AFAMRL-TR-82-57

SUBSIDIARY RADIO COMMUNICATIONS TASKS FOR WORKLOAD ASSESSMENT IN R&D SIMULATIONS: II. TASK SENSITIVITY EVALUATION

CLARK A. SHINGLEDECKER

AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY

MARK S. CRABTREE

SYSTEMS RESEARCH LABORATORIES, INC.
2800 INDIAN RIPPLE ROAD
DAYTON, OHIO 45440

SEPTEMBER 1982

AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY
AEROSPACE MEDICAL DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433

NOTICES

TECHNICAL REVIEW AND APPROVAL

AFAMRL-TR-82-57

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Chief
Human Engineering Division
Air Force Aerospace Medical Research Laboratory

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER AFAMRL-TR-82-57	2. GOVT ACCESSION NO. AD-A122 545	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) SUBSIDIARY RADIO COMMUNICATIONS TASKS FOR WORKLOAD ASSESSMENT IN R&D SIMULATIONS: II. TASK SENSITIVITY EVALUATION		5. TYPE OF REPORT & PERIOD COVERED Technical Report 1 July 1980-2 January 1982 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Clark A. Shingledecker Mark S. Crabtree*		8. CONTRACT OR GRANT NUMBER(s) F33615-79-C-0503
9. PERFORMING ORGANIZATION NAME AND ADDRESS AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY AEROSPACE MEDICAL DIVISION (AFSC) WRIGHT-PATTERSON AFB OH 45433		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 62202F; 7184-14-07
11. CONTROLLING OFFICE NAME AND ADDRESS AIR FORCE AEROSPACE MEDICAL RESEARCH LABORATORY AEROSPACE MEDICAL DIVISION (AFSC) WRIGHT-PATTERSON AFB OH 45433		12. REPORT DATE September 1982 13. NUMBER OF PAGES 44
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

*Systems Research Laboratories, Inc.
2800 Indian Ripple Road
Dayton, OH 45440

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Communications	Pilot Workload	Task Sensitivity
Communications Workload	Radic Communications	Workload
Embedded Tasks	Secondary Tasks	Workload Assessment
Embedded Secondary Tasks	Simulation	Workload Measurement
Operator Workload	Subsidiary Tasks	Workload Scaling

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Because of the rapid growth of the technological complexity of modern aircraft and weapon systems, the assessment of operator workload at each stage of system design is of increasing importance. Secondary task measures of workload sensitive, reliable indicators of how hard the operator is working. However, primary task interference, poor operator acceptance, and the requirement for additional instrumentation in the cockpit or at the operator's station are problems associated with secondary task measures of aircrew workload which limit their usefulness during later stages of system development. One solution

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

19. KEY WORDS (con't)

Workload Sensitivity

20. ABSTRACT (con't)

to these problems would be to employ secondary tasks which not only are an integral part of the operator's duties but also possess the properties of valid measurement tasks. A previous report by the authors and their colleagues described the concept of using realistic radio communications activities as subsidiary workload measurement tasks for R&D simulations and actual flight tests. In order to develop tasks with controlled levels of added workload which could be tailored to specific system and mission contexts, three scaling techniques, including information theoretic analysis, pilot opinions of workload produced by specific messages, and subjective workload estimates of complete communications tasks, were applied to 13 communications typical of those occurring in A-10 aircraft.

Nonparametric correlational procedures revealed considerable agreement among the results of the three scaling methods, and a plan was outlined for several experimental dual task performance studies for testing the sensitivity of the communications tasks to primary task workload and for evaluating the adequacy of the three a priori scaling methods.

The first in a series of such studies was performed and is described in this report. After extensive training on both single and dual tasks, six subjects were exposed to all possible combinations of eight communications tasks and two levels of single-axis tracking task difficulty. Dependent measures were the number of control losses on the tracking task and the accuracy and response times for the verbal and manual responses to the communications tasks. Results indicate that realistic radio activities can be used as secondary tasks to provide objective measures of workload. Four communications tasks produced significant dual task interference and were sensitive to tracking task difficulty. The finding that mutual task interference was correlated with results from previous a priori workload scaling techniques indicates that analytically derived estimates of workload can be used in developing additional communications tasks. A recommendation is made for further research in high fidelity simulations.

PREFACE

This report describes an experiment conducted to evaluate pilot radio communications activities as subsidiary workload measurement tasks. The report was prepared in part by Systems Research Laboratories, Inc. (SRL), 2800 Indian Ripple Road, Dayton, Ohio 45440, under Contract F33615-79-C-0503. The work was performed in support of AFSC Project 7184, Man-Machine Integration Technology for the Air Force, for the Air Force Aerospace Medical Research Laboratory (AFAMRL), Human Engineering Division (HE), Wright-Patterson Air Force Base, Ohio 45433.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Section 1
INTRODUCTION

WORKLOAD MEASUREMENT

The maintenance of an optimally effective military force requires the development of modern airborne weapon systems which incorporate the most sophisticated products that the engineering disciplines have to offer. The use of such advanced technology greatly enhances the potential capabilities of combat aircraft. However, technological improvements are often achieved with a simultaneous increase in the monitoring, supervisory, and decision making responsibilities of the aircrew. These information processing activities can jeopardize the quality of task performance by placing additional demands on the mental resources of the individual crewmember. Since the ability of an aircraft system to fulfill its mission objectives is a function of the performance of its human operator as well as that of its other components, serious consideration must be given to the management of operator workload throughout the system design process.

In order to insure mission effectiveness, a variety of accurate and reliable methods are needed to assess mental workload at various stages of system development. A widely accepted conceptual framework upon which such measurement techniques are based views the human operator as a limited capacity information processing device. According to this general model, workload may be defined as the degree to which the operator's processing capacity is occupied by mental activities. Overload, and resulting performance decrement, occurs when capacity is insufficient to meet task demands. Since the momentary capacity of the operator is unknown and submaximal workload levels cannot be inferred from his or her performance on the task of interest, an indirect measure can be obtained by evaluating the amount of spare capacity available under a given set of task conditions.

Although the general concept of spare capacity was derived from an early single channel model of the information processing system (Broadbent, 1958), this notion can also accommodate recent undifferentiated capacity or effort theories (Kahneman, 1973). Furthermore, more complex models which propose multiple channels or multiple capacities (e.g., Wickens, 1980) are also congruent with spare capacity measures of workload if individual processing resources are carefully defined.

Two common approaches to estimating spare capacity are available to the system designer (Wierwille and Williges, 1979). Analytical methods assume that all task components are performed serially and that each requires a specified length of time to complete. A measure of spare capacity is obtained by summing the required task times and using the maximum time available to compute the proportion of excess time. The validity of many of these methods rests on the tenuous assumption that capacity and the availability of time can be equated. Thus, if a change in workload is not accompanied by a reduction in task performance time, the analytical method will not provide a sensitive measure. A further limitation of time available versus time required approaches is that an extensive data base of time requirements is mandatory for adequate workload estimates in all but the simplest of task environments.

The behavioral approach to assessing spare capacity involves the use of the secondary task technique. In this method, operators are given an additional information processing task to perform in conjunction with the task of interest. The rationale underlying the use of secondary tasks is that by applying an extra load which produces a total information processing demand that exceeds the operator's capacity, workload can be measured by observing the difference between single task and dual task performances. As noted by Ogden, Levine, and Eisner (1979), secondary tasks can be employed in two ways. Used as a loading technique, the method requires subjects to perform the secondary task under all circumstances with the intent of displaying overload effects in primary task performance. When secondary tasks are used as a workload measure, performance on the primary task is emphasized and secondary task performance is observed as an index of the workload of the primary task. Although specific research questions may require a choice of one of these applications, combined task decrement may also be used as an estimate of mutual interference and workload (see Wickens, 1981).

Unlike analytical methods, the secondary task approach to assessing spare mental capacity has the potential for being sensitive to the degree of mental effort or attention devoted to information processing as well as to the temporal aspects of workload. The secondary task technique has the further advantage of producing a measure based on task performance, which is the variable that all workload measures ultimately must predict if they are to be of any value.

Although secondary task methodology has proven to be a useful technique for the investigation of cognitive processes, its practical application as a workload measurement tool has often been confined to the earliest stages of aircraft system design. As Schiflett (1976) has noted, most workload measures have been developed for, and are most applicable to, the laboratory environment in which highly controlled, part task studies of workload can be conducted. When subsystems are combined to evaluate mission performance in the context of high fidelity simulations or flight tests, many workload assessment methods become difficult to employ because they are impractical or present potential safety hazards. As a result, workload measurement at the critical later stages of system development is often performed using relatively informal and qualitative techniques.

