AD-A053 756 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 6/16
DETECTION OF AN OCTAVE BAND OF NOISE AS A FUNCTION OF STIMULUS --ETC(U)
NOV 77 P T CORNELL N00017-73-C-1418
UNCLASSIFIED ARL/PSU/TM-77-311 NL

1 OF 2
AD
A053756

(12)

AD A053756

DETECTION OF AN OCTAVE BAND OF NOISE AS A FUNCTION OF STIMULUS PRESENTATION

Paul T. Cornell

AD No.
DDC FILE COPY

Technical Memorandum
File No. TM 77-311
November 4, 1977
Contract No. N00017-73-C-1418

Copy No. 5

The Pennsylvania State University
Institute for Science and Engineering
APPLIED RESEARCH LABORATORY
Post Office Box 30
State College, PA 16801

DDC
RECEIVED
MAY 10 1978
RECEIVED
E

APPROVED FOR PUBLIC RELEASE
DISTRIBUTION UNLIMITED

NAVY DEPARTMENT

NAVAL SEA SYSTEMS COMMAND

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

14

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. ~~REPORT NUMBER~~ ARL/PSU/TM-77-311
2. GOVT ACCESSION NO.
3. RECIPIENT'S CATALOG NUMBER

4. TITLE (and Subtitle) 4 DETECTION OF AN OCTAVE BAND OF NOISE AS A FUNCTION OF STIMULUS PRESENTATION,

9

5. TYPE OF REPORT & PERIOD COVERED ~~MS Thesis, Psychology~~ ~~March 1978~~ Master's thesis,

6. PERFORMING ORG. REPORT NUMBER TM 77-311

7. AUTHOR(s) 10 Paul T./Cornell

8. CONTRACT OR GRANT NUMBER(s) 15 N00017-73-C-1418

9. PERFORMING ORGANIZATION NAME AND ADDRESS

Applied Research Laboratory
P. O. Box 30
State College, PA 16801

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS

11. CONTROLLING OFFICE NAME AND ADDRESS

Naval Sea Systems Command
Department of the Navy
Washington, D. C. 20362

12. REPORT DATE 11 ~~November 4, 1977~~ 4 Nov 1977

13. NUMBER OF PAGES 12 122 p ~~pages & figures~~

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)

15. SECURITY CLASS. (of this report) Unclassified, Unlimited

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release, distribution unlimited, per NSSC (Naval Sea Systems Command), 1/8/78.

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

18. SUPPLEMENTARY NOTES

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

hearing
Psychoacoustics
noise
stimulus
auditory detection
threshold

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

Two experimental methods were developed to examine the effects of time uncertainty and other variables on the detection of a 500 Hz centered octave band of noise presented in noise. In one method, denoted the fixed presentation method, the signal was presented at a sixed SNR, following a specified interval of time after the onset of an ambient noise stimulus. The other method, the modified threshold-forced response method, presented the signal with the noise at the start of the trial at a low SNR. During the trial, the SNR increased

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

391007

UNCLASSIFIED 12C

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. ABSTRACT (Continued)

at a constant rate of one-half dB per two seconds. Upon attaining a specified SNR, the stimulus was gated. Six levels of ending SNR were examined with each method. Four subjects responded on a six-point confidence rating scale.

Results indicated that with the variable SNR method, performance was much worse than in the fixed SNR method. Not only were confidences lower, but the probability correct was likewise lower. Results also indicated that subjects could maintain a fairly consistent set of criteria throughout the experiment, as rank ordered correlations of responses to identical tapes were generally high. Consistency was found to increase with SNR.

Comparisons to a 2AFC and Janota's (1977) modified threshold procedure were made. The 2AFC and fixed SNR methods resulted in nearly equal performance, the psychometric functions relating Green's d'_{opt} to P(C) were less than one dB apart. The function of the modified threshold-forced response method was shifted 5.0 to 6.0 dB to the right of those found with either of the above procedures. It was shifted 0.8 dB to the left of the modified threshold method. Predictions of the shifts based on the Stallard and Leslie (1974) hypotheses were good with the modified threshold-forced response method, but not with the fixed presentation method.

Performance differences obtained with these four procedures are discussed in terms of varying amounts of signal and time uncertainty. The two experimental procedures tested are used to examine the merits of Green's (1960) d'_{opt} model of auditory detection of noise bands. The model was found to fit the data fairly well. Implications for future psychoacoustical research and for predictive models are discussed.

ACCESSION for
NTIS White Section ☒
DDC Buff Section ☐
UNANNOUNCED ☐
JUSTIFICATION..............................
....
BY..............................
DISTRIBUTION/AVAILABILITY CODES

Dist.	AVAIL. and/or SPECIAL	
A		

ABSTRACT

Two experimental methods were developed to examine the effects of time uncertainty and other variables on the detection of a 500 Hz centered octave band of noise presented in noise. In one method, denoted the fixed presentation method, the signal was presented at a fixed SNR, following a specified interval of time after the onset of an ambient noise stimulus. The other method, the modified threshold-forced response method, presented the signal with the noise at the start of the trial at a low SNR. During the trial, the SNR increased at a constant rate of one-half dB per two seconds. Upon attaining a specified SNR, the stimulus was gated. Six levels of ending SNR were examined with each method. Four subjects responded on a six-point confidence rating scale.

Results indicated that with the variable SNR method, performance was much worse than in the fixed SNR method. Not only were confidences lower, but the probability correct was likewise lower. Results also indicated that subjects could maintain a fairly consistent set of criteria throughout the experiment, as rank ordered correlations of responses to identical tapes were generally high. Consistency was found to increase with SNR.

Comparisons to a 2AFC and Janota's (1977) modified threshold procedure were made. The 2AFC and fixed SNR methods resulted in nearly equal performance, the psychometric functions relating Green's d'_{opt} to P(C) were less than one dB apart. The function of the modified threshold-forced response method was shifted 5.0 to 6.0 dB to the right of those found with either of the above procedures. It was shifted 0.8 dB to the left of the modified threshold method. Predictions of the shifts based on the Stallard and Leslie (1974) hypotheses were good with the modified threshold-forced response method, but not with the fixed presentation method.

