NUWC-NL Technical Report 10,193
6 October 1992

(1)

Stimulus Presentation Formats and Measurement Techniques for the Quantification of Target Detection Performance

W. Ronald Salafia
Dino A. DaRos
Submarine Sonar Department

Naval Undersea Warfare Center Detachment
New London, Connecticut

PREFACE

This report was prepared under NUWC Project Nos. C15420 and C77400. The sponsoring activities were the Naval Sea Systems Command (NAVSEA, PMO 409) and Space and Naval Warfare Systems Command (SPAWAR, PMW 181). The technical effort was performed as part of the senior author's activities under an Intergovernmental Personnel Act (IPA) agreement for FY 1991 and FY 1992, which was set up to take advantage of the assignee's expertise in the areas of human factors engineering and research design and analysis for the benefit of NUWC's research and development efforts in display system engineering and system operability evaluation.

The Technical Reviewer for this report was Christopher Colby, Code 2112.

REVIEWED AND APPROVED: 19 October 1992

F. V. Kingsbury

F. J. KINGSBURY
HEAD: SUBMARINE SONAR DEPARTMENT

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	6 October 1992	Final

4. TITLE AND SUBTITLE
STIMULUS PRESENTATION FORMATS AND MEASUREMENT TECHNIQUES FOR THE QUANTIFICATION OF TARGET DETECTION PERFORMANCE

5. FUNDING NUMBERS
PR C15420 and C77400

6. AUTHOR(S)
W. R. Salafia and D. A. DaRos

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Undersea Warfare Center
Detachment New London
New London, CT 06320

8. PERFORMING ORGANIZATION REPORT NUMBER
NUWC-NL TR 10,193

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Naval Sea Systems Command
Washington, DC 20362

Space and Naval Warfare Systems Command
Washington, DC 20363

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

The forced-choice (FC) format for stimulus presentation and performance assessment has been gaining popularity over other formats in a variety of human performance tasks, such as target detection and acquisition. Laboratory experiments, comparison investigations, and system performance assessments that require statistical testing, for example sonar system MDL, have been reported in the literature with claims that use of the FC procedure leads to simplification of data handling and increased cost-effectiveness. These and other claims are examined in the present report and a number of concerns are raised about the nature of the information acquired when the FC method is used for the quantification of performance in tasks that primarily involve vigilance, monitoring, and search behaviors.

These concerns may be summarized as follows. First, the kinds of performance outcomes assessed using the FC format are often different from those assessed by continuous-search (CS) procedures. Second, the FC format holds some aspects of response bias constant, making it impossible to assess many variables that are of

14. SUBJECT TERMS
Forced Choice; Continuous Search; Vigilance; Psychophysical Methods; Stimulus Formats; Detection Probability; Threshold Recognition

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR

13. Abstract Continued:

paramount importance in sonar system development, such as the effects on performance of signal probability, motivation, training, and experience. Third, the procedures have sometimes been misunderstood, and the term "forced-choice" has been misapplied to situations that technically are not FC. Finally, although there may be some valid practical reasons for using the FC format for stimulus presentation in final system performance evaluation, there is no compelling evidence that the method is superior to others in terms of data analysis, programming, or cost efficiency for the controlled laboratory experiments and comparison investigations that constitute system development.

TABLE OF CONTENTS

Reverse Blank

FOREWORD

The forced-choice (FC) format for stimulus presentation and performance assessment consists essentially of presentation of two or more spatial locations or temporal intervals, one of which contains signal plus noise while the others contain noise alone. The task of the observer is to indicate which location or interval contains the signal. The FC format has been gaining in popularity over other formats, for example continuous-search (CS), in a variety of tasks involving human performance, such as target detection and acquisition. With the CS format, the observer's task is to monitor a display which may contain multiple targets with various characteristics and to respond in some specified way to each target. There are many issues concerning the use of these formats, not the least of which are the ambiguity of the boundaries between them and the fact that they are often seen as interchangeable. The present report attempts to define the boundaries of these stimulus presentation formats, as well as examine their advantages and disadvantages for the assessment of target detection performance, both generally, and with specific reference to sonar.

One concern is that the performance outcomes produced by FC procedures are different perceptually, psychophysically, and behaviorally from those produced by CS techniques. FC procedures systematically rule out the evaluation of phenomena that could be essential in the development of sonar display systems, including the effects of differences in response criteria resulting from manipulation of important variables such as signal probability, motivation, training, or experience. Furthermore, FC procedures may permit observers to shift detection strategies in a manner that could confound performance measurements. Another concern is that FC procedures have sometimes been misunderstood and terminology misapplied to situations that bear little resemblance to actual FC tasks. Finally, data-analysis, programming and cost efficiencies claimed for the FC method appear to apply only to a limited set of circumstances surrounding final system performance evaluation, but not to system development. For these reasons, it is suggested that continuous-search procedures be used for testing during research and development of advanced sonar systems, and that the use of FC procedures be limited to those situations during final system evaluation, in which it may be difficult or impractical to use the CS format.

STIMULUS PRESENTATION FORMATS AND MEASUREMENT TECHNIQUES FOR THE QUANTIFICATION OF TARGET DETECTION PERFORMANCE

INTRODUCTION AND DEFINITIONS

The purpose of the present report is to compare and contrast various methods for the quantification of target detection performance in system testing situations. Particular attention is paid to tasks involving vigilance, monitoring, and search behaviors, since these are characteristic of the types of tasks and behaviors engaged in by sonar operators, radar operators, air traffic controllers, and the like. An emphasis on these kinds of tasks mandates careful examination of two of the traditional stimulus presentation and performance assessment formats in particular, namely the FC and CS formats.

Before proceeding to an in-depth comparison of these formats, some background information is in order. This introductory section is divided into three parts. Because threshold is the basic measure of target detection performance, Part I describes the three classic psychophysical methods for measuring thresholds. Part II defines outcome measures common to all of the psychophysical methods. Finally, Part III describes standard formats for stimulus presentation and performance assessment, including especially the FC and CS formats.

I. PSYCHOPHYSICAL METHODS

Performance in many sensory, perceptual, and motor modalities is often described in terms of thresholds. A threshold is the stimulus level, expressed in energy, intensity, signal-to-noise ratio (SNR), or some comparable measure, of the point at which an observer detects the presence of a stimulus (absolute or detection threshold), or a change in the stimulus (difference or differential threshold). Thus, the threshold is a point-of-change, either from no sensation to sensation, or from one sensation to another. The concept of the threshold is not absolute, either within an individual or among individuals. Rather, it is a relative, statistical concept referring to the stimulus value at which observers can detect a change, on average, some specified percentage of the time.[1]

Thresholds are usually measured by one of a variety of procedures known as psychophysical methods. Psychophysics is a hybrid branch of science that attempts to express mathematically the relationship between the physical dimensions of stimuli and the psychological dimensions of sensation and perception. It was one of the antecedents of the science of Psychology, and is one of the major foundations on which modern research in perception and human performance has been built. Today, the classic psychophysical methods, and more modern adaptations of them, are used for the study of everything from basic sensory and perceptual processes to system development and testing in vigilance, monitoring, and search tasks. The present report will focus

on this latter use of the methods. The following discussion of psychophysical methods has been condensed to the essentials from one of the classic presentations on the subject by Woodworth and Schlosberg[1] and from an excellent update by Falmagne[2].

