio and video tapes using the TV monitor in the student's station; (b) design provisions (e.g., in the software and through sensing of the student manipulation of various panel controls and displays) can enable the student to learn his instrument check procedures without instructor intervention. The simulator can sense the student's actions and compare them with the proper sequence and the maximum allowable time. The results of his attempts can be presented to the student on his TV monitor or other cockpit display specifically provided for this purpose. The same display can direct the student to repeat selected portions of the sequence he is learning, based on his earlier mistakes. In this fashion the student is learning procedures under automatic adaptive control. The instructor will be furnished with a complete hard copy summary of the student's training when some given criterion of time and accuracy has been met.

The trainee, in like manner, can be automatically cycled through several specially controlled diagnostic instructional blocks, hard-copy results of which the instructor can consult in determining the entry level of the student. In this way a training program can be tailored for each student by the instructor directly or by means of a computer program which assembles a training program based on the student's demonstrated entry level. In addition, the student can be evaluated automatically during the course of training and the training program suitably revised in terms of his changing strengths and weaknesses.

Two important emerging principles that should govern training device station design are (a) the student station should be a modified version of the operator's position it represents, and (b) the instructor's station need not resemble the student's position at all. Figure 13–7 (Clausen et al., 1968) is an artist's conception of an instructor station for a flight simulator. Notice how completely it differs in appearance from a pilot's trainee station.

The first principle provides for incorporating in the trainee's position such features as:

1. The sequential illumination of instruments to guide the trainee through a fixed (or variable) optimum scan sequence.
2. A TV camera for viewing and recording the trainee's gross behavior.
3. A recording of the eye movements of the trainee to aid in improving search behavior.
4. A CRT (and/or an audio device) for quickly and automatically displaying information (e.g., feedback or knowledge of results) to the student about the adequacy of his performance.
5. Equipment to measure aspects of the student's physiological state (e.g., heart rate) as an indirect indication of his state of attention.
6. Rendering "touch sensitive" those instruments which must be visually inspected in a flight check. The trainee is instructed to touch each

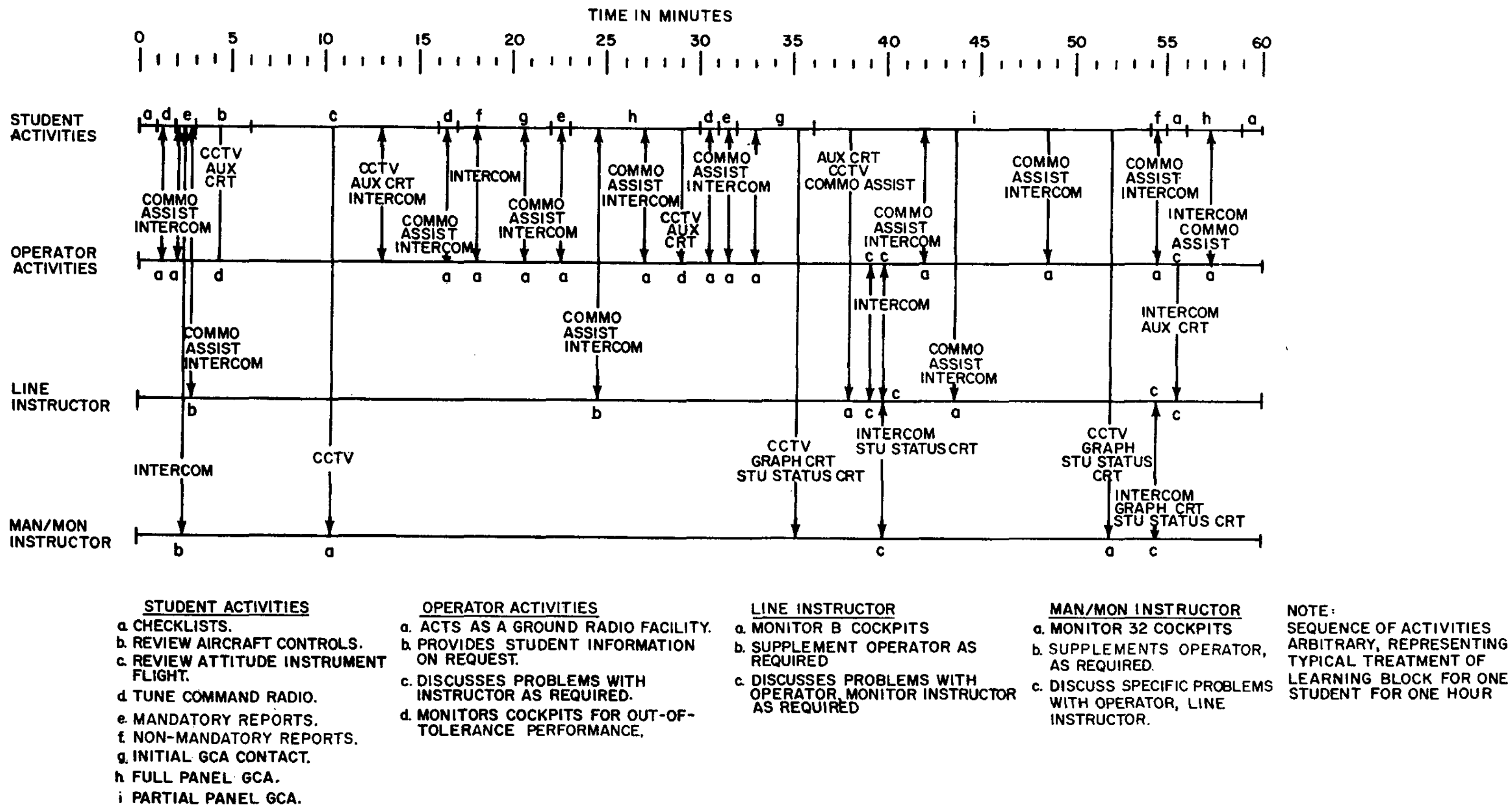


FIGURE 13–6. Radio voice procedures learning block (Flexman et al., 1968).

FIGURE 13–7. Artist's concept of an instructor control layout (Clauson et al., 1968).

instrument as he inspects it, thus allowing the computer to record automatically the completeness and accuracy of the procedure.

7. A display which may be the CRT described above which advises the trainee of impending out-of-tolerance conditions, thus implementing features of cuing and guided training.

The second principle suggests that, since the task of the instructor is different from that of the student, it can best be accomplished with displays and controls which are also different from those of the student. The instructor station should contain such provisions as:

1. Computer-generated displays which can sample the state of the system in a general fashion (e.g., ground tracks) and in detail (e.g., an instrument reading). On these displays the instructor can call up information in various combinations and at times of his choosing.

2. Pre-programmed automatic modes which sample data according to expected instructor need or only during out-of-tolerance conditions.

3. Computer-controlled slides which contain data about the training problem (e.g., the rules of a decision game) which the instructor needs only at particular times.

4. TV monitor for viewing trainee activity not easily observed otherwise.

5. Adaptive training scores and conditions summarized by the computer for rapid assessment.

6. Controls for sequencing performance information from a number of trainee stations on a given display thus contributing to an increase in the ratio of students to instructors.

Tasks, the learning blocks they lead to, and a tentative instructor/student interface have been selected. An instructor station layout must now be designed.

Figure 13–8 illustrates such a layout for the instructor station of a flight simulator. It includes:

Macro (graphic plotter) displays. These present computer-generated general information on the status of the trainee and the training problem. There is one for each trainee station monitored. In the layout shown it is assumed that the instructor will be monitor'ng four trainee positions, thus there are four macro displays. At the discretion of the instructor the display will present, among other items:

1. Trainee identification information.

2. Status of training, that is, the length of time the trainee has been in the trainer or the extent of progress that he has made.

3. Data on the trainee's present performance.

4. Ground-track plots with radio navigation facilities, terrain features, and other relevant background information added symbolically.

5. Out-of-tolerance information (e.g., off course).

6. A signal that the trainee has requested assistance, or that he is calling the ground radio station.

7. Advance information on the occurrence of preprogrammed malfunctions.

8. Indication that an automatic briefing or demonstration session is in progress.

9. Trainee's adaptive score.

Micro (cockpit instrument) display. (This single display serves the four trainee stations. The instructor calls up information on one

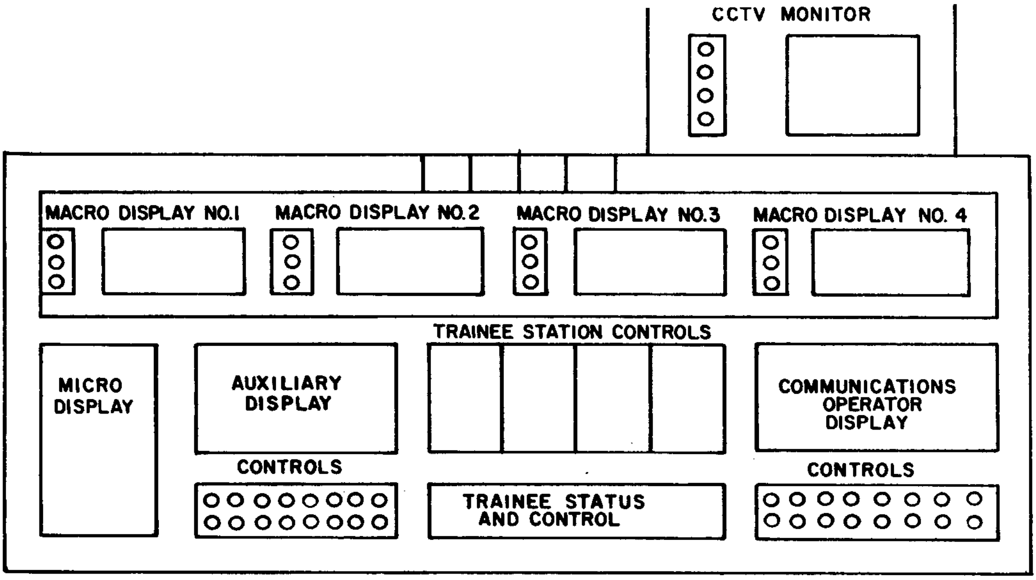


FIGURE 13–8. A preliminary instructor station layout for a flight simulator (Naval Training Device Center, 1968).

trainee position at a time.) The display presents detailed, computer-generated: (a) trainee instrument and meter readings; and (b) graphic presentations of the status of a variety of instruments. These graphic presentations include readily identified out-of-tolerance conditions and boundaries.

An auxiliary display panel which is a CRT with a series of pushbuttons which allows the instructor to call up individual pages of information that have either been preprepared or developed during the training session. This display includes auxiliary controls such as a light pen for manipulating the displayed information. The keyboard and light pen associated with the auxiliary display are used by the instructor to communicate with the computer for the purpose of modifying conditions of training such as turbulence or wind direction during the manual training mode or to select predetermined initial conditions for given training sessions.

A communications operator display. This display and its associated controls are used by an associate instructor who acts as a ground collector and in other ways simulates air traffic control operations. Provisions are made to record and play back all of the simulated radio transmissions.

A closed-circuit television monitor. This display allows the instructor to monitor certain of the activities of the students as seen by the TV camera in each trainee station. In addition, a video tape recording of selected portions of any trainee's activity will be available.

Limited portions of the traffic control communications operation (e.g., Ground Controlled Approach, GCA) can be automated by using computer-generated speech. For example. deviations from the flight path will automatically cause a standardized brief voice message to be sent to the trainee.

Adaptive modes of instruction are implemented in this flight trainer example by providing modifications in the aerodynamic simulation of a number of parameters as an immediate and automatic function of the changing tracking skill of the trainee. There are a number of parameters which can be varied as a function of trainee performance. The present state-of-the-art does not permit a precise specification of which parameters and values are most effective for which training situations. It is therefore necessary for the designer to specify a number of parameters and a range of values for each from which the operators of the trainer can select, through use, the most effective. For example, the system gain, damping, and frequency can be modified in various combinations (Matheny and Norman, 1968). Other characteristics of the system such as buffeting, control order, lags, and flight-path tolerances can be altered as the student progresses. The amount of buffeting encountered by the student may be a direct result of his increasing skill so that as he improves in controlling the flight vehicle, increased amounts of buffeting are supplied. If his performance deteriorates, the buffeting is reduced.

Another adaptive variable that might be included is the systematic (and automatic) incorporation, in addition to the main task of controlling the aircraft, of tasks such as the detection of targets, dead reckoning, dealing with radio contacts, and so on. Kelley and Wargo (1968) describe a number of adaptive variables and considerations that are directly applicable to this example.

The key of effective student/instructor station design, once the initial task descriptions and analyses are available, includes:

1. A listing of tasks to be trained,
2. The identification of learning blocks including student and instructor staff interaction,
3. Layouts of the stations based on considerations discussed above,
4. A detailed analyses of the methods for recording and analyzing trainee performance,
5. Provisions for automating as much of the training program as possible,
6. Establishment of a suitable ratio of students to instructors,
7. A systematic program for arranging (adaptively where possible) training blocks into a total program.

13.4 Designing the Utilization Guide

At one time the designer of training equipment had no responsibility for preparing the utilization guide to go along with his equipment. Or, if one was prepared, it would not presume to tell the future instructor how to instruct, but simply how to turn the equipment "On" or "Off." But the introduction of the systems approach has changed all of that. Competent overall system design concerns itself with the training use of the equipment after it has been constructed and installed. This section deals with some of the general considerations associated with the *use* of training equipment.

Figure 13-9 illustrates the importance of sound utilization procedures in the conduct of training with training devices. Using the same trainer, two equal groups were practiced daily for six days. The group (represented by a solid line on the graph) with planned utilization procedures improved significantly in their six days of practice, while the group (represented by a dashed line on the graph) with a disorganized training program actually became worse with practice within the same time period.

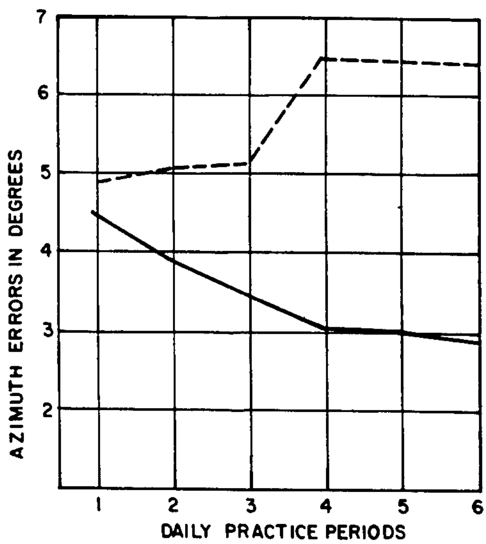


FIGURE 13-9. Learning curves for two groups of subjects trained for six days on the same trainer (adapted from Wolfle, 1945).

13.4.1 Establish Practice Procedures

The final step in designing a training system is the specification of practice procedures. Factors that contribute to increasing the effectiveness and reducing the cost of practice should be taken into consideration.

Distribution of Practice and Rest

Schedules for rest and practice must be such that they do not depress performance, or extend training beyond the minimum necessary to develop an appropriate degree of proficiency.

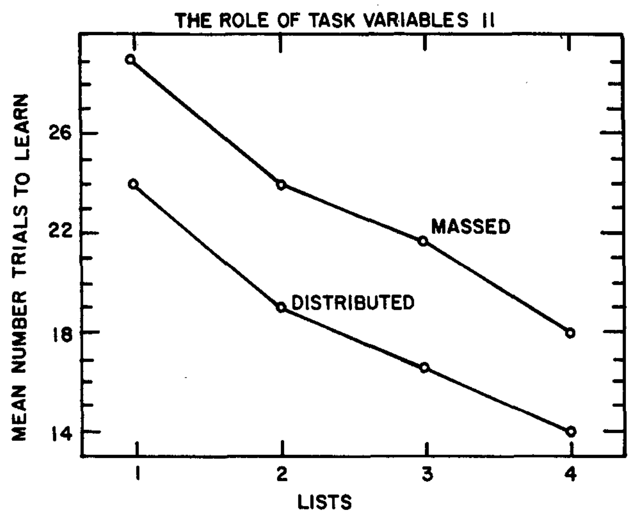


FIGURE 13–10. The effect of stage of practice on learning by massed and by distributed practice (adapted from Underwood, 1951a).

If the student is given little or no rest, his rate of improvement will be lowered. This will mean he will have to practice much longer than otherwise necessary to meet the appropriate standard of proficiency. On the other hand, if rest periods are too frequent or excessively long, time will be wasted.

A trial is defined as a continuous practice session, whether the task is practiced once or several times (in the case of short tasks). Typically, a trial will be followed by a short rest. (See Figure 13–10.)

While it is difficult to prescribe an optimum practice schedule, the following suggestions should help prevent procedural errors.

When trials are short,—a few minutes or less—assuming minimal demands on physical strength of the individual, a rest of 30 to 45 seconds should be allowed between trials, followed by a ten-minute break after every ten trials. If trials are longer, than a rest period of 35 to 45 seconds should be allowed between trials, followed by a ten-minute break after every few trials.

If performance on the task is being measured to determine when a student has attained satisfactory proficiency, measurements may be taken immediately after the ten-minute break.

Verbalization during practice. In procedural or problem-solving task practice, advantages of having the student "talk his way through the task" while practicing have been found (Esper and Lovaas, 1962; Gagne and Smith, 1962; Neumann, 1960; and Ray, 1957). This is one of the features of the LOCKON method of instruction (Woolman, 1960) developed by HumRRO for the training of Nike missile operators.

Pacing

Some tasks are externally paced, that is, the actions must be performed within time limits established by a cue sequence that is also timed. If the task is going to be externally paced, practice should also be externally paced. Research results show some advantage for groups practicing on externally paced tasks when they switch to self-paced conditions (Adams, 1954; Anderson et al., 1955; Nystrom et al., 1955 and 1956).

Overlearning

If retention of a skill is particularly important, the student should overlearn. Overlearning is additional practice after performance standards have been met. Important tasks should be overlearned, especially when there is little expectation for frequent review after initial learning. In addition to aiding retention, overlearning tends to prevent skill deterioration under stressful conditions. such as combat.

Mental practice. When individuals think their way through a task, perhaps making appropriate motions even though normal job cues are not present, they are performing mental practice. Shadow-boxing is an example of such practice. In one situation, the mental practice of the one-hand basketball foul shot was nearly as effective as physical practice (Clark, 1960).

Preparation for practice. A situation contributing to inefficiency frequently arises in practical exercises. This is the requirement that the student spend time preparing his materials before practice: breadboards of electrical components to be assembled, laboratory equipment to be set up, equipment to be checked out to supply, etc. This preparatory activity is frequently wasted time, and should be reduced to a minimum. If equipment cannot be set up in advance, clear instructions should permit quick preparation.

13.4.2 Special Methods of Use

Opportunities exist for building training simulators that will permit special methods of utilization. One attractive possibility is to design the simulator so that unusually good or unusually bad performances can be demonstrated easily to the student. The instructor should be able to set the equipment so that it would go through its paces in a fault-free manner for the student to see. It also should be able to demonstrate improper operation, common mistakes, and critical states to be avoided.

A closely related possibility is the use of video tape to record the student's performance so that it can serve both the student himself and (hopefully with identity withheld) future generations of students. It is particularly important for the student to have an opportunity to see how things look and how they work when all goes well. As a beginner, he is not likely to see this in his own performance.

A third important, though seldom employed, mode of trainer operation involves both diagnosis and drill. In this, an individual trainee (or a group) engages in a training practice sequence. Diagnosis of his performance reveals errors at selected points within the sequence. The trainer is designed to allow for the presentation of only those parts of the training task on which the trainee needs practice. For example, suppose a training problem involves a sequence during which the trainee must distribute target data in a specified way to three different stations. The trainee may have trouble making the proper distribution. The properly designed system permits the problem to start where the trainee is having trouble and gives him repeated opportunities to practice and be scored on that particular segment of the problem.

In complex trainers, which usually contain a number of trainee positions devoted to different tasks, a special problem can arise. For example, in a two-place fighter aircraft simulation, the objective may be to train the pilot and the radar operator to function smoothly as a team. At the outset of training both members of this team may be new to the situation for which the training is to be conducted. Since successful accomplishment of the task is dependent on the joint successful operation of both members of the team, it is unwise to place both novices in the trainer together. To do so would allow each to contribute unpredictable variance during the course of practice. There are several ways to solve this problem. One is to put one member of the team in the trainer and have the second trainee position taken over by an experienced instructor who is skilled in the manning of that position. The instructor will then be able to control the input and output activity characteristic of the position in a way that will optimize the rate of learning of the trainee. The procedure can then be reversed for training the other member of the two-man student team. In other cases it may be more efficient to simulate the activities of each station so that the member undergoing initial training can receive controlled information from the other position which is appropriate to his level of skill. In somewhat more complex situations it may be desirable to simulate the outputs from a series of operational positions which are not directly involved with the group or individuals to be trained, but whose activities influence the requirements imposed on the trainees. If one extends this mode of operation to a multi-position training complex, the design implications become significant since it may be necessary to simulate subteams or large blocks of individual operators during certain stages of training.

It is desirable, at some point in the course of training, to assemble the team of trainees and to permit free play among them in the course of solving the training problem. This interaction provides training for the unpredictable and the inaccurate inputs sometimes fed from one team member to another in an operational situation.

13.4.3 The "Preprograming" of Training

A significant problem in the design of trainers pertains to preprograming versus the free play of the training sequence. In addition to the usual question of training effectiveness, the problems in this area include large cost differences, matters of stimulus or problem fidelity, and the inflexibility of a training sequence that is completely determined before the trainee begins the exercise. An example of this difficulty is encountered in presenting auditory signals. In one instance the signals are recorded, edited, and played into the training system. In another case they are synthesized from an analysis of the physical characteristics of operational sounds. The advantages of synthesizing the sounds derive from the fact that they can be varied in location and in other respects as a function of decisions made by the instructor or the activity of those being trained. In general, a tradeoff is necessary between the high fidelity of the preprogramed, taped, stimulus material and the flexibility associated with the unprogramed, synthesized stimuli. The differences in cost, fidelity, and flexibility between programmed and unprogrammed presentations have been reduced somewhat by advances in engineering. In particular, the capability of mixing such presentations has been increased through the use of computers. When synthesized stimulus material is used because of its free-play characteristics, it is recommended that the designer also make provisions in the trainer for incorporating a programed presentation with high-fidelity characteristics. This combination of programed and unprogramed presentations can, under proper utilization schedules, provide the best of both worlds.

A similar problem exists in relation to training simulation of visual events. Simulation of the visual world, be it the view from an aircraft, a ship or a tank, presents momentous technological problems. The designer must, with proper help, determine what is perceived visually in the operational situation and how what is seen affects behavior to be learned. Finally, he must estimate how much of what is seen must be simulated for positive transfer between training and subsequent on-the-job performance.

13.4.4 Consumer Acceptance

Among other things, full and effective utilization of training equipment depends upon its acceptance by training personnel. In many instances, two hostile camps seem to spring up. The "design people" accuse the "field people" of under-utilizing their products. The field accuses the developing agency of over-design. Needless to say, instances of both types are not altogether unknown in the history of training system design and utilization. Certainly no attempt will be made here to affix the blame for those cases where users and designers end up by pointing accusing fingers at each other. Rather, let us simply agree that such a state of affairs is sufficiently undesirable that substantial efforts are warranted to avoid them. For it is clearly wasteful to design features which will not be used or to fail to use all of the capabilities of training systems once they are built. A widespread, costly example of this problem may be found in flight simulators. Expensive features of flight simulators are often not used adequately in the field (Smode et al., 1966). Rather, they are used as simple "procedures trainers." When this happens, instructors often ignore capabilities for conducting navigational exercises, instrument flight, pilotage, and control skills which are built into these simulators at great expense. Whether the flight simulator should, or could, be used to develop or maintain such skills is entirely academic in many cases since instructional staffs simply do not even try to use these features. When not used, the argument (frequently heard) that some particular multimillion-dollar trainer could work just as well if reduced to a one-thousand dollar "procedures trainer" is correct. This points up the importance of identifying attitudes of the potential user at the time a device is designed. Improved designer-user communication at this point will reduce the number of expensive misunderstandings. One mechanism for improving communication is for the two groups to work together designing instructional material to be played on the new device. Continued efforts of that sort assure that the new device will come equipped with appropriate support material that reflects the interests and concerns of both the designers of the instructional system and the instructors who are essential components of the system. It will also provide a basis for continued dialogue between the designers and the instructors during the utilization period, thus providing extremely important follow-up information which is so necessary if evolutionary improvements are to be made.

13.4.5 System Evaluation

A widespread shortcoming in the design and employment of training equipment is the failure to provide for an adequate evaluation of training systems from a training point of view. Three kinds of questions can be asked about a piece of training equipment in terms of its training usefulness: first, does an individual, who has experience in the controlled environment of a training device, learn? That is to say, is some unit of behavior measurably improved during the course of his training? For example, suppose the purpose of a trainer is to teach an individual to control a vehicle. Is he measurably better able to perform this task at the end of some period of training than he was at the beginning? If so, he has learned. Secondly, how much and how long will he retain what he has learned? Information can be obtained by returning the student to the training situation and measuring the difference between his performance upon his return and his performance when he left the trainer. Both learning and retention are important measures of training efficacy, but they are not important in and of themselves. However, if a training system gets good marks with respect to rate of improvement and proportion retained, it is now appropriate to raise the third question. It is this third question which is the "proof of the pudding." Stated simply, to what extent, is what has been learned and retained in the trainer reflected in some designated operational situation? Positive transfer will not always occur, and in some cases negative transfer will make the trainee worse in the operational situation (in some way) than he would have been with no training. In general, however, there is a gross positive relationship between the amount learned in the trainer and the amount subsequently retained and transferred to some other situation. This obscures detailed deficiencies which could be corrected, if detected. It is useful to have "fine-grain" measures of learning and retention and to design a capability for obtaining these measures into the trainer.

Overlearning is a related and important concept which refers to continued practice by a trainee after he has reached some predetermined criteria of performance. For example, suppose the operational situation demands that a controller maintain a vehicle on course with an average error of 4 or 6 ft. The trainee may be said to have learned his task in the trainer when he is able to maintain the vehicle track within these limits. If practice in the trainer continues beyond this point, the additional skill acquired is referred to as overlearning. In general, the advantages of overlearning take at least two forms: first, the trainee is likely to retain more for a longer period of time; second, he is less likely to experience performance disruption during conditions of overload and stress in the operational situation.

Evaluation, then, concerns itself with measures of human learning and transfer. The problem of transfer takes the designer beyond the immediate confines of a piece of training equipment. It requires him to arrange for comparisons between student performance on the trainer with subsequent performance. This comparison is more expeditious in those cases where planning has forged specific prior agreement regarding the measurements taken and the comparisons to be made.