Performance differences obtained with these four procedures are discussed in terms of varying amounts of signal and time uncertainty. The two experimental procedures tested are used to examine the merits of Green's (1960) d'_{opt} model of auditory detection of noise bands. The model was found to fit the data fairly well. Implications for future psychoacoustical research and for predictive models are discussed.

TABLE OF CONTENTS

PART A. REVIEW OF PSYCHOACOUSTICAL LITERATURE

TABLE OF CONTENTS (Continued)

Page

PART B. THE EXPERIMENT

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

The author acknowledges the significant contribution of Dr. C. P. Janota. The knowledge, cooperation and efforts extended by this individual were paramount to the completion of this thesis. All the hardware systems used in the recording and taping of the experimental sessions were designed by Dr. Janota. Recognition of the sophistication in design of these systems is duly warranted.

Furthermore, the author expresses his appreciation for the suggestions rendered by Drs. D. Trumbo and L. Kerr, and fellow graduate student D. Martin. Their inputs were instrumental in the refinement of the experimental design and in the analysis of the data.

The assistance of the Applied Research Laboratory of The Pennsylvania State University under contract with the Naval Sea Systems Command is also recognized. The research was conducted as part of a project funded by the Naval Air Systems Command (AIR 370).

CHAPTER I

INTRODUCTION

During the past several years, human performance in complex auditory environments has been studied at the Applied Research Laboratory at The Pennsylvania State University. Due to the nature of the problems being examined, traditional psychoacoustical procedures were untenable. As a result, a new means of assessing an individual's ability in an aural detection task was developed. This technique, the modified threshold technique, was unlike others in that the signal-to-noise ratio (SNR) of an auditorily presented signal increased during the stimulus presentation. In most psychoacoustical research, the signal is presented at a fixed signal-to-noise ratio.

A procedure similar to the modified threshold technique could not be found in the literature and, hence, comparisons could only be made to results obtained with different methods. This analysis indicated that performance was considerably worse when signals were presented via the modified threshold technique. Several questions concerning the determinants of auditory performance arose as a result of these findings.

This thesis delineates the existence and magnitude of some possible factors. The factor primarily addressed is the effect that certain parameters in the method of stimulus presentation have on performance. To accomplish this task, two experimental procedures are employed. One, the fixed presentation method, is a single interval, yes-no paradigm, in which the SNR of the signal is fixed. A derivative of the modified threshold technique, the modified threshold-forced response method, is the other. Uncertainty is less in this second procedure than in the modified threshold task; however, it is still greater than that in the fixed presentation task. One important distinction between these procedures and that of the modified threshold is that the subject is in a forced-choice situation rather than a free-choice situation.

In all psychological experimentation, many decisions have to be made regarding which variables should be manipulated and which should be held constant. These decisions affect the outcome of the experiment and the applicability of the results. Part A of this paper presents a review of variables and factors examined in other studies, discussion being primarily in terms of the variable's or factor's relevance to the present design, and the reasoning for manipulating or holding it constant.

In Part B, the actual design of the experiment and the procedures used are concisely set forth. The independent variables are discussed as well as the characteristics of the auditory stimulus and the method of response. Subsequent to this, the purpose of the experiment is explained as well as the expected results. The remaining three chapters consist of the Methods, Results, and Discussion sections of the experiment.

The results of this experiment should aid in the future design of man-machine environments, contribute to the prediction of human performance in certain settings, and assist in the understanding of the effects of specific parameters on the detection of signals in noisy backgrounds.

PART A

REVIEW OF PSYCHOACOUSTICAL LITERATURE

CHAPTER II

SIGNAL PARAMETERS

2.1 Stimulus Complexity

In a vast majority of psychoacoustical research, quite simple signals are used. The most frequently employed stimuli are pure tones, usually of 1 KHz frequency. The frequencies used as pure tone stimuli range from about 200 Hz to 2 KHz.

In other studies, quite different signals are used which are considerably more complex. Green (1958) examined subjects' performance in detecting single component signals and compared it with that for two component signals. His results indicated that the detectability of the two component signals was higher than that obtained by either component alone. Green (1960) also did a study in which the signal was a narrow band of noise presented with a continuous noise background. In this paper, Green tested the accuracy of an equation, d'_{opt}, used to predict detection of bands of noise for several signal durations, center frequencies, and bandwidths. He concluded that if a constant specific to the individual was added to the equation, then the predictions were accurate. In a study of consistency in auditory detection, Green (1964) again used a band of noise as a signal. He used three increments of power and determined the percentage of agreement in detection with successive exposures to the same stimuli.

In a study of complex noise signatures, Fidell (1974) examined the detectability of twelve synthetically produced signals. The stimuli were narrow bands of noise, broad bands of noise, or pure tone in character, and were presented with three different varieties of background noise. Signals varied in presence and amount of frequency modulation, width of the bands, and presence and amount of amplitude modulation. He concluded that the detectability of a signal was determined by its most detectable component, which was contrary to Green's (1958) results. Complex marine signatures presented with a background of ocean ambient noise has been examined by Janota (1977). Using the modified threshold technique which is discussed in a later section, it was found that some components of a complex signal had little effect on the detectability of that signal.

Schulman (1971) broke tradition and defined a signal trial as one in which a 1 KHz tone was not present. In this yes-no (YN) paradigm, coincident with a warning light, the tone was either removed or maintained in the continuous background noise. Data were also obtained in which subjects had to detect the addition of the tone to the background. Results indicated that in order to obtain similar detectability indices, the tone level in the removal condition would have to initially be 3.5 dB higher than in the addition situation. These results agree with the discrepancy found by MacMillan (1971) in his increment-decrement procedure. He defined a signal as an increase or decrease in the intensity of a 1 KHz tone. In both YN and two-alternative forced-choice procedures (2AFC), decrements were found to be more difficult to detect.

2.2 Stimulus Intensity

In many experiments, data are gathered using several signal intensities. Markowitz and Swets (1967), for instance, investigated six different signal intensities in a single and double interval task. They collected results using both binary and rating decision procedures. From a receiver operating characteristic (ROC) analysis of the two interval rating data, it was found that as the signal intensity decreased, so did the slope of the function. This indicated that intensity had an effect on the variances of the signal-plus-noise and noise distributions of the subject. With smaller signal intensities, it appears as though the observer's internal noise has a greater effect, resulting in an increase in the perceived variance of the distribution. Slopes obtained with the binary data remained close to unity, regardless of the intensity. Neither response method with the single interval procedure showed any dependence of slope upon intensity.