Three specific problems are encountered when traditional laboratory secondary tasks are considered for use during advanced development of aircraft. One practical consideration is the physical instrumentation of the secondary task. In a flight test environment, and to a lesser extent in a simulator, introducing or adding any extra equipment to the crewstation may be unacceptable (Wierwille and Williges, 1979). The space required for electronic data recording and experimental control equipment as well as display and input devices may not be available in an already crowded cockpit. Even when sufficient space can be reserved, the possibility of obstruction or distraction caused by the additional instrumentation can limit the feasibility of using a secondary task.

A second problem with the implementation of secondary tasks is the possibility of intrusion on primary flight duties. Although some performance decrement may be tolerable, task interference can easily complicate the interpretation of data in test environments where measures of all performance variables may be unavailable. A more serious consequence of primary task intrusion in the flight test environment is the potential for compromising flight safety.

The final factor limiting the use of secondary task measures is operator acceptance (Ogden et al., 1979). Whether used to induce stress or to measure reserve capacity, a secondary task is likely to produce misleading data if the operator fails to integrate it with his normal duties. Acceptance is a potential problem with all laboratory tasks because they are obvious, artificial additions to the crewstation and have little face validity or congruence with the general performance situation. Such test conditions can lead the operator to neglect the secondary task or, because of its novelty, allow it to assume an artificially high priority. Thus, lack of operator acceptance can become a major contributor to primary task intrusion as well as a source of measurement error.

EMBEDDED SECONDARY TASKS

An analysis of the problems associated with the practical use of traditional laboratory secondary tasks prompted the development of a program of research at the Air Force Aerospace Medical Research Laboratory to explore the feasibility of designing an embedded secondary task methodology for simulation and flight test environments. The concept of the embedded secondary task is based on the hypothesis that instrumentation limitations, task intrusion, and poor operator acceptance can be minimized by designing secondary tasks which are fully integrated with system hardware and with the crewmember's conception of his mission environment. By their nature, such tasks would be realistic components of crewstation activity, yet their performance could be manipulated and measured independently of the primary activities of interest.

Several classes of aircrew activity such as stores management or threat monitoring are potential candidates for isolation and use as embedded tasks. Another particularly promising approach would be to adapt radio communications tasks for this purpose. The radio communications which might be most useful as embedded tasks are those initiated by a message sent from another aircraft or a ground controller to a pilot whose workload is to be assessed. Upon detection and identification of a relevant message, the pilot must engage in a sequence of verbal responses and radio switching activities in order to meet the demands of the communicated request.

Such tasks closely resemble the nonadaptive discrete secondary tasks used in numerous workload studies and have many properties of good measurement tasks. Communications call upon a wide variety of information processing abilities and can be varied along several dimensions of complexity. Furthermore, no auxiliary crewstation equipment is necessary to control the experiment or to collect performance data. The opportunity for obstruction or peripheral interference is also minimized since the auditory channel is not shared by other tasks and verbal responses are generally unique to radio communications activities, while switch actions can be dealt with by the pilot's free hand. Most importantly, communications tasks are an integral part of a pilot's inflight duties. As a result, lengthy training requirements are eliminated and high face validity is achieved. Additionally, the realistic nature of the activity makes artificial task interactions improbable because the pilot has predetermined priorities assigned to communications and other cockpit functions. This feature makes communications activities especially suitable for use as secondary tasks since pilots consider them to be important, but will normally devote less attention to communications as more crucial tasks become difficult to perform.

Communications Task Scaling

In an initial evaluation of realistic communications for use as embedded secondary tasks, Shingledecker, Crabtree, Simons, Courtright, and O'Donnell (1980) interviewed operational A-10 pilots to obtain sample radio communications tasks from a typical air-to-ground attack mission. Each task was specified in terms of an input message and the detailed verbal and manual responses required of the pilot. Thirteen tasks associated with identification friend or foe (IFF) demand, threat alert, traffic control, waypoint passage, jammed communications, and strike clearance were selected for analysis.

Examination of these tasks revealed a problem with the embedded secondary task technique which does not exist in standard laboratory tasks. While the use of realistic tasks offers many advantages, this quality also makes the tasks unamenable to precise experimental control of task demand. Traditional secondary tasks are designed to impose constrained and highly describable requirements on the performer. Thus, task parameters are easily varied and can be selected to permit precise control of loading. In contrast, realistic communications are complex processing tasks which vary along multiple dimensions. As a result, no obvious scheme can be employed to choose sets of tasks with equivalent task demand characteristics. Furthermore, excessive use of repeated task presentations must be precluded since this method of controlling task demand could sacrifice face validity.

In order to resolve this dilemma, an attempt was made to scale the workload of the A-10 communications tasks. The purpose of the effort was to derive estimates of the loading associated with each task so that they could be combined in a realistic scenario in order to produce controlled levels of subsidiary task demand. Since no single a priori approach to workload evaluation was expected to produce a superior quantitative estimate, three techniques were used to provide alternative measures for later comparison to performance data.

Because workload associated with communications tasks was assumed to be partially determined by task information transmission requirements, the first scaling approach was based on information theoretical measures. The tasks were analyzed by subdivision into activities requiring perceptual decisions and those requiring manual and verbal action decisions. Each decision was then reduced to a bit measure under strict assumptions of equiprobability of alternatives and independence of sequential actions. Information transmission demands were then calculated for each task to obtain scale values.

A second scaling technique was used to generate a more comprehensive estimate of loading by deriving weights for information processing activities not accounted for in the first effort, and adding them to the information theoretical scale values. In order to derive weights for the demands of information gathering activities, memory requirements, and instructional complexity, 15 messages which varied along these dimensions were extracted from the sample tasks and arranged in a paired-comparisons format. Forty A-7 and A-10 pilots were asked to examine each of the 105 pairs and to indicate which of the two entailed the greater workload. An interval scale was derived from the data using Thurstone's Law of Comparative Judgment. The scale values for the 15 messages were then used to produce weights for the extra processing activities by generating a set of simultaneous equations where the summed effects of each activity were set equal to the total scale value. A second hybrid scale was derived by selectively adding the weighting factors to the normalized bit scores for each of the original tasks.

The final scaling approach tapped the subjective component of workload. Thirty pilots were asked to examine each of the 13 complete communications tasks and to rank them according to workload. An extension of Thurstone's technique was used to derive the third set of a priori estimates. The three scaling techniques were found to generate fairly consistent estimates of the workload produced by the communications tasks. Kendall's coefficient of concordance revealed a significant level of agreement among the information theoretical, hybrid, and subjective scale values ($W = .929$, $p < .01$).

Sensitivity Analysis

The development of the concept of using radio communications tasks for workload assessment and the use of a priori scaling techniques to estimate the demands associated with these tasks were preliminary steps toward the design of a viable workload assessment methodology. In order to validate this approach, two questions must be answered through behavioral research. The hypothesis that the use of realistic embedded radio communications tasks will minimize primary task interference is based upon the assumption that trained and experienced aircrew members have developed performance strategies which promote optimal performance on all crewstation tasks, and that, as workload increases, these strategies are biased in favor of higher priority functions at the expense of communications performance. To test this hypothesis, research is needed in which qualified flying personnel are required to perform secondary communications tasks in the context of a realistic full-mission scenario simulation.

However, a more basic question concerns the sensitivity of the workload measures that can be obtained from embedded communications tasks. Although the scaling methods employed by Shingledecker et al. (1980) revealed a wide range of estimated loadings associated with the sample communications tasks, these data cannot predict which, if any, of the tasks will yield sensitive measures when primary task workload is manipulated in a dual task environment.

The major concern of the present effort was the investigation of the ability of secondary radio communications tasks to produce useful measures of primary task workload. In the experiment described in this report, civilian subjects were asked to perform a subset of the A-10 communications tasks in a fixed-base mock-up aircraft cockpit. A unidimensional unstable tracking task was used to produce controlled dual task loading. The study was designed to serve three specific purposes. First, the experiment was intended to explore the range of time and accuracy metrics available for communications task performance in order to select one or more standard indices. Second, dual task demand was manipulated to assess the differential sensitivity of the communications tasks to workload. Finally, performance data were obtained as criteria for evaluating the predictive ability of the a priori scaling methods described by Shingledecker et al. (1980) in order to determine their value for future selection of appropriate subsidiary communications tasks.

Section 2
METHODS

SUBJECTS

Four male and two female paid civilian subjects, with ages ranging from 18 to 26 years, served in the experiment. Subjects were required to be right-handed. Their visual acuity was 20/20 corrected. All subjects had previously served in a variety of other experiments involving the assessment of human operator tracking performance under varying experimental conditions. None of the subjects had previously performed the radio communications activities evaluated in this experiment.