Traditional measures include time and error scores. In a discrete procedural task, time refers to the length of time from the initiation of a procedure to its conclusion. Errors are the number of steps incorrectly taken, omitted, or performed out of sequence. In a fixed procedure, the correctness of which is fully known, this is a straightforward measurement problem, which can be dealt with in several ways. A primitive method that tends to be unreliable is to have an instructor time the procedure and count the errors. Alternatively, the training device can provide for automatic timing and error counting. Smode (1966) lists a variety of recording devices which can be used for these purposes.

In large team trainers where many individuals are performing a variety of tasks, decisions concerning what to measure are much more complicated. It is customary to provide a measure of system performance in large complex trainers. In a tactical trainer where the team task is to acquire and process information and to select various response alternatives (e.g. in an Air Defense system), it is customary to define performance in terms of the number of enemy aircraft intercepted, the number of kills, and other such measures of systems performance. These measures alone are generally too gross from a training point of view because they are influenced by numerous man/machine considerations many of which are not under the direct control of the individuals being trained. In the air defense example, an individual or subgroup makes an incorrect identification, the tracking team tracks an incorrect target, and the fire-control system intercepts an unassigned target. All of these incorrect activities can, under a given set of circumstances, add up to a successful interception of a hostile target, and, of course, the reverse can happen. Sets of correct actions occurring along with one or more incorrect ones can also yield an unsuccessful outcome in system performance terms. Both the long time lag between execution and evaluation and the occurrence of compensating errors compound performance measurement. Research evidence indicates that when teams are being trained, desirable performance can be reduced or eliminated systematically by providing group feedback which is not representative of individual activity within the group. (Klaus and Glaser, 1968.) Thus, the designer must provide measurements of individual and subgroup activity which can be identified directly with the trainee in question.

For example, in an Air Defense trainer, performance of the individual operators, fliers, decision-makers, controllers and the like should be available both with respect to individuals and subteams. In any given case, what is to be measured has to be determined from the task analysis that precedes the system design. It is inadequate to rely on the instructional staff to observe and assess all of the group performance that is manifest in a complex training situation. On the other hand, the complete recording of all the activity, directly observable and otherwise, of the teams and individuals of large trainers yields data, the quantities of which are difficult to cope with. In addition, much of this data is of questionable or unknown value in assessing learning. Provisions should be made for sampling the data and for relating individual and group performance measures. Nonetheless, provisions for the reasonably complete recording of training activity should be made with the expectation that the trial use of the trainer will allow the instructional staff to develop appropriate amounts and kinds of activity to record and analyze.

The designer's job is not complete with the installation and check-out of a major piece of training equipment. He must maintain a continuing relationship with the trainer for some period of time. Since most large complex trainers are one or two of a kind, he will seldom have earlier copies of the same trainer on which to rely. Each major training system is a research and development effort in its own right. The designer, in the preparation of specifications and in the execution of the design, should anticipate a role in the trainer's evaluation. In fact, systems testing should include full-scale operation of the trainer over a reasonably long period of time with provisions for modifications, not only in the use patterns, but in the actual hardware configuration itself. This step can do more than any other to improve the quality of training equipment.

13.5 Conclusion

The use of training equipment has won increasing acceptance. Its use is becoming widespread and is relied on to do more of the training job than it did in the past. That training devices have been effective is clear. The authors suggest nonetheless that training equipment can be significantly improved in both design and employment through adopting a systems approach which will integrate engineering and behavioral science data. Recent developments in both these disciplines afford new opportunities for improvement, while the new world of educational technology is placing additional demands on training equipment. The following points of view if reflected in design and employment can help to meet these demands and take advantage of these opportunities.

Training and operational equipment, although they have much in common, also differ in important ways. This chapter has sought to emphasize these differences. They derive from differing objectives.

There are few simple rules which can be used across the range of design and use problems. At the present time most major training device design situations must be treated individually. Thus this chapter provides guidelines rather than prescriptions. The authors believe that if the guidance set forth here is followed, design solutions will be obtained which are different from (and superior to) those which result solely from an attempt to represent the operational environment physically.

A key step in the design process is to obtain specific statements concerning the behavioral objectives to be met by the trainer. What should the trainee (or team of trainees) be able to do,—at what level of accuracy, in what time frame, and under what circumstances—when he completes his training? All design should then be aimed at developing a trainer which will effectively meet these objectives. Concern should be with what the trainer *can do* rather than with what it *will be.*

The points of view, technical information, and behavioral data necessary to the design of effective training equipment are seldom available in a single person or discipline. It is important, therefore, for the designers to form a team which can bring this variety of skills to bear on the design problem.

The bibliographic section which follows cites a substantial portion of the literature related to the design and employment of training systems. A review of it will detail further the nature of the task faced by those who design training equipment.

References

Adams, J. A. The effect of pacing on the learning of a psychomotor response. *Journal of Experimental Psychology*, 1954, **47** (2), 101–105.

Anderson, N. H., Kresse, F. H., & Grant, D. A. Effect of rate of automatically-paced training in a multidimensional psychomotor task. *Journal of Experimental Psychology*, 1955, **49** (4), 231–236.

Bowen, H. M., Bishop, E. W., Promisel, D., & Robins, J. E. Study, assessment of pilot proficiency. Technical Report 1614–1, August 1966 Naval Training Device Center.

Brooks, F. A., Jr. Operational Sequence Diagrams. *IRE Transactions of the Professional Group on Human Factors in Electronics, HFE-1*, 1960, 33–34.

Chenzoff, A. P., & Folley, J. D., Jr. Guidelines for training situation analysis (TSA). Technical Report No. 1218–4, July 1965, Naval Training Device Center, Orlando, Fla.

Clark, L. V. Effect of mental practice on the development of a certain motor skill. *Res. Quart. Amer. Ass. Hlth. Phys. Educ. Recr. 31*. No. 4, 1960, 560–569.

Clausen, J. T., Curtin, J. G., Egler, J. F., Johnson, G. M., Nelson, T. M., Szczuka, A. E. & Voss, W. J. Concept formulation report for synthetic flight training system. Technical Report No. 68–C–0108–1, April 1968, Naval Training Device Center, Orlando, Fla.

Esper, E. A., & Lovaas, O. I. The effect of verbal symbolization on the serial position curve in motor perceptual maze learning. *Journal of General Psychology*, 1962, **66**, 47–51.

Fleishman, E. A. The description and prediction of perceptual-motor skill learning. In R. Glaser (Ed.), *Training research and education.* Pittsburgh: University of Pittsburgh Press, 1962.

Fleishman, E. A., & Hempel, W. E., Jr. The relation between abilities and improvement with practice in a visual discrimination reaction task. *Journal of Experimental Psychology*, 1955, **49**, 301–312.

Flexman, R. E., Jameson, W. T., Walsh, J. M., & Cohen, E. Concept formulation report for synthetic flight training system. Technical Report No. 68–C–0106, April 1968, Naval Training Device Center, Orlando, Fla.

Gagne, R. M., & Smith, E. C., Jr. A study of the effects of verbalization on problem solving. *Journal of Experimental Psychology*, 1962, **63** (1), 12–18.

Gilbert, T. F. Mathetics: The technology of education. *Journal of Mathetics*, 1962, **1**, 7–73.

Hickey, A. E. Computer assisted instruction: A survey of the literature. (3rd ed.) Technical Report No. 8, Contract No. N00014–68–C–0236. October 1968, Office of Naval Research, Washington, D.C.

Katcher, A., & Hunter, M. How to lead a debriefing—A guide for senior directors. Report No. TM–132, November 1957, System Development Corporation, Santa Monica, Calif.

Kelley, C. R., & Wargo, M. J. Adaptive techniques for synthetic flight training systems. Technical Memorandum, August 1968, Naval Training Device Center, Orlando, Fla.

Kinkade, R. G. Unpublished original data, 1969.

Klaus, D. J., & Glaser, R. Increasing team proficiency through training #8 Final Summary Report, AIR–B1–6/68–FR. May 1968. American Institutes for Research, Pittsburgh, Pa.

Lavery, J. J. Retention of simple motor skills as a function of type of knowledge of results. *Canadian Journal of Psychology*, 1962, **16**, 300–311.

Matheny, W. G., & Norman, D. A. The effective time constant in tracking behavior. Technical Report No. 67-C-0034–3, February 1968, Naval Training Device Center, Orlando, Fla.

Naval Training Device Center, Orlando, Florida, unpublished data, 1968.

Neumann, E. Frequency and usefulness of verbal and nonverbal methods in the learning and transfer of a paired-associate serial motor task. *Journal of Experimental Psychology*, 1960, **60** (2), 103–110.

Nystrom, C. O., Morin, R. E., & Grant, D. A. Effects of amount, rate and stage of automatically-paced training on self-paced performance. *Journal of Experimental Psychology* 1955, **49** (4), 225–230.

Nystrom, C. O., Morin, R. E., & Grant, D. A. Transfer effects between automatically-paced training schedules in a perceptual-motor task. *Journal of General Psychology*, 1956, **55**, 9–17.

Pask, G. Electronic keyboard teaching machines. *Education and Communication*, 1958, **24**, 336–348.

Ray, W. S. Verbal compared with manipulative solution of an apparatus-problem. *American Journal of Psychology*, 1957, **70**, 289–290.

Schrenk, L. R. Personal communication, 1969.

Skinner, B. F. The science of learning and the art of teaching. *Harvard Educational Review*, 1954, **24**, 99–113.

Smith, K. U., & Smith, M. F. *Cybernetic principles of learning and educational design.* New York: Holt, Rinehart & Winston, 1966.

Smode, A. F. Some engineering developments pertinent to pilot training programs (preliminary draft), September 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

Smode, A. F., Hall, E. R., & Meyer, D. E. An assessment of research relevant to pilot training. Report No. AMRL–TR–66–197, November 1966, Aerospace Medical Research Labs., WPAFB, Ohio.

Underwood, B. J. Studies of distributed practice: III. The influence of stage of practice in serial learning. *Journal of Experimental Psychology*, 1951a, **42**, 291–295.

Wolfle, D. The use and design of synthetic trainers for military training. Report No. OSRD–5246, July 1945, Applied Psychology Panel, NDRC.

Woolman, M. On-site training of guided missile operators. Technical Report No. HumRRO–64, August 1960, Human Resources Research Office, Alexandria, Va.

Selected Readings

Chenzoff, A. P., & Folley, J. D., Jr. Guidelines for training situation analysis (TSA). Technical Report No. 1218-4, July 1965, Naval Training Device Center, Orlando, Fla.

Holding, D. H. *Principles of training.* Oxford: Permagon Press, Ltd., 1965.

Kurtz, A. K., & Smith, M. C. Annotated bibliography of human factors laboratory reports (1945–1968). Technical Report No. IH–158, Feb. 69. AD 686–174. Naval Training Device Center, Orlando, Fla.

Miller, R. B. Psychological considerations for the design of training equipment. Report No. TR 54–563, December 1954, Wright Air Development Center, WPAFB, Ohio.

Smith, K. U., & Smith, M. F. *Cybernetic principles of learning and educational design.* New York: Holt, Rinehart and Winston, 1966.

Smode, A. F., Gruber, A., & Ely, J. H. Human factors technology in the design of simulators for operator training. Technical Report No. 1103-1 or NAVEXOS P–2766, December 1963. Naval Training Device Center, Orlando, Fla.

Chapter 14

Training Device Design

Robert G. Kinkade
George R. Wheaton

American Institutes for Research
Washington Office

With the development of complex systems that require highly skilled operators, the importance of training devices is increasing. Future developments in appliances, transportation systems, military systems, etc. are likely to require even higher skill levels. Acquisition of these skills is facilitated by effective design and use of training devices. This chapter discusses the nature of training devices, presents design recommendations for training aids and training equipment, provides guidelines concerning the fidelity of simulation required in training devices, and presents recommendations concerning the design of student and instructor stations.

CONTENTS

This chapter was reviewed by M. P. Ranc.

14. Training Device Design

14.1 Training Devices and Training Systems

A training device is only one part of a total system which guides, facilitates, and reinforces a learning process. The design of a particular training device must be viewed within the larger context of this total training system.

Inputs to the training system are students who lack specific types of knowledge or skill. As these students progress through the system they are exposed to a variety of new situations and are encouraged to develop and practice skilled responses. Exposure is provided by a variety of training devices which supply part of the conditions needed to guide and reinforce the learning experience. The training system's outputs are graduates who possess new knowledge and skills and who are expected to transfer these assets to an operational situation. If the graduates can perform at specified criterion levels on the job for which they have been trained, then the training system has been successful.

The training system factors that may influence the design of a training device are shown in Figure 14–1. These include: the training objectives, student population, other training devices, practice procedures, the training program, and evaluation requirements.

Training Objectives

Training objectives determine a number of design parameters. For example, if the graduates are expected to be truly proficient, provision must be made for extensive and intensive training on a wide spectrum of tasks in realistic situations. On the other hand, if the trainees are expected simply to become familiar with the general task situation, only limited exposure, practice, and personal experience may be required. To be most useful in the design of training devices, objectives should be (adapted from Eckstrand, 1967):

Relevant. Each training objective should be stated in terms of the precise knowledge and skills required for adequate job performance;

Complete. The objectives should account for all performance outputs under all anticipated situations; and

Measurable. The objectives should be stated so as to suggest ways for determining that they have been achieved. If the training objectives meet these criteria, meaningful inferences concerning training device design can be made from them.

Student Population

Student population characteristics should be reviewed for design implications. Student-population characteristics most likely to affect the design of training devices include:

Range of aptitudes and abilities. What are the capacities for learning or the natural abilities of potential students?

Range and level of prior education and experience. How difficult can the material be? At what level can it start? How much practice will be required? Are other training programs prerequisite to the one being entered?

Attitude toward the training and learning process. What special techniques will be required to attract and maintain student interest?

Anthropometric data. Will the height, weight, arm reach, leg reach, vision, hearing, etc., of students require special attention in the layout of student areas and in the design of the training device?

The design implications of anthropometric measures are obvious. Some student populations are selected on the basis of their height or weight (e.g., fighter pilots, submarine crew members, special forces, etc.), and these factors

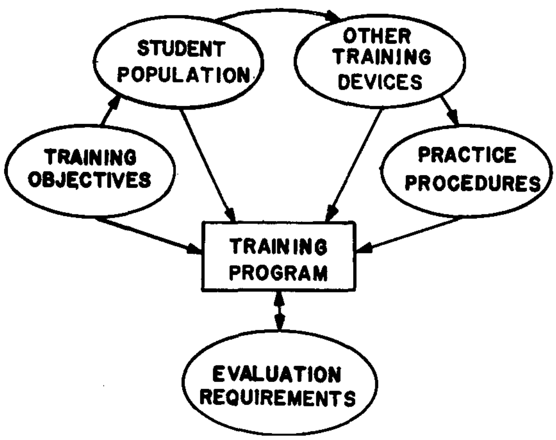


FIGURE 14–1. Training systems factors that could influence the design of a training device.

should be taken into account in designing the physical configuration of the training device. Educational experience and aptitude scores will help to determine the level of task specificity and operating simplicity that should be included in the training device. Figure 14–2 shows differences in the rate of skill acquisition on two tasks as a result of varying levels of aptitude for three student groups. Such differences should be reflected in the training device design if rate of skill acquisition is important. For lower aptitude student groups special consideration should be given to incorporating training device features which will increase the rate of acquisition.

Practice Procedures

Practice procedures which are to be used will influence the design of training devices. Will prompting and cueing be desirable during practice? What will be the instructor's role during practice? How many trainees are expected to practice on the device during a given period of time? The answers to these kinds of questions provide insights into specific design requirements.

Training Program

The training program, in terms of the amount and difficulty of the material to be learned, the number of trainees processed during a given period of time, flexibility in the training schedule, etc., should be reviewed for design implications. Such a review also contributes to fulfilling reliability and maintainability requirements of the equipment as well as requirements for rugged design and accessibility.

Other Training Devices

Other training devices used in the training system should be reviewed. The design of a new training device should be compatible with companion devices and should possess features that augment the training capability of the other elements in the training system.

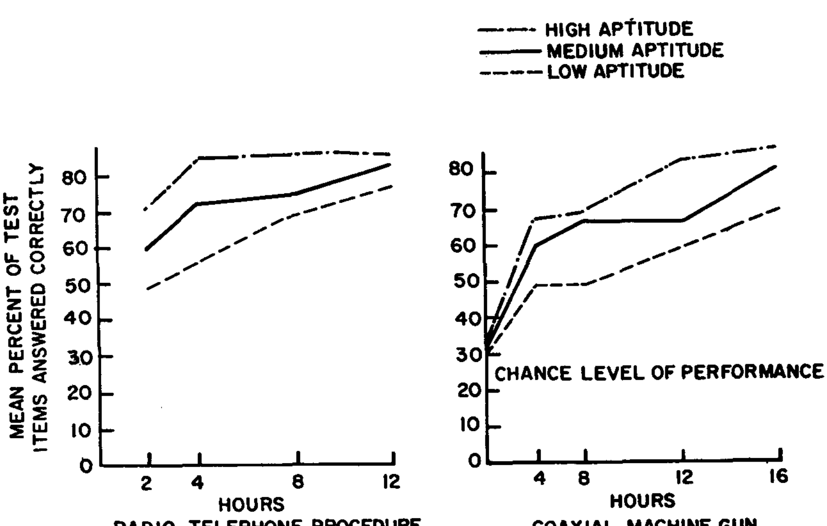


FIGURE 14–2. Differences in acquiring skills on two tasks as a function of different levels of aptitude for three student groups (Crawford, 1965).

Evaluation Requirements

Evaluation requirements also have important design implications. Such requirements are often overlooked or considered as an afterthought. Too frequently, it seems, design emphasis is placed on the engineering aspects of a piece of training equipment, sometimes at the expense of compromising its training function. If there are reasons for measuring trainee proficiency, and there usually are, means for such measurement must be incorporated, right from the beginning, into the design of the training device.

In summary, one must avoid a tendency to treat the design of training devices in isolation, as if the larger training system did not exist. In subsequent sections of this chapter, specific design guidelines and principles are suggested, but these should be viewed in the context of the training system. By proper analysis of the training system, design implications can be inferred which will produce effective, efficient training.

14.2 Nature of Training Devices

14.2.1 Classes of Training Devices

A training device is any arrangement of equipment, components, apparatus, or materials which provides conditions that help trainees learn a task (Lumsdaine, 1960). Training devices include two-dimensional displays (i.e., textual, symbolic, or pictorial material) and real or simulated three-dimensional apparatus.

A distinction should be made between two major classes of training devices: training aids and training equipment. Training aids are devices used by an instructor to help him present subject matter. They may be either two-dimensional aids (such as charts or slides), or equipment mockups or components. The term training equipment, on the other hand, refers to devices which provide for some form of active practice by the trainee.

There are two general classes of training equipment, part task trainers and whole task trainers. A part task trainer gives the trainee opportunity to practice selected aspects of the total task. For example, a navigation flight trainer allows the trainee to practice cross-country instrument flight skills, but does not permit him to practice other flight skills such as landing, complex maneuvering, etc. A whole task trainer provides an opportunity for the trainee to practice larger portions of the total task.

Although the distinction between part task and whole task trainers is questionable in certain instances, it has been helpful in determining training device design requirements. Trainers for a single task component can be less expensive than ones designed for practice on the entire task. But the decision of when to use part task trainers depends on whether transfer from the component to the total task will be expected. Usually, component skills can be practiced separately from the total task with considerable transfer if a task has been analyzed correctly. However, there are two exceptions. The first is time-sharing. When flying an aircraft for example, a pilot must control the vehicle while performing other tasks, such as establishing his ground position, communication with the ground, etc. Arrangements must be made to practice these together if satisfactory training is to be attained (Dougherty et al., 1957; Adams and Hufford, 1962; Briggs and Naylor, 1962). The second exception is the interaction between task components (Briggs and Waters, 1958). For example, making a turn in an aircraft involves the operation of the turn controls, but at the same time, the aircraft must be banked to an appropriate degree. If turning and banking are practiced as separate components, there will be little transfer of either skill to the whole task.

The advantages of using part task trainers have been pointed out by various people. (See Muckler et al., 1959.) Chief among these advantages are the following:

1. They are less expensive to build and to maintain than are whole task trainers. Thus, a greater number of training hours will be achieved for the same amount of money compared to whole task trainers.

2. Part task trainers may be made available at the time of, or preceding, the delivery of the operational equipment, but this is usually not feasible for whole task trainers.

3. They can be modified to meet changes in

the operational equipment more readily than a whole task trainer.

4. Since maintenance is not as difficult, fewer training hours will be lost in keeping the trainer in operation.

5. "Specialist" instructors can be utilized on part task trainers. This may mean that less time is required to train instructors, and that one instructor can instruct students in several trainers simultaneously.

6. Practice on part task trainers may be carried out in shorter periods, thus allowing for more frequent and extended training on the component task.

7. Training on part task trainers may be as good as or better than training on the whole task, since the student may concentrate on the learning of one particular skill without dividing his attention among several activities.

These advantages may lead to the conclusion that part task trainers are clearly superior to whole task trainers. This is not always the case (McGeoch and Irion, 1952). Arguments against part task trainers are:

1. While a whole task trainer usually represents a substantial cost investment, the number of related part task trainers necessary to achieve the same level of proficiency might cost as much or more than the whole task trainer.

2. The use of several part task trainers in a training school would involve extensive housing facilities.

3. If "specialist" instructors are required with each part task trainer, the instructional cost could increase far beyond that associated with a whole task trainer. The requirements of the training system must be analyzed before a decision can be made concerning number and type of component tasks to include in the training device.

A part task or a whole task trainer which (a) attempts to duplicate the essential features of a task situation and (b) provides for direct practice, is considered to be a simulator. Not all training equipment is designed to simulate essential task characteristics. For example, environmental training devices which attempt to teach the trainee how to withstand or recognize certain environmental conditions such as O_2 deficiency, extreme *G* forces, high psi, etc., are not considered to be simulators. In addition, scaled mockups are not usually considered to be simulators. The term simulator is most often applied to highly complex training devices, such as flight simulators, that attempt to reproduce a large number of the operational characteristics of a system and include the system's response to control actions.

The purpose of a training simulator is to provide conditions which facilitate learning. In some cases, a decision must be made between a simulator and the use of operational equipment to facilitate learning. A simulator should be used when:

1. It is less expensive than the actual equipment but still represents the essential task elements. The actual equipment may require the use of fuel or additional personnel not needed with the device.

2. It is the only feasible way to practice a task. It should be used when there is a possibility of damage to equipment, injury to students, or destruction of property when performing the task.

3. It is more reliable for practice purposes than the actual equipment. For instance, the launch operator's console in a missile system is connected to other parts of the system such as the missile handling subsystem. When the other parts are not functioning, either because of operator error or equipment malfunction, the launch operator's console may be affected. Through simulation, such problems can be eliminated.

4. It permits more effective control over the learning process than the actual equipment. For instance, in training a student to select and employ electronic countermeasures (ECM), a training device may permit greater control over the types of ECM, their frequency, and intensities, than is possible in the actual environment.

In summary, training devices are classified into two major classes: training aids and training equipment. Training equipment may be considered as a part task trainer or a whole task trainer, and most of these are referred to as simulators. The distinctions among training devices are shown in Figure 14–3.

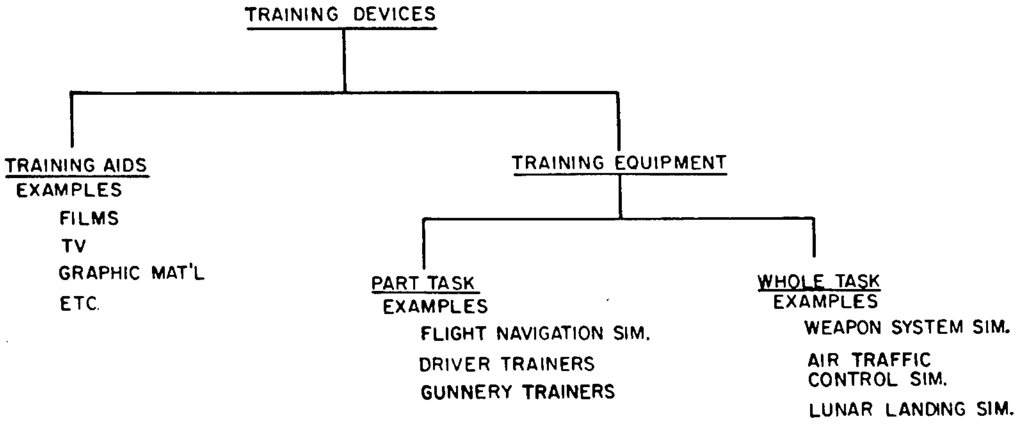


FIGURE 14-3. Summary of different classes of training devices.

14.2.2 Stages of Training

Traditionally, five types of training have been associated with different stages of learning: indoctrination, procedural, familiarization, skill training, and transition training. Indoctrination training occurs early in the learning process. The trainee learns what his task consists of and how he should go about performing it. Procedural training provides the trainee with the essential nomenclature and knowledge concerning the sequence for performing task elements. Familiarization training provides the trainee with an opportunity to practice task procedures and learn something about the task dynamics. Skill training allows the trainee to develop proficiency in performing the task. Finally, transition training is required when a person who is skilled in the operation of one model of equipment must learn to operate another model of equipment. Most of the skills and knowledge required to operate the earlier model transfer directly to the advanced model, but there may be minor differences between the two models that must be learned. A pilot who has learned to fly the F4B, for example, requires some transition training to learn to fly the F4J. The basic instruments, operating procedures, aircraft dynamics, etc., are similar for the two aircraft but the placement of specific instruments differs; the more advanced model possesses additional instruments, a limited number of operating procedures have been changed, and the advanced version is faster than the earlier model. Consequently, during transition training, the pilot must be given an opportunity to learn what the differences are and to practice the new set of appropriate responses.

For each of these stages of training, certain devices have been found to be appropriate:

Indoctrination training. Training films, instructional television, graphic materials, and functional mockups are appropriate.

Procedural training. Functional mockups, non-functional equipment, photographs, or even drawings are appropriate.

Familiarization training. The equipment should be functional; part task trainers are usually effective.

Skill training. The equipment should be functional and man-machine dynamics should be represented; either part task or whole task trainers are effective, depending on the task requirements.

Transition training. Part task training systems are appropriate. The aspects of the tasks that are different should be emphasized.

Certain classes of training devices are appropriate for each of these five stages as shown in Figure 14-4.