Using rating scales of three, five, and nine category size, Shipley (1970) found that intensities in the middle range of detectability result in smaller variances in the signal distribution. She also found that presenting a disproportionate number of weak or strong signals caused a shift in the observer's criteria towards those signals.

Traditionally, if several signal intensities are going to be used in one experiment, data are collected such that the level remains the same for each session, or block of trials. Emmerich (1968a), however, presented two different intensities within the same block of trials and then compared the ROC's with those obtained under homogeneous intensities. Using the area under the ROC as his detectability measure, Emmerich found very little difference in the two methods. These results are unexpected in that when the signal intensity varies from trial to trial during a session, the subject's variability should increase. The outcome would be a reduction in performance. If this hypothesis is incorrect, as Emmerich's results suggest, then this has certain ramifications on the design of future auditory experiments. One possible explanation for the results, however, is that the intensities were fairly close. If they were more widely separated, the expected outcome may have occurred.

Occasionally, stimulus intensities are varied only to insure that the obtained results are not contingent upon one particular level. Interactions may exist at some intensities which do not at others. Egan, Schulman, and Greenberg (1959) used three different signal levels in a rating task to check the hypothesis that the departure of the ROC curve from the positive diagonal is a direct function of the intensity of the signal. The hypothesis was supported, with the difference increasing with intensity. Watson, Rilling, and Bourbon (1964) used two signal intensities when comparing the detection results obtained with rating vs. binary procedures. Both intensities showed higher d' measures for the binary method. Three levels of intensity were presented to subjects by Green (1964) when he was determining the degree of consistency of his subjects. No substantial difference was found between any two levels. In relating d' to signal duration, Green, Birdsall, and Tanner (1957) found nearly parallel psychometric functions for four signal levels.

2.3 Stimulus Duration

When duration is not one of the variables to be investigated, the default value is usually 500 msec, but this varies. Most values fall within the range of 100 to 1000 msec, but some are as high as 4000 msec.

Some of the earliest work done on the effects of duration upon signal detection was carried out by Green (1960). In his equation for the optimal detectability of a waveform, T, the stimulus duration plays a prominent role in the value of d'_{opt} (Table 1). Green makes two assumptions with this model. One is that the "device" measures the power in two waveforms and selects the larger. The second is that the "device" knows the exact starting time and bandwidth of the stimulus. In essence, the human observer is viewed as an energy detector, and predictions can be made as to his performance based on this assumption. Green tested the model with a variety of durations from 3 to 5000 msec. Durations of 300 msec and less resulted in performance that varied from the predicted by a constant. The longer durations, however, did not match what was predicted, with the performances being inconsistent. Green and Sewall (1962) proposed that the reason for the difference between the predicted and observed was due to the non-ideal observer not knowing the exact starting time of the signal and its duration. By presenting subjects with signals sufficiently loud enough to specify stimulus onset, this problem was circumvented. To do this, the task was changed to detecting the larger of two waveforms in a 2AFC task; this also served to reduce the memory load on the individual. The obtained psychometric functions were quite similar to those predicted, only shifted to the right by a constant factor. The authors cited this as support for the model.

Green, Birdsall, and Tanner (1957) used durations from 250 to 3000 msec. The probability correct, P(C), was found to increase proportionately with increases in duration. With every increase, the psychometric function relating P(C) to duration was shifted to the of the groundwork has been laid for theories in perception using simple signals; further study of these theories should now be done using complex signals.

The experimental use of complex signals seems more logical in that single component signals and pure tones are often found outside an experimental situation. Hence, to have some applicability to human performance in natural environments, complex signals should be examined. More precise models of auditory processing can be constructed when complex signals are used because a better understanding of the interactions of various stimulus components is acquired.

In this experiment, complex signals, noise bands, will be used. (The signals will be "complex" in the sense that they are more than pure tones. Be it understood that noise bands do not, however, constitute a very high level of complexity.) In an attempt to examine performance under different levels of difficulty, several signal intensities will be employed. This will not only result in the construction of a psychometric function, but will allow for the examination of interactions between intensity and method of presentation as well. In this experiment, the duration of the signal-plus-noise stimulus will be dependent on the method used. In one condition, it will be constant, eight seconds, and in the other, it will vary from 38 to 45 seconds. This will be more adequately described later.

CHAPTER III

TASK PARAMETERS

3.1 Signal Uncertainty

Uncertainty in an experiment comes in two forms: time of onset of the signal, and the type of signal itself. Gundy (1961) investigated signal uncertainty when he studied the effects of feedback on the detection performance of subjects under specified and unspecified signal conditions. In the specified condition, the signal was presented three times before the beginning of a session. No presession exposure to the signal was given the other subjects. Both conditions were divided into feedback and no-feedback groups. One group of subjects was run using a signal-to-noise ratio (SNR) of 15.8 dB and another with a level of 25.1 dB. A value on a four-point rating scale indicated the subject's confidence in his response, and d'_e (see Section 5.2) was used as the measure of detectability. When the signal energy was low, feedback and signal specification gave best results with an average d'_e of 1.6. Without feedback, the average d'_e was 1.35. Both groups under the unspecified condition yielded d'_e's of about 0.6. When the signal energy was higher, the difference in the groups was less pronounced. The average d'_e after the third block of trials was: signal specified - feedback 2.3, no-feedback, 2.4; signal unspecified - feedback 2.0, no-feedback 1.9. Evidently, specifying the signal via a cue facilitated the subject's performance. Although performance in the unspecified conditions was poor initially, learning occurred quickly with d'_e increasing from 0.7 to 1.6 in just three blocks of 50 trials. No learning curve was found with the lower signal energy or specified signal conditions.

Green (1961) found that when a pure tone signal from a set of signals is presented, performance is worse relative to those situations in which only one signal may be presented. In all the sets, a center frequency of 2250 Hz was used, with ranges from 100 to 3500 Hz. Green measured the amount of signal energy needed to attain a P(C) of 0.75. Although higher energies were necessary with increasing ranges, the increase was relatively small. Green concluded that much uncertainty existed at the start, and this additional uncertainty did little to degrade performance further.