Traditionally, three Psychophysical Methods have been distinguished. In the *Method of Average Error (Adjustment)*, a comparison stimulus is continuously varied (adjusted up or down), either directly by the subject or by an experimenter under the direction of the subject, until a designated change in the stimulus is detected. This adjustment process is done repeatedly, and the mean and variability of the settings measure the accuracy of detection or discrimination. A number of problems and limitations in the use of this method have resulted in its being used less often than the others, and when it is used, more often than not, it is for pilot and exploratory studies, or for establishing a quick first approximation to the threshold, for use with another method. For these reasons, it will not be considered further in the present report.

On each trial of the *Method of Limits* (and its variant, the *Staircase Method*) the value of the stimulus is made to increase (ascending trial) and decrease (descending trial) in discrete steps, and the stimulus value is noted for the step on which the subject's response shifts from one category of response, e.g., "No," to another, e.g., "Yes." Each trial consists of noise alone or signal-plus-noise, and the subject's response indicates whether he or she thinks that a stimulus was present. With the method of limits, the stimulus value is recorded at the transition points and various statistical calculations made, e.g., Ascending Threshold (A = the mean of the ascending trials), Descending Threshold (D = the mean of the descending trials), Interval of Uncertainty (IU = A-D), and Point of Subjective Equality (PSE = [A+D]/2), which is usually designated as the threshold value. Analogous procedures are available for calculating thresholds with staircase methods, which are simply a more rapid methods for "zeroing in" on the threshold value.

The third and final of the basic psychophysical methods is the *Method of Constant Stimuli*. To implement this method, an approximation of the detection threshold must first be made, perhaps through the method of average error. The researcher then selects a number of stimulus values which bracket this approximate threshold. Stimuli at these values are presented many times in a random order. Each trial consists of the presentation of one of the stimulus values, to which the observer responds by indicating whether he or she perceives a stimulus. If used well, this procedure yields a sigmoid curve of performance, such that the threshold (usually defined to be the 50% point) may be located by inspection, interpolation, or statistical treatment.

II. OUTCOME MEASURES

The most important type of threshold for the purposes of the present report is the target detection threshold, i.e., the stimulus energy, usually expressed in units of SNR, at which an observer can detect a target in noise. (In the context of system development, the threshold concept

can be extended to include target recognition or target identification, but we will focus our discussion on target detection.) Regardless of which specific variant of psychophysical methodology is used, a variety of performance measures may be employed in the assessment of target detection thresholds. While these measures are grounded in the traditional measurement approaches of classical psychophysics, they are extensions of those approaches to encompass concepts needed for the development of signal detection theory. Presented below are three of the most frequently used measures, namely, probability of detection, probability of false alarms, and reaction time.[2, 3, 4, 5]

Before proceeding with these three measures, however, four additional terms must be understood, namely: hit, miss, false alarm, and correct rejection.[3, 5] If a target is present on a given trial, the observer may make a *hit*, i.e., detect it, or a *miss*, i.e., fail to detect it. On the other hand, if only noise is present on a given trial, the observer may recognize that state of affairs, i.e., make a *correct rejection*, or mistakenly report the presence of a signal, i.e., make a *false alarm*. These four outcomes are exhaustive of the possible occurrences on any given trial of a simple target detection experiment. The occurrence of these four outcome measures under various conditions of signal strength, noise level, number of signals, etc., comprise the basic data for assessment of target detection performance.

One of the most commonly used performance measures is the rate of correct detection of targets or hit rate, which is readily converted to the *Probability of Detection* or *P(D)*, that is, the number of hits divided by the number of instances where a signal was, in fact, present. Related to *P(D)* is *Minimum Detectable Level (MDL)*, which is usually defined as the SNR value at which there is a 50% *P(D)* of the signal. The *MDL* is the typical threshold measure referred to in the previous discussion of the method of constant stimuli, and it is analogous to the PSE in the method of limits. Furthermore, it is the measure that is used most consistently and is most informative in both the development and testing of sonar systems.

Perceptual sensitivity is a function not only of hits, but also of false detections or false alarms. The rate of false alarms, therefore, is the next most important measure of detection accuracy. In traditional signal detection theory, *Probability of False Alarms* or *P(FA)* is defined as the number of false detections divided by the number of opportunities to make a false detection. This is the measure that differs most across different tasks and measurement methods, and, as we shall see shortly, the technical meaning of *P(FA)* is sometimes determined not by the observer's behavior, but rather by the stimulus presentation format used to measure performance. For the present, however, it is important to note that one cannot make meaningful comparisons of performance in terms of *P(D)* or *MDL*, without having established that observers are comparable in terms of *P(FA)*.

Finally, a frequently used, although indirect, way to assess target detection performance is reaction time, which is the interval between onset of the signal(s) and the appropriate response(s). While this measure does not lead directly to the calculation of a threshold value, it can nevertheless convey a considerable amount of peripheral information about performance.

A Brief Note on Signal Detection Theory. In many current textbooks, signal detection theory (SDT) is presented as a replacement for classical psychophysics, since the underlying account of what constitutes a threshold is different for the two approaches.[5] We argue instead that SDT complements classical psychophysics. Specifically, the position taken in this report is that the "classic" methods of constant stimuli and limits (or staircase) are most useful in sonar system development and performance evaluation, and that the standard outcome measures, *P(D), P(FA), MDL,* response time, etc., are the most appropriate measures of target detection performance. Still, regardless of the psychophysical method or the performance measure used, SDT methods may be applied to performance data for the determination of observer sensitivity (d') and bias (ß), after which receiver operating characteristic (ROC) curves may be plotted.

One reason to use SDT is that it allows the investigator to separate the "true" effects of sensitivity to the targets due to signal and noise characteristics, from shifts in the observer's response strategies that may be a function of other variables.[5] On the other hand, in sonar system development, it is often desirable to assess the effects of those "other variables," such as fatigue, experience, etc., on the observers' response strategies. Two observers with the same d' could have different combinations of *P(D)* and *P(FA).* Accordingly, *P(D)* and *P(FA),* individually, may be the more critical indices of performance, while d' can provide useful additional information. A specific instance of this situation will be discussed shortly.[6]

III. FORMATS FOR STIMULUS PRESENTATION

To this point, we have seen that regardless of the psychophysical method used, thresholds are measured by presenting stimuli at different levels (energies, amplitudes, SNRs, etc., depending on the stimulus dimension being assessed) above and below an estimated threshold level, many times each, and tabulating the proportion or percentage of times the observer responds to the noise alone or to the stimulus-plus-noise at each level. From these tabulations, specific measures such as *P(D), MDL, P(FA),* and response time can be determined. Definitions of the methods and outcome measures do not, however, specify the exact format either for the presentation of stimuli on each trial or for the response of the observer. In this final section of the introduction, we consider stimulus presentation formats. Three basic techniques are distinguished, namely Yes/No, FC, and CS.[3, 4] Although the different stimulus presentation formats may be used with both the Method of Limits and the Method of Constant Stimuli, for convenience, the discussion of formats will focus on the latter method.