14.2.3 General Principles of Training Device Design

Each training device is unique and it is impossible to state specific design principles which will be applicable to all training devices. However, certain general training device design principles can be stated.

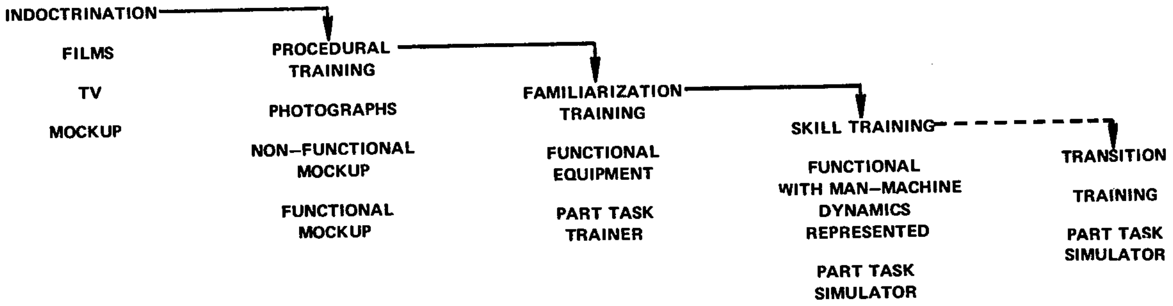


FIGURE 14-4. Types of training devices associated with stages of training.

Design for unusual use patterns. Trainees do not always know how to operate a training device when they first start. Someone will try to operate it incorrectly and, when this happens, both the trainee and the device should be protected.

Design for easy access. Frequently, trainees use a training device sequentially (i.e., after one has practiced, another takes his place). This means that a great deal of training time could be lost unless positions can be exchanged rapidly. In addition to body access, visual access for both the trainees and the instructor is important. Trainees may learn from watching others practice with the training device, and an instructor may be able to point out incorrect behaviors if he has visual access to the trainees.

Design to ruggedize the training device. Most training devices receive hard usage. A trainee is likely to drop components, to lean on surfaces, to "experiment" with and abuse a device in ways not used by an experienced operator. Considering that a training device is frequently exposed to a large number of trainees for long periods of time, it is essential that the device be as rugged as possible. Where careful handling is part of the training, "break-away" design is usually preferred rather than fragile construction. For example, in certain missile handling situations it is essential that the trainees learn that fins are not suitable handles, because they may break or be bent out of tolerance. Instead of designing the training device so that it will actually break or bend when inappropriate forces are applied, a slotted "break-away" design causes the fin to detach from the training missile when it is mishandled. The instructor can then reinsert the fin and the trainee can continue practicing.

Design for reliability and maintainability. Most training schools have very tight schedules. If a training device malfunctions and cannot be quickly repaired, trainees may not have an opportunity to practice on the device. This degrades their training since it is frequently impossible to alter the schedule. When a training device cannot perform its intended function, the designer has failed.

Design for simplicity. The instructor is usually an expert in the tasks to be trained, but he may not be an expert in operating training devices. For this reason, training devices should be designed to reduce the time necessary to train instructors in their use. Furthermore, a minimum amount of time should be required during the training period for explanation on how to use the device.

Design to provide efficient conditions for learning. Such techniques as guided practice, prompt correction of errors, cueing etc., should be taken into consideration.

Design to teach specific tasks. Only those features related to the tasks to be learned need to be represented. Distracting elements, such as irrelevant noise, flashing lights, equipment movement, etc., should be eliminated to enhance effective learning.

Design to provide proper feedback. The student should be given prompt and accurate information about his performance. He should be informed when his performance is acceptable, the direction of his errors if possible, and some overall measure of performance.

Design for practice of difficult components.

Training devices should be designed so that most of the student's time is spent practicing the things that are difficult to learn. That is, design should be based on application of psychological principles governing effective learning.

In this section, classes of training devices and the application of each class to different stages of training were discussed. In addition, general design principles were presented. In the following section, specific design recommendations for training aids are presented.

14.3 Training Aids

14.3.1 Training Films

A comprehensive series of studies on the use of films for training has been conducted by the Instructional Film Research Program of Pennsylvania State University (Carpenter, 1953; U.S. Naval Training Device Center, 1956). The following material is based primarily on these studies.

Films are useful during early training to teach factual information, to provide knowledge needed for developing procedural and motor skills, and to instill attitudes. They are most effective when used to accomplish specific training objectives (i.e., teaching Ohm's law), and when they are directed toward specific audiences with defined intelligence and aptitude test scores (i.e., high school graduates, 18 to 19 years old). They should be produced with the following principles in mind.

1. Repetition of material, with slight variations in the way the material is presented, increases learning. Concepts or factual material should be repeated, reviewed, and summarized throughout the film. Sequences of operations should be repeated, using different operators or different environmental conditions, if possible, to maintain interest.

2. Use of a clearly perceived organization in the presentation of written and pictorial material and the commentary increases the learning of separate facts. The organization should start with major ideas and then present more specific information. Written material should be in outline form and support the commentary. Pictorial material should start with an overview and proceed to a more detailed level.

3. Use of the second person in commentaries, i.e., speaking directly to the student, gives meaning to the material. Use of the passive voice should be avoided.

4. Film sequences should develop concepts or procedures slowly. Periods of time should be planned between film showings to permit practice of manageable segments of the task.

5. Techniques used in dramatic or entertainment films are not recommended. Special effects designed as attention-getting devices or optical effects, such as wipes, fades, and dissolves, should not be used.

6. Sequential procedures should be shown, such as learning to tie a knot, field-strip a weapon, performing a preventive maintenance task, etc., from the point of view of the performer, i.e., what he sees and what he does. An objective look at the procedure may be effective as an overview. But a subjective look is more effective when imitation is required.

7. Present information without introducing too many ideas during one period, and provide pauses in the sound track to allow the trainee time to absorb new ideas.

8. Do not introduce unnecessary names or technical terms in a film.

9. Avoid an overly brief treatment of a topic. Rapid coverage of the barest essentials is not an effective means of presentation.

14.3.2 Instructional Television

Instructional television should be compared with other media in the light of cost effectiveness and its capability to solve training problems. As a presentation medium television can expose a larger audience to the material than may be possible with lecture or film presentation media. It gives wide coverage to both good and poor presentations. It is an aid in training because material which may be difficult to view directly can be seen through closeups, and repeated presentations can be made quickly by use of tapes and kinescopes. (See the Instructional Television Research Reports of the U.S. Naval Training Device Center, 1956.)

Television is not applicable for training when two-way communication between instructor and student is required or when classified material is involved. Television is applicable in the following circumstances:

1. For demonstration;
2. For rapid dissemination of new or urgent information;
3. If there is a shortage of qualified instructors;
4. If the training situation involves physical risk or danger;
5. If training aids or actual equipment are in short supply;
6. If training aids and equipment are difficult to move because they are large, heavy, or unwieldy;
7. If closeups are necessary to help in training;
8. If making a record is desirable;
9. If much training time is lost in moving from one training area to another;
10. If weather interferes with other forms of presentation.

As with films, the use of dramatic special effects in television does not add to training effectiveness. Presentations are most effective where they are integrated with practice.

14.3.3 Graphic Material

Pictorial or graphic materials can be used as training aids. They provide a visual map of system flow and of the relationships among system components (Aukes and Simon, 1957). When a series of conditions or subsystems is to be described, aids presenting only the information which is relevant to a specific condition are preferred.

Graphic material used in conjunction with lectures can provide additional background for trainees. Some overlap with lecture material is often desirable. There is little difference in relative training effectiveness among operating mockups, non-operating mockups, cutaways, animated panels, or charts when they are used in conjunction with lecture material.

14.3.4 Tape Recordings

Tape recordings can be an acceptable substitute for a live lecture. Newman and Highland (1956) compared four methods of presenting a 21-hour radio course in four days including instruction by a lecturer rated above average in ability, recordings, workbook chapters in a notebook, and tape recordings with slides. No overall superiorities were found among these methods; mass media were as effective as the instructor for the first two-thirds of the course, while the instructor was superior for the last third. In two separate studies of college teaching, Popham (1961, 1962) compared tape recordings with live lecturers and found no difference. Follettie (1961), using military trainees and content from Army basic military training, found no difference between recorded and live lecturers.

14.3.5 Written Materials

Assuming reasonable reading proficiency, textbooks and other written materials can present a wide range of content. All training materials should be well organized. Ausubel (1960) and Ausubel and Fitzgerald (1962) have shown the training value of the "advance organizer," i.e., an introductory statement describing general concepts to be presented later in detail. Klare et al., (1955) compared the effects of underlining as opposed to not underlining key words, but did not tell their purpose of underlining. Able students were helped by underlining, the less able hindered, which indicates the importance of familiarizing students with any organizational system used.

In summary, this section has presented specific design recommendations for training aids, including films, television, graphic material, tape recording, and written materials. For further information about the design of training aids, refer to Smith, 1966; Lumsdaine, 1960; Hahn et al., 1966.

14.4 Training Equipment

Training specialists use a vast array of training equipment to help trainees acquire skills. Training equipment often reduces actual training costs and may be safer to use than on-the-job training with operational equipment. Figure 14–5 shows that almost half of the flight-training trials required to develop proficiency during elementary instruction could be saved by using training equipment instead.

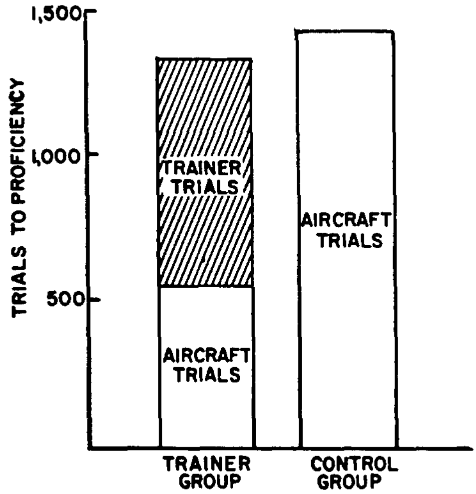


FIGURE 14-5. Trials required to reach standard proficiency in 13 exercises of elementary flying training. One group practiced each exercise in the training device first, and later in the aircraft. The control group began and continued practice in the aircraft. The treatment of the trainer group resulted in a saving of 100 total trials, or 874 aircraft trials (Williams and Flexman, 1949).

Training equipment ranges from relatively simple procedural trainers to full-scale simulation systems. Establishing design specifications for such equipment is difficult. One approach is to base these requirements on the training functions that the equipment is supposed to perform. If, for example, a training device is used only to facilitate learning of nomenclature and very simple switching procedures, functional displays and controls are not required. This is demonstrated very clearly by the results shown in Figure 14-6. These results show that increased display and control requirements have very little impact on the efficiency of learning a simple procedural task. On the other hand, a training device that is supposed to provide skill training has increased display-control requirements. It must be able to present dynamic situations and the control responses to these situations that are involved in the performance of the skill.

Minimum display-control requirements for different training functions are discussed below. The training equipments have been separated in accordance with three stages of training: procedural, familiarization, and skill acquisition. Training equipments are seldom used for indoctrination training, and the display-control requirements for transition training are defined primarily by the differences between models of operational equipments.

14.4.1 Procedure Trainers

Procedure trainers are designed for teaching nomenclature and procedures. A procedure is a step-by-step series of activities involving no special skill requirements. Display-control requirements for this type of trainer are minimal. Symbolic, non-functional representations will usually perform the required training functions effectively.

14.4.2 Familiarization Trainers

Familiarization trainers are designed to provide the trainee with an opportunity to practice procedures or techniques and to develop concepts during exposure to different situations. Display-control requirements are more complex than for procedure trainers, but high fidelity of representation for displays and controls is not required.

14.4.3 Skill Trainers

Skill trainers are designed to expose the trainee to a wide variety of situations and allow him to practice responses to these situations. Continuous tasks require complex displays with accurate control relationships. These trainers are usually more complex than familiarization trainers. Aircraft simulators, driver trainers, and radar tracking simulators are examples of this kind of trainer. Training time with this type of trainer may be reduced and, hence, fewer trainers of this type may be required, if the trainee has had a chance to practice on less expensive procedure and familiarization trainers before he is exposed to the more complex skill trainer.

Some skill trainers use operational equipment, with the training equipment providing dynamic signals to displays and proper responses for

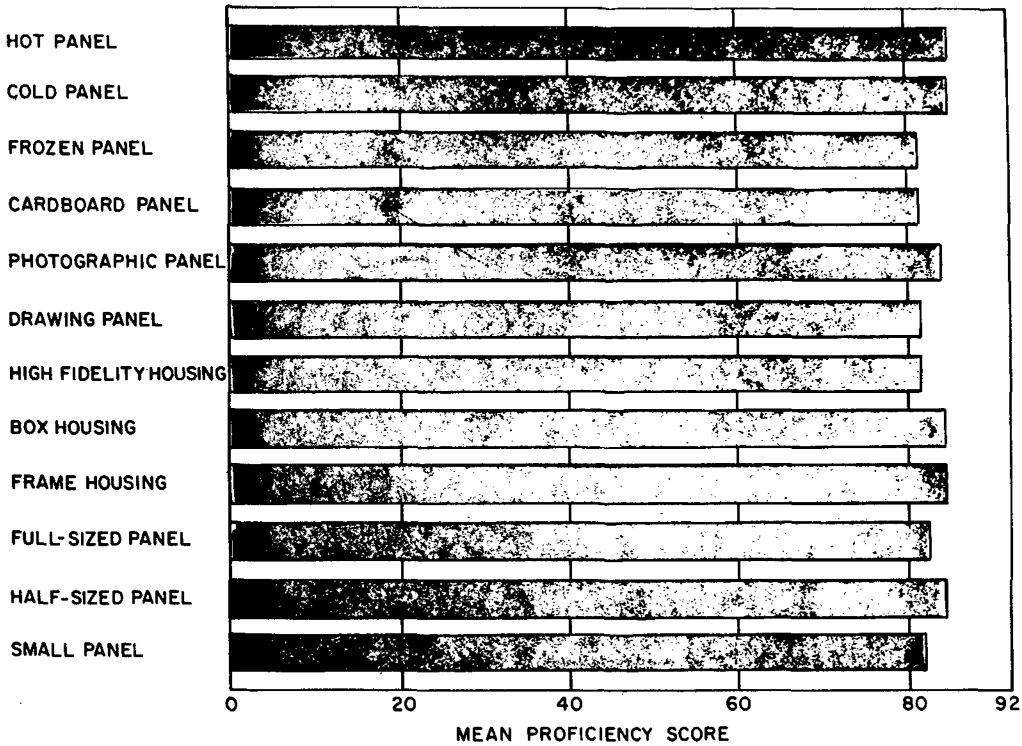


FIGURE 14–6. Comparison of average proficiency level produced with twelve training devices differing with respect to control-display requirements (Cox et al., 1965).

controls. One application of this approach which has proved to be very useful is in command and control systems, such as SAGE and BUIC. Operational display equipment in these systems is provided with simulated radar information, and operators are allowed to make responses to the displayed information. The consequences of these responses are subsequently shown on the displays through inputs provided by the trainer.

The use of obsolete equipment as a trainer has been shown to be effective. HumRRO Work Unit FORECAST showed how a combination of medium-fidelity mockups representing new equipment, plus obsolete equipment which allowed trainees to practice tasks common to both the old and new models, reduces the amount of practice required on the new equipment (Shriver et al., 1964).

In summary, design principles for training equipment should be relevant to specific training objectives. The designer must decide when symbolic representations, some fidelity of display and controls, or simulators and operational equipment will most efficiently facilitate learning. Table 14–1 summarizes the principles outlined in this section with specific examples.

14.5 Fidelity of Simulation

Fidelity of simulation refers to how realistically the task situation is represented in the training situation. A high degree of fidelity in simulation hasbeen thought to be be essential for two reasons: (a) to provide a practice situation, and (b) to gain acceptance and motivation on the part of the trainee to use the training device.

Research over the past several years has shown that these two reasons are not entirely justified. Practice on low-fidelity flight trainers has resulted in savings, in terms of flight hours, of up to 50% (Smode et al., 1962). In some cases, differing degrees of fidelity of simulation have had little impact on acceptance or motivation. The best overall design guideline appears to be that fidelity of simulation is required when the trainee must learn to make difficult discrimin-

TABLE 14–1. MINIMUM PRINCIPAL DISPLAY-CONTROL REQUIREMENTS OF TRAINER TYPES

Types of trainers	Essential form of display	Essential form of control	Training function
PROCEDURE TRAINERS			
Nomenclature and locations trainers.	Diagrammatic; non-functional mockup; miniatures.	None required; response is symbolic.	Basis for memorizing and executing job instructions.
Demonstrators.	Symbolic-diagrammatic in two or three dimensions.	None required; response is symbolic.	Motivation background for learning and performance; knowledge background may include some "nomenclature" and "locations" training.
FAMILIARIZATION TRAINERS			
Detection of condition trainers.	Simulation of displays in critical work environment; displays may be intermittent or continuous.	None required; response is symbolic.	Scanning techniques; perception through noise; detection of absolute values or relative changes in displays.
Identification of condition trainers.	(1) Symbolic or diagrammatic. (2) Simulation of critical displays plus work-context cues in later training.	None usually required; exception when controls must be used to get data; critical response is symbolic.	Response to patterns of display data; inference making about conditions; rapid perceptual response; short-term recall.
Problem-solving and decision-making trainers.	(1) Symbolic or diagrammatic. (2) Simulation of critical displays may be miniaturized. (3) Response correction should schematize correct decision process.	(1) None required; response is symbolic. (2) Nonfunctional or diagrammatic as part of display problem. May be miniaturized and/or schematized.	Variables in required decision; inference making; short-term recall; response alternatives and implications.
Instructed-response for procedures.	(1) Diagrammatic. (2) Discrete-valued displays; discrete-valued response to control activation. (Symbolize or represent conditions under which procedure itself should be initiated). May be miniaturized.	Simulated for relative location, direction of movement; control forces irrelevant; may be miniaturized.	Long-term conceptualization of steps in a procedure; precautions; threefold association of environmental stimulus, conceptual response, and overt response.
SKILL TRAINERS			
Tracking trainers.	Simulated tracking displays; simulated control-display interactions. Other displays to be scanned may be discrete-valued.	Tracking controls; in compensatory tracking, control forces and amplitudes well simulated; in pursuit tasks, less simulation required.	Anticipations of target motions and of "cursor" capabilities (i.e., display-control dynamic interactions).
Job segment trainers, simulators.	Full simulation (with some qualifications).	Full simulation (with some qualifications).	Integration of part tasks; time-shared or time-linked activities; automatized habits; proficiency evaluated and diagnosed.

Modified from Miller (1954).

ations between stimulus events, and where the responses are either difficult to make or are highly critical to system operation. For instance, fidelity of simulation would not be required in a situation where the trainee must learn to discriminate between a red light and a green light since marked deviations in the wavelengths represented in the simulator and the operational equipment would have no impact on his performance. However, if he had to learn to discriminate between two red lights, the importance of representing the actual wavelengths increases dramatically. The following sections attempt to provide more specific design guidance concerning the degree of fidelity required in training devices.

14.5.1 Fidelity of Simulation Concepts

Fidelity of simulation consists of three different components: (a) equipment fidelity, (b) environment fidelity, and (c) psychological fidelity. Equipment fidelity is the degree to which the simulator duplicates the appearance and "feel" of the operational equipment. For example, an actual automobile cab, with steering wheel, steering wheel feedback dynamics, speedometer, fuel gauge, etc., would have high equipment fidelity for a driver trainer designed to teach driving. Environmental fidelity is the degree to which the simulator duplicates the sensory stimulation (excluding control feel) which is received from the task situation. A driver trainer that provides motion cues and a three-dimensional, dynamic visual presentation of the external world (i.e., the road, trees, sky, etc.,) would have high environmental fidelity. Psychological fidelity is the degree to which the simulator is perceived by the trainee as being a duplicate of the operational equipment and the task situation. If a trainee has not had experience with a particular automobile model, for example, he may perceive a driving simulator as being highly realistic when, in fact, it deviates substantially from the model it is supposed to represent. It is important to note that minor deviations from the actual model (i.e., a slightly larger steering wheel, a slightly displaced speedometer) will probably not be noticed even if the trainee is familiar with the particular model being represented. Despite minor deviations, the device would still have high psychological fidelity.

Psychological fidelity is dependent upon the perception of the device by the trainee. This perception is affected by his experiences and limitations in his sensory-perceptual capabilities. In the case of visual perception, for instance, a driving simulator may yield low environment fidelity, but a reasonably high psychological fidelity, by taking advantage of limitations in the human visual response mechanism. Up to about 100 m.-L., visual acuity increases with luminance. Beyond this value, it increases very little. Thus, where extremely high ambient luminances occur in the real world, such as driving in bright sunlight, representation of visual fields in the training situation with more than 100 m.-L. luminance may not be necessary. A simulation display technique that is capable of generating more than 100 m.-L. is not likely to possess greater psychological fidelity than a technique which produces only 100 m.-L. luminance. Similarly, in the case of tactual perception, a simulator may possess low equipment fidelity, but a reasonably high level of psychological fidelity by taking advantage of the rather gross limitations in human ability to detect differences in size or shape of controls.

As a general rule in choosing training device designs, if the trainee cannot discriminate between different levels of equipment or environment fidelity because of perceptual limitations, the least expensive level is best. Care must be taken, however, to insure that psychological fidelity is achieved on the basis of perceptual limitations rather than on the basis of experiential limitations. This is especially the case if performance is likely to deteriorate when the trainee is exposed to the actual situation. For example, a control that is pulled up in the training device might be perceived as being realistic by a trainee. But, if the control is actually pushed down in the operational device, the results could be disastrous (i.e., an emergency handle).

The overall level of fidelity required in a training device is partially determined by the desired amount of transfer of training. High levels of fidelity for complex systems are often needed to provide transfers from the training situation to the operational situation. The training device provides a situation where the trainee builds up a set of expectancies and appropriate responses to both the actual equipment and to critical environmental conditions (George and Handlon, 1957; Lybrand et al., 1957). Where these expectancies are met in the operational situation, positive transfer of training will occur. If, however, the expectancies are incorrect, negative transfer of training will occur. If critical environmental conditions occur in the operational situation and the trainee has not had a chance to build up an appropriate response repertoire to these conditions, his performance may suffer.

While assessing the overall level of fidelity, it is important to remember that there is so much variability in equipment characteristics between and within systems that the concept of absolute fidelity is questionable (Miller, 1954). It is easy to fall into a dangerous conceptual trap where the training system is perceived as though such an entity (i.e., absolute fidelity) really existed, and as if the actual system is identically represented by its different models. For example, the BUIC system (a command-control system with sites located throughout the U.S.) which exists at one location is different from the BUIC system which exists at a different location. The number of radars, the radar types, the weapon capability, etc., are unique to each representation of the system. This situation is not peculiar to the Air Force. Military systems are not like automobiles where a model is duplicated a number of times. They differ in several critical dimensions from one to the other (Kinkade et al., 1969). Because of this variability between and within systems, it is virtually impossible to construct a training device exactly like the operational equipment to which a particular student might be assigned.

All factors that affect transfer from a simulation to an operational situation have not been explored, but the following discussion provides information on a few of these factors.

14.5.2 Equipment Fidelity

In order to duplicate the appearance and "feel" of the operational equipment, the designer should be aware of the following aspects of equipment fidelity: cabin realism, location of instruments and controls, and control feel.

Little has been reported on the significance of providing a realistic cabin or cockpit frame, but such factors as accessibility, trainee observation, and instructor participation should influence design requirements. The cockpit design should provide efficient conditions for learning. Care should be taken to minimize wasted time and effort for both students and instructors. Where simulation is not likely to enhance training, and may distract the trainee from the task, high equipment fidelity is not desirable. For instance, the constraints of an F4 fighter cockpit in terms of normal entrance and exits might seriously hamper the training situation for a group of students. Getting into the cockpit would be time-consuming, and other students, as well as the instructor, may not be able to observe the participating trainee. But, in certain applications, such as "ditching" operations by Navy pilots, learning to cope with cabin or cockpit constraints is part of the training. Adequate representations of cabin constraints should be included in those situations where the trainee must learn to cope with them, but are not required when the constraints interfere with the training process.

In addition to general cockpit requirements, designers should assess equipment fidelity with respect to the location of instruments and controls. The relative location of instruments appears to be very important. If the trainee develops scanning techniques and learns where to look for certain information in the simulator, deviations in the relative placement of the instruments can affect transfer. Deviations in size and, to a limited extent, shape, can be tolerated without affecting transfer significantly.

In the simulator, the trainee will develop habitual reaching patterns. These patterns will be established for most controls in the forward plane, for those controls that are frequently used, and for critical controls, such as the ejection control in an airplane. Precise location of these controls is advisable. For other controls, visual guidance will be required and minor deviations in location between the simulator and the operational equipment will have little effect on transfer.

Finally, the designer should address himself to the question of how much fidelity of control feel is required for effective training.

Control feel is a function of both the force and displacement that are applied to a control device to obtain a required system response such as turning a car, banking an aircraft, or tracking a target on radar. A number of studies have considered this question, and the results show that transfer depends more on a correspondence between the patterns of control forces than on the absolute amount of force or displacement required (Muckler and Matheny, 1954; and Matheny et al., 1953). In addition, Briggs et al. (1957) showed that varying the control amplitude and spring tension does not affect transfer. It may be concluded that reasonable attention to factors affecting control feel will be sufficient in a simulator for training purposes, even though psychological fidelity may be low. That is, even though the operator perceives a substantial difference between the simulator and the actual device, it will have little effect on his training.

The control-display ratio in the simulator may differ substantially from that in the operational system and high transfer will still be obtained (Rockway, 1955). Again, the critical factor learned in the training situation is the necessary pattern of responses. The operator can then quickly adjust the degree of his responses once he finds himself in the operational situation.

In conclusion, design recommendations concerning equipment fidelity are:

1. Adequate representation of cabin constraints should be included in those situations where the trainee must learn to cope with them, but is not required when the constraints interfere with the training process.

2. Maintaining the relative position of instruments with respect to each other and controls is important in the design of a training simulator, but minor deviations in size and shape of instruments have little effect on transfer.

3. Care should be taken in the placement of controls located in the forward plane, frequently used controls, and critical controls, but minor deviations in control location for controls requiring visual guidance will have little effect on transfer.

4. Reasonable attention to factors affecting control feel is sufficient for training simulators.

5. Minor deviations in control-display ratios have little effect on transfer.

14.5.3 Environment Fidelity

In the design of training simulators, the design issues concerning environment fidelity are : (a) displaying an abstract representation of the external visual world for those systems where the operator perceives the environment directly, (b) duplicating the effects of the environment on system displays, and (c) duplicating the sensation of motion for those systems where the operator is subjected to movement. The area of auditory simulation has not received much attention except in sonar training situations.