Three different intensities, crossed with four levels of pretrial cueing, were examined by Emmerich (1971). Cueing was absent, the same intensity, or three or six dB higher than the intensity of a 500 Hz signal. The results of this 2AFC task indicated that, at low intensities, any cue improves performance, but at the high intensities, it may actually degrade performance. Emmerich found this with a 4 KHz tone as well. His results support those of Gundy for low signal intensities. Robinson and Watson (1972) say that pretrial exposures to the signal not only reduce uncertainty, but they reduce variability in performance as well. If variability could be reduced, then fewer trials would be required to attain an accurate assessment of the detectability of a signal.

Using a 2AFC task, Pastore and Sorkin (1971) studied signal uncertainty by presenting their subjects with either a 500 or 2000 Hz tone. Samples of each signal alone or in noise were presented to subjects before each block of trials. Feedback was given as to which interval contained the signal, but did not specify which signal it was. They found that performance was worst, in terms of P(C), when the signal was different from the one presented on the previous trial. P(C) was greatest when there was a run of one signal. They cited this as support for the single-band model in that, when the signal was switched, the subject would be attending to the wrong band, and thus his chance of detecting that signal would be reduced.

The single-band and multiple-band models of processing were discussed by Swets (1963) in regard to signal uncertainty. He states that both theories predict decrements in performance in uncertain situations. Single-band processing claims that the subject cannot listen to more than one band. Thus, if he was not attending to the band with the signal, then his performance would be less than optimal. In the multiple-band model, many bands are processed simultaneously. This theory accounts for decrements in performance in that more noise is processed under unspecified conditions. This leads to a reduction in the efficiency of the observer.

3.2 Time Uncertainty

The effect of time uncertainty is another concern in analyzing results. The time of signal onset has little influence when the signal is loud enough to attract the attention of the observer. Of course, it has to be of sufficient duration also. At low signal intensities, or short stimulus durations, subjects may not be able to attend to the signal adequately. Stallard and Leslie claim that this results in a 3.0 dB decrement in performance. Time uncertainty may be reduced by providing indicators, usually lights, that either warn the subject that a stimulus is about to be presented, or that the stimulus is being presented. Robinson and Watson cite this technique as a good means for avoiding the effects of time uncertainty.

In a reexamination of the d'_{opt} model, Green and Sewall (1962) attributed Green's (1960) earlier results to uncertainty of signal onset. To eliminate that effect, they presented the signal at a much higher level, leaving no doubt in the subject's mind that the signal was being presented. The predicted and observed measures were found to be much closer with this technique than with the technique used earlier by Green.

Time uncertainty was systematically studied by Egan, Greenberg, and Schulman (1961). Subjects were instructed that a signal might occur after a variable period of time following a warning light. The interval between the light and stimulus was zero, one, two, four or eight seconds. Subjects responded on a four-point scale, indicating the degree of confidence they had that a signal was presented. As the interval of time uncertainty increased, d'_e decreased monotonically. When the signal energy was increased, the curve relating d'_e to time uncertainty shifted upwards, showing that time uncertainty still influenced performance even at high intensities. This result is confounded with memory in that increases in time uncertainty might be accompanied by short term memory decay. It is likely that poorer performance is attributed to both of these factors.

3.3 Comments

The potential effects of uncertainty in a task must be considered prior to conducting any study. Stallard and Leslie hypothesize that signal or time uncertainty will each reduce detectability by one-half. Varying amounts of these two factors may affect detectability differentially. Every subject's performance is influenced to a large extent by the uncertainty of the task, so in comparing the results of different procedures, the influence of time and signal uncertainty must be evaluated.

In the present experiment, signal uncertainty will be reduced in two ways. First, only one signal pair will be used throughout the experiment and all the subjects will have had some experience with this particular pair in previous experimentation. Uncertainty will be reduced further by presenting subjects with an exposure to the signals prior to each trial.

Time uncertainty, on the other hand, will not be minimized. In the fixed intensity method, the signal will be added to the noise stimulus after a fixed period of time has elapsed from the start of the stimulus. Subjects will be somewhat uncertain as to the exact time of onset, but, with training, will gain some feeling as to when to expect the signal to appear. In the other method, the signal will always be mixed with the nose, but the SNR will initially be low and increase with time. There will, however, be more uncertainty as to the time of offset of the stimulus, as the length of the trial and starting SNR will vary from trial to trial.

CHAPTER IV

PROCEDURAL CONSIDERATIONS

4.1 Feedback

When an observer is provided with information regarding the presentation of a stimulus in an a posteriori fashion, then he is being given feedback (FB). This may exist in the form of lights, numbers, explanations, or tones, and may inform the subject as to his performance in a direct or indirect way. When direct, FB is given after every trial and the individual knows immediately whether or not he was correct on a specific trial. Indirect FB would be given after a group of trials, when subjects get an indication of their overall performance.

Auerbach (1971) proposes a learning model for frequency discrimination which incorporates the effects of FB. Auerback states that through FB, a subject learns to attend to a particular aspect of the signal and his performance improves. Without FB, learning still occurs, but at a much slower rate. FB plays an important role in recognition according to Sandusky and Ahumada (1971). They say that it encourages sequential assimilation of trials. This is evidenced by a subject responding in a similar manner on trial "n" when he was informed he was correct on trial "n - 1." Stimulus configuration and "tonal position" were both found to influence the effects of FB in Snelbecker and Fullard's (1972) research. They used three levels of FB with two sets of eight tones. Subjects were required to respond with the ordinal number of the tone (one through eight) from the set after it was presented. A triple interaction between FB, spacing of the tones in the set and the position of the tone was found. Thus, stimulus characteristics and FB influence performance simultaneously.

Other studies show little if any influence of FB. Emmerich (1968b) examined FB and no-FB conditions with both 2AFC and SI (single interval) tasks, and found no-FB groups did slightly better in both conditions. Gundy's (1961) study showed inconsistent effects of FB. Regardless of whether FB groups did better or worse in either condition, the difference in performance was always minimal.