Yes/No Format

With the Yes/No method, the observer is required to report (verbally or by some mechanical or electronic means) whether the signal in question is judged to be present or absent on each trial. That is, the observer must say whether the display on a given trial was perceived to be signal-plus-noise or noise alone.[1,3] A variant of the Yes/No procedure, called the Rating procedure, permits subjects to be more flexible in their responses by rating the likelihood that the signal was present.

On the signal-plus-noise trials, as the energy level of the signal moves below the minimum energy required for detection, the proportion of "Yes" responses approaches zero. Thus, a psychometric function, often in the form of a cumulative normal curve, may readily be developed, relating correct detections (hits) to signal level. Since the lower limit of performance with the Yes/No procedure is zero, it is conventional is to use 50% *P(D)* as the threshold criterion *(MDL).*

Forced-Choice (FC) Format

The FC format differs from the Yes/No format in that each trial consists of more than one observation interval or spatial location preceding the response.[3] One interval or location contains the signal-plus-noise, while all others contain only noise. The observer is instructed to select the interval or location most likely to have contained the signal and to respond accordingly, e.g., by pointing or by pushing a button correlated with the correct interval or position. The number of alternative locations or intervals in which a target could appear is usually designated as "m," hence the term "m-alternative forced-choice (mAFC)." Two to four intervals or locations have been used most often in psychophysical research.

In the mAFC procedure, as the energy level of the signal moves below the minimum energy required for detection, the proportion of correct detections approaches 1/m, that is, the proportion of hits that would be expected by chance. Thus, the threshold is usually placed at the halfway point between chance performance and perfect performance. For example, in the 2AFC procedure, the lower limit of performance (the *P(D)* due to chance) is 50%, so that the threshold would usually be placed at $P(D) = 75\%$.

Continuous-Search (CS) Format

The defining characteristic of the CS format is that different numbers of targets, at different SNRs, as well as different spatial locations, may be presented on a single trial. The observer's task is to detect or identify each target, e.g., by moving a trackball-controlled cursor over the position of the target. (Recall that, while the FC format may also involve the presentation of multiple spatial locations, only one of these locations will contain a signal on any given trial.) Within the general framework of CS procedures, there are three important concepts to consider,

namely, "Vigilance," "Monitoring," and "Search." While it will not be necessary for the purposes of the present report to differentiate precisely among the three concepts, an initial distinction will be made for the sake of clarity, after which the terms will be combined.

According to Parasuraman,[7] the term *"vigilance"* has some physiological overtones in that it usually refers to the state of readiness of the central nervous system presumed to mediate performance on tasks requiring prolonged attention. One of the principal aspects of vigilance is temporal uncertainty; the signals for which the observer is vigilant tend to appear relatively infrequently and/or unpredictably. Recent research has begun to probe this state of readiness, using a variety of electrophysiological and brain-imaging techniques. Some of these studies, which are relevant to the discussion of presentation formats and performance measures, will be presented later.

When the task of the observer is to attend actively to one or more sources of stimuli in order to detect a signal or to respond in some other specific way to the information presented, we refer to the task as *monitoring*. While monitoring and vigilance are closely related, monitoring often refers to tasks with relatively more complex signals, and/or to continuous rather than discrete tasks. In *search* tasks, observers attempt to locate a stimulus or some aspect of a stimulus characterized by various amounts of spatial uncertainty.[7] Again, it can be seen that there is a considerable amount of overlap in terms of the observer behaviors required for search tasks and both vigilance and monitoring tasks.

Many target detection tasks, including most of those involving sonar, entail a combination of these three tasks, in that there may be temporal and/or spatial uncertainty, various levels of signal strength and/or signal probability, etc. It will not be necessary for present purposes to distinguish precisely among the three tasks, so the combined abbreviation "VMS" will be used to refer to tasks that fit the general vigilance/monitoring/search classification.

Theoretically all three of the stimulus presentation formats presented could be used to assess VMS performance. However, as Parasuraman[7] has pointed out, the yes/no and FC formats, while useful for establishing absolute and differential thresholds, are usually inappropriate for assessing the complexities of VMS performance. In general, researchers seem to have heeded this advice regarding the yes/no format, so we will not devote further attention to these procedures except in so far as they have implications for CS tasks. On the other hand, FC techniques seem to be used with increasing frequency in VMS research. For the remainder of this report we will argue, in agreement with Parasuraman, that CS procedures are generally better-suited to the assessment of VMS performance than are FC procedures.

COMPARISON OF STIMULUS PRESENTATION FORMATS

Now that the relevant terms and concepts have been presented, the various methods can be compared to determine which ones are optimal under what circumstances. As indicated above, some concerns are raised regarding the use of the FC format for the investigation of performance in tasks that primarily involve VMS behaviors. This discussion will revolve around three overlapping topics, namely, a) semantic issues and misapplications of terminology, b) perceptual, behavioral, and methodological differences, and c) efficiency and cost effectiveness. Because the focus of the present report is sonar system development and performance evaluation, examples of the use of these measures have been taken from a number of recent sonar research reports.[6, 8, 9, 10]

I. SEMANTIC ISSUES: COMMON TERMS, DIFFERENT MEANINGS

Hits, Misses, and False Alarms. The terms "hit," "miss," and "false alarm," are really quite different in FC and CS tasks. On a given FC trial, a response that a signal occurred in the temporal or spatial interval that actually contained the signal, is called a hit, while a response that the signal occurred in an incorrect temporal or spatial interval is called a false alarm.[3] But it must be remembered that these designations are, at best, only analogous to those in a CS task in which a hit would often be designated as movement of a cursor to the specific location of a target and depression of an enter key. Furthermore, a miss would be the failure to locate a target that was actually present, and this is distinguished from a false alarm, which would result from entering, as a target, a location that did not actually contain one. A CS trial, then, may result in both hits and false alarms. In other words, in FC tasks, hits, misses, and false-alarms are perfectly correlated with one another and are therefore only one measure, whereas in CS tasks, hits and misses are correlated, but hits and false alarms are independent measures of performance.

Probability of Detection. Recall that *P(D)* is the number of hits divided by the number of target opportunities, i.e., instances in which there was a signal. In FC tasks, *P(D)* is simply the number of hits divided by the number of trials, whereas in CS tasks, the number of hits is divided by the number of targets presented, because the number of target opportunities is usually different from the number of trials. Given the large amount of information contained in each CS trial, CS experimental sessions can be divided into successive time periods or blocks, and the detection rates in each time period can be compared with one another to evaluate sequential effects, such as those due to practice, fatigue, or other such variables. In the less complex FC technique, such variables are presumed not to influence performance, a feature that will be discussed later.

Probability of False Alarms. In signal detection theory, where the term has its widest application, *P(FA)* has been defined as the number of false detections divided by the number of opportunities to make a false detection. Though FC and CS procedures both use the term *P(FA)*,

only CS tasks strictly adhere to the SDT definition. In these tasks, the number of opportunities to make a false detection is measured in terms of sites, beams, bins, etc., that could contain targets, and *P(FA)* is a function of the observer's behavior. On the other hand, in the FC procedure, *P(FA)* is simply (m-1)/m. In other words, *P(FA)* is fixed by m, the number of alternative locations or intervals that may contain a target, even though the target could appear at different positions within each location or interval.