Perception of the visual world provides sensory cues for orientation and vehicular control in some systems, such as driving an automobile or flying an aircraft. Because of optical and engineering complexities, however, it is only in recent years that attempts have been made to provide simulation of the external visual environment. The impact of representing the external visual world on transfer of training in these systems has not been evaluated sufficiently to provide specific design principles, but it appears that the relationship between displaying environmental factors and psychological fidelity is important to transfer of training. From the standpoint of psychological fidelity, gross deviations from the display of environmental factors may substantially degrade transfer of training, although this may only be temporary.

Hammerton (1963) investigated transfer of training when a real-world visual display was simulated on a cathode ray tube (CRT) display. In one condition, subjects controlled the movements of a trolley moving along a miniature railway. They had direct visual access to the environment. In another condition, subjects used the same control dynamics and size of visual stimulus to control a display presented on a CRT. After training with the simulated presentation of the environment, i.e., the CRT display, performance with visual access to the actual environment was poor. However, after a pronounced initial decrement in performance, recovery was rapid. There are, of course, practical situations where, if the first transfer trial is poor, the shape of the rest of the learning curve is academic. This study indicates that gross deviations from environment fidelity can have a substantial impact on initial transfer, but these effects may not be lasting.

The degree of environment fidelity that is required to represent the external visual world may be determined on the basis of psychological fidelity. In attempting to establish the requirements for star field simulation, Kindake et al. (1965) found that maximum perceived realism is obtained from background, size, and brightness values which differ substantially from actual values in the real world. However, these values were considered to be sufficient for simulation purposes.

The conclusion that psychological fidelity is important when representing environmental parameters in a simulator is supported by research in driver trainers (Wheaton, et al., 1966). Behavior in an instrumented car, where the actual external world is available, differs appreciably from behavior in a situation where representation of the visual world is provided by a display, CRT, film, or TV. Although the physical dynamics of the visual situation, i.e., angular velocity, perspective, texture, etc., are represented in the display, responses to the simulated visual world are different from the responses made to the real world. Since the responses are different, it may be concluded that transfer of training will be less than perfect in a complex task, such as driving a car, when an abstract representation of the visual world is provided.

The question of environment fidelity where there appears to be a need to duplicate the effects of the environment on system displays has been investigated for many years. Principal areas of investigation have dealt with radar displays and sonar displays. The consensus of these studies is that a high degree of fidelity is required when the operator in the actual system has difficulty in distinguishing between various stimulus events. In those systems, for example, where the operator cannot easily detect the presence of a signal in noise, or where differences in the characteristics between two or more signals are minor, high fidelity of simulation is beneficial. The trainee learns to make these fine discriminations in the simulator. However, the dimensions used by the operators to make these discriminations are the only ones that must be represented in the training situation if they are to be unambiguously represented (Klipple, 1965). Hillix (1960), for instance, showed that effective transfer of training could be obtained in a bombadier-navigator radar discrimination task by adequately representing the degree of "fill" and certain proportions of characteristic patterns. Minor deviations in characteristic patterns had little effect on transfer. The dimensions attended to by the operator may not be apparent from a casual observation of his performance. In a radar air traffic control situation, for example, it was found that controllers derive a considerable amount of information concerning airspeed and heading from the "trails" left on the radar scope from one sweep to another. (For a discussion of display requirements in flight simulators, see Whittenburg and Wise, 1963.)

There does not appear to be a strict requirement for high environment fidelity where easily perceived and interpreted information from the real world is provided on visual displays in the operational situation. This is particularly the case when high environment fidelity implies presenting noisy signals on visual displays. Presentation of noises has little effect on transfer of training, except when the object of the training is to teach the trainee to discriminate the signal through the noise (Briggs et al., 1956).

A certain degree of environment fidelity is required when the information presented on displays conflicts with other sensory information in the training situation. Difficulties arose in a study by Rathert et al. (1961) when subjects performing in a moving-base simulator were presented with visual cues on their displays that conflicted with vestibular cues. In this type of situation, simulation is not merely inadequate, but can in fact introduce false cues.

The importance of motion simulation as another aspect of environment fidelity has been questioned for many years. Fraser (1966) provides an excellent review of the importance of this issue for space training application, and much of the following material is based on his review.

With respect to psychological fidelity, the use of the word "motion" is misleading, because a human does not perceive motion directly except by his visual senses. The body cannot perceive motion; it is sensitive only to higher derivatives such as acceleration and jolt. Thus, in regard to body response, it is the rate of change of motion that must be simulated.

A motion can be applied to a simulator at a rate of change and for a duration that is readily perceptible to the trainee and representative of the motion being simulated. The motion is thereafter removed or "washed-out" at a rate below the trainee's vestibular and proprioceptive threshold of sensitivity, although its apparent effects are continued on appropriate instruments. An illusion of continued motion is provided. Such an illusion, of course, does not produce the physiological effect of sustained motion or Coriolis acceleration, although the effects of buffeting and moderate intensity impacts can be reproduced with small excursions of the simulator.

The effect of motion on the transfer of training may be more important initially due to the trainee's perception of psychological fidelity. But the actual benefits of simulated motion depend on the degree and kind of response to motion cues demanded by the task. This is not a clear-cut issue as the following summaries of studies in this area indicate.

In a study of landing approach simulators, it has been found that for representation of longitudinal dynamics in which short-period frequencies are moderate (about 0.6 Hz, with good damping), fixed-base simulation is adequate and is perceived to be realistic. With respect to roll-damping and roll control power, in the region representative of normal aircraft operations, no difference in performance is observed between a fixed-base simulator, a motion simulator, and the actual aircraft. At higher rolling acceleration, however, the fixed-base simulator is found to be unrealistic. At very low rolling rates, fixed-base simulators do not have high psychological fidelity, because of the anticipation provided by the motion cues. In this case, the motion simulation is found easier to fly than the fixed-base simulator (Rathert et al., 1959; and 1961).

In a series of studies concerning flight simulators (Sadoff and Harper, 1962; and Sadoff, 1965), it was found that simulator motions are generally required where the pilot's ability to cope with stability augmentation failures are involved in training. However, Brown and Johnson (1959) found that pitching cues were not important where the oscillations encountered were within an acceptable range.

In a helicopter trainer, Fedderson (1961) found that the group with motion cues learned to hover faster and, when transferred to the helicopter, performed better initially in the air than the group trained in the static simulator. The difference, however, disappeared by the end of the 12-min. flying session. Because the difference disappeared so rapidly, it was concluded that use of a motion simulator could not be justified.

Buckhout et al. (1963) found that simulated motion during the learning of tracking skills in high-speed, low-altitude flight contributes to more effective performance because motion cues play an important role in this situation.

Motion simulation contributes little to the effectiveness of learning procedural tasks in flight simulators. Wilcoxin and Davy (1954) found that rough-air simulation had no effect on the transfer of training from simulator to aircraft in the performance of basic instrument and radio range procedures, although the students believed it added realism.

For space simulators, Urmer and Jones (1963) suggest that full simulation of motion on the ground might lead the pilot to expect vestibular and other proprioceptive feedback from gravitational forces which would be absent in a null gravity situation. Because of this, they recommend that it may be better to train without added motion cues, relying on instruments and, if applicable, on exterior visual display.

Motion cues are necessary in those situations where they contribute to improved control of the vehicle or where movement interferes with satisfactory performance. Motion cues will contribute to improved control of the vehicle in situations where visual information is degraded or inadequate, or where motion cues provide orientation cues. In situations where the operator must learn to compensate for motion to perform satisfactorily, motion should be represented, but high fidelity in representing the motion is not required.

In general, design recommendations concerning environment fidelity are:

1. The required degree of environment fidelity may be determined on the basis of psychological fidelity.
2. In a complex task, transfer of training may be degraded by an abstract representation of the visual world.
3. High environment fidelity is required when the actual task demands a difficult distinction be made between different stimulus events.
4. Where the operator must learn to compensate for motion in the actual task, motion cues should be provided although high fidelity is not demanded.

14.5.4 Levels of Fidelity and Stages of Learning

Although as a general guideline, the designer should consider environment fidelity as being more important than equipment fidelity of simulation, different levels of environment fidelity are appropriate at different stages of learning. For example, at certain stages of learning, deviations from absolute fidelity are suggested.

At early stages of training on a complex system, the student may be confused if provided with realistic presentation of environmental conditions. During later stages of learning, the student may become bored with a "normal" representation of environmental conditions. In either case, very little will be learned. The designer should seek to establish an appropriate level of problem difficulty within the following context:

1. Problem difficulty should be adjusted to the students' skill level, rather than to an arbitrarily chosen representation of realism.
2. Problem difficulty should be adjusted by varying different dimensions. For example, in an airborne Radar Intercept Officer training situation, one ECM-emitting target could be used during one exercise and two non-ECM-emitting targets could be used in another exercise. In subsequent exercises, target speed, maneuverability, target crossing angle, etc., could also be varied. The student should be exposed to variations in each parameter rather than to a limited set.
3. The equipment should be capable of presenting problems representing the worst possible environmental conditions that could be encountered, rather than restricting the representation to expected or normal environmental conditions.

Figure 14–7 depicts general relationships between environment fidelity, equipment fidelity, and stages of learning. Early in the training program (procedures training), the trainee cannot benefit from high degrees of either type of fidelity. However, as skill is acquired (familiarization training), there are requirements for increases in both environment and equipment fidelity, with the requirements for greater environment fidelity increasing at a faster rate. During later stages of training (skill training), increases in both types of fidelity are desirable, but there is a requirement for higher levels of environment fidelity. It should be pointed out again that these are general guidelines and that each device will have to be designed on the basis of specific task requirements.

14.5.5 Time Deviations From Realism

Deliberate deviations from fidelity in the time domain may facilitate learning. Designers should consider the advantages of compressing or expanding time.

In some systems the occurrence of events is very slowly paced. For example in Hunter-Killer submarine operations, it may take several hours to complete an exercise if it is conducted in real time. Studies have shown that training is enhanced if such exercises are conducted in fast time rather than real time.

In other cases, the occurrence of events may be very fast. For example, the acquisition, lock-on, fire, and track sequence in firing a REDEYE missile occurs rapidly. An inexperienced student may benefit from going through the sequence in slow time before he is exposed to events in real time.

Concerning time compression and expansion, the following should be taken into consideration:

1. The time scale used in training simulators should be selected on the basis of student skill

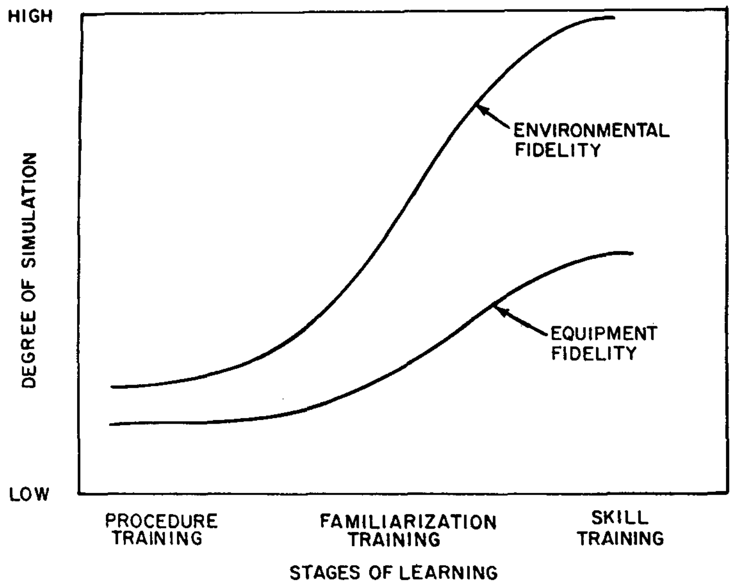


FIGURE 14–7. Relationship between environment fidelity, equipment fidelity, and stages of learning.

level and training objectives rather than a time scale dictated by realism.

2. If critical events are slowly paced in the operational system, time should be compressed. Care should be taken, however, not to time-stress the student. A variable time scale may be desirable for some applications.

3. If critical events are fast paced in the operational system, time should be expanded during initial exercises. Care should be taken to remove the time-expanding feature as soon as the student has acquired the proper skill.

4. In systems where speed of operator response is critical, it might be advisable to allow for some time compression capability as a means of increasing problem difficulty.

5. In some training systems, it is desirable to include a "freeze" (i.e., stop-action) capability. This feature is helpful when an instructor is present.

14.5.6 Interrelationship Between Cost, Fidelity of Simulation, and Training Effectiveness

Figure 14–8 is a simplified illustration of the relationships between fidelity of simulation and the amount of training value to be derived from increased fidelity of simulation.

The line marked "costs" shows that as the degree of fidelity in simulation increases, costs go up at an increasing rate. The line marked "transfer" is a hypothetical relationship between degree of fidelity and the transfer value which may arise. At low levels of fidelity, relatively small gains in transfer value can be expected with increments in fidelity. However, a point is reached where large gains in transfer can be expected for small increments in fidelity. The gains then diminish with further increments in fidelity. There is a point of diminishing returns, where transfer value gained is outweighed by higher costs.

It should be understood that these are hypothetical relationships. It is not possible, under the current state of the art, to quantify the scale representing the degree of fidelity.

14.6 Scoring Systems

Scoring systems are used in training in order to: (a) shape individual performance, (b) facilitate training effectiveness, (c) measure proficiency, and (d) evaluate the training system. The training system designer should provide for these uses wherever possible.

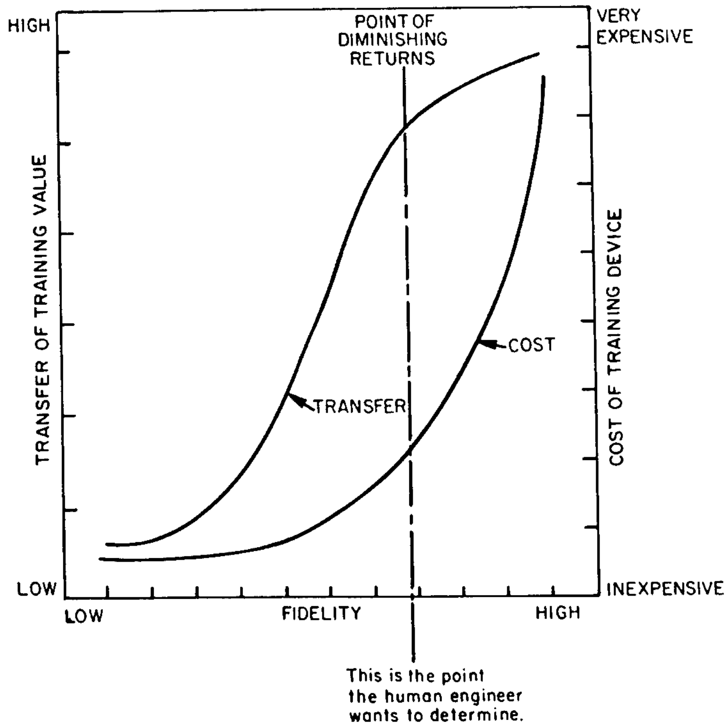


FIGURE 14–8. Interrelationship between cost, fidelity of simulation, and training effectiveness (modified from Miller, 1954).

14.6.1 Shaping Individual Performance

Scoring used to shape individual performance is applicable to early stages of training and is called knowledge of results (KOR), or feedback. There are two types of KOR: natural and artificial. Some tasks are generous in providing natural, or "built-in," feedback to the trainee.

Driving a car and keeping the vehicle on the roadway is an example of a task with abundant natural feedback. When a driver approaches the edge of the roadway, he can see that his vehicle is not traveling along the desired path. He has made a steering error. As he corrects this error, he can see whether he has compensated correctly, overcompensated, or undercompensated. Feedback is also provided by the feel of the automobile as it deviates from the pavement. Hitting a curb or a soft shoulder causes the steering to feel differently and will change the overall ride of the vehicle. The sound of a soft shoulder or the thud of hitting a curb provides additional feedback when the driver has made an error in steering. Feedback continues through the correction process, making the driver aware of reduced or increased error.

In contrast to the act of driving a car which is rich in natural feedback, there are many tasks where this type of feedback is scant or ambiguous. For these tasks, the system designer may supply artificial feedback in the training situation to expedite learning. Artificial KOR, or augmented feedback, is usefully applied when the trainee has hypotheses (or guesses) about what he should do and how he should do it, in a task situation which does not provide information that confirms or denies these hypotheses.

Artificial KOR should be employed with reservations. If it is provided during practice of a task, it may raise the level of performance during practice but, since the KOR is not natural to the task, performance may drop when the student goes to the job and the artificial KOR is no longer present. A number of experiments have shown that artificial KOR raises the level of performance, but that performance drops when it is withdrawn (Bilodeau, 1952; Bilodeau, et al., 1959; Goldstein and Rittenhouse,

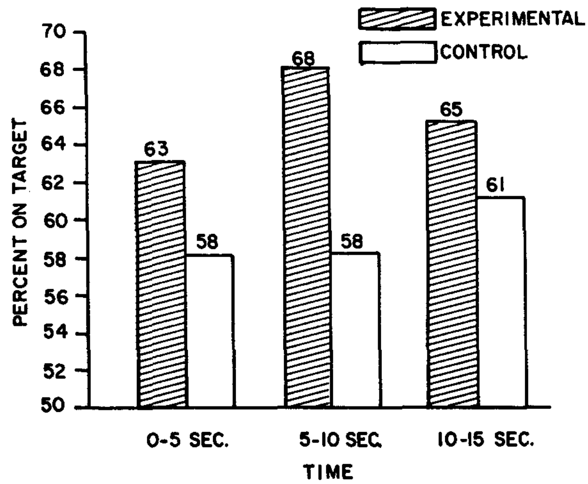


FIGURE 14–9. Comparison of tracking performance in a field setting for experimental (KOR) and control groups (NO KOR) by 5-second periods (data from Houston and Kinkade, 1966).

1954). This finding is not a universal one, however. Other studies have shown no reduction in performance when artificial KOR is withdrawn. Kinkade (1963) found that artificial KOR favorably affected learning on a permanent basis when stimulus conditions in the operational environment are clearly perceived. On the other hand, when the stimulus conditions are not clearly perceived, performance level is reduced when artificial KOR is withdrawn. Trainees use the augmented feedback as a "psychological crutch" to help reduce ambiguity concerning the stimulus conditions. These findings have been confirmed by other experiments (Micheli, 1966).

Artificial KOR should be specifically employed to increase the rate of improvement and skill level. If the task permits the student to make aiming errors, such as undershooting, overshooting, right and left errors, the KOR should indicate the direction of the error. The inclusion of a directional component in the artificial KOR will be more effective than merely indicating to the student that he was off-target or on-target.

Figure 14–9 illustrates the beneficial results obtained from applying artificial KOR in an Army artillery training device. In this field experiment, gunners with KOR, the experimental group, acquired the target faster and tracked the target more accurately for longer periods of time than gunners who were not trained with artificial KOR, the control group.

14.6.2 Facilitate Training Effectiveness

Scoring used to facilitate training effectiveness during intermediate stages of learning is called rehearsal feedback. There are two types of rehearsal feedback—self administered or instructor administered. When an instructor administers rehearsal feedback it is usually referred to as a briefing. The essential characteristic of rehearsal feedback is that it permits the re-creation of critical actions and consequences for review purposes.

In some training situations, the student must perform a series of actions and make a number of decisions before the outcome of his performance can be evaluated. For example, a Radar Intercept Officer directing an interception must acquire the target, determine how he will approach the target, be alert for evasive maneuvers, direct the pilot to a desirable target crossing angle, and be prepared for a re-attack if his first weapon misses. The adequacy of his performance can be evaluated only after the interception is completed. There are few interim, absolute standards against which his performance can be measured. For this kind of task, it is desirable to provide means for the student to review and evaluate his actions in the light of resultant consequences.

General recommendations for supplying rehearsal feedback are:

1. Rehearsal feedback should be provided as soon after a training exercise as possible.
2. Rehearsal feedback should indicate clearly where an incorrect action was taken and what the correct action should have been.
3. Irrelevant information should be filtered from the rehearsal feedback.
4. Rehearsal feedback should be easily interpreted and require minimum training for complete comprehension.
5. Rehearsal feedback should be at least semipermanent so that it can be reviewed several times.

14.6.3 Measure Proficiency

Scoring used to measure proficiency applies during later stages of learning. The purpose of proficiency measurement is to provide information about a student's performance for administrative and planning purposes. Based on this information, a student may be assigned further training, or the scope of his responsibilities may be made commensurate with his proficiency.

General recommendations for applying proficiency measurement are:

1. The proficiency measure should summarize a series of critical system actions in a single, quantitative score.
2. The proficiency measure should be directly related to system objectives, i.e., hit-kill, fuel consumption, system response time, etc.
3. Proficiency measures should encompass all system objectives related to a job segment.
4. The proficiency measure should be related to an "acceptability index" so that a particular score can be judged to represent at least five categories of proficiency: excellent, superior, average, poor, unacceptable.
5. The proficiency measure should be completely objective and "fake-proof."
6. Proficiency measures should be supplied with minimum delay.

14.6.4 Training System Evaluation

Scoring used to evaluate the training system is applicable near the period of graduation. A training system cannot remain static and continue to be effective. Operational requirements change as a function of new equipments or new missions. The training system must be responsive to these changes and, thus, it is necessary to provide measures which indicate where training is adequate and where increased emphasis is required.

General recommendations for applying measures of training effectiveness are:

1. They should reflect task performance rather than system performance.
2. They should reflect the effectiveness of specific training course segments or training devices.
3. They should be quantitative, objective, and "fake-proof."
4. There should be some method of relating the training effectiveness measures to criteria of acceptability.

14.7 Design of Student and Instructor Stations

14.7.1 Training Stations

In every training context there is an area set aside for the student. This area and all of the equipment located within it represents the student station. It is here that the student is to "learn" or practice new responses. Most training situations also provide a separate and distinct area for the instructor. This area, and the complex consoles, displays, and controls which may be situated within it, constitute the instructor station. From this station the instructor will induce problem variations, monitor and evaluate student responses, provide feedback, and generally exercise administrative and executive control over the training process.

Both stations can assume many forms depending on what is being trained and how training is being conducted. The possible variety of stations becomes evident when one realizes that they are present in classroom settings, flight trainers, weapon system simulators, and many other training environments. In each instance a specific design and layout of the student or instructor station will be dictated by the particular training context which is being developed. What constitutes ideal station design in a weapon system simulator (e.g., a particular workspace, size, shape and arrangement, specific displays, controls and their relative locations) may not be ideal in a flight trainer. Optimum student and instructor station design will usually be system specific.

The variability among training stations from one context to another reflects differences in the final design configuration which is adopted. In other words, student and instructor stations will differ in size, shape, and arrangement of the workspace area. They will incorporate different types and amounts of equipment. But the same general design principles are applicable to all training stations. Regardless of the particular configuration which is eventually adopted, certain basic human engineering requirements must be considered. In an analogous sense, the Piper Cherokee and F–104 cockpits differ dramatically and, yet, the same basic design guidelines can be applied to both.

14.7.2 Basic Instructional Functions

In the design of any student station, the workspace and equipment are devised to help someone *learn* to operate other equipment in another setting. In designing an instructor station, the workspace and equipment are devised for an actual operator whose task it is to *train* (Miller, 1954). Considered jointly, both training stations must be designed so that a specified set of training functions can be carried out efficiently and effectively between them.

In every training context, certain basic activities take place. These activities represent the major functions comprising the training process. Good human engineering design of training stations should reflect the following activities or functions:

Orientation. The instructor provides the student with information, principles, and ideas about the task to be learned, how it is to be performed, and how training will be conducted. Orientation is accomplished by direct conversation, by remote communication, by formal and informal demonstrations, dry runs, etc.

Stimulus presentation. The instructor selects, provides, controls, and monitors stimulus inputs to the student by operating controls and monitoring displays. Specific stimuli, problems, programs, or scenarios may be presented and monitored. Control may be exercised over the rate of presentation, sequencing of problems, problem difficulty, etc.

Response production. The student acquires input data and responds to them. The instructor monitors student responses and evaluates them.

Feedback. The instructor provides the student with timely and appropriate feedback based on student responses. Knowledge of results is supplied and suggestions are made about ways in which performance can be improved. The student monitors this feedback and attempts to adjust his responses accordingly.

Control. The instructor assumes general control of training. He determines conditions of practice (how much, when, and what intervals), establishes practice sequences, integrates component skills, etc. He effects control by operating (or monitoring the operation of) salient equipment needed to vary the content and conditions of training.

These functions emphasize that training is a dynamic process. It consists of a carefully guided and monitored dialogue between the student and instructor. The complex interaction between these participants is illustrated in Figure 14–10 (Miller, 1954).

At least indirectly, the functions described above are reflected in general design requirements of the student and instructor stations. Analysis of the training process into major instructional functions suggests that we are dealing with a specialized data transmission system. This system must be designed to permit and facilitate an exchange of information between two terminals (the training stations). Consequently, human engineering considerations should focus on the two major aspects of the system. Consideration must be given to the design and layout of the training station work area. Similarly, consideration should extend to the equipment at each station which makes possible the dynamic interplay between student and instructor.