Carterette, Friedman, and Wyman (1966) studied the effects of no-FB, and 100, 75, and 50 percent accurate FB in a 2AFC task. The no-FB and 100 percent groups resulted in higher P(C)'s than the other FB groups, but the difference was insignificant. The authors concluded that FB, correct or otherwise, causes the subject to question the adequacy of his criterion. This eventually leads to a reduction of the detectability index. McNicol (1975) contends that FB is a cause of criterion instability. In his study five FB conditions were used: none, 100, 50, and 20 percent accurate and a special 100 percent group. The no-FB group was found to produce the highest d' measure. Campbell (1965) found that the SNR necessary to maintain P(C)'s of 0.88, 0.75, and 0.62 were not contingent on FB. Irrespective of the experience of the individual, FB groups required no smaller an SNR than no-FB groups. Robinson and Watson conclude that FB does little to improve performance, saying that the only time it is advantageous, or even necessary, is during training.

Despite substantial evidence that FB does little to facilitate performance, it is still frequently utilized. Of the 22 studies reviewed that mention FB, one-half made use of it during the experiment. Incorporating an FB system into an experiment surely requires considerable effort. Its advantages should be accurately assessed before such efforts are expended.

4.2 Memory Load

Contingent upon the procedure itself, the memory capabilities of an individual may play a large or insignificant role in his performance. Complex signals, such as those in Janota's and Fidell's studies, place a larger demand on memory than pure tone signals. In a task where two similar sounding signals have to be discriminated, and the signals are presented in a noise background, the memory demands are considerably higher than when a pure tone signal has to be detected in a noise. A more exact memorial image is required in the former case, where the signals differ in only one or two ways.

Jesteadt and Bilger (1974) suggest that one reason for superior performance in multiple interval tasks is the reduction in memory requirements. In multiple interval tasks, both signal-plus-noise and noise-alone intervals are presented and the subject is only required to detect the difference in the two. In an SI task, the subject has to compare the presented stimulus with some image he has in memory. Faulty memory reduces the accuracy in detection. Green and Sewall cite that another reason for the failure of the d'_{opt} model is the observer's insufficient memory of the frequency spectrum of the signal. This nonsensory confound can be a serious deterrant to the validity of any study if not accounted for.

4.3 Consistency

A factor that bears heavily on the implications of all results is the consistency of the observers. If subjects are inconsistent in their level of performance, some question arises as to what it is that is really being measured and as to how reliable the obtained measure is.

Green (1964) and Bell and Nixon (1971) both examined subjects' responses to identical stimulation. Using a noise increment as a signal, Green had his subjects listen to the exact same stimuli six times and found a 65 percent agreement in responses. Pure tones were used in a second experiment, and the subjects listened to the same tape four times. This resulted in an average 70 percent agreement score. Bell and Nixon created 50 noise and 50 signal-plus-noise trials and arranged them randomly on 10 tapes. Responses were in terms of a five-point rating scale. Using four subjects, the correlation coefficients calculated between the responses obtained from separate presentations of the same signal-plus-noise stimuli were 0.33, 0.56, 0.67, and 0.81. The correlations obtained with the noise stimuli were -0.14, 0.12, 0.34, and 0.51. Atkinson's (1963) variable sensitivity theory is supported by these results. He says that the activation and decision processes of an individual are dynamic and result in changes in the subject's sensitivity within and between sessions. This was found by Binford and Loeb (1966) when the hit and false alarm rates shifted during experimental sessions. An obtained measure of detectability in any experiment thus appears to lack a high degree of reliability. With such inconsistency, an experimenter must be wary of his conclusions.

Inconsistent criteria are discussed by Shipley (1970). She shows that the variances of the criteria determine the shape of the ROC, and the inferences drawn about the variances of the noise and signal distributions. When the criterion variability is high, the slope of the ROC will approach unity regardless of the underlying distributions. Furthermore, as the criterion variability increases, detectability decreases. In a visual signal detection task, Wagenaar (1973) found results in agreement with Shipley. He calculated that shifts in his subject's criterion caused underestimates in P(C) by 0.08 to 0.10.

Hammerton (1970) studied subject's abilities to maintain consistent criterion by presenting random numbers sampled from distributions of known mean and variance, and hence, known d'. The noise distributions had a mean of 40 and the three "signal" distributions had means of 43, 47, and 50. The variance of all distributions was ten. One group of subjects responded in a YN manner, and another with a five-point rating scale. The subjects were to state from which distribution they felt the presented number was sampled. The resultant d's obtained with these subjects were all less than the actual d', suggesting that subjects were unable to maintain an optimal criterion, or, in some instances, optimal criteria.

4.4 Type of Response

Another consideration in the design of an experiment is the means by which the subjects are to respond. In responding in an experiment, a subject can be required to maintain a single criterion or several criteria. Single criteria, or binary responses, usually come in the form of a "yes" or "no" decision. In rating experiments, subjects maintain several criteria simultaneously. After a stimulus presentation, the subject responds by noting the highest criterion that was exceeded by the stimulus. This is analogous to, but not the same as, requiring subjects to respond with an estimation of the confidence they have in their decision.

Three to nine criteria are often used in studies, but as many as 36 have been reported. The criteria are usually accompanied by verbal descriptions such as "strict," "certain," "lax," and "possibly." These serve to aid the subject in determining what is actually meant by a criterion. Occasionally, probabilities of false alarms and hits are specified by the experimenter to further clarify the criteria. These descriptors are often used in single criterion experiments as well. In all procedures, the meaning of a criterion can be manipulated to some extent by the experimenter.

The advantage of a multiple criterion paradigm is that a ROC curve can be obtained in a shorter amount of time. Also, more information is acquired from each trial since the subject's response is based on a continuum. In one of the earliest experiments comparing binary and rating methods, Egan, Schulman, and Greenberg (1959) stated that the two procedures yielded similar results and that the rating was better because fewer trials were needed. A binary response procedure was conducted three times, each with a different definition of what the criterion should be. This definition was in terms of allowable false alarms. The rating procedure employed a four-point scale. The obtained d's differed by no more than 0.14. Binford and Loeb's results comparing binary to three-point ratings indicated that the multiple criterion group did a little better, with fewer false alarms and a greater number of hits. The slope of the ROC in a rating task was found to decrease with signal strength in a study by Markowitz and Swets (1967), whereas the slope under binary conditions remained near unity. Like Egan et al., they found little difference in their detectability measure. A pushbutton slider with 20 discrete categories was used by Emmerich (1968a). Using the area under the ROC, he found little difference between the binary and rating techniques.