Reaction Time. In CS tasks, reaction time is not nearly so "pure" a measure of performance as simple reaction time can be in absolute and difference threshold tasks. For instance, in a CS task, a signal may be detected, but not marked (responded to) until later, because of the direction of attention and behavior to other signals. In waterfalling sonar displays, for example, detection speed can be indexed automatically for each target by counting the average number of lines of data (average line count, or ALC) that were presented before the operator's response to that target. Such variants of reaction time provide excellent indirect measures of VMS performance, and consistently have been found to be correlated with signal strength.[6, 10]

Misuse of the Term Forced-Choice. This section could be subtitled: "When is the FC technique not the FC technique?" Here we are concerned with the fact that, at some point, different measuring techniques often blend into one another. The question is, where does one technique end and another begin?

A task that obviously fits the definition of FC, might entail presentation of two spatial locations, one of which contains a target at a specified SNR. The observer's task is to identify which of the locations contains the target and respond appropriately, e.g., by pressing one of two response keys. Target placement would be randomly distributed between the locations to minimize positional biases, and any advantages of the FC procedure could be realized. Now suppose that we increase the number of locations to four. Clearly, this can still be defined as an FC task. We could again double the number of possible target locations to 8, then 16, etc. Now, suppose we continue to increase the number of locations that might contain a target, say to 1800, as in the case of one study.[11] At what point would one stop calling the task "FC" and call it a "CS" task? Clearly there must be some point at which the term "FC" no longer applies to the situation.

Furthermore, the usual response in a FC task is for the observer to behave in a manner that identifies the location or interval judged to have contained the target. Typically, though not necessarily, this would entail pressing one of a specified number of buttons, corresponding to the number of alternatives. One might even slew a trackball-controlled cursor to the location out of m locations that contains the target. But, does slewing a trackball-controlled cursor to the target itself and "clicking on" that target constitute an appropriate response for an FC task? These questions are neither trivial nor easy to answer, but they are issues that must be considered in determining what to call the task and how to analyze the resulting performance data.

In one recent active sonar test, 20 operators were used to test the effects of certain color manipulations on target detection performance.[12] They were told that there were six contacts (targets) on each of ten display presentations and that their task was to identify them by moving a cursor on them and marking their most likely positions. Actual targets consisted of a random mix of contact locations, contact Doppler, and SNRs that varied within 3 dB of the expected minimum detectable SNR. This procedure was called an FC technique and it was specifically stated that the procedure fixed the false alarm level. However, as nearly as can be determined from the description of the testing method, it involved a typical CS format in which multiple targets were presented and located. As stated previously, the point at which the FC format evolves into the CS format may not always be obvious, but the methods used in this experiment make its designation as FC at least debatable.

In another recent study involving sonar display sensitivity, it was stated that the FC method was used.[13] Actually, each trial of the procedure consisted of inserting a signal (target) into one display beam in each of four display panels in a 2 x 2 panel array on the CRT. The number of beams per panel was not specified, but the observer's task was to point to the beam in each panel which contained the target, using a mouse which controlled a cursor. Feedback was presented after each set of four responses. The task appears, in reality, to have been a straightforward VMS task, and the stimulus presentation format appears to have been a CS format. The logic could be stretched to make it approximate an FC format, by assuming that the number of beams per panel constituted *m*, the number of FC alternatives or spatial locations, but this is not the customary meaning of the term FC technique.

Finally, even in a study that was generally critical of the FC format for passive sonar display studies,[14] one experimental option, referred to as the Fixed Number of Targets (FNT) option, was said to fall "under the general classification of forced-choice experiments" (p. 22), because the operator "has precise knowledge of the number of targets present and the time at which they will appear" (p. 22). In this experiment, the operator was required to mark where he believed that each of three targets were located every observation interval. Again, this appears to have many characteristics of a VMS task in which the number of targets to be located was specified as three. Knowing the number of targets in advance undoubtedly influences the perceptions and reactions of observers. Nevertheless, this knowledge merely defines a CS task with a specified number of targets, not a FC task, except perhaps in the generic, dictionary definition of the term.

Proposed Operational Definitions. Throughout this section, we have noted differences between FC and CS formats that are often masked by similar terminology. In some cases this overlap in terminology reflects a real ambiguity about where the boundaries between the formats should be. It seems that many of the problems outlined in the present report could be ameliorated if proper operational definitions, replete with suggested boundary conditions were set up. Then researchers could more easily determine when it was appropriate to use one format and when it

was not, or at least when the alternate technique would be more appropriate. With this aim in mind, we offer the following operational definitions.

Clearly, characteristics of the stimulus presentation form the primary basis for distinguishing between the formats. Each FC trial must consist of at least two spatial locations or temporal intervals, in one of which there is signal-plus-noise, while in all others there is noise alone. If more than one location or interval contains a signal to be detected, the format is CS. On the other hand, even if only one location or interval contains a signal, the task is not unambiguously an FC task. The grey area revolves around the total number of intervals or locations used. In this regard, there is no basis for choosing a precise numerical boundary between FC and CS formats, but as the number of locations increases, the observer must engage in the type of scanning behavior common to VMS tasks and the CS format, suggesting that such tasks should no longer be labeled and analyzed as "FC." (Note that if there is only one location or interval which either may or may not contain a signal, the format is neither FC nor CS, but is instead a Yes/No format.)

The second basis for drawing distinctions is the response required of the observer. A defining characteristic of the FC format is that there is one response per trial, whereas, using the CS presentation format, there may be many responses in each trial. The response in an FC situation may be verbal (e.g., "upper left" or "interval three") or manual (e.g., touching the target location on a touch sensitive screen or pressing a button corresponding to the spatial location or temporal interval containing the target). On the other hand, the task of observers in a CS procedure is to locate and respond to the specific targets themselves, rather than merely to identify the general location that may contain the target.

Using such operational distinctions to characterize FC and CS tasks is a first step in sorting out some of the problems in the literature. The problems, however, are not merely semantic. The two tasks are different perceptually, behaviorally, and methodologically, although, as we shall see, these differences are not always acknowledged by researchers.

II. PERCEPTUAL, BEHAVIORAL, AND METHODOLOGICAL DIFFERENCES

Perceptual and Behavioral Differences

The test situation using the FC format for stimulus presentation is quite different, *perceptually, psychophysically,* and *behaviorally,* from measurement of accuracy and speed of target detection in CS tasks (see, for example, Thomson[14]). Searching for and detecting the presence of one or more targets in noise, and responding in specific ways to those targets, is very different from picking the spatial or temporal alternative out of m possible alternatives, somewhere in which there may be a target. The two detection tasks are inherently different and call for different perceptual and cognitive processes.

For example, in a direct comparison between visual search tasks using the 2AFC format and stimulus presentation similar to what we defined above as a CS format, substantial differences in performance were reported by Holmgren.[15] These differences were complex and their specifics need not concern us here. What is important is that they could not be explained by simple processes such as the use of a more time-consuming or more exhaustive search with the FC procedure. Holmgren argued, on the other hand, that his findings reflected fundamental differences in the processes underlying visual search in the two tasks, rather than different stages of the same basic process.