In the sections which follow, general human engineering recommendations are discussed for four areas in which design decisions are often required. These areas include: general housing arrangements for training stations, work-sta-

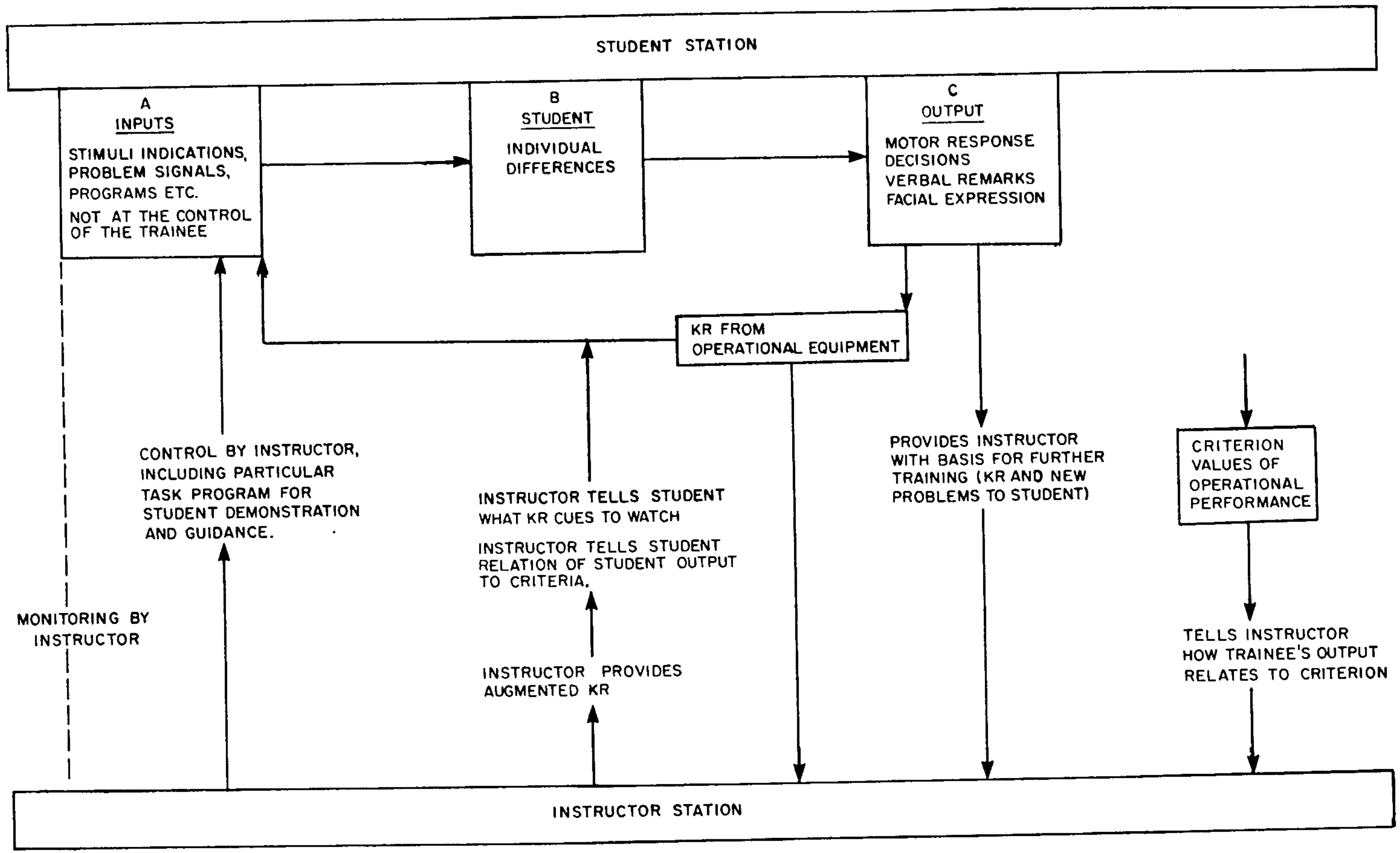


FIGURE 14–10. Student feedback loop as it is monitored and controlled by instructor. (Note: Monitoring and control may be vocal or mechanical. KR is short for Knowledge of Results or feedback, based on Miller, 1954).

tion layout, environmental controls, and equipment design. Two sets of general recommendations are provided. The first set refers to the "physical" layout and design of training stations. The second set relates to the "functional" design of training stations.

14.7.3 Physical Design Considerations

One aspect of good training station design involves consideration of "physical" design principles. Stated simply, these are the general human engineering principles and practices which should receive consideration in the design of any system, work area, or equipment. They may be employed just as readily in the design of an aircraft cockpit, computer consoles, or training stations.

No attempt will be made in this chapter to restate the many basic principles which have been treated in other chapters of the *Guide* or which appear in other standard reference documents. Instead, we wish to provide the reader with a summary checklist of human engineering considerations. The checklist will provide a convenient point of departure when design efforts are contemplated or are to be evaluated.

This checklist is a condensed version of one previously developed by Smode. et al. (1963). It is reasonably comprehensive, but by no means exhaustive. In using the checklist, the reader should pinpoint those considerations important to his design problem. By consulting the present Index of the *Guide*, he will be able to retrieve more detailed information from other chapters relating to specific design practices and recommendations.

Checklist of Physical Design Considerations

Design Area 1—Housing Arrangements

Size and shape of student and instructor areas. Sufficient space should be allocated to provide for variation in human body sizes. Arrangements should reflect the need for movement between and within areas, the size and method of operation of equipment, and communication requirements.

Traffic flow. Utilization of floor areas should follow from traffic requirements. Sufficient aisle and corridor space should be provided, and entrances, stairs, and ladders should be designed according to accepted standards.

Maintenance. Arrangement of men and equipment should consider the need for and location of maintenance areas. Sufficient space must be provided for maintenance to be performed; the use of hatches, modules, etc., should be considered in order to conserve space and promote ease of maintenance.

Safety. Facilities should be arranged with consideration for the location and movement of men relative to potentially dangerous equipments (high voltage, dynamic) and housing features (steps, ladders, etc.).

Design Area 2—Workstation Layout

Principles and criteria of good workplace design. Arrangement of equipment and associated controls and displays should take the following into account: frequency of operation, sequences of operation, functional relationships, the importance to presentation of and practice on the task problem.

Workplace dimensions. Displays should be located within the optimum viewing envelope and manual and foot controls should be within the optimum reach envelope. Seated versus standing operation should be considered with sufficient space allotted for the selected mode of operation.

Location of displays. Visual displays should be located to promote the speed and accuracy of seeing. Consideration should be given to the positioning of two or more operators who can share a display so as to facilitate multiple seeing. The viewing requirements of a mobile observer should also be considered.

Location of controls. Controls should be located to promote speed and accuracy of operation and adjustment. Consideration should be given to the location of controls shared by two or more operators to facilitate multiple use.

Design Area 3—Environmental Controls

Lighting. Quantity and quality of lighting should be consistent with general standards. General illumination should be considered with respect to: size of detail to be discriminated, brightness contrast and ratio, time available for viewing, and glare effects.

Acoustics and noise. Noise levels should be within recommended levels. The effect of noise on performance, and noise suppression devices, such as baffles and partitions, should be considered.

Temperature, humidity, and air flow. Provision should be made to maintain temperature and humidity within recommended tolerances for the range of operating conditions as well as for adequate ventilation.

Vibration and acceleration. Provision should be made to hold direction, frequency, and amplitude of mechanical vibration and forces of acceleration within recommended tolerances. The need for simulation of vibration and acceleration should be considered.

Design Area 4—Equipment

Displays and indicators. Symbolic and pictorial displays should promote interpretation of information and increase reading speed and accuracy. Displays should contain appropriate direction-of-movement relationships, be compatible with direction and amount of control movement, and be consistent with other displays. The following display elements should be considered: display scales, zone markings, labeling (position and color coding), and design of alphanumerics. Warning and caution devices for gaining operator attention and indicating the nature of malfunction should also be considered. (See Smode et al., 1963).

Controls. Selection of controls should consider the characteristics of each control type and their suitability for given applications. The following elements of controls should be considered: size, placement, direction-of-movement relationships, compatibility with display movement, control coding requirements, and methods of preventing accidental operation.

For many of the physical design considerations outlined above, specific design recommendations have been developed and can be easily referenced. In other instances, however, only general and suggestive principles are available. In these latter cases the human-factors specialist must bring his experience, ingenuity, and common sense to bear on the design problem.

14.7.4 Functional Design Considerations

The second aspect of good training station design involves "functional" design principles. These consist of guidelines devised to deal with the special problems and requirements peculiar to training systems. They include consideration of such factors as location of the instructor relative to the trainee, training station complexity, function allocation among instructor personnel, and methods of inter-station communication.

Functional design considerations are dictated primarily by experience and common sense. There are few hard and fast rules or principles for decisions about the functional design of any particular training station. To develop good functional designs, one must be familiar with the general conditions which facilitate learning. (See Chapter 13.) Careful study and analysis of how the major training functions are implemented in a particular system are also necessary. Based on knowledge of the conditions for learning and the activities or functions in training, design considerations will become evident.

In this section some major considerations are briefly discussed. In each case the attempt has been made to synthesize and integrate general principles and recommendations suggested by others. More detailed information and, in some cases, different viewpoints can be obtained by consulting Swain (1954), Miller (1954), and Smode et al. (1963).

Relative location of stations. Many situations arise in which a variety of housing arrangements are feasible with respect to the relative location of training stations. As a general rule of thumb, the student and instructor stations should be in fairly close proximity.

This arrangement is particularly desirable during the early stages of training. The instructor will often interact directly with the student during these early stages to provide orientation, demonstration, and general guidance about the task and training procedures. Even during later stages, however, this arrangement is useful. The instructor should have visual access to the student station so that he can anticipate scheduled changes in problem materials, practice schedules, etc. In this way, he can also provide very effective feedback by monitoring the student directly as well as monitoring responses which are displayed within the instructor's station.

However, student and instructor stations should not be too close together. The student should not be able to see or hear what is going on at the instructor station. These undesirable inputs represent noise to the student and may interfere with learning or practice. Similarly, if the stations are too close, the instructor may be tempted to interact too frequently or at inappropriate moments. Such interactions may disrupt learning and practice rather than facilitate them.

Student station displays and indicators. Consideration should be given to providing the student with additional displays during early stages of learning. The role of feedback and knowledge of results in the training process are vital. Consequently, the desirability of additional displays to provide added feedback should be considered. Such displays, however, should not become so numerous or be so positioned as to interfere with the visual demands of the training task. This is particularly important during the early stages.

Although the use of additional displays is often beneficial during early stages of training, this is usually not the case later on. As training progresses, the displays at the student position should gradually approximate and eventually duplicate the displays found in the operational setting. As was pointed out earlier in this chapter, it will usually pay to gradually eliminate artificial knowledge of results in attempting to facilitate transfer. Consequently, the displays provided in the final stages of learning should not condition the student to unreal feedback or expectancies.

The use of indicators or signaling devices should also be considered in designing the student station. It is generally advisable to include indicators which signify problem start and stop to the student. Another extremely useful indicator is one which signals that a communication from the instructor is imminent. If communication can be anticipated it will prove less disruptive to practice, since the student can finish his current task but will not start another complex practice sequence before the message is delivered. In addition to these uses, indicators can also be employed to provide various types of feedback such as the occurrence of a particular error; the correcting, timing, or sequencing of a response; or the correct sequencing of controls. Such aids, however, should be gradually eliminated during later stages of training if they are not found in the operational setting.

The type of indicator to employ should also be considered. Indicator lights, tel-lights, direct voice communication, as well as auditory signals (tones, buzzers, alarms) and special kinesthetic signals (i.e., automatic pilot capabilities built into the accelerator control of many automobiles) should be considered. Choice of which modality should be used is important. It is often advantageous to present special indications to the student through a modality which is not being heavily exercised already. In training tasks with primarily visual demands, effective use can be made of auditory indicators of some type.

Instructor station displays and indicators. Three types of information should customarily be displayed at the instructor station: information on the status of the training problem, information on the student's activities, and information on the student's performance (scoring and evaluation).

Information on the status of the training exercise should be displayed to the instructor. As Smode et al. (1963) point out, such information allows the instructor to maintain control of the exercise, to insert unprogrammed signals or malfunctions, and to foresee the approaching occurrence of a particular problem sequence or training equipment malfunctions. Presentations of this information to the instructor should be considered even when the training exercises have been programmed through the extensive use of automatic equipment.

Other displays are usually desired in which information about the actions being taken by the student and about task performance are presented. In many cases, it will be possible to locate the instructor station so that he acquires these types of information by directly viewing student activities and displays. When direct viewing is not feasible, important student displays should be repeated at the instructor's station. Repeater displays can be used to indicate problem status, student actions, and the effects of student actions on the task.

In some training systems the use of "observer" personnel should be considered in monitoring events occurring at the student station. Use of an observer can often take a great burden off the instructor by reducing the number of information channels to which he must attend simultaneously. If use of an observer is contemplated, however, provisions must be made for him from the start. He will require his own training station, displays, and communication links with the instructor.

Instructor station controls. It has been suggested that the most frequently encountered problem in instructor station design is that it is often too complex. On the one hand, the instructor is often faced with a vast array of displays providing information on training system status and the responses of the student to that status. On the other hand, he is often required to control and operate a tremendous amount of equipment. Extremely complex instrumentation is often involved in providing feedback to the student or in generating the training exercise.

It is essential that the instructor's station be kept as simple as possible. Attempts should be made to maximize the instructor's training and guiding activities and to minimize activities involving the monitoring and activating of the training itself. It is difficult for the instructor to effectively monitor and interact with the student if he must also devote much of his time to selecting problem sequences, varying problem difficulty, and presenting problems by manipulating and monitoring a substantial number of controls. Two general practices can be recommended. Insofar as possible, the training equipment should be automated. Second, in very complex systems, consideration should be given to the inclusion of additional training personnel.

Control of the training exercise can be accomplished automatically or manually. Automatic control offers the advantages of preprogramming syllabus items in terms of their difficulty, sequencing, and rate of presentation. It also provides for automatic recycling of the entire scenario of problems. By calling upon a particular routine (which may include malfunction checks) the instructor's work load may be reduced considerably. If provisions are made for manual overrides, for "freeze" or stop-action controls, or for the periodic insertion of special or emergency problems, flexibility can be added to the programmed system. When manual control is to be exercised, as in the presentation of an emergency problem, care should be taken to make this simple. In some systems, emergency problems can only be presented by operating a variety of controls. Insofar as possible, it is advisable to treat special system interdictions as subroutines which can be activated through a single control.

Even when extensive use of automatic equipment has been made, the instructor may still have to spend a great deal of time operating the training device. In these instances, a partitioning of the total task should be considered. Just as observers can be employed to aid in the process of instruction, so technicians can be used to aid in operation of the device. Again, however, provisions for technicians' work areas, and displays and controls must be made. Especially important are communication channels between the instructor and technician in the form of indicators, voice channels, and repeater displays. Operation of the system must remain synchronous with what the instructor is attempting to train.

Methods of inter-station communication. Careful consideration of methods of inter-station communication is essential to effective station design. There are four possible instructor communication conditions. These are shown in Table 14–2 along with the most comfortable way and the type of equipment for satisfying each condition.

In summary, good student and instructor station design involves consideration of both physical and functional requirements. The primary requirement is that the instructor is relatively available to the student. This cannot

TABLE 14–2. FOUR POSSIBLE INSTRUCTOR COMMUNICATION CONDITIONS AND ALTERNATE METHODS OF SATISFYING THESE CONDITIONS

Condition	Most comfortable way of satisfying condition	Equipment which satisfies condition
1. Instructor must be free to move about and there are extraneous interfering sounds.	a. Instructor wears mike close to mouth. b. Instructor wears receiver on ear.	Telephone-type headset.
2. Instructor must be free to move about and there are *no* extraneous interfering sounds.	a. Instructor wears a mike. b. Speaker is mounted on cabinet.	Chest mike and speaker in cabinet.
3. Instructor is not expected to move about his station much and there are extraneous interfering sounds.	Depends upon noise level.	a. If the noise level is high, a telephone-type headset should be used. b. If the noise level is not too high, earphones with the microphone mounted on the cabinet will work.
4. Instructor is not expected to move about his station much and there are *no* extraneous interfering sounds.	Instructor wears none of the equipment.	Speaker and mike mounted in cabinet.

Edgerton et al. (1953).

be the case if he must devote his energies to operation of the system itself.

In conclusion, the following steps should be undertaken in designing training stations (Miller, 1954). Implementation of these steps involves resolution of the design considerations which have been presented in this section.

Step 1. Conduct an analysis of the training goals of the projected system insofar as this information pertains to layout and instrumentation of the student and instructor stations.

Step 2. Design the trainee's station. This design includes the inputs and outputs of the trainee's station. Inputs consist of variables and channels, and programs of stimuli.

Step 3. Decide which inputs (task stimuli and knowledge of results) to the trainee should be merely monitored by the instructor and which should come under his active control.

Step 4. Decide which outputs (responses) of the trainee should be monitored by the instructor and which are not necessary for him to monitor.

Step 5. Make a time chart of the instructor's monitoring and controlling tasks based upon the time chart of the trainee's inputs, outputs, and response feedback requirements.

Step 6. Decide where the instructor should sit in relation to the trainee's station.

Step 7. Decide which monitoring and controlling aspects of the instructor's job will be done via the instructor's station equipment, which will be done via direct observation of the trainee or trainee's station equipment, and which will be accomplished by voice communication with the trainee.

Step 8. Decide what specific displays and controls are needed in the instructor's station.

Step 9. Lay out the instructor's displays and controls according to optimum placement for his required functions and sequence of activities.

Step 10. Evaluate and modify the tentative design.

References

Adams, J. A., & Hufford, L. E. Contributions of a part-task trainer to the learning and relearning of a time-shared flight maneuver. *Human Factors*, 1962, 4(3), 159–170.

Aukes, L. E., & Simon, G. B. The relative effectiveness of an Air Force training device used intact vs. with isolated parts. Research Report No. TN-57-77, June 1957, Maintenance Lab., Air Force Personnel and Training Research Center, Lowry AFB, Colo.

Ausubel, D. P. The use of advance organizers in the learning and retention of meaningful verbal material. *Journal of Educational Psychology*, 1960, **51**(5), 267–272.

Ausubel, D. P., & Fitzgerald, D. Organizer, general background, and antecedent learning variables in sequential verbal learning. *Journal of Educational Psychology*, 1962, **53**(6), 243–249.

Bilodeau, E. A. Some effects of various degrees of supplemented information given at two levels of practice upon the acquisition of a complex motor skill. Research Bulletin No. 52–15, April 1952, Perceptual and Motor Skills Lab., Human Resources Research Center, Lackland AFB, Tex.

Bilodeau, E. A., Bilodeau, I. M., & Schumsky, D. A. Some effects of introducing and withdrawing knowledge of results early and late in practice. *Journal of Experimental Psychology*, 1959, **58**(2), 142–144.

Briggs, G. E., & Naylor, J. C. The relative efficiency of several training methods as a function of transfer task complexity. *Journal of Experimental Psychology*, 1962, **64**(5), 505–512.

Briggs, G. E., & Waters, L. K. Training and transfer as a function of component interaction. *Journal of Experimental Psychology*, 1958, **56**(6), 492–500.

Briggs, G. E., Bahrick, H. P., & Fitts, P. M. The influence of force and amplitude cues on learning and performance in a complex tracking task. Report No. AFPTRC–TN–57–133, March 1957, Air Force Personnel and Training Research Center, Lowry AFB, Colo.

Briggs, G. E., Fitts, P. M., & Bahrick, H. P. Learning and performance in a complex tracking task as a function of visual noise. Report No. AFPTRC–TN–56–67, June 1956, Air Force Personnel and Training Research Center, Lowry AFB, Colo.

Brown, B. P., & Johnson, H. I. Moving-cockpit simulator investigation of the minimum tolerable longitudinal maneuvering stability. Report No. NASA–TN–D–26, 1959. National Aeronautics and Space Administration, Washington, D. C.

Buckhout, R., Sherman, H., Goldsmith, C. T., & Vitale, P. A. The effect of variations in motion fidelity during training on simulated low altitude flight. Report No. WADD–AMRL–TDR–63–108, 1963, Wright Air Development Division, WPAFB, Ohio.

Carpenter, C. R. Instructional film research reports (rapid mass learning). Human Engineering Report No. 269–7–36, January 1953. Special Devices Center, Port Washington, N. Y. (Contractor: University of Illinois.)

Cox, J. A., Wood, R. O., Jr., Boren, L. M., & Thorne, H. W. Appearance fidelity of training devices for fixed-procedures tasks. Technical Report No. 65–4, June 1965. Human Resources Research Office, Alexandria, Va.

Crawford, M. P. Concepts of training. In R. M. Gagne, (Ed.), Psychological principles in system development. New York: Holt, Rinehart, & Winston, 1965.

Dougherty, D. J., Houston, R. C., & Nicklas, D. R. Transfer of training in flight procedures from selected ground training devices to the aircraft. Technical Report No. 71–16–16, September 1957. U. S. Naval Training Device Center, Port Washington, N. Y.

Eckstrand, G. Current status of the technology of training. In J. E. Uhlaner, (Ed.), Psychological research in national defense today. Technical Report No. S–1, June 1967. U. S. Army Behavioral Science Research Lab., Washington, D. C.

Edgerton, H. A., Heinemann, R. F. D., & Barrett, R. S. Human engineering considerations in the design of the instructor's station of trailerized operational flight trainers. Technical Report No. 1042–00–1, August 1953. Special Devices Center, Richardson, Bellows, Henry & Co., N. Y.

Fedderson, W. E. Simulator research: Validation and motion studies. Presented at the 7th annual Army Human Factors Engineering Conference, 1961, Ann Arbor, Michigan.

Follettie, J. F. Effects of training response mode, test form, and measure on acquisition of semi-ordered factual materials. Research Memorandum, 1961, HumRRO Division No. 4 (Infantry), Fort Benning, Ga.

Fraser, T. M. Philosophy of simulation in a man-machine space mission system. Report No. NASA SP–102, 1966, National Aeronautics and Space Administration, Washington, D. C.

Gagne, R. M. Simulators. In R. Glaser (Ed.), *Training research and education.* Pittsburgh: University of Pittsburgh Press, 1962.

George, F. H., & Handlon, H. A language of perceptual analysis. *Psychological Review,* 1957, **64,** 1.

Goldstein, M., & Rittenhouse, C. H. Knowledge of results in the acquisition and transfer of a gunnery skill. *Journal of Experimental Psychology,* 1954, **48**(3), 187–196.

Hahn, C. P., Dayton, C. M., McNelis, J. R., & Potash, L. M. A guide for the consideration of new training devices and procedures. Report No. AIR–E16–2/66–RP, February 1966. American Institutes for Research, Washington, D. C.

Hammerton, M. Transfer of training from a simulated to a real control situation. *Journal of Experimental Psychology,* 1963, **66,** 4.

Hillix, W. A. Visual pattern identification as a function of fill and distortion. *Journal of Experimental Psychology,* 1960, **59,** 192–197.

Houston, J. A., & Kinkade, R. G. TOW conduct of fire trainer. Report No. ER 4447, July 1966. AAI Corporation, Cockeysville, Md.

Kinkade, R. G. A differential influence of augmented feedback on learning and on performance. Technical Report No. TDR–63–12, February 1963, Aerospace Medical Research Labs., WPAFB, Ohio.

Kinkade, R. G., Snyder, H. L., & Greening, C. P. Simulation of a star field. In C. A. Baker (Ed.), *Visual capabilities in the space environment.* New York: Pergamon Press, 1965.

Kinkade, R. G., Stoloff, P. H., & Theologus, G. C. Application and evaluation of Bayesian computer aiding in operational decision systems. Report No. AIR–F–81–5/70–FR. February 1970, American Institutes for Research, Washington, D. C.

Klare, G. R., Mabry, J. E., & Gustafson, L. M. The relationship of patterning (underlining) to immediate retention and to acceptability of technical material. *Journal of Applied Psychology,* 1955, **39**(1), 40–42.

Klipple, A. G. An exploratory investigation of the effects of ambiguity in training materials on sonar cue judgments. Technical Bulletin No. 65–13, April 1965. U. S. Naval Personnel Research Activity, San Diego, Calif.

Lumsdaine, A. A. Design of training aids and devices. In J. D. Folley, (Ed.), *Human factors methods for system design.* Pittsburgh: American Institutes for Research, 1960.

Lybrand, W. A., Avron, M. D., Gartner, W. B., & Scarr, H. A. Recommendations on simulation of extra-cockpit visual cues in contact flight transitional trainers. Report No. 57–4, 1957, Psychological Research Associates.

Matheny, W. G., Williams, A. C., Jr., & Dougherty, D. J. The effect of varying control forces in the P–1 trainer upon transfer to the T–6 aircraft. Report No. AFPTRC–TR–53–31, September 1953. Air Force Personnel and Training Research Center, Lowry AFB, Colo.

McGeoch, J. A., & Irion, A. L. *The psychology of human learning* (2nd ed.). New York: Longmans, Green, 1952.

Micheli, G. Augmenting feedback and transfer of training. Technical Report No. NAVTRADEVCEN IH–41, March 1966. U.S. Naval Training Device Center, Port Washington, N. Y.

Miller, R. B. Psychological considerations in the design of training equipment. Technical Report No. 54–563, December 1954. Wright Air Development Center, WPAFB, Ohio.

Muckler, F. A., & Matheny, W. G. Transfer of training in tracking as a function of control friction. *Journal of Applied Psychology,* 1954, **38,** 364.

Muckler, F. A., Nygaard, J. E., O'Kelly, L. I.,

& Williams, A. C. Physiological variables in the design of flight simulators for training. Report No. WADC–TR–56–369, 1959. Wright Air Development Center, WPAFB, Ohio.

Newman, S. E., & Highland, R. W. The effectiveness of four instructional methods at different stages of a course. Research Report No. 56–68, June 1956, Training Aids Research Lab., Air Force Personnel and Training Research Center, Chanute AFB, Ill.

Popham, W. J. Tape recorded lectures in the college classroom. *A–V Communications Review,* March-April 1961, **9**(2), 109–118.

Popham, W. J. Tape recorded lectures in the college classroom—II. *A–V Communications Review,* March–April, 1962, **10**(2), 94–101.

Rathert, G. A., Jr., Creer, B. Y., & Douvillier, J. G. Use of flight simulator for pilot-control problems. NASA Memo 3–6–59A, 1959, National Aeronautics and Space Administration, Washington, D. C.

Rathert, G. A., Jr., Creer, B. Y., & Sadoff, M. The use of piloted flight simulators in general research. AGARD–365, Advisory Group for Aeronautical Research and Development, NATO, Paris, 1961.

Rockway, M. R. The effect of variations in control-display during training on transfer to a "high" ratio. Report No. WADC–TR–55–366, October 1955, Wright Air Development Center, WPAFB, Ohio.

Sadoff, M. The study of pilots' ability to control during simulated stability augmentation system failure. Report No. NASA–TN–D–1552, 1965. National Aeronautics and Space Administration, Washington, D. C.

Sadoff, M., & Harper, C. W. A critical review of piloted flight simulator research. Report No. IAS–62–186, Institute of the Aerospace Sciences. Presented at IAS Man-Machine Competition Meeting in Seattle, Wash.: New York, 1962.

Shriver, E. L., Fink, C. D., & Trexler, R. C. FORECAST systems analysis and training methods for electronics maintenance training. HumRRO Research Report No. 13, May 1964. Human Resources Research Office, Alexandria, Va.

Smith, R. G., Jr. The design of instructional systems. Technical Report No. 66–18, November 1966. Human Resources Research Office, Alexandria, Va.

Smode, A. F. Learning and performance in a tracking task under two levels of achievement information feedback. *Journal of Experimental Psychology,* 1958, **56**(4), 297–304.