APPARATUS

Subjects were seated in a simplified simulated fighter aircraft cockpit (Figure 1). The instrument panel (Figure 2) was a wooden mock-up containing a Panasonic Model WV-5200 5-inch black-and-white television monitor, a digital clock with a lighted, 0.5-inch liquid crystal display, and a simulated threat warning display consisting of eight subminiature light bulbs at 3, 6, 9, and 12 o'clock on each of two concentric circles that correspond to ranges of 3 and 6 miles. The ninth bulb was located at the center of the display (Figure 3).

Communications panels from an actual A-10 aircraft were mounted in the rails located on the left side of the pilot's seat (see Figure 4). The panels were back-lighted with the bulbs supplied. The panels consisted of the IFF panel, the VHF radio, the UHF radio, the FM radio, the INTERCOM panel, and the antenna switch panel (see Figure 5). Switch positions were recorded for 15 switches on the IFF, UHF, and INTERCOM panels. On the IFF panel, these included the M-1, M-3/A, IDENT, and MODE 3/A thumbwheel switches. On the UHF panel, the switches recorded were the UHF PRESET CHANNEL SELECT, the UHF MANUAL TUNE, and the UHF MODE switch. The INTERCOM MODE SELECT switch was the only recorded switch on the INTERCOM panel.

Figure 1. Simulated Fighter Aircraft Cockpit

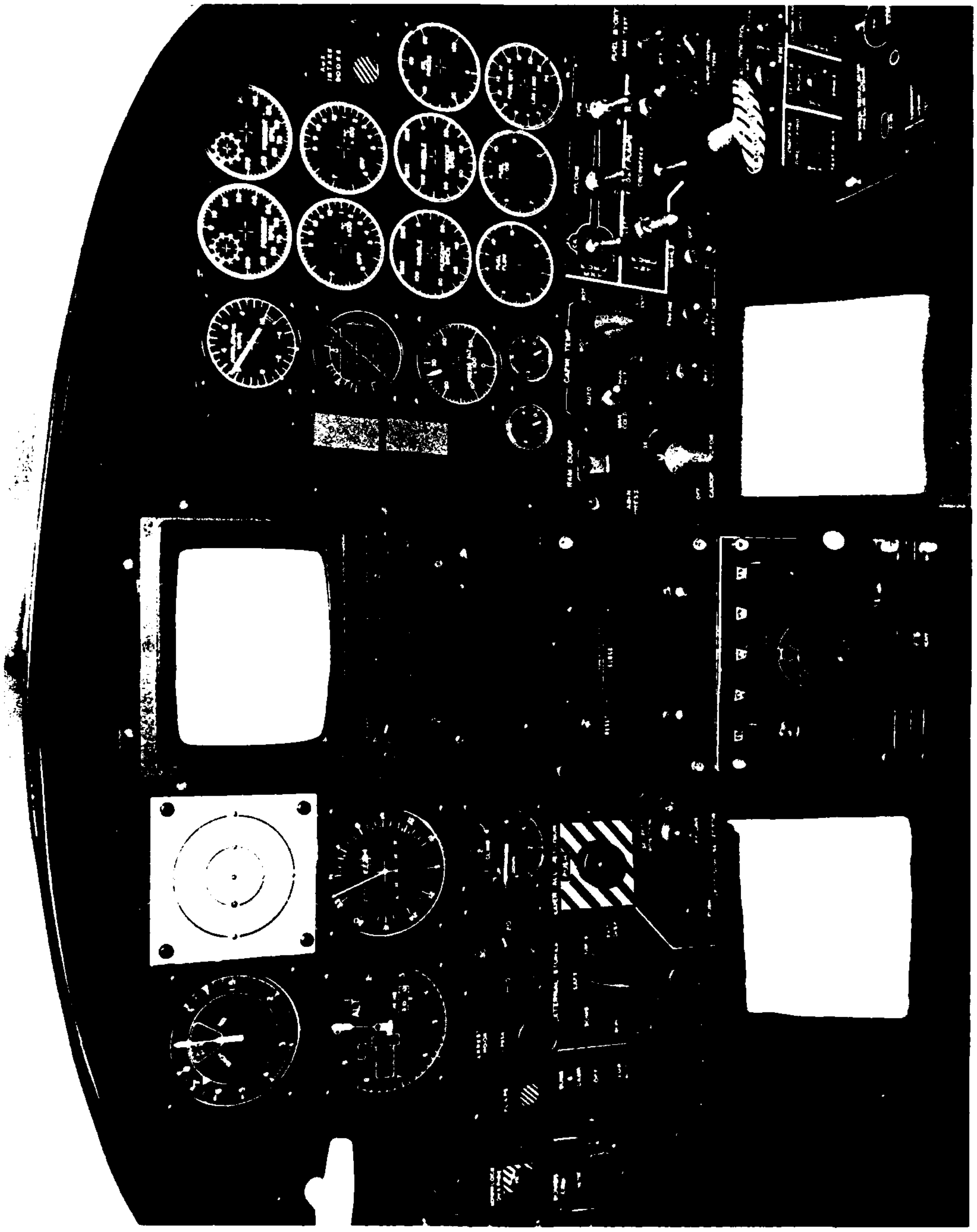


Figure 2. Instrument Panel

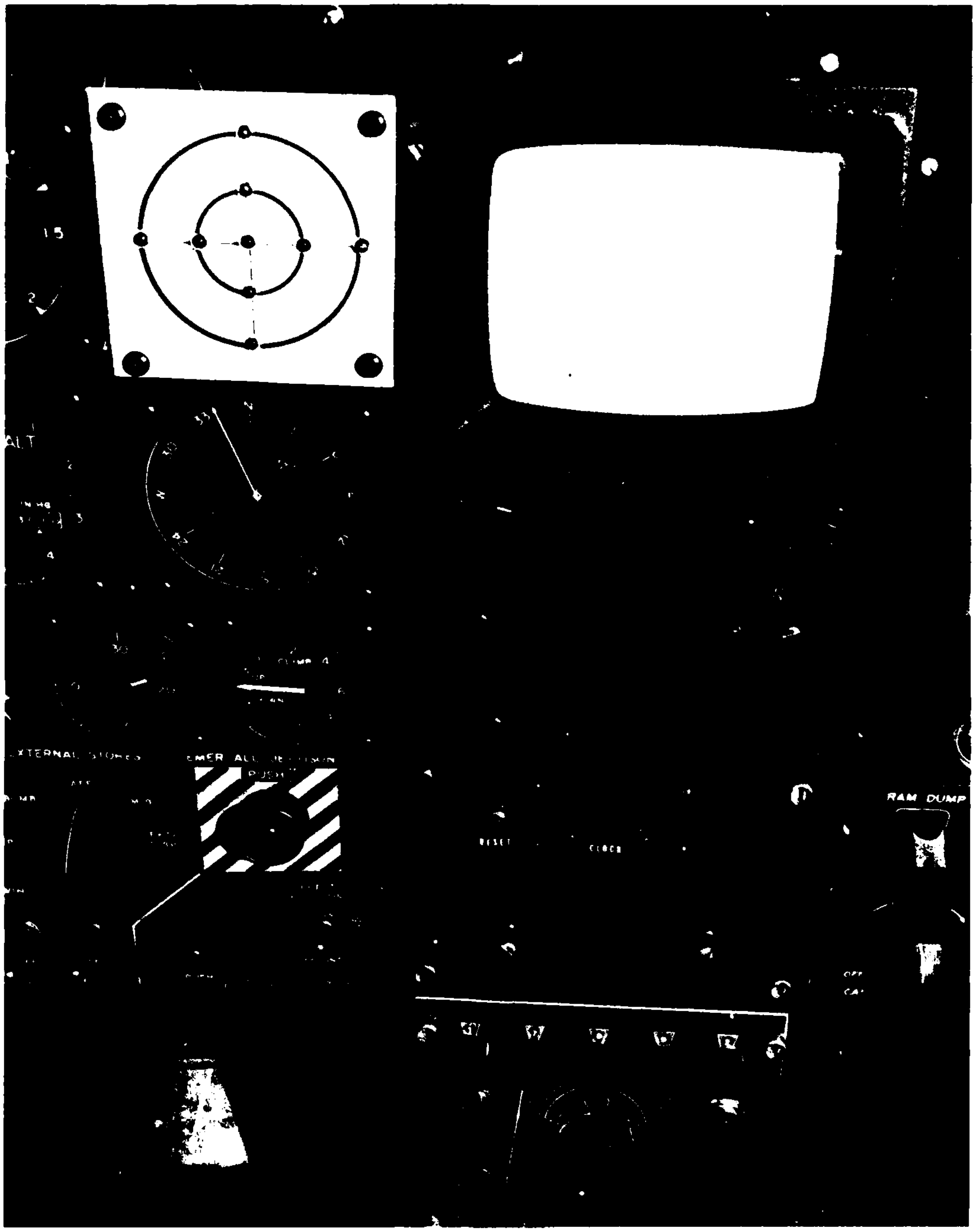


Figure 3. Close-Up of Threat Warning Display, Target Display, and Clock

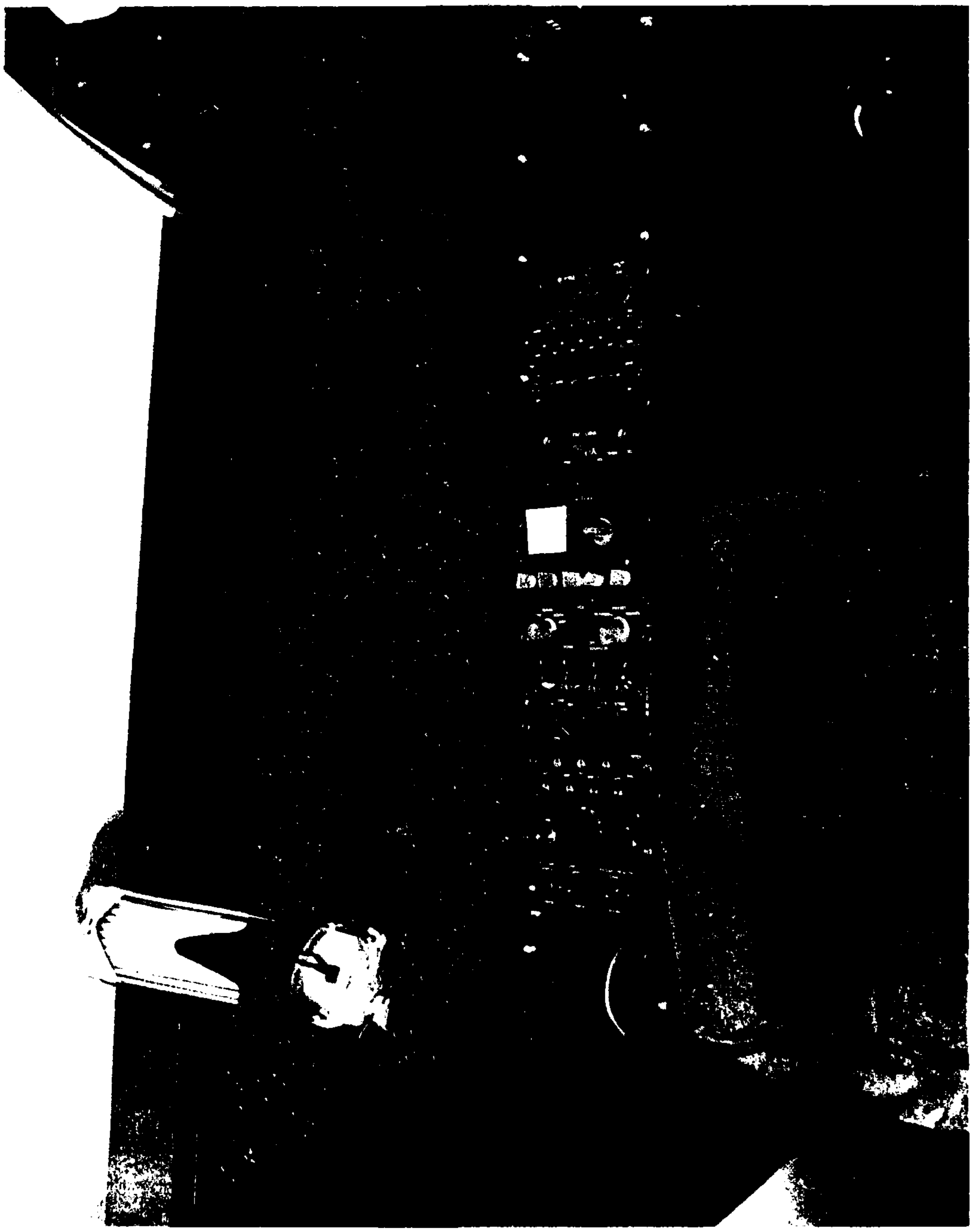


Figure 4. A-10 Communications Panels Mounted in Simulated Cockpit

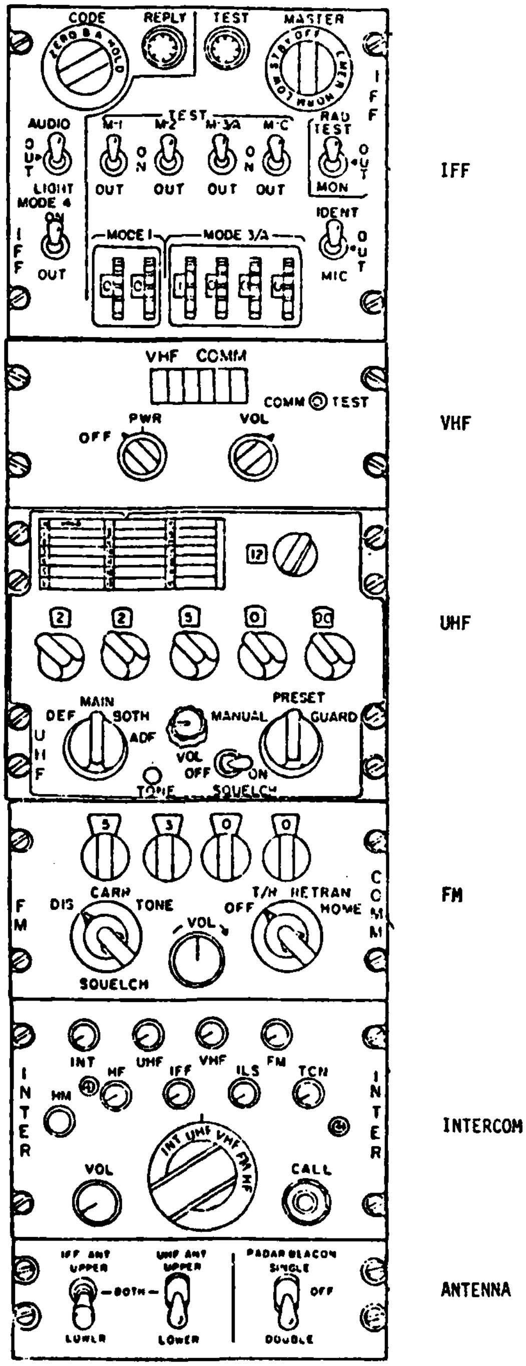


Figure 5. Line Drawing of A-10 Communications Panels

A simulated stationary throttle was located to the left of the pilot's seat and just in front of the radio panels. A push-to-talk, three-position microphone switch was located in the end of the throttle grip so that the subject could operate it with the thumb of his/her left hand (see Figure 4). Subjects wore a standard Air Force headset with boom microphone. The headset was connected to a communications system that permitted two-way conversation between the experimenter and the subject, and mixing of realistic background communications with the messages that