Despite these results, much evidence exists showing multiple criterion methods yield smaller measures of detectability. As remarked in the secion on consistency, a subject's criterion is not necessarily stable. Shipley points out that a criterion has a variance also, and that this reduces the detectability of the signal. When many criteria are used, their individual variances are increased. Shipley compared rating scales of three, five, and nine items, and used d'_e as her measure of detectability. The five-point scale resulted in the highest value, with the three-point scale close to it, and the nine-point a distant third. Shipley concludes that her hypothesis was supported by these data. McNicol (1975) agrees with Shipley, saying that the results obtained under multiple criterion procedures are less accurate due to their variability. Watson, Rilling, and Bourbon (1964) also found detectabilities with binary responses to be higher. They used a 36-point rating scale with d'_e as a the measure of detectability. For the binary condition, the d' measure was used. Consistently, d' was greater than d'_e. Some question exists as to the legitimacy of comparing two different detectability measures when each is computed in a different manner. Generally, however, binary and rating responses yield different measures of detectability, but the quantitative extent of these differences is yet to be determined.

4.5 Comments

An important decision facing every experimenter is that regarding the use of FB. Most systematic studies indicate that FB does little in improving an observer's performance, yet many researchers still utilize it. Nonetheless, the difficulties encountered in implementing FB into an experiment overshadow the few advantages it may present. Hence, it will be omitted from this experiment.

Efforts should be made to facilitate a subject's memory in those situations where it may be unduly taxed. Changing the task, e.g., Green and Sewall, is one means of accomplishing this. Another would be the method used to reduce signal uncertainty, presenting a cue or pretrial exposure of the signal to the subject. By using one signal pair and a cue in this experiment, it is felt that the demands upon a subject's memory will be minimal.

An observer's performance in a detection task varies considerably. This inconsistency can be reduced through practice and training, but will still be high. Indications are that large individual differences may occur between the subjects of any study. Such factors must and will be considered in the analysis.

Requiring subjects to maintain multiple criteria will yield smaller detectabilities than if they use a single criterion. This is due to the subject's inconsistency, which increases as the number of criteria available increases. Some studies do not show this effect, however, and since multiple criterion paradigms require fewer trials overall, it is a more economical arrangement, and will be used here.

CHAPTER V

MEASURES OF DETECTABILITY

5.1 Difference in Means

The best known index of detectability was a result of the work done in the mid to late 1950's by Tanner, Green, Birdsall, and Swets. This led to the now well-known theory of signal detectability. The theory describes the human observer as one in which the magnitude of a sensation elicited by a stimulus is a normally distributed random variable. This is similar to Thurstone's notion of discriminal dispersion. This random variable has a mean and variance which is relative to some arbitrary point along a psychological continuum. The detectability of a stimulus pair is defined as the normalized difference of the means of the two stimuli and is denoted d'. To determine this, one of the stimuli is arbitrarily given a mean of zero and a variance of one. Under the assumption that the ratio of the variances of the two distributions is one, d' can be calculated by obtaining the percentage of hits and false alarms and using a table of normalized values.

By varying the characteristics of a stimulus, one can determine the detectability of various components of a signal by examining differences in the resultant d' values. The effects of other variables can also be assessed in this manner.

The biggest fault with the d' measure is the assumption of equal variance. The variances of the underlying distributions are difficult to determine, and the assumption of equality makes calculation easier. Experimental results, however, suggest that this assumption is not valid; the variance of a signal-plus-noise distribution is usually found to be larger than that found for a noise-alone distribution.

Green's (1960) d'_{opt} measure, on the other hand, accounts for unequal variance. The variances of the two distributions are included in the calculations. These variances are determined by measuring the stimuli with an instrument, such as a real time analyzer. Once performance levels are obtained for various measures of d'_{opt}, a psychometric function can be derived. Using the psychometric function, performance at other values of d'_{opt} can be predicted.

It should be kept in mind that the underlying assumption in the d'_{opt} model is that the human observer is an energy detector. Thus, signal-plus-noise distributions determined by various instruments are assumed to be identical to those that are detected, or that arise within the individual. Such might not be the case, but Green and Sewall's results suggest that the model is indeed a good one.

5.2 Intersection with the Negative Diagonal

Egan, Greenberg, and Schulman (1961), and Egan and Clarke (1966) discuss a different method of determining detectability. It is still essentially a difference in means of two distributions, but the measure is independent of distribution variances. Using a ROC, detectability is calculated by taking twice the normal deviate of the point of intersection of the ROC with the negative diagonal. When two distributions are transformed such that the mean of one has a value zero, then the point of intersection represents the midpoint between the means. Doubling this gives the distance between the means. This measure is termed d_s. It is identical to another measure found in the literature called d'_e. The d'_e value is calculated by subtracting the ordinate value from the abscissa. In all cases, these two values are identical, so, for brevity's sake, only d_s will be discussed further.

Changes in the ratio of signal variance to noise variance will cause the slope of the ROC to change. The d_s measure is hypothesized to be the "pivot" about which the ROC, when drawn on normal-normal coordinates, rotates. Thus, d_s is not affected by differences in the ratio of the variances of the two distributions, and the d_s measure is less arbitrary than d'. When the variances are equal, and the slope is one, d' and d_s are identical in value.

5.3 Orthonormal Distance

A measure which is similar to the above in many respects is the length of the orthonormal from the ROC to the origin. The ROC would be drawn on normal coordinates and appears as a straight line. If the slope of the ROC is one, the variances are equal and the orthonormal distance d_{gm} is equal to d_s. As the variances become more divergent, d_{gm} and d_s become more and more dissimilar. Increases in the ratio of variance of stimulus one to stimulus two cause decreases in d_{gm} and have little influence on d_s, provided the ROC rotates about the point of intersection with the negative diagonal. Like the aforementioned measures, d_{gm} is also an estimation of the difference in means of the distributions.