While the processes underlying each task may be interesting in their own right, CS procedures are more likely to tap those processes which actually underlie operational situations, such as sonar target detection. It should be emphasized that favoring CS over FC procedures is not merely an issue of "face validity." That is, the concern is not simply that the test situation "look like" the reference situation that it purports to test. All experiments are artificial and many valid experiments do not possess face validity. The point is that in the compromise between artificiality and control, every attempt should be made to introduce as many characteristics of the reference situation as possible into test situations, while still retaining substantive control.[16]

In another CS experiment where the test situation was constructed to have as many features as possible, in common with the reference situation (a waterfalling, passive sonar display), Daggett and DaRos[6] identified differences in performance strategies that would not have been revealed with an FC format. The two of their 14 subjects who were experienced sonar operators a) detected targets significantly faster, b) had a slightly, but not significantly higher *P(D)*, and c) had a significantly higher *P(FA)* than inexperienced observers. This combination of results taken together with the results of interviews with the observers, indicated that different strategies, related to implicit payoff functions, were used by the two groups.

Inexperienced observers treated the experiment as a test in which the task was to perform as errorlessly as possible. They acquired targets more slowly than experienced operators and missed some targets that experienced operators detected, but they had fewer false alarms. Experienced operators, on the other hand, have been trained to acquire targets rapidly in order to monitor them. Ordinarily, the operators would "report" targets only if they were validated after continued monitoring or if they possessed some significance, e.g., were a potential threat. From a functional point of view, it is less important that sonar operators have a low false alarm rate than that they achieve a high and fast hit rate. Thus, the differences in performance between experienced and inexperienced observers are most likely to have resulted from the different strategies used by the two groups. The important point, however, is that this information could not have been ascertained if the FC format had been used, because, in that format the *P(FA)* is constrained by

the number of alternatives rather than by the characteristics of the task or the behavior of the observers, i.e., because the test situation bears little resemblance to the reference situation.

Methodological and Procedural Considerations

Recall that there are four basic outcome measures in target detection tasks, namely, hits, misses, false alarms, and correct rejections. In FC tasks, these measures are so thoroughly intercorrelated that any one of them will suffice to indicate the observer's response. A hit is by definition not a miss and not a false alarm, and implies correct rejection of the noise-only intervals or locations, etc. On the other hand, the four outcomes lead to two independent measures in CS tasks; the number of hits and the number of misses total to the number of targets, while the number of correct rejections and the number of false alarms total to the number of potential target locations. In other words, hits and misses are perfectly correlated with each other, but are independent of false alarms and correct rejections which are perfectly correlated with each other.

The two measures obtained from CS tasks allow perceptual sensitivity to be distinguished from observer's decision criterion. For example, a high hit rate accompanied by a low false alarm rate suggests high sensitivity, while a high hit rate accompanied by a high false alarm rate simply suggests a bias toward responding. These measures are not independent in an FC task because observers are not given the opportunity to exercise a bias toward or against responding, i.e., they must make a choice on each trial. Theoretically, that choice should reflect only their perception of the signal. Advocates of the FC format cite the elimination of the observer's response biases as one of the format's major advantages.[3, 17, 18, 19]

While it is clear that the FC format leaves no room for biases for or against responding, it is not clear that other forms of response bias have also been eliminated. One such bias is a tendency to choose a particular location or interval independent of the target's location. That is, with only chance operating, an observer should pick each available alternative an equal number of times (symmetrical response criterion). Unfortunately, the assumption of response symmetry may not always be valid.[20] Holloway[18] pointed out that "the assumption of lack of appreciable response bias will be more or less valid depending on the S's previous familiarity and expectations with regard to the response categories offered" (p. 175), and then went on to discuss three response bias correction procedures for FC tasks. The very existence of such procedures calls into question the frequent assumption that response biases have been eliminated from FC tasks.

Additionally, with relatively few trials, substantive departures from response symmetry could easily occur due to chance, in spite of the application of random procedures. For example, in the 2AFC case, Irwin[20] found that some observers will pick one alternative over the other substantially more than 50% of the time by chance, even though there is no actual difference between the alternatives. As usual when dealing with probabilities, only in the long run, with a relatively large

number of trials, would chance performance be expected to converge reliably around the 50% point. In most well-controlled laboratory experiments in signal detection, hundreds or even thousands of trials are routinely used, so that there should be no problem. In system evaluation, however, there may only be a few trials per stimulus condition and/or only a few stimulus conditions tested. These limitations increase the likelihood that asymmetrical responding might affect performance evaluation.

Besides assuming response symmetry, the FC procedure also assumes that observers will not be differentially sensitive to the different locations or intervals. In other words, if signals and signal intensity are equally distributed across locations, subjects should not only choose each location with roughly the same frequency, they should also have a comparable rate of correct detections in each location. However, Johnson, Watson, and Kelly,[21] using a variety of 3-interval FC auditory detection and discrimination tasks, frequently found differences in performance among the intervals. Proportion of correct responses tended to be highest in the last interval and lowest in the first. These investigators argued that the differences might reflect actual differences in sensitivity to the signal when it is presented in the various intervals.

Observer Strategies Specific to FC Procedures

Behavioral Evidence. There is strong evidence that observers behave differently in the context of different presentation formats. In one elegant study, Treisman and Leshowitz[22] distinguished among different modes of presentation of the alternatives in either a spatial or temporal 2AFC task. For example, both noise-alone and signal-plus-noise alternatives may be either a) superimposed on a continuously presented background (continuous FC), b) presented against a dark background field (pulsed FC), or c) presented against a background which is incremented slightly on each trial, over and above whatever increment results from either the noise-alone or signal-plus-noise increment (pedestal FC). The finer details of these presentation modes need not concern us, since it is not these details, but rather the resultant behavior that is of concern.

Basically, Treisman and Leshowitz found evidence that subjects use different strategies under different presentation modes. For example, they distinguished between what they termed the "differencing" strategy and the "double detection" strategy. With the differencing strategy, the observer mentally records two inputs, one from each alternative, without specifically making a decision regarding either input. The difference between the inputs is then used to determine the location or interval that contains the target. This is thought to be the "usual" strategy employed by observers in the majority of FC trials.

With the double detection strategy, on the other hand, the observer makes a covert determination as to whether the target is or is not present at each of the two locations or intervals. If the covert judgments are in agreement, i.e., a covert positive judgment for one alternative and a

covert negative judgment for the other, then the overt response, specifying the target location or interval, would be made. If covert detection occurred at both locations/intervals, or no detection at either location/interval, then a random response is assumed, provided of course, that there is no positional or interval bias. Such a strategy might have advantages in temporal FC tasks, especially when the time interval between presentation of alternatives is relatively long.