were read by the experimenter. The background activity was recorded from the radios of actual fighter aircraft during practice missions. The system also provided the subject with audible jamming signals and tones that identified waypoints. In addition, the communications system permitted the recording of all signals heard by the subjects and all of their responses.

A control stick was centered on the floor of the cockpit 12 inches in front of the pilot's seat. The control stick was 24 inches high and had a useful travel of approximately 45 degrees from side to side. No attempt was made to simulate the control stick resistance found in actual aircraft. However, mild spring loading was used to provide self-centering of the stick.

A version of the critical tracking task (Jex, 1966), which is a single-axis compensatory tracking task, was presented on the video monitor. At a viewing distance of approximately 65 cm, the subject saw a 1.0 x 2.0 cm fixed target centered on the screen. The cursor, which moved laterally from the center of the screen, was identical to the target. The subject attempted to keep the cursor centered over the target by making inputs with the control stick. An input by the operator resulted in a voltage which was combined by weighted summation with positive feedback voltage from the system output. The weighted sum of the two voltages was then multipled by predetermined values (k) before being fed to an integrator. The integrator had an RC time constant associated with it which had the effect of multiplying the integral of the input voltage by 1/RC. The output of the integrator (system output) was then fed back through a potentiometer, whose setting remained fixed for the duration of the experiment, and was combined as described with the stick input voltage.

Because the feedback was positive, the system being controlled was inherently unstable; and the operator had to compensate for the instability. An initial error in cursor position was negated by the operator control inputs. The operator control input included some error which, in turn, was negated by further control inputs, which contained more error, etc. The operator was essentially in the position of having to control his own remnant (nonlinear and noise inputs).