5.4 Non-Parametric

Hammerton and Altham (1971) proposed an index C which makes no assumption about the distributions of the stimuli. A rating scale with "r" criteria, "1" being sure signal "A," "r" being sure signal "B," is necessary in this calculation. First, mean ratings for either signal are obtained. Detectability is then defined as the mean rating of "B" minus that of "A" divided by (r-1). Perfect detectability equals one, when the difference in the mean ratings equals the difference between the highest and lowest ratings. This C measure is monotonic with d' and takes on the values zero to one.

Sakitt (1973) suggested that Hammerton and Altham's index was insensitive to differences in the distributions of the ratings. Sakitt proposed a measure D (A,B), which, like C, is based on multiple criteria response procedures. The numerator is the same as that used in C, the difference in the mean rating given each stimulus. The denominator, however, is the square root of the product of the standard deviations of the ratings. Thus, changes in the variance of a subject's responses affects the calculated value. The less variable the subject's ratings, the higher his detectability.

The area under the ROC is another often used measure of detectability. In this paper, this measure will be referred to as Ar. When performance is at chance level, the area equals 0.50, and when perfect, it equals 1.00. Green and Swets (1966) prove that the area under the ROC is the percent correct detections by the subject in a 2AFC task. This is an attractive measure in that no assumptions are made at all, other than that the ROC is accurate.

5.5 Comments

In the majority of studies in which it was possible to determine the ratio of the variances, which is simply the slope of the ROC on normal-normal coordinates, it was evident that the variances of the stimulus distributions are not equal. These results are usually obtained when one of the stimuli is one of signal-plus-noise, and the other is noise-alone. In view of this, d' is not a good measure to use in situations where the distributions are unknown. Due to the insensitivity to changes in the variances of ratings, the C index seems inappropriate. Another disadvantage to the C measure, and to the D(A,B) measure as well, is its dependence on rating responses. This negates the possibility of comparing detectabilities obtained with different procedures. Another complication with D(A,B) is that interpretation becomes difficult when the variances become small. Perhaps it is too sensitive to this factor, for detectability will approach infinity as the standard deviations approach zero.

Of the three remaining indices, the area under the ROC appears to make the fewest assumptions. In a study by Simpson and Fitter (1973), it was found that Ar varied the least with changes in the variances of the stimulus distributions. Comparisons between d_{gm}, d_s, and Ar were obtained by setting the standard deviation of noise to one, and then varying that of the signal-plus-noise distribution from 0.25 to 4.0. The orthonormal measure showed the greatest fluctuation. Pollack and Hsiah (1969), in a computer simulation study, examined Ar and d_s using Gaussian, rectangular, and exponential distributions. Using these measures to compute P(C), the authors found a difference of no more than 0.06 under all the sampling conditions. Although Ar was best, d_s seemed comparable.

Green's measure appears to be most appropriate in the present situation. One of the reasons is that it can be used with noise signals, and noise signals of different center frequency and bandwidths. Also, this measure takes into account the variances of the two distributions. By obtaining P(C), which is the same as Ar, for several values of d'_{opt}, psychometric functions can be obtained and compared to those of other experimental procedures. By using d'_{opt}, insight can be provided concerning the legitimacy of Green's energy detector model of the human.

As a point of interest, d', d_s, and d_{gm} will be computed for each signal intensity. A comparison of these three measures with themselves and with the d'_{opt} should prove interesting.

CHAPTER VI

METHODS OF STIMULUS PRESENTATION

6.1 Modified Threshold Procedure

The modified threshold procedure, developed by Janota (1977), was born out of a need to assess the detection capabilities of observers outside the laboratory. The signals that were used were similar to some of those found in marine environments. Changing signal strength, common in many "real life" surroundings, was assimilated into this procedure. Overall, the approach was more an applied than an experimental one, primarily because an actual problem was being investigated.

As outlined by Janota, each trial began with a brief exposure of two signals, arbitrarily labeled "A" and "B." After the exposure set, one of the signals was presented with an ambient noise background at a very low signal-to-noise ratio. At this value, the signal was well below detectable levels. As the trial proceeded, the SNR was incremented by one-half dB steps every two seconds. The signal strength relative to the noise, therefore, increased with the passage of time, becoming more and more detectable. The subjects' task was to respond with their decision regarding which signal was being presented with the noise. They were instructed to respond as soon as they were "reasonably certain" of their decision. The subject was assumed to have a criterion threshold and, once the signal strength exceeded this value, the subject responded. Hence, the procedure can be considered one of establishing where this criterion threshold is. The procedure is called modified threshold to distinguish it from the threshold techniques of classical psychophysics.

Responses were made by pressing one of two response keys marked "A" and "B." Once a response was made, the entire stimulus was gated, and there was silence until the start of the next trial. The trials were arranged such that the stimulus would terminate at a certain level automatically if no response was made by the subject. This cutoff was variable from trial to trial.

The signal pairs were marine in origin, and a total of 16 pairs were investigated. Several of the signal pairs were actual recorded marine sounds. Others were laboratory-generated signals resulting from the combination of several components. Among these components were narrow and broadband noise, and amplitude modulation. The signals of each pair were similar except for one specific feature. A dichotomous feature would be introduced by omitting or deleting one of the components from one of the signals.

In each session, four groups of six trials were run for a total of 24 trials per session. A different signal pair was used with each group. The trials lasted approximately two and one-half minutes apiece. The session was recorded on tape and played back over headphones to the subjects. When the subject responded, his decision, as well as the SNR of the signal, were both recorded on a cassette tape. No feedback was given to the subjects regarding their performance.

Results were analyzed in terms of the SNR necessary to reach a decision and the observed P(C). Due to the small number of data points, results were pooled across subjects, obtaining an average P(C) and SNR for each of the signal pairs. P(C) and SNR differed considerably from pair to pair, depending on the nature of the dichotomous feature. Some features were easily detectable, with P(C) being in the neighborhood of 0.97, but others were as low as 0.57. Further analysis indicated that, in some situations, it was more difficult to detect the absence of a feature than its presence.