Smode, A. F., Gruber, A., & Ely, J. E. The measurement of advanced flight vehicle crew proficiency in synthetic ground environments. Report No. WADD–AMRL–TDR–62–2, 1962. Aerospace Medical Research Labs., WPAFB, Ohio.

Smode, A. F., Gruber, A., & Ely, J. E. Human factors technology in the design of simulators for operator training. Report No. NAVTRADEVCEN–1103–1, 1963, U.S. Naval Training Device Center. Port Washington, N. Y.

Swain, A. D. Guide for the design and evaluation of the instructor's station in training equipment. Report No. WADC–TR–54–564, 1954, Wright Air Development Center, Air Research and Development Command, WPAFB, Ohio.

Urmer, A. H., & Jones, E. R. Criteria for spacecraft simulation. Presented at AIAA Simulation for Aerospace Flight Conference, August 1963. Columbus, Ohio.

U.S. Naval Training Device Center. Instructional film research reports (rapid mass learning), Vol. II, June 1956. Human Engineering Report No. 269–7–61, Naval Training Device Center, Orlando, Fla.

U.S. Naval Training Device Center. Instructional television research reports. Human Engineering Report No. 20–TV–4, June 1956, Naval Training Device Center, Orlando, Fla.

Wheaton, G. R., Kinslow, W. E., & Krumm, R. L. Validation of a part-task automobile driving simulator, Final Report. Prepared for Division of Accident Prevention, USPHS, HEW, Contract No. 108–64–167, 1966, Health, Education, and Welfare, Washington, D. C.

Whittenburg, J. A., & Wise, J. E. Feasibility for research application of visual attachments for dynamic flight simulators. Report No. 2: Requirements and state-of-the-art evaluation. Report No. HSR-RR-62/10 Mk-X, February 1963. Human Sciences Research, Inc., Arlington, Va.

Wilcoxin, H. C., & Davy, E. Fidelity of simulation in operational flight trainer, Part I. Effectiveness of rough air simulation. Report No. TR SPECDEVCEN 999-2-3a, January 1954. The Psychological Corp., New York, N. Y.

Williams, A. C., Jr., & Flexman, R. E. Evaluation of the school link as an aid in primary flight instruction. *Univ. Illinois Bull.* June 1949, **46** (71).

Chapter 15

Human Engineering Tests and Evaluations

Alphonse Chapanis

Department of Psychology
The Johns Hopkins University
Baltimore, Md.

Harold P. Van Cott

American Institutes for Research
Washington, D.C.

Testing and evaluation are indispensable to systems development. The sophistication and complexity of the tests that a human engineer may be called upon to perform range from simple evaluations of blueprints and drawings to dynamic simulations, laboratory experiments, and field trials. Each method has its own peculiar advantages and disadvantages, its own set of utilities and limitations. This chapter gives some procedures and principles that are useful in selecting, designing, and reporting human engineering tests and evaluations.

CONTENTS

An earlier version of this chapter was reviewed by Martin I. Kurke.

15. Human Engineering Tests and Evaluations

15.1 The Place of Testing in Technology and Systems Development

Testing and evaluation are almost synonymous with manufacturing and technology. Most manufacturing enterprises have their own quality control or quality assurance departments to monitor the products they turn out. Although most of this work is done to insure that products meet engineering specifications, many industries are broadening their concepts of inspection to include human considerations as well. In addition, thousands of consumer products and individual items of equipment are compared and evaluated against each other yearly by hundreds of government and private testing laboratories. Among the criteria used in these tests are some directed toward the needs of human users: utility, ease of use, convenience, safety, comfort, and cost. Finally, many laboratories are engaged in the human engineering evaluation of equipment and devices as part of basic or applied research programs.

No good evaluation can entirely ignore (a) the people who will most likely be using the product, or (b) the larger system into which the item will be fitted. Nonetheless, the emphasis in a great deal of testing is on the evaluation of specific pieces of equipment, unassociated with any particular system.

However important it may be to test individual products, devices, or gadgets, some of the most critical tests and evaluations are those conducted as part of the development of a larger, integrated system. Developing a new system is a dynamic and complex process. In the first place, systems engineering is always concerned with novelty and so involves elements of invention and creativity. The systems engineer never merely duplicates exactly a system already in operation. Copying another system is a manufacturing job. Systems under development are systems that have never been built before; to some extent, there is even some uncertainty about whether they can be built.

In the second place, many modern systems are large, so large that they are seldom designed and constructed in one place. Usually, different parts of the system are developed simultaneously by subcontractors who may be physically separated from each other and from the prime contractor. Although this kind of parallel development generally shortens the time needed for the design and construction of a new system, it involves certain risks—among them the risk that the different parts of the system may not function harmoniously when they are brought together.

These characteristics of systems development create the need for a carefully planned program of tests and evaluations. Such a program serves three general functions:

1. To Improve the Quality of Design Decisions. Systems development involves a series of choices and decisions. In order for the parallel development of different parts of a system to be successful and to reduce the total time needed for design and construction, the decisions that are reached in the design of parts of the system have to be firm. Once made, these decisions are often irrevocable. One important function of systems tests is to provide a sounder basis for making such design decisions.

2. To Integrate Hardware and Personnel. Testing is required throughout the development of a system to get better matches between hardware and the personnel who will operate, control, and maintain it. Since we do not know enough to forecast precisely the best ways of

integrating man and hardware, the only reliable way of validating designs is by "trying them out."

3. To Correct Design Deficiencies Early. Testing throughout all phases of systems development helps to detect design deficiencies before they become frozen. Once the detailed design of a system has been completed and a prototype is under construction, it is usually expensive and time-consuming to make major design changes. Design deficiencies must be detected early to avoid costly errors.

Test and evaluation programs carried out in conjunction with systems development can be divided roughly into two major classes: one directed primarily toward the hardware components of the system, the other toward the personnel components of the system. This distinction is by no means a hard and fast one. It is rather a way of conveniently indicating where the major emphasis lies in that part of the program.

15.1.1 System Test Programs

System test programs place their primary emphasis on the hardware components of the system, or as it is sometimes referred to, the hardware subsystem. It is often convenient to break these test programs into three phases, corresponding roughly to stages in the development of the system itself.

1. Subsystem Development Test and Evaluation. This part of the test program consists of testing individual components and subsystems as they are being developed. The primary purpose of these tests is to determine whether the product or the process meets the requirements that have been established for it. When deficiencies or faulty designs are discovered, there is an opportunity for redesign at an early stage in development and for reevaluation of the altered components or subsystems. Some of the specific human factors objectives that can be met by these tests are:

a. To determine how well component or subsystem design conforms to good human engineering practice.

b. To evaluate provisions for life support, escape, survival, and recovery of personnel when applicable.

c. To determine the reliability and maintainability of components or subsystems from a human factors standpoint.

d. To identify skills that will be required by the personnel in the system.

e. To evaluate and refine initial requirements for personnel and training.

f. To determine whether equipment being developed for training meets performance or procurement specifications.

g. To allow training specialists to become familiar with elements of the system so that they can better develop instructional materials for later training programs.

2. Systems Development Test and Evaluation. This part of the testing program focuses on subsystems as they are integrated to form a complete system. Some of the human factors objectives of these tests are:

a. To determine how well the integrated system design conforms to good human engineering practice.

b. To evaluate provisions for life support, escape, survival, and recovery of personnel.

c. To evaluate in realistic environments how well the entire system can be maintained with the resources and personnel allocated to it.

d. To determine if the product or process complies with specifications.

e. To evaluate possible new design changes before they are incorporated into the production model.

f. To evaluate the system's capabilities and limitations under actual or simulated climatic conditions.

g. To evaluate under realistic conditions whether the system is operable, effective, and compatible with other systems and supporting equipment.

h. To determine if the system is capable of performing its intended mission.

i. To familiarize personnel with the developing system and to get limited training on it.

3. Operational Tests and Evaluations. This part of the test program is done with an operational model of the system and is more concerned with the use of the system rather than its design. Some of human factors objectives that can be met in this phase of testing are:

a. To determine operational difficulties with the system and to identify ways to improve deficient products or processes.

b. To determine the operational usefulness of the system and to develop effective methods and standards.

c. To refine maintenance requirements in terms of personnel, skills, and training, and to establish performance standards for carrying out maintenance tasks.

d. To collect data on organizational and personnel skills and on training to refine requirements established earlier in the program.

e. To evaluate the adequacy of the manning structure, the table of organization, and the training program.

f. To provide final training for operational crews and to integrate the entire system into the framework of existing systems and operations.

15.1.2 Personnel Subsystem Test Programs

Since good system test programs cannot be divorced from personnel considerations, much of the data that comes out of such programs can be used in personnel subsystem design. The personnel subsystem test program, however, places greater emphasis on personnel than on equipment. Some of the objectives of this program are:

1. To evaluate whether the system can be operated, maintained, and controlled by the personnel assigned to it.
2. To determine the effect of human performance on system performance and vice versa. This objective is aimed at discovering critical inadequacies in the man–machine relationship and changes that will improve man–machine compatibility.
3. To develop valid qualitative and quantitative personnel requirements, selection procedures, manning documents, and organizational tables.
4. To evaluate individual, team, operational readiness, and other training programs.
5. To evaluate training equipment and supporting materials.
6. To evaluate job aids, technical publications, and other tools for training and for assisting on-the-job performance.

15.2 Overview of Human Engineering Testing

. What is meant by a human engineering test? How does one evaluate the adequacy of human engineering design in a piece of equipment or a system? What are some of the dangers to be avoided in this kind of work? These are some of the main questions to which we now turn.

15.2.1 Products and Processes Tested in Evaluation Programs

It is virtually impossible to formulate a concise set of rules to guide one through all the intracacies of human engineering tests and evaluations. As we have already seen, human engineering tests conducted as part of a program of system development cover a wide variety of products and processes: individual pieces of hardware or components, workplaces, entire systems, training equipment, job aids, publications, selection procedures, training programs, manning tables, tables of organization, and tactics and strategies for the employment of systems. Rather than attempt the impossible task of covering all these applications, this chapter concentrates on test and evaluation procedures for the first three products—those with which the practicing engineer is most likely to become directly involved.

Even so, the products that remain are still diverse. Equipment comes in an infinite variety of shapes, sizes, and forms. It ranges from lip microphones no larger than a postage stamp to worldwide communication networks containing subsystems that are literally out of this world. At one time an engineer may be required to evaluate something as simple as a drafting table, at another time something as complex as a computer-based tactical operations center. Details of the methods to use for tests such as these are almost as diverse as the equipments and the human tasks that are associated with them.

15.2.2 Reference Books and Sources

Many good books describe designs that can be used for planning experiments and the precautions that should be taken in applying these designs (for example, Campbell and Stanley, 1963; Cochran and Cox, 1950; Edwards, 1968; Kempthorne, 1952; Lindquist, 1953; Myers, 1966; Natrella, 1963; and Winer, 1962). But in the practical planning of human engineering tests, books like these are not enough. Writing down a few key words, looking them up in an index, and then picking out a test plan will not do the job. The principal difficulty with most books on this topic is that they are concerned only with the design of laboratory experiments. They seldom consider the many limitations (time, money, complexity of the problem or equipment, and restrictions in the number and availability of subjects) that confront the human factors specialist who tries to conduct carefully planned tests in this area. Nor do they consider many issues that are largely ignored in laboratory experimentation but are highly relevant for the practical man—issues, such as:

1. The representativeness of the subjects who will participate in the tests;
2. The fidelity with which the test equipment mimics the equipment it is supposed to simulate;
3. Statistical versus practical significance; and
4. The relevance of the criteria that are used in the tests.

Finally, most books on experimental design deal only with that topic. They do not mention other methods—questionnaires, accident data, observational methods, and so on—that can be used in the evaluation of hardware and systems.

Two books that help satisfy the practical needs of the human factors specialist are Meister and Rabideau (1965) and Chapanis (1959). The former, however, provides only a scanty introduction to the more precise techniques of experimental design. The latter lies somewhere between the one by Meister and Rabideau and those that deal with experimental design as an abstract statistical problem. To sum up, no one book provides a thorough and complete coverage of the field of human engineering tests and evaluations.

Collectively, the books referred to above describe certain general rules, certain basic tactics, that apply to all test methods. This chapter tries to (a) identify and list the most important of these and (b) outline some of the more important strategies that should be followed.

15.2.3 The Essence of a Human Engineering Test

What do we mean by a human engineering test? When we test an item of equipment, we try to measure or to estimate its value, its quality, or its worth under a certain set of conditions. When we speak of a *human engineering* test we refer specifically to the value or worth of an item of equipment when it is used by an operator or a maintenance technician. A human engineering test tries to answer such questions as: "Can an operator use it? If so, how well? Why is it good or bad in terms of the performance required of operators who are part of a system?"

A piece of equipment, or an equipment system, is always designed and built to serve some human purpose. In the case of highly automatic systems, the human being may interact with the system only as a designer, builder, maintenance man, or user. Even though these are more restricted roles than man plays in less highly automated systems, man is still an essential part of the system. The basic premise underlying all human engineering testing, therefore, is that equipment cannot be evaluated independently of the human beings who will work with it.

15.2.4 The Predictive Power of Various Test Techniques

When we make a test or an evaluation, what we hope is that the outcome of the test will enable us to make valid statements about how the system will perform in the real world. The principal purpose of any test is as a predictor, a way of forecasting what will happen when the equipment or system is actually put to work. Test and evaluation techniques differ greatly in their predictive power. Some give very good indications of how equipment and systems will perform in the real world; others give results that have to be interpreted with great caution.

The variable that largely determines the predictive power of a test is the fidelity with which the test conditions mimic or match those in the real world. The situation can be roughly represented as in Figure 15–1. The real world is represented at the left of the diagram. That, of course, is what tests try to predict. The most direct way of finding out about the behavior of systems in the real world is to make observations and measurements on systems that are at work. Unfortunately, sensors, recording devices, and observers are unnatural, and the very process of making observations means that we distort conditions somewhat and so lose something in fidelity. If concealed recorders, one-way vision mirrors, and similar devices are used, workers and operators may not suspect that they are being observed and may, under these circumstances, behave normally. As soon as people know they are being observed, however, behavior is no longer spontaneous and natural. In addition, if recording devices and observers actually interfere with operators, the amount of distortion and the loss in fidelity may be considerable.

Field studies have less fidelity than do observations and measurements made on functioning systems primarily because field studies are contrived or artificial situations.

Simulations, models, and games cover a wide range on the fidelity scale. Some simulations are highly realistic. At the other extreme, some models, for example, mathematical models (shown as a separate category in Figure 15–1), are so highly abstract that they bear little resemblance to the real world they are supposed to represent.

Laboratory experiments also cover a wide range along the fidelity scale. Highly realistic laboratory experiments seldom approach a good simulation in fidelity, primarily because a laboratory experiment is, by its very nature, an artificial and constrained situation. Still, laboratory experiments can be made so realistic that the subjects in them become almost as involved in the experiment as they would in a real-life situation. At the other extreme, laboratory experiments may become highly abstract and remote from the real world they are supposed to test.

The fidelity of test conditions is, unfortunately, directly related to cost and inversely related to the ease and flexibility with which tests may be made. A highly realistic simulator is an expensive item, and the greater the fidelity of the simulator the greater the cost. At the extreme, when the fidelity of a simulator approaches that of the real world, its cost approaches that of the real world system too.

The flexibility with which tests can be made is inversely related to the fidelity of the test technique that is used. In general, it is very dif-

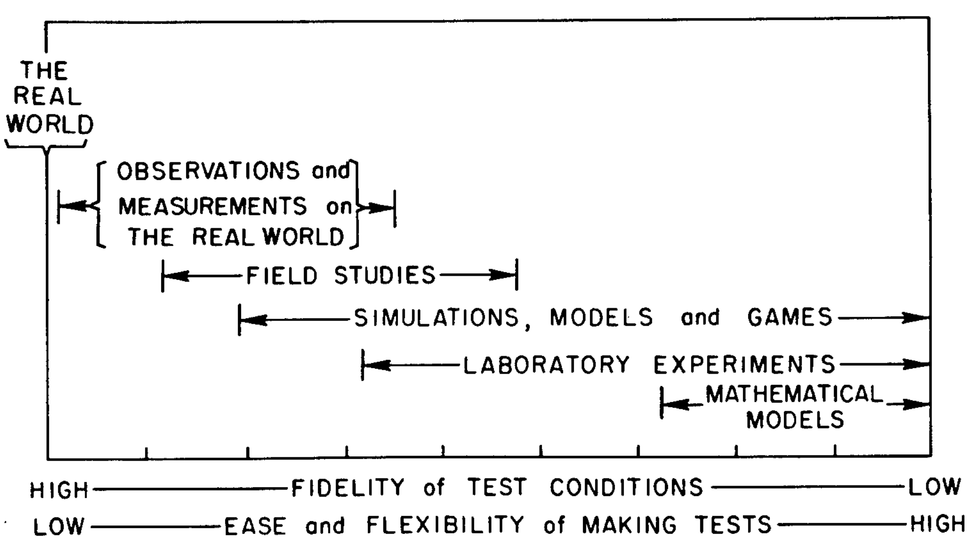


FIGURE 15–1. The fidelity of various test techniques and the flexibility with which they may be applied.

ficult to manipulate systems that are operating in the real world and to observe what happens as they are manipulated. Laboratory experiments permit much more flexibility in this respect and mathematical models can be manipulated even more freely and at less cost. However, there is a real dilemma here. Test and evaluation techniques that permit the greatest flexibility are those that have low fidelity. Techniques that have high fidelity, and so greater validity for some real-world application, are generally more constrained and restricted. Arriving at the correct balance between these two conflicting requirements can be a difficult decision.

15.2.5 Variables that Contribute to the Fidelity of a Test

Since it is not practical to include every system input, output, and performance characteristic in a test, an important part of designing a human engineering test is deciding which real-world features should be reproduced in the test situation. In general:

1. It is not necessary to include in tests those input-output variables that have a negligible effect on overall system performance.
2. It is necessary to include those variables that have an important effect on system performance.
3. Human engineering tests should make use of equipment, personnel, procedures, environmental conditions, and feedback stimuli which duplicate those that will eventually appear in the operational system and environment. The validity of any test procedure is determined to a considerable extent by how successfully the human engineer has been able to identify and include such critical variables in his test and evaluation. The importance of this principle will appear again and again in the material that follows.

15.2.6 Some Purposes of Human Engineering Tests

Human engineering tests serve a variety of purposes, and test techniques should be selected accordingly. Those listed below cover the full range of research and development activities that human engineers might engage in, not simply those associated with systems development.

Exploration. One important function served by human engineering tests is exploration, to discover things that are wrong, and to formulate hypotheses and suggestions for further work. Tests of this kind are usually done on operating systems or prototypes, although they may also be done on simulators and even functional mockups.

It is sometimes obvious that an operational system is not performing well. The purpose of the human engineering evaluation at this point is to try to find out what seems to be wrong with it. Examples might be an air traffic control system, the highway traffic system, or a post office. Before it is possible to do any constructive redesign of such a system, one has to know first what the difficulties are.

Exploratory evaluations ask open-ended questions. In a sense, the evaluator is not exactly sure what he will find out, or even what questions to ask. Techniques that are useful in these unstructured situations are operator opinions, formal questionnaires, interviews, accident data, critical-incident studies, and some methods of direct observation (see Chapanis, 1959). In general, mathematical models are not helpful for this kind of work. Simulators and functional mockups can be used but only if they have high fidelity and match some system closely both in form and function.

Measurement of behavioral constants. Another important purpose served by human engineering tests is the determination of behavioral constants, that is, numerical values needed for system design or for the construction of a mathematical or other kind of model. Examples of questions might be, "What's the probability that a well-trained operator will make an error in using this particular data-input device?" "On the average how long will it take for experienced operators to respond to this particular warning light?" Or, "On the average, how long will it take experienced commanders to make a decision about whether to proceed on course or to change course?"

Data of this kind are best obtained from measurements on operational systems in actual situations. One difficulty with this approach is that the data are usually needed before a system has been constructed. The next best alternative is to try to collect such data from systems that resemble the one undergoing design and development. Less satisfactory alternatives are to collect data from simulators, functional mockups, and laboratory experiments. As indicated earlier, such substitute procedures give biased data. Subjects who are being tested on a simulator or in a laboratory experiment know that they are being tested in an artificial situation and their responses are not typical of what would be obtained in a real environment. A simulator, or an experiment, presents a relatively pure or abstract situation uncontaminated by all the stresses, interferences, and complexities of a real work situation. For example, reaction times of automobile drivers to traffic lights during real driving are typically quite different from reaction times obtained on driving simulators in the pure context of a laboratory. Similarly, the pressures that people can exert on brake pedals when they are driving real automobiles are different from the pressures that are exerted on brake pedals when these same people are tested with realistic functional mockups in a laboratory. To sum up, behavioral constants obtained on simulators, functional mockups, or in laboratory experiments must be interpreted with great caution.

A final consideration is that specific questions of fact can often be put into a broader context and so yield much more useful data than the answer to the question as it was originally posed. For example, instead of asking, "What's the probability that a well-trained operator will make an error in using this particular data-input device?" we might ask, "Is the probability of error higher on data-input device A or B?" Or, instead of asking, "On the average, how long will it take for an experienced operator to respond to this particular warning light?" we might ask, "What's the average reaction time we should expect to this warning light if we systematically change its intensity? Its color? Its location?" Dependable answers to more general questions of this type are usually easier to obtain even on simulators or from laboratory experiments. Illustrations of these usages are discussed below.

Comparison against a standard. Another purpose of human engineering tests is to discover whether a piece of equipment, or a system, meets some standard. The standard may have been arbitrarily established, or may have been established by the customer. Standards that may be specified for a piece of equipment or a system cover a wide range. At one extreme, the standard may be an overall specification of system performance, such as, "The vehicle must be capable of flying at Mach 3," to very specific and detailed design considerations, such as, "The design of controls and displays must conform to MIL-STD-1472." Questions that the human engineering tests try to answer are: "Will the equipment do this well?" or "Does the equipment meet the specifications?"

Devices that may be used to assist in arriving at answers to such questions are drawings and diagrams, static mockups, functional mockups, simulators, prototype models, and actual functional systems. Some of the techniques used with these are discussed later.

Comparisons of two or more items of equipment. A common kind of human engineering test asks, "As it is used by the typical operator, is this piece of equipment better or worse than some other piece of equipment?" The precise wording of this question may, of course, take any of an infinite variety of forms:

1. Can one track a meteorological balloon more accurately with a new recording theodolite than with an existing standard theodolite?
2. Can two soldiers erect portable mast X more quickly than portable mast Y?
3. Can one hear better with this earphone or that one?
4. How do these six absolute altimeters compare in terms of speed and accuracy of reading?

Equipment comparisons may end with simple appraisals of relative rankings, or may carry some quantitative statement about the size of the difference between equipments. In the former case, the results say simply that A is better (or worse) than B, but not by how much. If several alternative equipments are involved in the comparison, the results may enable the evaluator to say that A is better (or worse) than B, that B is better (or worse) than C, and so on. Occasionally, one may want to know only that A is not distinguishably worse than B. This happens, for example, when A and B, representing two alternative ways of doing something, have greatly different costs associated with them. If tests show that operators can get approximately the same performance out of A and B, this simple answer might save a considerable amount of money.

Sometimes, however, one needs to know not only that A is better than B, but by how much it is better. There are two basic ways in which differences may be quantified. One is to say that A is so many percent better than B. The other is that it is better than B in terms of some numerical index or score. The former is meaningful only when the data are measured on what is known as a ratio scale (see Stevens, 1951). There are very few ratio scales in behavioral measurement and those in common use are based on such fundamental physical scales as, for example, time, length, weight, density, and resistance. Unless one is sure that he is dealing with such a scale of measurement, he should be cautious about stating results in terms of A being so many percent, or so many times, better (or worse) than B.

Examples of quantitative statements resulting from human engineering tests might be:

1. Operators made 0.015 errors per trial on this console and 0.004 errors per trial on that one.
2. The mean-time-to-repair for equipment A is twice that for equipment B.
3. The articulation index (AI) for this radio receiver is 0.70, for that one 0.35.
4. Twenty-three percent more users preferred water canteen A over canteen B.

As a general rule, it is simpler to devise human engineering tests to yield comparative statements of relative value between equipments than to devise tests to give estimates of an absolute quantitative difference. The strategies for the former are less demanding because the measurement techniques can be cruder. The correct estimation of differences between equipments usually requires more attention to experimental design, more rigor in measurement, more trials, and more sophisticated analyses of the data. This increased attention to technique in establishing exact differences is needed to get the most efficient test possible, that is, to get data with whatever precision is needed but with the fewest possible trials.

Establishing performance characteristics. In the use of completely new types of equipment, there may not be any other similar type of equipment available for comparison and there may not even be a performance standard against which to evaluate the equipment. Under these circumstances, the primary purpose of the evaluation is to find out how well people can operate or maintain the equipment. The question asked under these circumstances may take such forms as, "What can we expect of this item?" or, "Can the operator use it?"

Somewhat similar conditions hold when human factors specialists are engaged in research on human performance under new and exotic circumstances. An example is the work that was done when explorations were first begun on human performance under conditions of weightlessness. Now that we know some answers, it is difficult to remember the uncertainties that preceded that early work. But, before man had ever experienced orbital flight, we had almost no basis for predicting how he could perform in zero-G. The purpose of the research that was initiated at that time was to discover what man can do under these conditions.

With a system having new applications, it is usually better to describe performance in terms of one or more functional relationships. An example is a curve showing the amount of information that can be correctly received by an operator over a new communication system as a function of the distance between transmitter and receiver. Other curves might show losses in visual acuity as a function of the frequency and amplitude of vibratory forces applied to the body. The performance represented by such curves begins to show what the potentialities are for the equipment or the system. In addition, functional relationships reveal the combination of operating conditions under which one can expect best performance. In this way, tentative scores or performance criteria are eventually generated.

In many types of new equipment that essentially represent an extension of human senses (e.g., rangefinders, detection devices, acoustic magnifiers, etc.), an appropriate but seldom-used standard of comparison is the average capability of an unaided person using the most relevant sense modality. For example, a meaningful "cost effectiveness" ratio for a new target detection device might take the following form:

$$\text{Gain per dollar} = \frac{\left[\begin{array}{c}\text{Targets detected}\\ \text{with device}\end{array}\right] - \left[\begin{array}{c}\text{Targets detected}\\ \text{with no}\\ \text{performance aid}\end{array}\right]}{\text{Cost of device in dollars}}.$$

15.3 General Principles of Human Engineering Test Design

Human engineering tests and evaluations come in so many forms and serve so many purposes that it is difficult to set up a highly detailed set of steps that will be valid for all of them. Nonetheless, there are certain general principles that apply to most human engineering tests and evaluations.