As shown in Figure 6, the pertinent system parameters are the respective summation weights (G and F) of the stick and feedback voltages, respectively; the multiplication factor (K) applied to the weighted sum before being input to the integrator; and the time constant (RC) of the integrator.

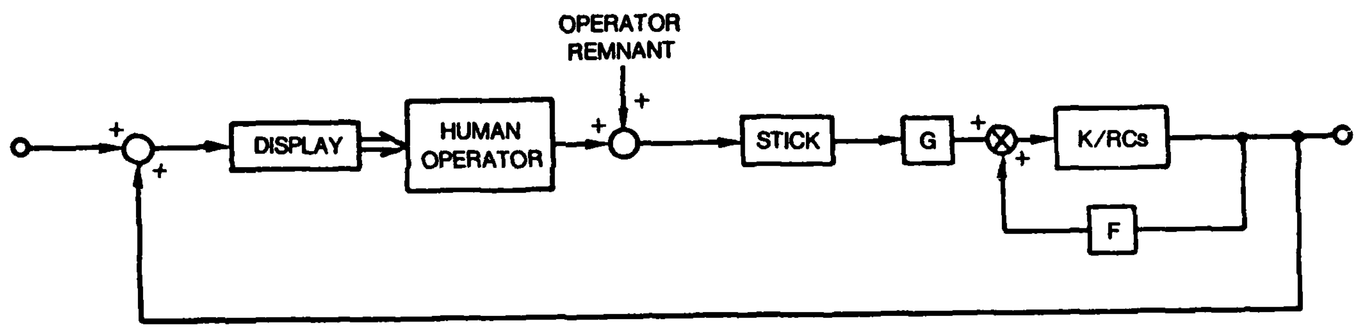
The system transfer function was:

$$\frac{GK/RCS}{1-{}^{FK}/RCS} = \frac{GK/RC}{S-{}^{FK}/RC} = \frac{GF^{-1}\lambda}{S-\lambda}$$

where $\lambda = {}^{FK}/RC$ or about 6 K, varying from 0 to 6 rad/sec.

Cursor movement dynamics were defined by the transfer function 33 $1/3\lambda/S-\lambda$. This expression represents an unstable system for which the response increases exponentially in magnitude for virtually any input that is not dependent upon the output. Lambda (λ) is the rate of exponential increase of the output. That is, a component of the response was always increased in proportion to $e\lambda t$ making the response unbounded over time. As λ was increased, the cursor became more difficult for the operator to control as he/she was increasingly forced to respond to the velocity of the cursor movement in addition to its position. Since there are inherent limits on the ability to respond in this situation, the operator lost control at some point. When the cursor reached either edge of the display area, it immediately reset to the center position and the operator continued the task.

A box containing the circuitry for presenting the tracking task was located on the experimenter's console. Digital displays for presenting integrated



G: GAIN APPLIES TO STICK VOLTAGE $G = 2$

K: EXTERNAL ADJUSTABLE POT SETTING $0 \leq K \leq 1$

RC: INTEGRATOR TIME CONSTANT $RC = .01$

F: INTERNAL FIXED POT SETTING $F \approx .06$

Figure 6. Tracking Task System Diagram

error and control losses, and switches for starting and stopping the task were provided to the experimenter. The experimenter selected the threat locations on the threat warning display by switching to the appropriate bulb when needed. A control box on the experimenter's console permitted selection of individual bulbs.

The communications system controller was also located at the experimenter's console. The controller permitted the experimenter's microphone to be turned on or off, the presentation of jamming signals and waypoint identifications, and starting and stopping of the two tape recorders that recorded subjects' responses and presented the background radio chatter. A clock controller was available to the experimenter for resetting the subjects' clock to 00:00 and for controlling the clock's illumination.

Because of the need for automatically timing and recording the positions of the switches on the radio panels, and for providing a script for the experimenter, the simulator system was designed around a PET 2001 microcomputer. All data entry by the experimenter was through the system keyboard and all commands were entered after system prompts. The computer system was interfaced to the radio switches by a National Semi Conductor 16 channel multiplexer and A/D converter. All of the switch change data were collected through the interface.

Most of the radio switches toggled from 0 to +5 VOC when actuated, although some switch settings were divided into several voltage levels from 0 to 5 by voltage divider networks. The IFF code select switch signals were provided by using four digital thumbwheel switches. Their outputs were connected to a D/A converter and transmitted to the PET via one of the 16 A/D lines. Binary values of each digit displayed on any thumbwheel were presented to the D/A converter which transformed it to an analog voltage between 0 to 5 volts DC. When the A/D converter on the PET read this channel, the voltage was transformed by software into the numerical value on the thumbwheel switch. The four thumbwheel switches were connected to four channels of the PET A/D converter. This design required the minimum number of interface wires between the simulator and the PET.

All operations software was written in BASIC provided with the computer. This included the experimenter's input and most data output and formatting routines. The interface software drivers between the PET and the simulator were written using the 6502 assembler. All assembly language routines were called from BASIC, and data along with controlling parameters were passed using the protocol provided for in BASIC. Switch positions and time of change were displayed to the experimenter through a 300-baud serial interface between the PET and a Decwriter line printer.

DESIGN

An 8 x 3 repeated measures design was used to investigate the effects of communications task characteristics and tracking task difficulty on subjects' tracking performance and on their verbal and manual responses to communication requests. The dependent measures for the communications tasks were verbal and manual response times and accuracy. Tracking performance was assessed by the number of control losses that occurred in an experimental trial.

The 24 experimental conditions were composed of single task performances on 8 communications tasks and on low and high difficulty tracking tasks, and of a full factorial combination of each communications task performed in conjunction with the two tracking tasks.

PROCEDURE

Each of the six subjects attended four 3-hour sessions run on consecutive days. The first 3 days were devoted to familiarization and practice while data was collected during the final session. On the first day subjects were introduced to the simulated cockpit, the communications system, and both of the experimental tasks. Following familiarization with the various switches and displays to be used in the experiment, invididual single task practice was initiated.

An adaptive training regime was used during tracking task practice in order to determine the maximum level of instability at which each subject could maintain control of the task. Both trial duration and task instability were varied during the training sessions which took place on the first and second days of the experiment. Early in practice, subjects tracked for 30 seconds and task instability was raised in large increments. For later sessions, task duration increased to 150 seconds and instability was varied in smaller increments. In all cases, a criterion of five or fewer control losses was used to raise or lower task instability in an adaptive manner. At the end of the second day, subjects had received a total of approximately 45 minutes practice on the tracking task. Tracking difficulty criteria were then established for individual subjects by multiplying the maximum level of instability achieved by .95 for the high difficulty condition and .60 for the low difficulty condition.

Complete scripts for the eight communications tasks used in the experiment are shown in Table 1. Communications task practice was interspersed with tracking task practice during the first two experimental sessions. While the same communications tasks were used in both practice and test sessions, channel and code values were varied randomly throughout the experiment to prevent memorization. The variable components of each of the tasks are underlined in Table 1. Communications tasks practice consisted of 24 trials with each followed by feedback from the experimenter.

On the third day, subjects again practiced the communications and tracking tasks singly followed by two practice replications of each of the 16 dual task combinations.

The final experimental session was devoted entirely to testing and data collection. Following a brief warm-up period on the tracking and communications tasks, subjects performed each of the communications and tracking tasks both singly and in all dual task combinations in a randomized sequence. During dual task trials, subjects were instructed to begin tracking as soon as they heard background radio chatter over the headphones. After a randomly determined period of 10 to 20 seconds, the experimenter read