Of course, either of these strategies could be used on any trial of any FC task, regardless of the presentation mode. That is, Treisman and Leshowitz[22] utilized presentation modes that fostered the use of one strategy over another in order to investigate the nature of the strategies. In fact, however, either strategy might be used in any FC situation, depending on other aspects of the task, such as signal probability, signal intensity, or confidence in the decision by the observer. Obviously, this situation could cause major problems for the use of FC formats in system development research. It implies that the cognitive processes that go into signal detection and identification in such tasks are, at least in part, a function of observer strategies, which are themselves a) indeterminate at any given time or for any given trial, b) changeable over the course of the test, and, most importantly, c) not a part of the "real world" search tasks that the system tests purport to evaluate. Note that these strategies are a function of the FC format and would be meaningless in CS tasks.

Electrophysiological Evidence. At first glance, it may seem that electrophysiological research would be far afield from the present considerations. Surprisingly, however, a number of studies have investigated electrical activity of the brain in the context of psychophysics and signal detection. In one attempt to elucidate the nature of the observer's strategy in FC tasks, Sutton, Ruchkin, Munson, Keitzman, and Hammer[23] monitored event-related potentials while subjects participated in a 2-interval FC auditory detection task.

Event-related potentials (ERPs) are specific modes of electrical activity of the brain, recorded on an electroencephalograph and analyzed by computer, that are correlated with the cognitive tasks in which the subjects are currently engaged. Event-related potentials, sometimes referred to as "cognitive evoked potentials," are distinguished from the more common evoked potentials (EPs) in that ERPs reflect neural correlates of cognitive activity, rather than neural activity evoked more directly by external stimuli. One of the most frequently observed ERPs is the P300 (also called P3), which is a distinctive positive wave, occurring at a latency of about 300 ms after presentation of a decision stimulus.

Part of the rationale for the research of Sutton et al. was that the different strategies identified by Treisman and Leshowitz should yield different patterns of ERPs, and this would suggest different sorts of cognitive activity accompanying the strategies. Indeed this is exactly what they found, using an auditory detection task with two different signal intensities which yielded mean accuracies of 82% and 98%. A critical element of the study was the report by the subjects of their

degree of confidence in their judgments.

If subjects were highly confident, the P300 tended to be large for the observation interval which actually contained the signal. This type of brain electrical activity indicates that a double detection (called "serial independent" by Sutton et al.[23]) strategy was being used. Furthermore, if subjects were highly confident and the signal appeared in the second interval, the latency of the P300 tended to be shorter than it was when the signal appeared in the first interval. This was interpreted as further evidence for the serial independent strategy in that it suggests that the absence of the signal was indeed noted in the first interval. On the other hand, at low confidence, the P300 was either low amplitude or absent for both intervals, suggesting a deferred decision and the use of a differencing strategy.

Clearly, discretion is required in extending the results of experiments using interval FC tasks to those involving spatial FC tasks. In the research of Sutton et al.[23], temporal intervals had to be used in order to permit electrophysiological measurements to be made. Careful arrangement of alternative spatial locations and fixation points, combined with sensitive monitoring apparatus, perhaps including magnetoencephalographs or other brain imaging devices, might permit the identification of similar processes in spatial FC tasks. Additionally, although a number of studies, exemplified by Smid et al.[24], have investigated electrophysiological aspects of visual search tasks, the tasks and procedures employed do not permit generalization to the target detection situations that have been central to this report. Until the gaps in the research are filled, however, it would seem reasonable to exercise caution in extending the results of FC performance in which different strategies may exist as a function of the testing format itself, to CS-type reference tasks (e.g., sonar) in which these strategy variations have no counterpart.

But there are other implications of this research. If, indeed, the findings on different strategies are generally applicable to FC tasks, then there may be a subtle and potentially misleading interaction at work. The argument goes as follows. If observers use one detection strategy for low signal intensity/low confidence trials and a different strategy for higher signal intensity/higher confidence trials, then it is possible that a new sonar technique or device that produces a modest enhancement or decrement in target detection performance might appear to have produced a more substantive improvement or decline, because of a change in the detection strategy. Thus the performance change could be a function not only of the new technique or device but also of the difference in strategy produced by the stimulus presentation format. Since these different strategies have no counterpart in CS tasks, it seems reasonable to conclude that performance differences in such tasks would more unambiguously reflect changes attributable to the new technique or device. Although this argument involves some amount of inference, researchers should, nevertheless, be aware of the potential problem when choosing a method for system performance testing.

Other Variables Affecting the Decision Criterion

As discussed earlier, one of the advantages touted for the FC format is that it prevents operator variables from interfering with measures of target detection sensitivity. We already raised questions about the validity of this assumption. Even in cases where the assumption is valid, however, it should not necessarily be viewed as an advantage. The primary purpose of system testing is evaluation of system performance. Since the operator is part of the system and contributes to overall system performance, procedures that permit the evaluation of the complete system, including the operator, would seem to be preferred to those in which the operator is treated essentially as a null instrument. These comments should not be misinterpreted as an argument for a confounded measure over a pure measure; rather it is an argument for a measure which allows both sensitivity information and operator performance variables to be extracted from the data. We now turn to some important operator performance information that would be overlooked by the FC format.

Changes in P(D) and P(FA) over Time. One of the more consistent findings in tasks requiring sustained attention, has been the so-called "vigilance decrement," which is the deterioration over time of the ability to remain vigilant.[25] This decrement is indicated primarily by a lowering of the rate of correct detections (and therefore an increase in the rate of misses), and is often accompanied by a change (usually a decrease) in the false alarm rate.[7] While the extent of the decrement in complex tasks has been debated (see Parasuraman[7] for an extensive analysis of the debate), the preponderance of evidence points to the existence and importance of the vigilance decrement in simple as well as complex tasks. What concerns us here is that using the FC format for stimulus presentation in this type of target detection situation, with a fixed *P(FA)* and no distinction between false alarms and misses, a decrement would be difficult if not impossible to assess in a meaningful way.

Signal Probability. Using the standard FC format, signal probability cannot be varied; regardless of the number of intervals or locations, signal probability is 1.00 on each FC trial. However, SDT suggests that differences in signal probability will affect target detection performance. Consistent with the predictions of SDT, the decision criterion in CS tasks varies inversely with the signal probability.[7, 26] More probable signals tend to be responded to with a lower decision criterion which, all else being equal, produces an increase in both the *P(D)* and *P(FA)*. In most reference tasks that are of concern to this report, e.g., sonar system development, signal probability varies over a wide range, so its effects should not be ignored.

Instructions and Practice. Instructions to observers, e.g., to adopt criteria that are more or less "risky" or more or less "cautious," can also change decision criteria, and, thereby, affect target detection performance. These manipulations are even capable of producing differences in performance that rival in magnitude those produced by differences in signal probability.[7]

Additionally, practice, especially if accompanied by adequate feedback, can produce substantive changes of performance. The general finding is that there tends to be a decline in false alarm rate over the first few trials in CS tasks, followed by relative stability thereafter.[7, 27] Moreover, as we have seen, experienced observers often differ in performance from less experienced or inexperienced observers, presumably because of the effects of both practice and instructions.[6] Again, the FC procedure, with its fixed *P(FA)*, rules out the precise assessment of the effects of such variables.