15.3.1 Determining the Purpose of the Evaluation

The first, most important, and most exacting part of preparing a human engineering evaluation is defining the "test objective." One thing that must be clearly established at the start is the general purpose of the test and evaluation. Is the test's purpose exploratory, or merely open-ended interrogation, or is it probing? Is the purpose of the test the determination of some behavioral or performance constants? If so, what constants are being sought? Is the test intended to compare the equipment or the system with a set of specifications, or against some standard? To compare two or more items of equipment? To evaluate the functional performance of the system? Deciding on the purpose of the test helps to narrow the choices from among all the possible kinds of test and evaluation techniques that are available.

Deciding on the *general* purpose of a test and evaluation is still not enough, however. One also needs to prepare a specific statement of what the test should be designed to accomplish. Unfortunately, many human engineering test objectives are too vague. An objective which states, "To evaluate the performance and operational accuracy of the equipment, and to determine its suitability for field use," is only a standardized introductory statement. It needs to be made much more precise. Even if the purpose of a test is exploratory, one still needs to know within what areas to concentrate his efforts. Will probes be specifically directed towards the displays, controls, communication problems, adequacy of training, maintenance problems? No test and evaluation can hope to examine the whole world. The more precisely one can specify what segment of the world the test is going to examine, the better that test can be planned.

For example, the question, "How do you evaluate the human engineering design of an automobile in terms of its performance?" can be interpreted in many ways. Performance can mean the quality of the ride on turnpikes, on dirt roads, or on mountain curves; the road-handling qualities of the car in dry weather, or in rainy weather; the traction delivered by the rear wheels in dry weather, in damp weather, in snow, or on ice; the ease with which the car can be parked along a street by a large man, a small woman, or an elderly person; the fuel economy of the car during normal city driving, or high-speed turnpike driving. And so on. An imaginative person can ask questions about hundreds of different ways in which the performance of an automobile can be evaluated. Moreover, the tests needed to answer these questions are all different and not necessarily related to one another.

Rarely, of course, is a single kind of performance tested. Decisions have to be made among all the measurable kinds of performance. Although there are no infallible rules for telling anyone how to select performance questions that are most relevant and sensible, a good strategy is to concentrate on those that are critical to the performance of the system. First, prepare as exact a list as possible of all the performance requirements for the equipment or the system. Sources of information for the construction of this list are written statements, past experience, conversations with operational personnel, interviews with customers or purchasers of the equipment, and tactical doctrine. The list should consider all major equipment items and all system and subsystem activities, such as scheduled and unscheduled operations, maintenance, and emergency situations.

When the list of performance requirements has been assembled, the various entries should be rated. Critical performances are those that have to be performed in a certain way to avoid serious adverse effects on the functioning of the system as a whole. Such performances should be given high priority in the evaluation. Critical performances include:

1. Time-critical performances which directly affect the timing and successful completion of operations and cycles that follow.
2. Accuracy-critical performances which affect the behavior of the system.
3. Cost-critical performances which affect the cost of system operation or the efficiency with which it operates.
4. Vital, non-redundant links in an operational cycle, such as an order to prepare a missile for launch. If such an order is not passed, or if it is garbled, the entire system may fail to attain its mission. Vital, non-redundant steps in operational cycles must always be considered critical.

Throughout, one important rule applies:

The more precisely one can state test objectives, the better.

15.3.2 Consulting the Literature

Having defined the test objectives, the next step in setting up a test plan is to consult the literature. The value of doing a literature search varies greatly with the nature of the proposed human engineering test. In general, the more basic the question, the more likely it is that a literature search will turn up material of value. Conversely, literature searches are less likely to be of value for evaluations of specific pieces of equipment or systems.

It is often possible to find in the literature useful suggestions about ways of carrying out the test, the approximate number of subjects to use, variables that may be appropriate to investigate, and ways of collecting meaningful data. The human factors literature is a large reservoir of accumulated facts, data, and experience. A few hours of library work often pay off handsomely in planning a human engineering test.

Sources of information. This *Guide* contains a valuable list of selected references. Another good approach is through the journal, *Psychological Abstracts,* copies of which are to be found in all major libraries. The *Abstracts* give complete citations and abstracts of all relevant articles published in scientific and technical journals, all relevant books published through regular publishing channels, plus some documents that appear through irregular channels. A major section of the *Abstracts* is called "Military and Personnel Psychology." Citations in this section appear under such subheadings as "Training," "Task and Work Analysis," "Performance and Job Satisfaction," "Special Environments," "Engineering Psychology," and "Driving and Safety."

Certain journals, for example, *Human Factors* and *Ergonomics,* are entirely concerned with human factors. Other journals, such as the *Journal of Aviation Medicine* and the *Journal of Applied Psychology,* contain a substantial number of articles on human factors and should be scanned for appropriate material. All these journals are published through regular media and can usually be found in any large college or university library.

In addition, there is a vast document literature in human factors consisting of memorandum reports, technical reports, technical notes, technical documentary reports, and various other kinds of reports issued irregularly and with limited distribution from government laboratories and other organizations under contract to government agencies. To these must be added a large number of company reports prepared by departments or laboratories within particular industries and circulated generally only within the industry. Because all these reports appear irregularly and in limited quantities, they seldom get into regular library channels and so are difficult to find. The document literature also lacks continuity. Government laboratories change names often and laboratories outside of the government come and go depending on the availability of funding. Most of the document literature is not subjected to the normal critical editorial supervision that is applied to articles and books. As a result, the document literature is often poorly written, badly edited, and sometimes technically suspect.

For all that, there is a great amount of useful information contained in this document literature if one can locate it and screen it appropriately. The best single source of information about it is the series of *Human Engineering Bibliographies* prepared by the Human Engineering Information and Analysis Service of Tufts University under contract first with the Office of Naval Research and later with the U.S. Army Human Engineering Laboratories at Aberdeen Proving Ground, Maryland. Tens of thousands of reports, articles, and books are listed in these bibliographies and are abstracted and indexed by subject matter, author, and issuing agency. Unfortunately, the series was terminated in 1966. The more recent document literature can now be located only through government abstracting agencies.

15.3.3 Determining the Kind of Evaluation that is Most Appropriate

The next step in setting up a test plan is to select a test procedure that will meet the test objectives by providing the kinds of answers that the test objectives require. The possibilities here are many. Some of the more common ones are:

1. Opinions. Informed opinions by experienced engineers and operators can be an invaluable source of data. The engineer often knows about some of the kinds of difficulties that have been encountered in the past, and special problems that he might anticipate in the future. The operator who has used similar equipment often may have ideas about the kinds of problems, difficulties, and malfunctions that are likely to occur in the field. For this reason, the novice, or the outside consultant, because of his lack of familiarity with the equipment, may easily overlook faults that are apparent to a more practiced eye. On the other hand, the man who works with a system day-in and day-out may become adapted to the shortcomings of his equipment and so take them for granted. In these instances, an outsider, with a fresh viewpoint, can sometimes see difficulties with a system that are often overlooked by the people who are close to the operation.

Despite their value, opinions can be misleading and dangerous, because they are often strongly biased, subjective, and sometimes even whimsical. They should be substantiated with observable data and well documented reports of failures, errors, and malfunctions. Opinions bolstered by observations gain in credibility. Opinion data can also be refined through the use of rating scales, forced-choice comparisons, rankings, and semantic differentials. Applying these more sophisticated opinion collection techniques requires the help of an expert since they are too complicated to summarize adequately here.

Opinion data, however collected, can only be applied to certain kinds of evaluations, namely those in which there is an equipment, a system, a model, or a prototype, on which people have been able to acquire some experience and about which they may have opinions. Opinions are generally of little value for the evaluation of novel systems, for systems that will be used in strange and exotic environments, for the determination of behavioral constants, and for the determination of performance characteristics.

2. Accidents, Failures, and Breakdowns. A second major source of data for human engineering tests comes from reports of mistakes, accidents, near-accidents, failures, and breakdowns that occur during the operation and maintenance of equipment. For these purposes, human engineers make no distinction between accidents and near-accidents. Both identify potentially serious or hazardous situations. Accident, or near-accident, data can be collected only from systems, prototypes, or models. They cannot, in general, be collected from earlier stages of systems development and have limited usefulness for the design of radically new equipment and systems. Accident and near-accident data also suffer from the limita-

tions that they are generally infrequent events, they occur at unexpected times when people are not prepared to observe them, and they are strongly dependent on human memory. Nonetheless, carefully collected accident data can provide some useful clues to the improvement of equipment and systems. Well-established procedures for collecting data of this type can be found in a number of sources (for example, Chapanis, 1959).

3. Checklists. A human engineering checklist is similar in some ways to a questionnaire. It contains a list of items pertaining to the human engineering aspects of equipment and one or more evaluators rate the adequacy of the equipment on each item. Checklists are most useful in the early phases of systems development where concern is primarily with static human engineering design properties such as the placement of displays and controls for adequate viewing, the labeling of instruments, and other features not directly related to man–machine performance. An example of such a checklist is one developed for the evaluation of early engineering plans (Berkun and Van Cott, 1956).

Checklists have limited usefulness in the evaluation of dynamic human interaction with equipment. No one has yet constructed a checklist that can determine how well a pilot uses a joystick, or how well a teletype keyboard operator performs. Measurements of this kind need to be observed and recorded directly.

4. Other Observational Techniques. Useful data for evaluations can be collected by observing a system while it is operating under near realistic conditions. A number of different observational techniques are available. Some of these, for example, activity sampling, micromotion study, and process analysis, have been borrowed from industrial engineering. Others, such as link analysis and some varieties of activity analysis, have been devised and modified by psychologists, systems engineers and others (see, for example, Chapanis, 1959). Each of these methods serves a different function. Each is good for collecting certain kinds of data and for answering certain types of questions. The method has to be appropriate to the problem.

Methods of direct observation are best applied to systems in operation. When systems are under development, direct observations may be applied to prototypes, models, and in some cases, mockups and even drawings. When it is undirected and without specific purpose, however, direct observation can be costly and wasteful. To be most useful, specific objectives should be set up in advance, special sheets should be designed for the easy collection and objective recording of data, and the observers should be trained or, at least, well rehearsed in what is to be observed and how they are to observe it.

5. Mockups. Mockups are both a system product and a method for evaluating systems in certain stages of development. Because of their importance, mockups are considered later, in Section 15.5.

6. Simulation, Modelling, and Gaming. Being simpler, cheaper, and easier to construct than the systems they mimic, simulators and models are often used to arrive at predictions about the behavior of a system before the system is actually constructed. This is such an important class of techniques that they are discussed later in Section 15.6.

7. Experimental Methods. Laboratory experimentation is generally conceded to be the most precise and powerful method of collecting data for purposes of human engineering tests and evaluations. The power of experimental methods comes from the fact that variables are deliberately manipulated under controlled conditions. In addition, the experimenter sets up his conditions of observation so that he is prepared to observe what happens when it happens. This embraces such an important class of methods that they are discussed separately.

Measurement of the usual dependent variables (errors, accuracy, speed, etc.) is not feasible during field trials of many equipment items. This frequently is the case when such measures would themselves bias the dependent variables or when no meaningful measurements are possible (e.g., troops testing two rucksacks during routine field exercises; troops confined in vehicle carriers with no tasks to perform). In this type of test the pre–post test design has proven useful. Prior to the equipment test, experimental subjects are measured on a series of relevant measures (e.g., physical coordination, vigilance, cognitive tests, etc.); following the equipment tests, subjects are again measured. Decrements in performance represent the effects of the equipment item. The use of a control group usually strengthens the conclusions of such a test.

15.3.4 Determining the Setting of the Evaluation

Simpler kinds of evaluations, for example, the evaluation of drawings, are usually easy to arrange and to conduct. More complex tests and evaluations, on the other hand, require more careful planning and involve a number of administrative questions.

1. Should the test and evaluation be done in-house or contracted out to some other agency?
2. Where should the test be conducted? At the work site? In a laboratory? In an office? In the field?
3. What department, group, or agency will be responsible for seeing to it that the test is supervised and run, and that a report is written?
4. Who will be in charge of the test and evaluation? What will be the composition of the evaluation team and who will be the experimenters, evaluators, and observers?

Although these are all largely administrative questions, they must be answered before any successful test can be undertaken.

15.3.5 Defining the Full Range of Operating Conditions

A good human engineering test must take into account the environmental and operational conditions under which men and equipment will work. Trucks which can easily be driven in temperate climates sometimes cannot be used in the arctic where drivers may wear several layers of protective clothing, thick, heavily insulated shoes, and bulky gloves. Electronic gear that might be suitable for moderate climatic conditions may be impossible to handle under a hot desert sun. In composing precise questions that the test will try to answer, the planner needs to set forth the full range of operational conditions under which the equipment must be tested. These include the temperatures, humidities, altitudes, terrain, maintenance conditions, degradation due to wear and tear, and so on, that are relevant to the performance of the equipment and the humans who will operate and maintain it.

Having defined the full range of environmental and operational conditions under which the equipment or system will perform, the test analyst must select a representative sample of them to include in the evaluation. These become some of the independent variables of the test. At the very least, test conditions should include values near each extreme and one value in between. For example, if a system is designed to operate at temperatures ranging from −25°C to 45°C, tests should be conducted at −25°C, 45°C, and say, 15°C. Although it is better to take more, rather than fewer, values of a variable, one has to compromise with reality. Taking too many values might make the tests so lengthy and costly that they cannot reasonably be carried out. If a system performs satisfactorily at both extremes of an operating condition and at some point in between, it is generally safe to conclude that the system meets that environmental test. If the equipment or system fails at one or the other extreme, however, it may then be worthwhile making tests at other values to discover at what point the performance of the system begins to deteriorate.

15.3.6 Selection of Test Personnel

At this point the human engineer must make decisions about the personnel who will serve as test subjects. Two kinds of decisions are involved: (a) How many people should be tested? and (b) what kinds of people should be tested?

On the number of test subjects. The issue here is that one should use enough test subjects to get dependable results, but not so many subjects that he increases the length and cost of the tests unnecessarily. If dependable results can be obtained with 10 test subjects, it is wasteful and inefficient to test 15.

There is, unfortunately, no easy way of deciding in advance how many test subjects should be recruited for a particular test. Although statistical formulas can be derived to make such forecasts, they are often more of theoretical rather than practical interest. The difficulty with most such equations is that they require certain numerical values—for example, an estimate of the so-called population variability in performance on some measure—that are seldom known in advance, at least for tests and evaluations involving human subjects. If the data of a test will be analyzed with certain simple, so-called non-parametric tests of significance—for example, a sign test or a binomial test—then it is possible to make some reasonably good forecasts about the minimum number of subjects that will be required. Even these predictions, however, involve some fairly intricate arguments about acceptable levels of Type I and Type II errors, and the probable true differences in the population. These are too complicated to try to present in succinct form here. The test planner is advised to consult a statistician for advice on this aspect of his plans.

In the absence of more precise quantitative guides, the best approach is to get advice from someone experienced in the design and conduct of human tests. Veteran experimenters, in the course of years of experimentation, learn that some kinds of experiments require more subjects than others. Moreover, they learn "about how many" subjects will be required for this or that kind of test. Although these are, to be sure, subjective impressions, they often turn out to be reasonably good predictions.

On the kinds of test subjects. Unfortunately, human engineers rarely have an opportunity to do much selection of personnel for tests and evaluations made early in the systems development program. In these early stages of system design, subjects are usually drawn from contractor technicians, engineers, and other company personnel. Contractor personnel are generally more skilled and experienced than the people who will ultimately use the system, and this disparity often leads to erroneous conclusions. For example, highly skilled personnel may not even be aware of difficulties that would be severe if less highly skilled personnel were using the equipment. Although the practice of using contractor personnel as test subjects is common, simple, and expedient, it almost invariably results in biased outcomes. Efforts should be made to find more typical people as test subjects, even if this involves considerable trouble and expense.

The importance of using representative personnel as test subjects increases as a system nears completion. In fact, *all operational tests, whether conducted in the field or the laboratory, should be done on personnel from the user population or one closely resembling it.* Some of the human characteristics that often interact with equipment evaluations are:

1. Age,
2. Sex,
3. Ethnic origin,
4. Body dimensions, such as height, weight and arm reach,
5. Sensory characteristics, such as visual acuity, auditory acuity, or color perception,
6. Psychomotor characteristics, such as strength, reaction time, or handedness,
7. Intellectual characteristics, for example, general intelligence, and specific aptitudes,
8. Personality, attitudinal, and motivational characteristics, for example, cooperativeness, initiative, or persistence, and
9. Training and experience, for example, level of general education, amount of specialized training, and specialized experience.

To a large extent the validity of human engineering tests depends on how well the human factors engineer has been able to match his test subjects to the characteristics of people who will ultimately use the equipment he is testing.

15.3.7 Selection of Test Variables

A careful selection of test variables insures that all important and relevant conditions are tested so that the purposes of the evaluation can be properly met. The facility with which this step can be carried out depends on how carefully the human factors engineer has gone through earlier steps in the planning process. A precisely stated and detailed set of objectives suggests a number of relevant variables. A good list of environmental conditions under which the equipment or system must operate defines another set of variables. Careful definition of the characteristics of the operator population usually identifies still other variables that need to be tested. And so on.

In general, variables fall into three major classes: independent, dependent, and controlled variables.

Independent variables. Independent variables are the *causes* of the cause-effect relationship. They are the variables that influence or have a significant effect on the performance of the system. When an experimenter can pick and choose among variables, he deliberately designs certain independent variables into the test because he wants to see what effect they will have. In the case of certain other independent variables, however, it is not so much a matter of designing them into the test as being sure that they are represented in the tests. Independent variables may be of four different types:

1. Environmental Variables. These are variables that describe the conditions under which the man–machine system will operate or describe influences originating from outside the system. Examples of such variables are: altitude, temperature, humidity, and speed. This category also includes input variables to the system, for example, variations in the number of planes around an air-traffic-control tower.

2. Personnel Variables. These are characteristics of the user population that might be presumed to interact with the performance of the system. Since automobiles are designed to be used by a broad spectrum of people, brakes, steering wheels, and other controls should be designed so that they can be used by both men and women (the personnel variable of sex), by people of different sizes (another personnel variable), and by people who are old as well as those who are young (still another personnel variable).

3. Mission Variables. These are variables defined by the way in which the system is to be used. An automobile may, for example, be used for in-town shopping and commuting, or it may be used for long, cross-country trips. It may be used to carry great numbers of children, dogs, and other paraphernalia, or may be used as a one-man personal conveyance. A vehicle that is satisfactory for one use may be totally unsatisfactory for another. In this context, standby, normal, and emergency operations may all be considered mission variables.

4. System Variables. These are variables that are due primarily to the characteristics of the system itself. A particular radar, for example, may have target scan rates of 2¼, 4½, 9, and 18 per sec. These are built-in characteristics of the system that may, however, affect the performance of the radar.

Dependent variables. Dependent variables are measures of the performance outcomes of the system. They are the *effects* of the cause-effect relationship. They are also sometimes called criterion measures, or, more simply, criteria. Examples of dependent variables are:

1. The number of words heard correctly over a communication system.

2. The average number of errors made by operators in calibrating an instrument.

3. The time it takes an operator to detect a target on a radar scope.

Since the overall performance of large man–machine systems is composed of many different activities, there are many different kinds of dependent variables that one could measure in evaluating systems. To a considerable extent, the usefulness of any test depends on how clever the human engineer has been in selecting appropriate criteria. Some measures of system activities are significantly related to overall system performance; others are not.

The ideal, or ultimate, criterion is one that measures the performance of a system under completely operational conditions. For example, the ideal measure of the performance of a fighter aircraft is the total number of enemy aircraft that are destroyed by it. All other dependent measures (speed, rate of climb, maneuverability, number of rounds of ammunition fired, etc.) are only approximate criteria. Unfortunately, one can rarely measure the performance of systems with ideal criteria. In the first place, it is difficult to make measurements under operational conditions. In the second place, most tests have to be made on systems before they become operational. Under such circumstances, the goal of the human factors engineer is to select dependent variables that will most likely correlate highly with ideal, or ultimate, performance criteria. There are no infallible guidelines for doing this, and much depends on the ingenuity and insight of the human factors engineer who is designing the tests.

Controlled variables. The third class of variables are those that are essentially nuisance variables. They are not of immediate interest in the test, that is, it may be presumed that they have no direct causal relationship to the effects that are being studied. However, if they are left uncontrolled, they increase the error variability in the measurements and so make the tests less efficient. For example, having several children and a dog in an automobile may disrupt a driver's performance for reasons that are entirely unrelated to the road-handling qualities of the vehicle. If the number of passengers is left uncontrolled throughout a series of tests, it will almost certainly result in greater variability in the performance of drivers than if this factor were to be controlled.

Keeping variables pure. When statisticians talk about this problem, they refer to it as the problem of *confounding*, and the recommendation they make is: *Don't confound variables.* What they mean is that variables should be kept unmixed. Only in this way can you be sure about what causes what.

The following illustration of confounding is taken from a human engineering test report of a certain kind of flight instrument. Two independent variables were tested: Altitude and airspeed. One part of the final report reads: "The greatest deviations were encountered at minimum altitudes." However, if we look at the details of the test plan, we find that these tests were made at three airspeeds (100, 300, and 600 knots) and at three altitudes (ground level, 25,000 ft. and 50,000 ft.). The following table shows the conditions that were tested:

	Altitude		
Airspeed (knots)	Ground level	25,000 ft.	50,000 ft.
100	X	......	......
300	......	X	......
600	......	......	X

Of the nine possible combinations of these two variables, only three (identified by X's in the table) were tested, and these were tested in such a way that the two variables are completely confounded! This means that there is absolutely no way to sort out what caused what. Were the large errors in performance due to the low airspeed, to the low altitude, or to a combination of both? This question is unanswerable from these conditions of testing.

As a general rule, all combinations of the independent variables should be tested if there is any reason to suppose that they will exert an influence on the outcome of a test. The difficulty with this rule is that with a large number of independent variables, test conditions multiply so rapidly that one quickly reaches an experiment of impossible size. For example, if we were to add three different levels of instrument illumination to the evaluation described above, a complete design would call for testing 27 different combinations (3 illuminations $\times$ 3 altitudes $\times$ 3 airspeeds). If one more variable, vibration, were added and three levels of vibration were tested as well, there would be 81 test conditions. And so on.

One way of keeping evaluations down to manageable proportions is to combine certain values of variables in such a way that the essential information can still be recaptured from the data. This is called partial confounding. The knowledge and skill of a competent statistician or experimenter are needed to do this, however. If such a person cannot be found, it is better to use fewer variables and to leave them unconfounded.

15.3.8 Planning for Data Collection and Analysis

Any evaluation, even if it consists only of opinion or questionnaire data, will involve some sort of statistical summary and analysis. The statistical treatment may be relatively simple, involving no more than some summary tables with perhaps an average and some other simple descriptive statistic, or it may be quite complex, involving least-squares, curve fitting, or the analysis of variance. Whatever its nature or form, a statistical analysis can be no better than the data that enter into it.

Statistical analyses also have to be matched to data. Some forms of statistical analysis cannot be carried out with certain kinds of data. Many times a statistician will be approached with the statement: "Here are some data that I have collected. Now what can I do with them?' This is putting the problem the wrong way around. If the data were not collected properly in the first place, it may not be possible to calculate the kinds of statistics he wants. Suppose, for example, that one collects data on the times to repair 100 electronic components, but terminates his observations at 8 hours. A sample set of data might look like this:

Time to repair (in hours)	Number of successful repairs
0-1	3
2-3	18
4-5	38
6-7	24
8 or more	17

With the data collected and reported in this way it is impossible to compute the kind of average known as the *arithmetic mean*. It is, moreover, impossible to calculate either of two conventional measures of variability known as the *standard deviation* and *average deviation*. The reason that these measures cannot be computed is the open-ended "8 or more" category. To avoid difficulties of this kind, data analyses have to be anticipated at the time that data collection procedures are set up.

On the amount of data to collect. A critical question that has to be answered in planning any evaluation is, "How much data should we collect?" The alternatives are fairly straightforward. To collect too few data may result in conclusions that cannot be supported by the observations. It may be impressive to read that 75% of A receivers failed within the first 100 hours of operation whereas only 50% of B receivers failed during the same period of time. However, if these percentages are really based on four A receivers and two B receivers, one would be foolhardy to base any large scale action program on such scant data.

On the other hand, the ready availability of automatic recording equipment makes it easy for an experimenter to collect too many data. The disadvantages of collecting too many data are not only that test and data collection resources are needlessly squandered but also that the analysis of the data becomes time-consuming and expensive. In any case, it is important to remember that large masses of data never compensate for defects in experimental plans.

It would be convenient if there were some simple and infallible way of predicting exactly how much data one should plan to collect in an evaluation. Unfortunately, there is not. The problem is closely related to that of the number of subjects one should test—a problem considered in an earlier section. The solution depends on several answers:

1. How sensitive must the tests be, that is, how small à difference do you want the tests to reveal?

2. If there is a difference between the equipments, or systems, with what certainty (what probability) do you want to be able to say that the difference is genuine? Or, to put it the other way around, what risk are you willing to take in arriving at a wrong decision, that is, of saying that there is a genuine difference when in fact there is none?

3. If the analysis shows that there is no genuine difference, with what certainty do you want to be able to conclude that there is no difference? Or, conversely, what risk are you willing to take that you will conclude there is no difference when in fact there really is?

4. How much inherent instability, that is, random error, is there in the measurements?

These are questions that the *task manager*, or the *test engineer*, must answer. They are not questions for a statistician to decide. However, is he is given good approximate answers to these questions, the statistician can give a close estimate of the amount of data that are required. Unfortunately, these questions are easy to ask, but difficult to answer. A test manager can usually arrive at answers to questions 2 and 3. He may occasionally have some basis for giving an answer to question 1, but he rarely has any way of estimating an answer to question 4. So, for most practical purposes, one is forced to rely on the judgment of sophisticated experimenters, which, fortunately, is often reasonably good.

15.3.9 The Experimental Design

Human engineering tests are, unfortunately, much more complicated than simple engineering or physical tests primarily because people, the subject matter of human engineering tests, are so variable. Moreover, people are constantly changing. They learn, they become bored, and they are influenced by what has just happened to them. To get dependable results in the face of incessant change, the human factors engineer has to resort to techniques that are often unfamiliar to the physical scientist and engineer. Some of these techniques are the use of control groups, counterbalancing and randomizing. In essence, all these methods consist of arranging trials and sequences of trials, so that changes in the behavior of test subjects will not systematically bias the outcomes of the tests.