TABLE 1. RADIO TASKS USED IN EVALUATION

TASK A1	SUBJECT'S MANUAL RESPONSES	VERBAL COMMUNICATIONS
NAIL:		Nail to Tiger 2, call 1 minute out of Bravo
TIGER 1:	(Pushes MIC down) (Resets CLOCK to 0:00 when beep is heard in headphone)	Roger, Nail
	(1 minute later, pushes MIC down)	Nail, Tiger 1 out of Bravo now
NAIL:		Roger, 1
TASK A2		
NAIL:		Nail to Tiger 1, call Friendly 1 minute out of Bravo on UHF 232.100
TIGER 1:	(Pushes MIC down) (Resets CLOCK to 0:00 when beep is heard in headphone, sets UHF MANUAL TUNE switches to 232.100, sets UHF MODE switch to "manual")	Roger, Nail
	(Pushes MIC down 1 minute after resetting clock)	Friendly, Tiger 1 out of Bravo now
FRIENDLY:		Roger, Tiger 1
TIGER 1:	(Returns UHF MODE switch to "preset," pushes MIC down)	Nail, Tiger 1 back on your frequency
NAIL:		Roger, 1
TASK B1		
NAIL:		Nail to Tiger 1 (squeal)
TIGER 1:	(Hears squeal and changes UHF PRESET CHANNEL SELECT to 2) (Pushes MIC down)	Nail, unable on Channel 1. Do you read 2?
NAIL:		Roger, 1, remain this frequency
TASK B2		
NAIL:		Nail to Tiger 1 (squeal)
	(Pushes MIC down) (Hears squeal and changes UHF PRESET CHANNEL SELECT to 2)	Nail, unable on channel 1. Do you read 2?
		(Squeal)
TIGER 1:	(Continues to hear squeal, changes INTERCOM MODE SELECT switch to VHF, pushes MIC switch up to call POUNDER)	Pounder, unable to read Nail on 1 and 2. Do you have another frequency?
POUNDER:		Roger, Tiger, go UHF 9
TIGER 1:	(Pushes MIC up) (Changes UHF PRESET CHANNEL SELECT to 9, changes INTERCOM MODE SELECT switch back to UHF)	Roger, Pounder, UHF 9
	(Pushes MIC down)	Nail, Tiger 1. Do you read?
NAIL:		Roger, Tiger, remain this frequency
TASK C1		
NAIL:		Nail to Tiger 1, report SAMs
TIGER 1:	(Looks at THREAT WARNING SYSTEM display, pushes MIC down)	Tiger 1 has SAMs at 6 o'clock and 3 miles
NAIL:		Roger, Tiger 1
TASK C2		
NAIL:		Nail to Tiger 1, go UHF 5, report SAMs to Dogbone
TIGER 1:	(Pushes MIC down)	Roger, Nail
	(Changes UHF PRESET CHANNEL SELECT to 5, looks at THREAT WARNING SYSTEM display, pushes MIC down)	Dogbone, Tiger 1 has SAMS at 6 o'clock and 3 miles
DOGBONE:		Roger, Tiger 1
TASK D1		
PARADISE:		Tiger 1 from Paradise, squawk ident
TIGER 1:	(Pushes IDENT switch to "ident," changes INTERCOM MODE SELECT switch to VHF, pushes MIC up)	Roger, Paradise
TASK D2		
PARADISE:		Tiger 1 from Paradise, squawk 30400
TIGER 1:	(Changes M-3/A switch to "on," sets four MODE 3/A thumbwheels to 0400, changes INTERCOM MODE SELECT switch to VHF, pushes MIC up)	Paradise, Tiger 1 squawking 30400
PARADISE:		Roger, 1

the initial message for the communications task. Tracking task performance was recorded only while subjects were performing a communications task (approximately 10 to 100 seconds). Baseline tracking performance was recorded over a period of approximately 60 seconds.

Section 3
RESULTS

COMMUNICATIONS PERFORMANCE MEASURES

In order to select an optimal standard index of communications task performance, several individual measures were collected under single and dual task conditions and assessed for variability as a function of loading. The raw data were transcribed from the switch data acquisition system output and from the taped records of verbal performance to derive response time, switching error, and verbal error scores.

The potential variety of time-based events that could be examined varied considerably depending on the number of logical segments into which a particular task could be separated. However, since statistical comparisons would be facilitated by using common measures that were available from all communications tasks, the following three response time measures were chosen for analysis.

1. Time to first switch action: the reaction time (RT) to the receipt of a relevant message. RT was calculated as the elapsed time from message onset to the occurrence of the first switch action. Normally, this action was the microphone switch response to confirm the receipt of the message. However, in four of the tasks where no confirmation was required, the initial responses were the times taken to tune to an alternate channel because of jamming (tasks B1 and B2), to depress the ident switch (task D1), and to activate the mode 3A switch on the IFF panel (task D2).

2. Request to response time: the time taken to carry out any information gathering or switching activities needed to respond to an input message. Response times were calculated from the termination of an instruction to the completion of the requested activities as measured by a final switch action or verbal statement.

3. Total task time: the total time to complete a communications task measured from the onset of an input message to the final switch action.

Switching errors were assessed by summarizing any functionally grouped series of switch activities as a single instance of manual behavior. Thus, if changing a radio frequency involved the manipulation of more than one switch, only one error was recorded regardless of whether one or more of the digits was dialed incorrectly. Errors included instances of reversal and overshooting of switch positions, as well as incorrect final positions. Verbal response errors were assessed for each task by tabulating the number of times subjects failed to respond to an input message, used an inaccurate radio call name, failed to confirm a message, or required message repetition. Since the frequencies with which manual and verbal errors occurred were very low, all error measures were combined for purposes of analysis.

Evaluations of the four time and error scores discussed above were performed using separate single factor analyses of variance (ANOVAs) for each communications task in which baseline single task and dual task performances were compared. Inspection of Table 2 shows that the error count and the time to first switch action measures were relatively insensitive to manipulation of dual task loading. However, in six of the eight tasks, measures of request to response time and total task time varied significantly with concurrent task load. Since total task time is a more generalized common measure of performance and its pattern of significant findings was redundant with the request to response time measure, this index of communications task performance was selected for sensitivity analysis.

SENSITIVITY ANALYSIS

The primary purpose of the study reported here was to assess the sensitivity of workload measurement obtainable with the secondary communications task methodology. Accordingly, the eight communications tasks with unknown relative workloads were performed in conjunction with a tracking task which had two fixed levels of loading. In this experimental design, sensitivity can

TABLE 2. INDIVIDUAL COMPARISONS OF CANDIDATE COMMUNICATIONS TASK MEASURES

COMMUNICATIONS TASK	First Switch Time				Request/Response Time				Total Time				Verbal and Manual Errors			
	*B	DL	DH	P<	B	DL	DH	P<	B	DL	DH	P<	B	DL	DH	P<
A_1	5.00	4.45	5.51	ns	4.17	6.25	9.91	ns	72.38	73.88	78.48	ns	0	0	0	ns
A_2	10.00	11.00	11.83	ns	6.81	15.31	13.57	.01	88.51	109.48	107.90	.01	.33	1.00	1.33	ns
B_1	4.45	7.76	6.86	.01	3.42	7.58	6.92	.01	8.67	12.69	11.44	.01	0	.50	.50	ns
B_2	4.27	7.09	6.47	ns	9.50	17.42	18.50	.05	31.13	42.59	43.36	.01	1.00	3.00	3.50	0.5
C_1	5.00	4.43	5.17	ns	3.58	3.83	5.50	ns	9.19	8.61	8.99	ns	0	0	0	ns
C_2	7.23	9.89	21.52	ns	10.75	18.42	21.17	.05	20.80	27.92	31.86	.05	.33	.83	.67	ns
D_1	5.68	8.93	7.94	ns	6.85	10.58	10.33	.05	11.34	14.58	15.71	.05	0	.50	.83	ns
D_2	9.04	13.65	20.33	.01	12.83	21.72	32.42	.01	21.79	33.06	39.26	.01	0	.83	.33	ns

*B: baseline single task

DL: dual task low difficulty tracking

DH: dual task high difficulty tracking

P<: statistical significance

be defined as the existence of a significant positive change in a workload measure as a function of an increase in workload. Since the secondary task paradigm assumes that mental resources can be shared among tasks in a multiple task situation, the most accurate measurement of loading effects should be obtained when combined performance on the dual tasks is used as an index of sensitivity.

In order to assess mutual performance on the communications and tracking tasks, dual task scores were expressed as decrements from single task performance baselines. For each of the communications tasks, performance decrement was measured as the percent increase (or decrease) in total task time from the baseline to the dual task conditions. Performance on the tracking task was measured as the ratio of the number of control losses to the total time during which subjects were required to track the target. Because the experiment was designed so that subjects were able to control the tracking task perfectly under both levels of difficulty when performed singly, dual task decrement was measured as the absolute number of time-averaged control losses that occurred in the combined task conditions. A common scale was derived to express dual task decrements on the communications and tracking tasks by applying separate normal score transformations to each dependent variable. The scales were then aligned by adding mean "zero decrement" normal scores to each individual score.

Using scores obtained by the above method, the combined dual task decrement for each communications task-tracking task pair can be illustrated as a single point in a mutual interference space. As noted by Wickens (1981), a dual task observation expressed in this way is equivalent to a point on a performance operating characteristic curve (Norman and Bobrow, 1975) and represents the decrement in two combined tasks relative to their respective single task performance levels. The data obtained from the present experiment for each communications task under low and high difficulty dual tracking task conditions are plotted in this fashion in Figure 7. For purposes of illustration, theoretical performance operating characteristic curves on task D2 are also shown in Figure 7.