Payoff Function. The costs and values associated with the detection outcomes (payoff function or payoff matrix) also affect signal detection performance. Surprisingly, studies in which the payoff function has been varied in CS tasks have found relatively small and variable effects.[7] This finding may be due to the difficulty of simulating in a laboratory the kinds of payoff functions that exist in the real world, or to the difficulty of separating the explicit payoff function from implicit payoff functions that operators bring to the situation. As described above, experienced and inexperienced operators were found to have different patterns of performance on a sonar target detection task.[6] It was proposed that these findings were due to the different implicit payoff matrices of experienced and inexperienced operators as a result of direct or perceived differences in training and instructions. Much more research is called for in this area. But again, since errors in a FC task do not differ intrinsically in cost, the FC technique would be the least useful of the available techniques for this research.

III. EFFICIENCY AND COST-EFFECTIVENESS OF PROCEDURES

The thrust of this report has been to argue that for assessing target detection performance in complex situations, such as radar, sonar, air-traffic control, and the like, the CS format is more appropriate and informative than the FC format. We have questioned the assumption that the two techniques are essentially interchangeable, and we have demonstrated that characteristics generally portrayed as advantages of the FC format might actually be limitations under some circumstances. A final, alleged advantage often cited for the FC technique is its efficiency and cost-effectiveness. For instance, Buratti and Trezona[28], implicitly accepting interchangeability, argued that, since a number of procedures are available for the actual determination of performance outcomes, the choice of which procedure to employ can often be based on indirect considerations, such as testing time, cost-effectiveness, etc. Benefits claimed for the FC format were that it minimizes test monitoring requirements, permits real time monitoring of test results, and utilizes less test time. The ability to develop automated software routines for implementation of the method, was also seen as a particularly attractive aspect of the FC format. While there can be no doubt that test monitoring requirements are simple for the FC format, as we have indicated throughout this report, this simplicity is gained at the expense of amount of information gathered. Unfortunately there is no compelling evidence for the rest of Buratti and Trezona's claims regarding the FC format.

If it were true that the characteristics of efficiency and programmability applied uniquely to the FC format, then there might be reasonable justification for its use to quantify performance in VMS tasks. In fact, however, the CS format for stimulus presentation, including target designation and presentation, data gathering, and analysis, can be and has been computerized, as exemplified by some of our recent research.[6, 9, 10] Furthermore, even the supposed advantage of greater simplicity may be overstated. As Kaernbach[29] has pointed out, tasks that require presentation of multiple intervals or locations prior to a response cannot be optimally efficient. A 2AFC task, for instance, requires two presentations to obtain one bit of information. Even the Yes/No task obtains one bit of information per presentation, while a single trial of the CS format may elicit numerous bits of information. Of course, more performance information could be obtained in FC tasks by giving the subject less information about the task, e.g., that there might not be exactly one signal per two presentations. But the penalty for this modification is a reduction of the much touted analytical simplicity and efficiency of the procedure.

DISCUSSION AND CONCLUSIONS

The FC technique has been gaining in popularity among some researchers for the study of performance in target detection and acquisition tasks. The rationale for the use of the procedure is that it a) eliminates differences in response bias on the part of subjects, b) leads to simplification of data handling, and c) is less costly than other procedures. In the present report, the FC format was compared to the CS format for the investigation of target detection performance in vigilance, monitoring, and search tasks. One concern noted in the report was that the boundaries between FC and CS tasks were not as clear as they could have been, so an attempt was made to fine-tune those boundaries. A summary of the proposed distinctions follows.

1. Each FC trial consists of at least two spatial locations or temporal intervals, exactly one of which contains a signal. If more than one interval or location contains a signal, the task is a CS task.

2. In a detection task, as the number of intervals or locations increases, the task progressively elicits more of the scanning behavior typical of a CS rather than an FC format and should be labeled accordingly.

3. Each FC trial demands one response from the observer, while each CS trial may permit many responses.

4. In the FC format, responses entail indicating the interval or location which may contain a target, whereas in the CS format, responses entail locating the target itself.

Having drawn these distinctions, we went on to argue that there are a number of often unacknowledged, but substantive, differences between the formats, and, furthermore, that the CS format is more appropriate to assessing performance in VMS tasks than is the FC format.

1. There is substantial evidence that different perceptual and cognitive processes underlie performance in the two tasks.

2. Because CS tasks are so much more similar to the reference tasks which concern us, e.g., sonar operation, the processes they tap are more likely to be relevant than are the processes tapped by FC formats.

3. The FC format claims to obtain a pure measure of target detection sensitivity, uncontaminated by response biases.
   a. There is evidence that this claim may be excessive and that some response biases may still influence performance.
   b. Because real world performance does not reflect sensitivity alone, operator variables and biases should be assessed rather than eliminated. The CS format is structured to assess sensitivity along with other variables which may influence performance, such as response biases, signal probability, instructions, payoff functions, etc.

4. There is suggestive evidence that the FC format may elicit cognitive strategies that have no counterpart in other formats or in real world applications, which raises questions about the generalizability of some FC findings.

5. Finally, data-analysis, programming, and cost-efficiencies that are often claimed for FC methods, seem overstated in most instances.

Although our recommendation - that the CS format be used for the controlled laboratory experimentation involved in sonar system development - can be made with confidence, further consideration may have to be made in the case of total system performance evaluation, since elaborate measuring hardware and software often cannot be built into the completed system. That is, there may be situations in which display capabilities limit stimulus presentation in a way that is more amenable to FC than CS format. For example, when the display interface contains a number of discrete sections (different grams) that could be used as locations for target presentation, and/or when the number of targets that may be presented is limited, the FC format may be the best alternative. Nevertheless, since today's complex systems are architected for computational flexibility, incorporation of a non-intrusive, removable, CS-format test package (typically software only) into the product under test does not seem unreasonable.

In closing, two major conclusions, one theoretical and one practical, may be drawn from the evidence presented in this report. On the theoretical side, it appears that some methods and formats for assessing detection, recognition and similar kinds of human performance may be so different from others that attempts to equate the outcomes or compensate for differences may be useless or inappropriate. Specifically, it is suggested that FC formats should be considered as a different class of performance measure from CS formats for the investigation of target detection performance in VMS tasks. The fact that these procedures deal with the same concepts, namely "threshold," "detection," "recognition" and the like, and the fact that they often yield similar performance outcomes, may be misleading and may obscure the fact that they are actually assessing different cognitive strategies and, hence, different perceptual and neural processes.

On the practical side, the overall conclusion is that multi-target, multi-level, CS procedures, built around the classic method of constant stimuli, should be the methods of choice for performance evaluation in VMS tasks such as radar, sonar and the like. If FC procedures are used out of necessity or for convenience, it should be understood that the surface similarities in performance may mask underlying differences in strategies and processes.

REFERENCES

1. R. S. Woodworth and H. Schlosberg, *Experimental Psychology*. Holt, Rinehart, and Winston, New York, 1954.

2. J. C. Falmagne, "Psychophysical Measurement and Theory," in K. R. Boff, L. Kaufman, and J. P. Thomas, eds., *Handbook of Perception and Human Performance, Vol. I, Sensory Processes and Perception*, John Wiley & Sons, Inc., New York, 1986.

3. D. M. Green and J. A. Swets, *Signal Detection Theory and Psychophysics*, Krieger Publishing Co., Huntington, New York, 1966.