The foregoing implies that human engineering tests cannot be conducted by simply testing people in any order and without any forethought. On the contrary, good human engineering tests are planned in advance, through a careful and deliberate assignment of people to trials and to test conditions. A plan that shows how subjects will be tested is called an *experimental design.* The varieties of experimental design are many and the selection of a valid design is best done with advice from an expert.

The design in Table 15–1 illustrates some of the basic ideas involved. That experimental design is for some articulation tests on three microphones. It makes use of four talkers, and tests are conducted on two separate days. The remaining variable is the word list used on each test. The experimental design has been set up so that, for example, each microphone is tested once and only once with each word list. In addition, no talker ever uses the same word list twice.

From Table 15–1 one can then arrive at a testing order, such as is illustrated in Table 15–2. This is a detailed plan listing the order in which combinations of variables will be tested.

The illustration given here is for a relatively simple kind of evaluation. In the case of elaborate field experiments, experimental designs and plans may become quite complicated, including, for example, the disposition of men and equipment, the movements of men and equipment, and the assignment of test subjects or teams of various equipments or systems. The plans for a systems experiment, or a field experiment, resemble in some ways the scenario for a movie film in that all the actions are deliberately plotted in advance. In any case, the validity of most human engineering tests rests very largely on how carefully the test manager has planned tests in advance.

Experimental design and statistics. In planning an experimental design, the test manager has to consider the way in which he hopes to analyze the data, once he gets them. Some experimental designs yield data that cannot be analyzed legitimately by any form of statistical analysis. The rule here is one that has been stated earlier: Experimental designs have to be constructed with particular forms of statistical analysis in mind. In constructing an experimental design, the test manager must always have in mind a particular form of statistical analysis. Only in this way can he be sure that he will be able to get the answers he needs.

The experimental design should include procedures for estimating random error. Variations that cannot be systematically accounted for by independent variables in a test are called *random error.* Random error provides the yardstick by means of which one can decide whether the differences among two or more items are "real" differences, "significant" differences, or accidental ones. A difference between performances on two pieces of equipment, or under two procedures, is said to be significant or real if the difference is greater than can be accounted for by random error. If the difference is not appreciably greater than can be accounted for by the random error, one is forced to conclude that the difference is an accidental one without any genuine significance.

To make valid tests of significance, one needs to get a good estimate of the random error in a test. This has to come from the same trials that are used to collect the main data of interest. In human engineering tests, random error is estimated from data taken on different people tested under identical conditions. To get such estimates of random error, at least two people must be tested under each of the separate conditions, or combinations of conditions, designed

TABLE 15–1. A SAMPLE EXPERIMENTAL DESIGN FOR 24 ARTICULATION TESTS ON THREE TYPES OF MICROPHONES

Day	1				2			
Talker	A	B	C	D	A	B	C	D
Microphone 1	3 (1)	9 (2)	4 (3)	10 (4)	22 (5)	19 (6)	16 (7)	23 (8)
Microphone 2	12 (2)	2 (3)	6 (4)	8 (1)	15 (6)	18 (7)	21 (8)	13 (5)
Microphone 3	5 (3)	7 (4)	11 (1)	1 (2)	17 (7)	20 (8)	24 (5)	14 (6)

Note: The Arabic numbers show the order in which successive tests are made. The numbers in parentheses identify the word list used in each test.

TABLE 15–2. A TESTING ORDER FOR 24 ARTICULATION TESTS ON THREE TYPES OF MICROPHONES

Day 1				Day 2			
Test number	Talker	Micro-phone	Word list	Test number	Talker	Micro-phone	Word list
1	D	3	2	13	D	2	5
2	B	2	3	14	D	3	6
3	A	1	1	15	A	2	6
4	C	1	3	16	C	1	7
5	A	3	3	17	A	3	7
6	C	2	4	18	B	2	7
7	B	3	4	19	B	1	6
8	D	2	1	20	B	3	8
9	B	1	2	21	C	2	8
10	D	1	4	22	A	1	5
11	C	3	1	23	D	1	8
12	A	2	2	24	C	3	5

Note: This table is derived from the entries in Table 15–1.

into the test. Although there are some exceptions to this rule, the exceptions are complicated ones best left to experts.

15.4 Equipment Drawings and Diagrams

15.4.1 Equipment Drawings

Equipment drawings are pictorial drawings that show the external and internal features of a piece of equipment in full or partial scale. They are often in blueprint form. Even a single piece of equipment may require a large number of blueprints and a simple system may require hundreds, if not thousands, of them. Blueprints are usually drawn in layers with the top drawing showing the external configuration of the equipment. Successive drawings become progressively more detailed as they expand components down to the assembly and part level. Equipment drawings usually give the following information:

1. The general configuration of the equipment.
2. Detailed spatial and dimensional relationships between parts.
3. The shapes and sizes of components.
4. Packaging details, such as the position of handles, fasteners, and screws.
5. The nomenclature of components and parts.
6. A list of all the hardware needed to fabricate the equipment.

Although the ultimate purpose of equipment drawings is to enable production personnel to construct and assemble the equipment, they obviously serve many other functions. Engineers use them to understand what the equipment will look like and how it will be put together. Designers use them to design related equipment, and human factors engineers use them to evaluate how easily the equipment can be operated and maintained.

Equipment drawings are prepared early in the design process and are normally revised frequently until they are finally completed and released to production. Several reasons account for the frequency with which they are revised. Drawings may be changed because they do not account for newly discovered equipment or human requirements. In addition, any change in the overall configuration of the equipment will require changes in more detailed drawings.

The fact that drawings are revised frequently creates serious difficulties for the human engineers who attempt to use them for evaluations. If human engineering personnel are in a department separate from that in which design takes place, they may not know whether they are working on the latest revision of the drawings. In fact, newer revisions in progress may cancel out the value of a human engineering evaluation. Further, because of the pressures of time, drawings may have to be completed and released before human engineering personnel are able to do a thorough evaluation of them. For these reasons, the first and most important requirement for an effective human engineering evaluation of drawings is: Keep up-to-date.

The human engineering evaluation of drawings. The top drawing in a set is generally the most useful one for human engineering evaluations. This is the drawing that shows the external packaging of the equipment, control-display panels, and workspace arrangements. More detailed drawings may have to be examined for certain aspects of the evaluation, for example, the evaluation of a design from the standpoint of its maintainability. The latter would show details about the dimensions and locations of accesses and hatches.

Drawings contain a great deal of information, but for purposes of human engineering evaluations, most of the relevant information has to be arrived at by inference. For example, drawings may show clearly the locations of controls, but will not reveal functional aspects of their design, such as display-control ratios and direction of movement relationships. The latter have to be deduced from other sources.

In performing his evaluation, the human engineer is, of course, most directly concerned with the activities and tasks of the operator who will use the equipment represented in the drawings. His evaluation is usually performed by inspection using some sort of a set of human engineering standards or checklist. Standards commonly used for this purpose can be found in this book and in Woodson and Conover (1966). Checklists can also be found in Berkun and Van Cott (1956), Krumm and Kirchner (1956), and U.S. Navy Electronics Laboratory (1960).

Under special circumstances somewhat more sophisticated techniques can be used for the evaluation of drawings. One of these is the method of link analysis, a detailed description of which can be found in Chapanis (1959) and in Chapter 1 of this *Guide*. In a link analysis the human engineer counts the sequences of movements that an operator would make, or the sequences of steps that he would follow, in operating the equipment. These are then plotted as a series of lines on the diagram of the equipment. The links may also be weighted differentially for importance. A link analysis, therefore, shows the frequency (and/or importance) with which the various sequences of operations are made in using a control panel. High linkages show those sequences that are made frequently or are very important. In general, controls or displays should be grouped so as to reduce the lengths of such linkages. In addition, controls or displays connected with high link values are those that merit greatest attention in the design of the panel.

Still another technique for evaluating diagrams is to construct several alternative control-display layouts and to ask experienced personnel to rate or rank them in terms of their acceptability. Raters may also be asked why they find particular layouts satisfactory or unsatisfactory. Such ratings should generally be interpreted with caution since experienced operators often rate high those layouts they are familiar with, irrespective of their inherent merits.

Although these techniques are crude, they are often sufficient for the human factors engineer to discover glaring design deficiencies before any equipment has been constructed.

15.4.2 Equipment Diagrams

In addition to the equipment drawings discussed above, two other kinds of diagrams are often of interest to the human factors engineer. These are:

1. *Functional block diagrams* which show systems, subsystems, and equipments in block or diagrammatic form.
2. *Schematics or circuit diagrams* which show functional relationships and connections among electrical and mechanical components by means of graphical symbols.

Neither block diagrams nor circuit diagrams show the correct physical and spatial relationships between the components in a system. For this reason they are usually not evaluated by the human factors engineer but are rather studied and analyzed for the information they contain. Such diagrams are useful in showing:

1. The principal functions and operations that will be performed by the various parts of the system.
2. How the inputs to and outputs from various components are related logically and functionally.
3. What performance requirements have implications for the operator.
4. At what points man will interact with machine elements.

15.5 Mockups

Mockups are specialized forms of simulators. When they are constructed to assist in the development of a system, they are not evaluated as products themselves, but are rather tools used to evaluate equipments or systems before the equipments or systems are actually constructed. Once hardware has been built, it would be a needless expense to develop and construct a mockup since evaluations are better done on the equipment itself than on a model of it. Mockups designed and constructed for purposes of training, however, are evaluated as products, in terms of their fidelity of simulation and the amount of transfer of training that can be made from the mockup to the equipment it models. Mockups come in two principal forms: static and functional mockups.

15.5.1 Static Mockups

A static mockup is a three-dimensional, full-scale model of a piece of equipment. It is usually made of inexpensive materials such as cardboard, fiberglass, or plastic. All major internal elements should be represented either by actual controls, displays, and other small components, or by cardboard cutouts, drawings, photographs, or sketches of them. The external dimensions of the mockup are usually not critical for human engineering evaluations. On the other hand, internal dimensions—those having to do with the workplace, component panels, controls, and displays—should be reasonably exact. Tolerances, however, need only be approximate.

Purposes of static mockups. Static mockups serve a number of functions (Seminara and Gerrie, 1966):

1. Design Integration. The design of most complex systems is of necessity a piecemeal affair. Parts of a system are normally developed by different groups within a company and oftentimes by groups in different companies. A mockup is frequently one of the best ways of integrating the work of many, often scattered design groups contributing to a development program.

2. Design Verification. It is usually hard to visualize from a reduced scale drawing what a product will look like. It is even harder to visualize from a drawing how a man will interact with equipment. A full-scale, three-dimensional mockup permits the human factors engineer to evaluate designs by having operators of various sizes go through the actions and motions that they will have to make in carrying out their duties. A mockup is also useful for verifying the adequacy of design for operators who will be encumbered with specialized personal or protective equipment such as pressure suits and arctic clothing.

3. Design Experimentation. A mockup is usually made of inexpensive materials and is

held together simply with nails and glue. This makes it a flexible tool for evaluating alternative design configurations. Various workplace arrangements and crew compartments can be quickly disassembled, reassembled, and tried out with suitable test subjects. In this way a mockup can be used for: (a) The identification of workplace difficulties by simulating operational tasks; (b) The discovery of accessibility problems by simulating maintenance operations; (c) The evaluation of various locations and routings of wiring, harnesses, cabling, and piping; and (d) The evaluation of various locations for junction boxes, connectors, and other auxiliary items of equipment. A mockup often permits the human factors engineer to identify unworkable arrangements that might be overlooked completely if the problem were to be attacked with drawings or expert opinion.

4. Design Conceptualization. It is difficult to visualize three-dimensional objects from scaled-down, two-dimensional drawings. Fabricating a simple full-scale, three-dimensional mockup often helps greatly in giving the designer a greater sense of realism and "feel" about how the final product will look.

5. Documentation. Some companies use photographs of successive mockups as documentation to show the evolution of design configurations. If necessary, earlier configurations can easily be recreated and reconsidered as part of this process.

6. Training. Because training for the operation of most systems must begin before the system is actually constructed, a mockup can serve as a valuable training aid by allowing the trainee to get a greater degree of familiarity with the equipment than he could get from a training manual.

7. Presentation. Although this function is not directly related to the design of a system, a mockup can serve as a valuable public relations tool. It reveals current thinking and progress on the design of a system far better than does a series of drawings.

Evaluations made with static mockups. Two types of human factors evaluations can be made with static mockups: observational and operational. In the former the human factors engineer uses the mockup as a highly sophisticated drawing. He uses a checklist to record his judgments about what he observes. Estimates of the adequacy of various layouts and arrangements are made by comparing them against tabled values, criteria, and recommendations given in the checklist.

Operational tests with static mockups are made by having subjects simulate movements that they will have to make in operating the equipment. It is useful to make a list of representative tasks in advance and to have each of a number of subjects perform these tasks. The evaluator judges the ease with which movements are made and notes any difficulties experienced by the subject. These observations are usually supplemented with opinions from the subjects themselves. Examples of the kinds of tasks that can be used are:

1. Reaching for and operating simulated controls and other equipment.
2. Measuring the reach envelope of the subject when he is seated or standing at his normal work station.
3. Finding various components in response to instructions.
4. Connecting and disconnecting cables or tracing harnesses.
5. Climbing to reach a component that must be checked or removed.
6. Removing and replacing test cover plates.
7. Recording and measuring the sequences of movements an operator must go through in performing various operations.
8. Grasping and using handholds for ingress and egress.
9. Measuring fields of view available through windows or openings.

15.5.2 Functional Mockups

A functional mockup is a full-scale, three-dimensional model that can function in a quasi-operational manner. Although they are represented in static mockups, the controls control nothing and the displays never change. A functional mockup, by contrast, has displays that move in response to control actions and in response to simulated outside environmental influences. The number of operations that may be built into a functional mockup covers a wide range. In the simplest of such mockups only normal operations are modelled. More complicated functional mockups may include certain emergency operations, and still more complicated ones may add maintenance operations. Operations are controlled mechanically in simpler functional mockups; more complex ones may be programmed and operated by computers.

As has already been said, the most complex functional mockups are indistinguishable from some so-called simulators. Functional mockups are complex and expensive and they require a considerable amount of time and talent for their construction. For these reasons, the designer and the human factors engineer must always be sensitive to the question of whether it is really worthwhile to allocate valuable resources to the design and construction of a functional mockup. There are really two questions involved:

1. Is it worthwhile constructing a functional mockup at all?
2. If it is worthwhile, how much realism should be designed into it, that is, how complicated should it be?

These involve delicate questions of tradeoffs. The gains to be anticipated from the mockup must offset the cost of its construction if valuable resources are not to be squandered.

Evaluations made with functional mockups. Functional mockups provide all the information that is provided by static mockups, and more besides. They may be used, first, to evaluate equipment in much the same way as static mockups are used. In addition, a functional mockup makes it possible to study the performance of personnel in simulated operational situations. The human factors engineer can now evaluate operating characteristics of equipment in terms of human performance. He can also compare various design configurations, or verify the superiority of some one design, in terms of human performance. When used in this way, functional mockups become the apparatus on which more-or-less conventional laboratory experiments are carried out. The rules of sound experimentation (discussed elsewhere in this chapter) apply here just as much as they do in the conduct of more conventional kinds of investigations.

15.6 Simulation, Modelling, and Gaming

A considerable amount of systems and human engineering testing is done with techniques that are variously called simulation, modelling, or gaming. Although all three terms are used widely, there is fair agreement only about the definition of the last. Gaming is the simulation or modelling of a contest in which there are at least two antagonists or opponents. So, for example, a war game is a model of a battle or a military encounter between two adversaries, and a business game is a model of an economic contest between a person, or an enterprise, and competitors.

Simulation and modelling, however, are not so easily defined and differentiated. To an operations researcher, a simulation means a mathematical model—a set of equations—describing some system or process. To a training specialist, on the other hand, simulation means the design and construction of a device, a piece of equipment, that in some way mimics another equipment or system for which an operator is supposed to receive training. In short the terms, simulation, modelling, and gaming, cover a wide variety of devices and techniques, ranging from sets of abstract mathematical equations to, small- or full-scale replicas of ships, aircraft, space capsules, and factories. In what follows we make no distinction among simulation, modelling, and gaming, but regard them all as minor variants of the same basic technique. In the remainder of this section, therefore, the word *simulation* should be understood to include *modelling* and *gaming*, and the word *simulator* should be understood to include *models* and *games*. For textbook discussions of simulation, gaming, and modelling, see Churchman et al. (1957), Flagle (1960), and Bekey and Gerlough (1965).

It is logically difficult to make a sharp distinction between some kinds of simulation and experimentation. An experiment is also, in a manner of speaking, a simulation of a part of the real world. The distinction becomes even more blurred when a simulator is used as the apparatus for a laboratory experiment. The primary way in which simulation and experimentation can be differentiated is in terms of their goals, or intents (Knowles, 1967). Most simulation is done to evaluate and demonstrate the application of specific procedures and equipment to specific operations. Most experimentation is done to describe and measure relationships between operator performance and machine or system variables.

By the same token, it is difficult to make a sharp distinction between some simulators and some mockups, particularly functional mockups. In fact, it would be correct to say that mockups are specialized simulators. Mockups are almost invariably full-scale simulators in which emphasis is placed on the faithful representation of spatial dimensions and arrangements. In any case, much of what is said here about simulation applies equally well to experimentation and to the use of mockups.

Perhaps the safest way of defining simulators is to say that they are analogies. They are representations, or likenesses, of certain aspects of complex events, structures, or systems, made by using objects or symbols that in some way resemble the thing being modeled (Chapanis, 1961, and Obermayer, 1964). By this definition, simulators may be physical (for example, a Link aircraft trainer, a small globe representing the earth, or a war game played with real men and equipment) or symbolic (for example, a set of full non-linear coupled equations describing the aerodynamic characteristics of a high-performance vehicle, a Monte Carlo model of a traffic queue, or a computer simulation of trading activities on the New York Stock Exchange). All three may contain both physical and symbolic elements, of course. A globe of the earth is a physical representation in the sense that it is round, like the earth. It is symbolic in that land masses, oceans, countries, and states may be represented by colors and words.

15.6.1 The Basic Philosophy of Simulation

The basic philosophy underlying simulation is simple. If one can get a good simulation of some part of the real world, he may then manipulate, study, and measure the model instead of the real world. The purpose of this activity, of course, is to find out something about the real world by studying the simulation. So, for example, aerodynamic engineers study airflow characteristics around a model of an aircraft in a wind tunnel as a function of the shape and contour of the model, air speed, air density, and so on. The purpose of studying the model, however, is to try to find out something useful about the flight characteristics of full-scale aircraft.

15.6.2 Reasons for Using Simulation

The reasons for resorting to simulation are several:

1. Simulators are usually smaller, cheaper, and easier to construct than the systems or processes they simulate.
2. Simulators can be instrumented to collect data that would be difficult, or impossible, to get from real systems and processes.
3. Simulators can be manipulated more easily than the systems they mimic.
4. Simulators may be put through maneuvers or exercises (for example, crashes and accidents) to which one would not want to subject real systems.
5. Simulators may be used to study systems and processes that have not yet been constructed or put into operation. Simulators may be used to test the characteristics of space vehicles that have not yet been constructed, and a war game may be used to study unfought battles.

15.6.3 Simulation and Experimental Design

Once a good simulation has been constructed, alternative system configurations, methods of use, tactics, and strategies, may be tested by manipulating the model. In so doing, the precise conditions and the ways in which the model is manipulated should conform to one or more of the experimental designs that are discussed elsewhere in this chapter. In this kind of work, the human factors specialist conducts, in a manner of speaking, a laboratory experiment on a model. The rules of good experimentation apply here just as much as they do in the conduct of conventional laboratory experiments.

15.6.4 Some Requirements for Good Simulation

The most important requirement of a good simulation is that it should be an accurate representation of some segment of the real world. Critical variables in the real world should be correctly duplicated in the simulator. Unfortunately, there are no completely trustworthy guides about which aspects of the real world need to be duplicated in a simulator and which do not. The problem is not as simple as that of deciding on the proper amount of real-world detail to incorporate into a simulator (Obermayer, 1964). There is evidence, for example, that relatively simple simulators may provide a good level of training for the operation of certain complicated systems. Increasing the fidelity of the simulator by adding still further detail may even decrease the effectiveness of the training that the simulator gives. Furthermore, in some cases, simulators must actually introduce deliberate distortions if they are to yield measurements that are valid for the real world. The problem of the fidelity of simulation is a complex one in which there is still almost as much art as science.

Once variables have been selected for inclusion in a simulation, the next problem is that of deciding how these variables are connected. This is an especially difficult problem for all simulations, especially mathematical ones, that try to model the human operator and his behavior. Our knowledge about human behavior is still sufficiently incomplete that we cannot always say what forms of relationship connect stimuli and responses in human activity. This is the primary reason why so few mathematical simulations of human operators have yielded anything of value for systems design work. To get good mathematical representations about human behavior for a model, one usually has to do an experiment first to find out precisely what these relationships are.

Finally, one usually needs numerical values, constants, coefficients, and exponents, to insert into the mathematical equations of the simulation. The numerical data needed for one man-machine model include such things as average subtask execution times, distributions of subtask execution times, probabilities of success for various subtasks, sequences of operations together with average waiting times and idle times, and so on. Data such as these needed for a simulation must first come from careful measurements of human tasks and operations. They cannot be made up.

15.6.5 Some Difficulties with Simulation

Since a simulation tries to model only a part, and usually only a small part, of some larger system, every simulation is incomplete and, in a sense, wrong. One of the principal dangers of simulation is that it tends to invite overgeneralization. Because a simulator, or a model, mimics some aspect of a system faithfully, it is easy to fall into the trap of supposing that everything the simulation tells you can be applied to the real world.

Simulations are often wrong because relationships that have been made to hold between the variables in the model are incorrect (Chapanis, 1961). Simulations may also be wrong because the values assumed for constants may be incorrect. In addition, simulators often add ingredients of their own that are peculiar to themselves and that do not appear in real systems. A particularly striking illustration is the fact that many highly realistic driving simulators induce subjects to get sick. Even people who never experience motion sickness in driving real automobiles often become sick in these simulators. The simulation has clearly added something extra and that something extra produces an important, unwanted kind of behavior. The curious thing is that all attempts to isolate and remove the causes of sickness in driving simulators have been largely unsuccessful.

To sum up, the results obtained from simulations have to be interpreted and extrapolated to real-world situations with great caution. In the final analysis, the validity of a simulation has to be proven experimentally. It cannot be taken for granted, no matter how impressive, internally consistent, or elegant the simulation.

15.6.6 Man-in-the-Loop Simulation versus Simulation of the Human Operator

The simulations discussed above may be characterized as man-in-the-loop simulation. The equipment, system, or process is simulated but real operators are used to interact with the simulator. There are a number of disadvantages to this kind of simulation. For one thing, tests with man-in-the-loop simulators must be done in real time. When entire processes are simulated, on the other hand, they can often be run in fast time with tremendous reductions in the total time needed to arrive at solutions.

For these and other reasons, there has been a considerable amount of effort devoted recently to the construction of mathematical and other kinds of models for simulating the human operator. In general these models are adequate for representing the input-output characteristics of human operators doing certain simple tasks, for example, tracking low-frequency signals in one dimension (Bekey and Gerlough, 1965). Aside from these restricted applications, however, it is fair to say that the use of human-operator models is limited at the present time to certain preliminary investigations. If simulation is used as a tool, the vast majority of detailed man–machine design problems requires real-time simulation using actual human operators.

References

Bekey, G. A., & Gerlough, D. L. Simulation. In R. E. Machol (Ed.), *System engineering handbook.* New York: McGraw-Hill, Inc., 1965.

Berkun, M. M., & Van Cott, H. P. Checklist of human engineering evaluation factors (plans inspection) (CHEEF), Report No. AF–WP–(B)–0–23 Nov 56 150, September 1956, Wright Air Development Division, WPAFB, Ohio.

Campbell, D. T., & Stanley, J. C. *Experimental and quasi-experimental designs for research.* Chicago: Rand McNally, 1963.

Chapanis, A. *Research techniques in human engineering.* Baltimore: The Johns Hopkins Press, 1959.

Chapanis, A. Men, machines, and models. *American Psychologist,* 1961, **16** (3), 113–131.

Churchman, C. W., Ackoff, R. L., & Arnoff, E. L. *Introduction to operations research.* New York: John Wiley and Sons, 1957.

Cochran, W. G., & Cox, G. M. *Experimental designs.* New York: John Wiley and Sons, 1950.

Edwards, A. L. *Experimental design in psychological research* (3rd ed.). New York: Holt, Rinehart, and Winston, 1968.

Flagle, C. D. Simulation techniques. In C. D. Flagle, W. H. Huggins, & R. H. Roy (Eds.), *Operations research and systems engineering.* Baltimore: The Johns Hopkins Press, 1960.

Kempthorne, O. *The design and analysis of experiments.* New York: John Wiley and Sons, 1952.

Knowles, W. B., Jr. Aerospace simulation and human performance research. *Human Factors,* 1967, **9**(2), 149–159.

Krumm, R. L., & Kirchner, W. K. Human factors checklist for test equipment, visual displays and ground support equipment. Report No. AFSWC–TN 56–12, February 1956, Air Force Special Weapons Center, Kirtland AFB, N. M.

Lindquist, E. F. *Design and analysis of experiments in psychology and education.* Boston: Houghton Mifflin Co., 1953.

Meister, D., & Rabideau, G. F. *Human factors evaluation in system development.* New York: John Wiley and Sons, 1965.

Myers, J. L. *Fundamentals of experimental design.* Boston: Allyn and Bacon, 1966.

Natrella, M. G. *Experimental statistics.* National Bureau of Standards Handbook 91. Washington, D.C.: U.S. Government Printing Office, 1963.

Obermayer, R. W. Simulation, models, and games: Sources of measurement. *Human Factors,* 1964, **6**(6), 607–619.

Seminara, J. L., & Gerrie, J. K. Effective mockup utilization by the industrial design—human factors team. *Human Factors,* 1966, **8**(4), 347–359.

Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of experimental psychology.* New York: John Wiley and Sons, 1951.

U.S. Navy Electronics Lab. Suggestions for designs of electronic equipment (rev. ed.). San Diego, Calif., 1960.

Winer, B. J. *Statistical principles in experimental design.* New York: McGraw-Hill, 1962.

Woodson, W. E., & Conover, D. W. *Human engineering guide for equipment designers* (2nd ed.). Berkeley, Calif: University of California Press, 1966.

Index

☆ U.S. GOVERNMENT PRINTING OFFICE: 1972 O—414-256