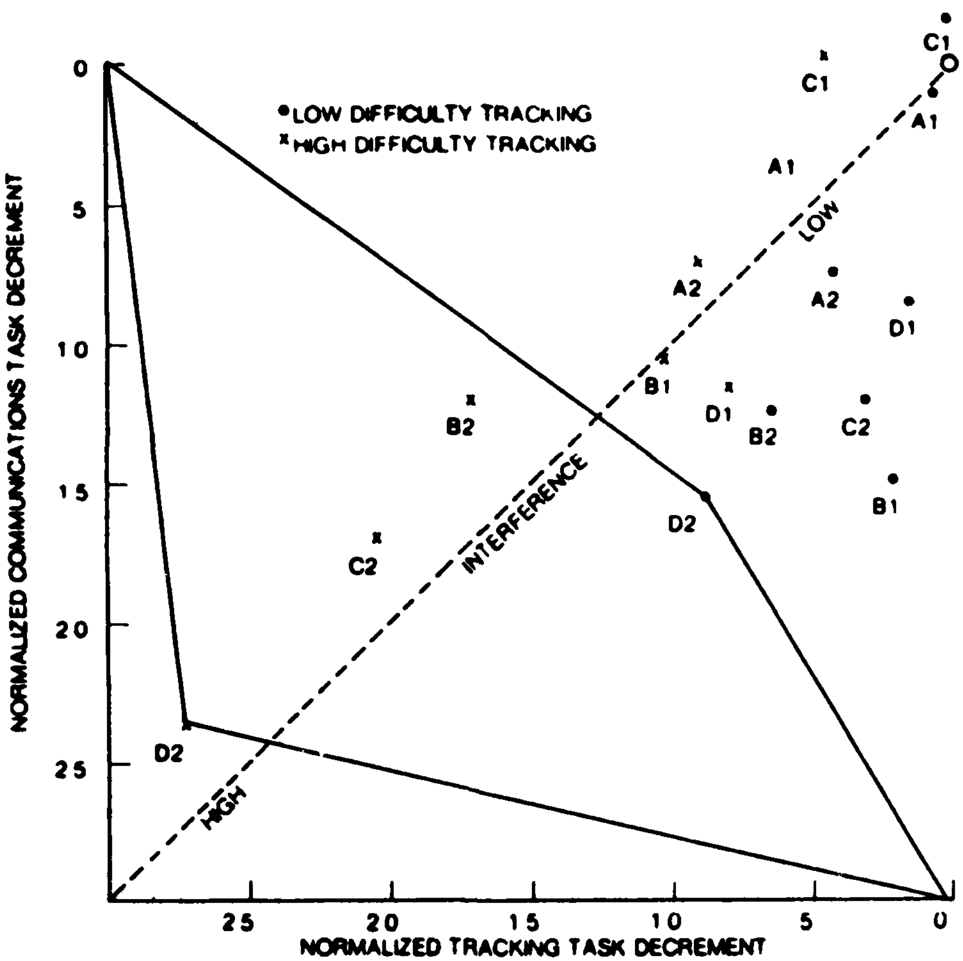


Figure 7. Combined Dual Task Performance Decrement

Two dimensions of variation can be observed in Figure 7. The relative locations of data points along the negative diagonal reveal the resource allocation policies adopted by the subjects in time sharing the two tasks. A point located toward the vertical axis indicates an allocation of resources favoring the communications task while a point located nearer to the horizontal axis indicates a bias toward the tracking task. Although of secondary importance to this experiment, Figure 7 shows that, in general, the strategy used by subjects to distribute attention in the dual task conditions varied as a function of the difficulty of the tracking task. In each case, subjects tended to allocate more attention to communications task performance when tracking task demand was high than when it was low.

A second dimension of performance variability displayed in Figure 7, which is of primary importance to the assessment of sensitivity, is the level of mutual interference observed under each experimental condition. Positions of points along the positive diagonal are representative of the degree to which two tasks can be efficiently time-shared. Since the secondary task technique relies on a decrement in one or both of the coincidentally performed tasks to generate a measure of workload, dual task data points located near the upper right corner of the space which show little combined decrement are associated with communications tasks that are unlikely to produce useful measures. Conversely, as mutual interference increases, data points are shifted toward the lower left corner of the space and communications tasks associated with them can be expected to provide reliable workload measures.

An overview of the method for interpreting dual task data described above suggests two criteria for judging the sensitivity of workload measures available for the communications tasks under evaluation. First, a sensitive task must generate sufficient interference to produce an observable performance decrement in one or both tasks. Second, those tasks which do result in mutual interference must also be differentially sensitive to changes in tracking task loading. In order to test the communications tasks against these criteria, a series of statistical analyses were performed in which dual task interference for each task pair was treated as a bivariate observation consisting of normalized decrements on the two tasks from their respective single task performance levels.

An overall hypothesis test was accomplished using a three way (subjects x communications task x tracking task difficulty) multivariate analysis of variance. The results of this analysis yielded significant effects of tracking difficulty (approximate $F = 23.88$, $p < .01$) and of communications task (approximate $F = 9.24$, $p < .01$). The interaction was not significant ($p > .10$). Thus, both the level of demand on the tracking task and the loading differences among the eight communications tasks were reliable contributors to dual task performance decrement.

Inspection of Figure 8 shows that the various dual task combinations produced a variety of interference effects ranging from an average improvement in communications task performance and a small tracking decrement on task C_1 when combined with the low difficulty tracking task, to large decrements in both performances when task D_2 was combined with the high difficulty tracking task. In order to discrimnate among the levels of interference produced by the communication tasks, multivariate post-hoc comparisons were performed on the effects averaged over the two tracking demand levels. No significant differences were detected among tasks A_1, A_2, B_1, B_2, C_2, and D_2. However, all of these displayed significantly greater interference effects than task C_1. Since all of the tasks that produced statistically equivalent effects yielded mean positive interference on both performance measures while task C_1 actually improved during one dual task condition, tasks in the former group were selected for further analysis to determine their sensitivity to differential tracking task loading.

Planned multivariate comparisons between the low and high tracking demand conditions for each communications task yielded four significant effects. Of the seven tasks which produced measurable dual task interference, tasks B_1, B_2, C_2, and D_2 also generated significantly greater interference when performed with the high demand tracking task than with the low demand

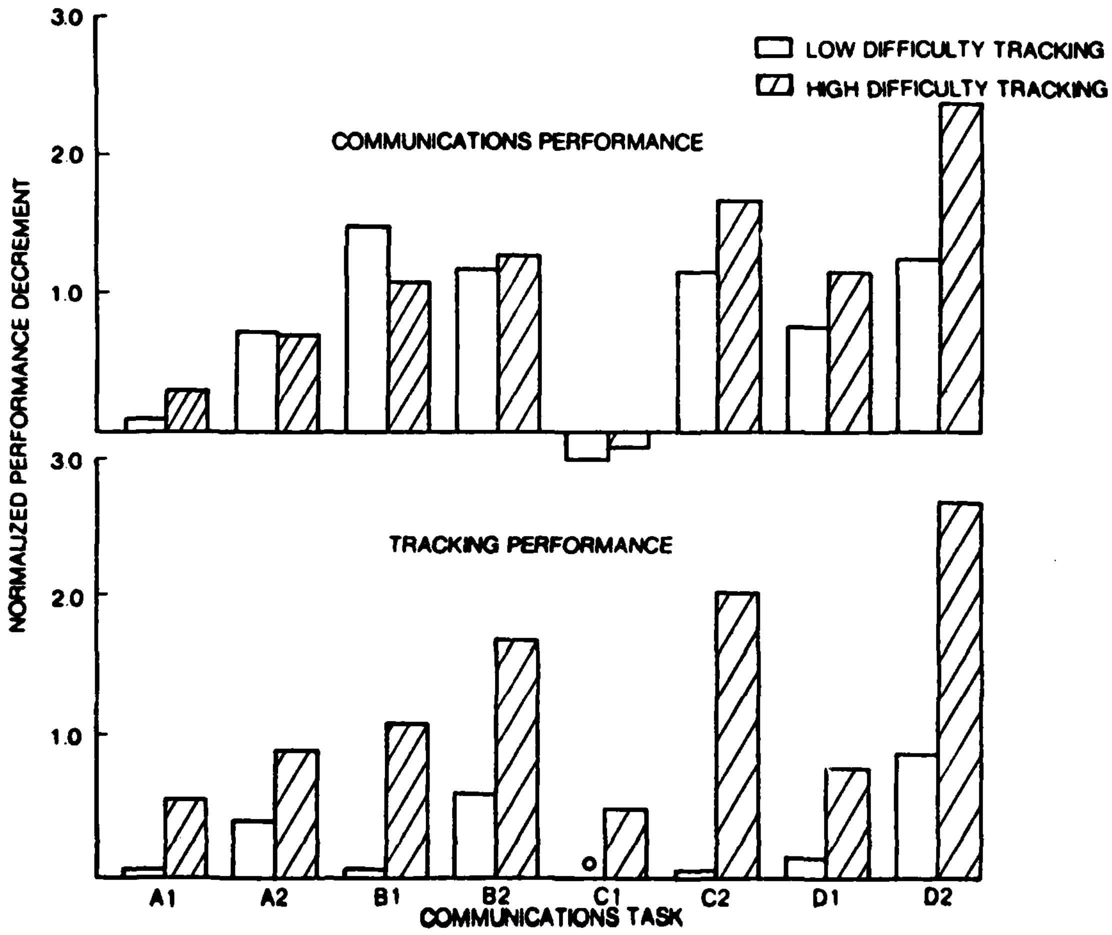


Figure 8. Dual Task Performance

tracking task ($F_{2,79}$: $B_1 = 5.52$, $p < .01$; $B_2 = 3.47$, $p < .05$; $C_2 = 11.53$, $p < .01$; $D_2 = 8.86$, $p < .01$).