4. K. R. Boff and J. E. Lincoln, eds., *Engineering Data Compendium: Human Perception and Performance*, AAMRL, Wright-Patterson AFB, Ohio, 1988.

5. K. T. Spoehr and S. W. Lehmkuhle, *Visual Information Processing*, W. H. Freeman, San Francisco, CA, 1982.

6. T. A. Daggett, and D. A. DaRos, *"Improved Noise Recognition Differential Performance in Sonar Displays Using Image Processing Techniques,"* NUSC Technical Report 8345, Naval Underwater Systems Center, New London, CT, October 1988.

7. R. Parasuraman, "Vigilance, Monitoring, and Search," in K. R. Boff, L. Kaufman, and J. P. Thomas, eds., *Handbook of Perception and Human Performance, Vol. II, Cognitive Processes and Performance*, John Wiley & Sons, Inc., New York, 1986.

8. G. Volkov, *"Effects of Shadow Mask Color vs. Monochrome CRTs on Detection of Sonar Signals,"* Paper presented at the Society for Information Display International Symposium, Orlando, FL, April 1985.

9. W. R. Salafia, D. A. DaRos, and P. R. Boivin, *"Color Coding of Amplitude Data as a Means of Improving Target Detection in Passive Sonar Displays,"* NUSC Technical Report 8277, Naval Underwater Systems Center, New London, CT, June 1988.

10. W. R. Salafia, *"Target Detection Performance as a Function of Color in Sonar Gram Displays,"* NUSC Technical Report 8647, Naval Underwater Systems Center, New London, CT, June 1990.

11. A. Burgess and H. Ghandeharian, "Visual Signal Detection. II. Signal Location Identification," *Journal of the Optical Society of America*, vol. 1, no. 8, 1984, pp. 906-910.

12. R. J. Buratti, J. Rio, and M. N. Witlin, "Use of Multicolor Displays for Sonar Detection," *Proceedings of Underwater Acoustic Data Processing Symposium*, NATO Advanced Study Institute, 18-29 July 1988.

13. R. L. Hershman, J. L. Kaiwi, and E. Pilmore, *"Tests of Sonar Detection Display Sensitivity,"* NPRDC Technical Note 87-15, Naval Personnel Research and Development Center, San Diego, CA, January 1987.

14. W. G. Thomson,*"Passive Sonar Display Studies,"* NUC TP 436, Naval Undersea Center, San Diego, CA, June 1975.

15. J. E. Holmgren, "Visual Search in a Forced-Choice Paradigm," *Perception and Psychophysics*, vol. 16, no. 2, 1974, pp. 253-258.

16. D. Meister, *Conceptual Aspects of Human Factors*, Johns Hopkins University Press, Baltimore, MD, 1989.

17. C. M. Holloway, "The Effects of Response Bias in Two-Alternative Forced-Choice Discrimination Matrices," *Behavior Research, Methods, and Instrumentation*, vol. 1, no. 5, 1969, pp. 175-178.

18. C. Auerbach, "Correcting Two-Alternative Forced-Choice Data for Response Bias," *Perceptual and Motor Skills*, vol. 32, 1971, pp. 533-534.

19. P. L. Emerson, "Observations on Maximum-Likelihood and Bayesian Methods of Forced-Choice Sequential Threshold Estimation," *Perception and Psychophysics*, vol, 39, no. 2, 1986, pp. 151-153.

20. R. J. Irwin, "The Psychophysics and Norms of the Seashore Measures of Musical Talents," *Educational and Psychological Measurement*, vol. 44, 1984.W. G.

21. D. M. Johnson, C. S. Watson, and W. J. Kelly, "Performance Differences among the Intervals on Forced-Choice Tasks," *Perception and Psychophysics*, vol. 35, no. 6, 1984, pp. 553-557.

22. M. Treisman and B. Leshowitz, "The Effects of Duration, Area, and Backround Intensity on the Visual Intensity Difference Threshold Given by the Forced-Choice Procedure: Derivations from a Statistical Decision Model for Sensory Discrimination," *Perception and Psychophysics*, vol. 6, no. 5, 1969, pp. 281-296.

23. S. Sutton, D. S Ruchkin, R. Munson, M. L. Keitzman, and M. Hammer, "Event-Related Potentials in a Two-Choice Detection Task," *Perception and Psychophysics*, vol. 32, no. 4, 1982, pp. 360-374.

24. H. G. O. M. Smid, W. Lamain, M. M. Hogeboom, G. Mulder, and L. J. M. Mulder, "Psychophysiological Evidence for Continuous Information Transmission between Visual Search and Response Processes," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 17, 1991, pp. 696-714.

25 N. H. Mackworth, "The Breakdown of Vigilance during Prolonged Visual Search," *Quarterly Journal of Experimental Psychology*, vol. 1, 1948, pp. 6-21.

26. A. D. Baddeley and W. P. Colquohoun, "Signal Probability and Vigilance: A Reappraisal of the 'Signal Rate' Effect," *British Journal of Psychology*, vol. 60, 1969, pp. 165-178.

27. J. R. Binford and M. Leob, "Changes Within and Over Repeated Sessions in Criterion and Effective Sensitivity in an Auditory Vigilance Task," *Journal of Experimental Psychology*, vol. 72, 1966, pp. 339-345.

28. R. J. Buratti and G. E. Trezona, "Utilization of a Forced-Choice Technique to Verify Statistical Performance Requirements for Detection Systems," *Journal of the Acoustical Society of America*, vol. 57, no. 4, 1975.

29. C. Kaernbach, "A Single-Interval Adjustment-Matrix (SIAM) Procedure for Unbiased Adaptive Testing," *Journal of the Acoustical Society of America*, vol. 88, no. 6, 1990.

INITIAL DISTRIBUTION LIST

External Distribution

Addressee		No. of Copies
Space and Naval Warfare Systems Command (SPAWAR)		
PMW 181-T	Dr. L. Parish	1
Department of the Navy Program Executive Officer		
PMO-425	Capt. G. Kent	1
	N. Cook	1
	W. Johnson	1
	J. Smerchansky	1
	M. Basilica	1
PMO-SCWS-X2	S. Lose	1
OPNAV 224C	CDR. R. Brandhuber	1
Office of Naval Technology (ONT)		
ONT 20	E. Wald	1
Naval Sea Systems Command (NAVSEA)		
SEA-06U24	E. Hale	1
SEA-06UR	CDR T. Mason	1
	G. Kamalakis	1
Naval Research Laboratory (NRL)		
Code 5530	Dr. H. Gigley	1
Naval Command, Control and Ocean Surveillance Center		
NRaD Code 402	R. Wasilausky	1
Defense Technical Information Center (DTIC)		6
General Electric Co.		
Syracuse, N.Y.	E. Brennan	1
Moorstown, Pa.	Dr. J. Hassab	1
IBM Federal Systems Division		
Manassas, Va.	W. Gross	1
	D. Bresson	1
	M. Witlin	1
Raytheon Submarine Signal Division		
Portsmouth, R.I.	M. Cohen	1
Fairfield University		
College of Arts and Sciences	D. Danahar	1