Subsequent to the determination of multivariate signficance, separate univariate analysis of variance were performed on the communications task and tracking task performance variables to evaluate their relative sensitivities to loading. Mean tracking control losses were shown to be significantly affected only by communications task demand differences ($F_{7,7} = 9.23$, $p < .01$). However, dual task decrement in communications task duration proved to be statistically sensitive to both factors (tracking demand $F_{1,7} = 43.87$, $p < .001$; communications demand $F_{7,7} = 5.71$, $p < .01$). Thus, although both of the individual measures contributed to the estimation of workload, communications task performance tended to be more sensitive to manipulations of task loading. This finding is of particular importance for applied workload assessment since it indicates that performance on secondary communications tasks can provide a common univariate index of workload in cases where detailed measures of primary task performance are difficult to obtain.

CORRELATIONS

In order to assess the concurrent validity of the secondary communications task measure of workload, a subjective rating scale was used to obtain estimates of workload after each experimental trial. The instrument used for this rating procedure is known as SWAT (Subjective Workload Assessment Technique) and is composed of three, three-point scales on which time loading, mental effort, and psychological stress are assessed (Reid, Shingledecker, and Eggemeier, 1981). Conjoint scaling procedures are used in SWAT to derive an interval scale which represents a multidimensional rating as a single workload value.

A coefficient of correlation computed between mean communications task performance decrements and obtained SWAT scale values revealed a significantly high level of agreement on the two independent measures of combined task workload ($r = .703$, $p < .01$).

Measures of correlation were also used to assess the degree to which the a priori estimates of communications task loading obtained in a previous effort (Shingledecker et al., 1981) were predictive of performance decrements when the tasks were used as workload measures. Separate coefficients of correlation were computed between mean dual task communications performance and the information theoretical, analytical, and subjective scale values described in the introductory sections of this report. Positive correlations were obtained with each of the estimation techniques ($r = .64$, .56, and .42, respectively). However, statistical tests showed that only the information theoretical analysis provided a significant level of prediction of criterion secondary task performance ($p < .05$).

Section 4
DISCUSSION AND CONCLUSIONS

The results of the experiment described in this report offer empirical support for the hypothesis that realistic radio communications activities can be used as secondary tasks to provide objective measures of workload. Four of the eight aircraft communications tasks examined in this study produced significant dual task interference when combined with a compensatory tracking task, and generated measures which were statistically sensitive to variations in control difficulty. Furthermore, these metrics were shown to be significantly correlated with a subjective measure of the workload obtained for each dual task combination, thereby providing additional corroborative evidence for their utility as workload measures. The finding that mutual interference measures obtained with the communications tasks were significantly and positively related to one of the a priori methods used to scale the workload of the communications tasks is also of considerable importance. This result indicates that a simple, analytically derived information theoretical estimate of workload can be used to select or construct additional communications tasks for assessing the workload associated with a variety of other Air Force systems.

Further direction for future use of the embedded communications task technique can be obtained from a review of the individual tasks which failed to produce sensitive measures of workload in the present study. Although such a post hoc analysis cannot be scientifically rigorous, a few practical guidelines are suggested by the results. Originally it was hypothesized that a memory demand factor would contribute to the workload of communications tasks and increase their sensitivity to primary task loading. However, performance on tasks A1 and A2, which required subjects to report waypoint information following a one-minute delay, did not significantly reflect control task workload. A possible explanation for this finding is that the memory load induced by these tasks was quite low and that, in fact, the retention interval provided subjects with additional time to perform any needed radio switching actions. Thus, it appears that an embedded secondary task should be selected or designed to induce rapid, continuous performance in order to provide effective workload measurement.

A second feature of the communications tasks which were insensitive to the workload produced by the primary tracking task was that they were relatively simple tasks requiring only one or two switch actions and, in some cases, a report of information displayed in the cockpit. Conversely, the tasks which provided useful workload measures were characterized by more verbal exchanges and more complex switching activity. This difference raises the possibility that the tasks that generated significant mutual interference measures did so because they placed demands on structure-specific motor output resources in common with the primary tracking task. Such a conclusion would be congruent with the multiple resource model of workload proposed by Wickens (1980). However, since the manual response component was often confounded with other dimensions of complexity in the communications tasks used in the present study (as it is in most other actual aircraft radio communications tasks), it is impossible to draw any firm conclusions in this analysis. As a result, further research is required to address the issue of multiple capacities in which input, central processing, and output demands of communications tasks are carefully controlled as they are performed together with various piloting tasks.

Section 5
RECOMMENDATIONS

The successful demonstration of the potential usefulness of embedded radio communications tasks for workload measurement clearly indicates that further investment in the development and evaluation of the methodology is warranted. Beyond the additional refinement that might be pursued in the laboratory work suggested above, a simultaneous effort should be made to evaluate the secondary communications task technique in high fidelity aircraft mission simulation. Two specific validation criteria would be met in this research. First, prior to its implementation as a fully operational workload assessment tool, the measurement sensitivity obtained with a laboratory manual control primary task should be replicated with actual combined aircrew tasks including flight control, systems management, and other supervisory and decision making functions. Second, in order to assess the hypothesis that embedded tasks will produce less primary task intrusion than traditional secondary tasks, operational aircrew personnel must be tested in the context of a normal mission environment.

If the results of these simulation studies confirm the findings of the experiment reported here and demonstrate improved operator acceptance as well as reduced task intrusion, formal guidelines will be developed for the construction and use of radio communications workload measurement tasks. These materials will then be delivered to human factors specialists at the field level in order to permit optimal tailoring of the methodology to specific systems, missions, and crew stations.

REFERENCES

Broadbent, D. E., 1958, Perception and Communication, Pergamon Press, NY.

Jex, H. R., J. D. McDonnell, and A. V. Phatak, 1966, Critical Tracking Task for Manual Control Research, IEEE Transactions on Human Factors Engineering, HFE-7, 138-145.

Kahneman, D., 1973, Attention and Effort, Englewood Cliffs, NJ: Prentice-Hall.

Ogden, G. D., J. M. Levine, and E. J. Eisner, 1979, Measurement of Workload by Secondary Tasks, Human Factors, 21, 529-548.

Norman, D. A. and D. J. Bobrow, 1976, On the Analysis of Performance Operating Characteristics, Psychological Review, 83, 508-519.

Reid, G. B., C. A. Shingledecker, and F. T. Eggemeier, October 1981, Application of Conjoint Measurement to Workload Scale Development, Proceedings of the Human Factors Society 25th Annual Meeting, Rochester, NY, 522-525.

Schiflett, S. G., December 1976, Operator Workload: An Annotated Bibliography, Patuxant River, MD: U.S. Naval Air Test Center, SY-257R-76.

Shingledecker, C. A., M. S. Crabtree, J. C. Simons, J. F. Courtright, and R. D. O'Donnell, December 1980, Subsidiary Radio Communications Tasks for Workload Assessment in R&D Simulations: I. Task Development and Workload Scaling, Wright-Patterson AFB, OH: Aerospace Medical Rsearch Laboratory, AFAMRL-TR-80-126.

Wickens, C. D., 1980, The Structure of Attentional Resources, in R. Nickerson (ed.), Attention and Performance VIII, Hillside, NJ: Erlbaum.

Wickens, C. D., S. J. Mountford, and W. Schreiner, 1981, Multiple Resources, Task-Hemispheric Integrity, and Individual Differences in Time-Sharing, Human Factors, 23, 211-229.

Wierwille, W. W. and R. C. Williges, September 1978, Survey and Analysis of Operator Workload Assessment Techniques, Blacksburg, VA: Systemetrics, Report S-78-1010.

*U.S. GOVERNMENT PRINTING OFFICE: 1982—559-007/6018