Standard deviations in the SNR at response ranged from 3.37 to 4.99. Unfortunately, it is not known how much of the variability was due to within- vs. between-subject factors, but it appears as though both factors contributed significantly to the variability. There were indications that certain samples of subjects avoided no response trials even at the cost of a higher error rate. Zero response trials should have occurred when the initial SNR was low and the feature was difficult to detect. In these situations, the SNR should not have exceeded the criterion before the stimulus was terminated automatically. To avoid zero response trials, these subjects responded at a lower SNR, knowing that the trial was soon to end. Such behavior increased the variability by an undeterminable amount.

Stallard and Leslie (1974) predicted that detectability would be 5.4 dB less in passive sonar environments than in a 2AFC detection task such as Green's (1960). The modified threshold procedure bears some similarities to this environment, so Janota used his results to check Stallard and Leslie's predictions. Using the d'_{opt} equation, the detectability of the dichotomous feature of four signal pairs was calculated. With a correction factor of 5.4 dB, the obtained detectabilities were 0.3, 0.5, 1.3, and 2.2 dB from the predicted. This discrepancy could have been due to the novelty of the task; very few trials, 24 to 96, depending on the individual, were collected per subject. With experience, the actual values may have approached those predicted.

Analysis of the data in terms of practice effects, sequential responding, and individual differences were not possible due to the small number of recorded trials. Statistical tests were largely out of the question for the same reason.

Although no conclusive results were provided, Janota did provide psychoacoustics with a new method of signal presentation, one in which the dynamic characteristic of the signal intensity more closely approximates that found outside the laboratory. The modified threshold research conducted thus far is not without its faults, but it does provide some new information regarding how individuals respond in a free choice, dynamic SNR situation. It also provides an excellent approach for an examination of the reliability of a subject's criterion. Further research with this method should help delineate some of the parameters of signal detection.

6.2 Multiple and Single Interval Tasks

Most detection tasks can be classified as either single or multiple interval. In single interval (SI) tasks, the observer receives one presentation of a stimulus, after which he makes some kind of response. In contrast, two or more stimuli are presented to the subject in multiple interval tasks. These presentations occur in close temporal proximity to one another, and are separated by some specified length of time. After the last interval, the subject makes his response. The use of as many as eight intervals have been reported in the literature, but only two interval tasks will be discussed here.

These are commonly called two-interval forced-choice procedures and are denoted 2AFC.

Based on the theory of the ideal observer, Green and Swets (1966) predict the relationship between the detectabilities obtained with 2AFC and SI procedures to be:

$$d'_{2AFC} = \sqrt{2} \times d'_{SI} \quad .$$

This hypothesized ratio is accepted by many, but research does not always support it. Stallard and Leslie use this ratio in their calculation of reduced detectability in the sonar environment. The Stallard and Leslie predictions fit Janota's data rather well.

The hypothesized ratio was tested by Schulman and Mitchell (1966). Data were collected under both procedures with a six-point rating scale. ROC lines drawn on normal-normal axes indicated that slopes of 2AFC results were steeper and closer to unity than slopes from the SI task. Using the ratio given by Green and Swets, predictions could be made of either detectability based on the detectability obtained with the other procedure. Schulman and Mitchell found the ratio of d'_{2AFC} and d'_{SI} to be approximately 1.46, which is close to Green and Swet's prediction of 1.41.

Jesteadt and Bilger (1974) argue that since the SI task requires memory for the signal, performance will be reduced by more than that predicted by Green and Swets. They state that the task in the 2AFC condition is one of detecting a difference in energy, which does not make as great a demand on memory.

Swets (1959) compared an SI task with a 2AFC and 4AFC and found no difference in performance except that the results of the SI task were more variable. Four different signal intensities were used and in none of them did the SI consistently show smaller d's.

Emmerich's (1968b) results comparing different combinations of interaural stimulus presentation showed no difference in performance under either multiple or single interval tasks. In the SI task, a pushbutton slider was used by the subject to rate "signal likeness" [see Watson, Rilling, and Bourbon (1964)]. In the 2AFC task, a YN response was recorded under different criteria. The difference in the methods of response could account for the unexpected result; however, detectabilities obtained under rating procedures are usually smaller than with YN responses, so this should increase the difference, not diminish it. No explanation was given as to why these results occurred.

6.3 Comments

Janota has provided psychacoustics with a more life-like procedure in which to determine signal detectability. This is not to take away from the groundwork laid, and being laid, by research performed using other techniques. But results of modified threshold work can be applied more readily to natural environments.

A comparison with 2AFC procedures showed a marked reduction in detectability with the modified threshold technique. The hypothesized difference of 5.4 dB, from Stallard and Leslie, was close to that observed. The primary difference between these procedures lies in the amount of uncertainty found with the modified threshold method; this seems a likely candidate as the cause of the reduced performance. Reductions in uncertainty should diminish the differences between these methods. The question then becomes one of how much must uncertainty be reduced, and which form of uncertainty, time or signal, contributes the most to reduced performance.

As mentioned, no conclusive results have yet been obtained with the modified threshold technique due to individual differences, small amounts of data, subject variability, and lack of adequate subject experience with the task prior to data collection. These problems will be addressed and, hopefully, resolved in the present study.

PART B

THE EXPERIMENT

CHAPTER VII

RESEARCH PLAN

Two different methods of signal presentation are utilized in this study. One method is a derivative of the modified threshold technique, and is called the modified threshold-forced response procedure (MTFR). As in the modified threshold technique, after an exposure set, one of the signals from a pair is presented at a low SNR. As the trial proceeds, the strength of the signal relative to the noise is increased in one-half dB steps every two seconds. This continues until a final SNR is reached. The signal remains at this final SNR for a period of time, after which the signal is gated. In the silent interval that follows, the subject responds. The primary difference between this procedure and that of the modified threshold is that the observer is no longer in a free response situation; he has no control over termination of the signal. Another difference is that the length of the stimulus presentation varies from 45 to 55 seconds rather than from 76 to 95.

The second method is an SI task which presents the signal at only one SNR. This method is called the fixed presentation (FP) technique. After the exposure set, a noise-alone stimulus is presented for a period of 15 seconds. This noise-alone period is presented to make the memory demands more comparable to those in the MTFR method. After this interval, one of the signals is added to the noise at a fixed SNR. It remains at this level for a period of time and is then terminated. The observer responds in the interval which follows.

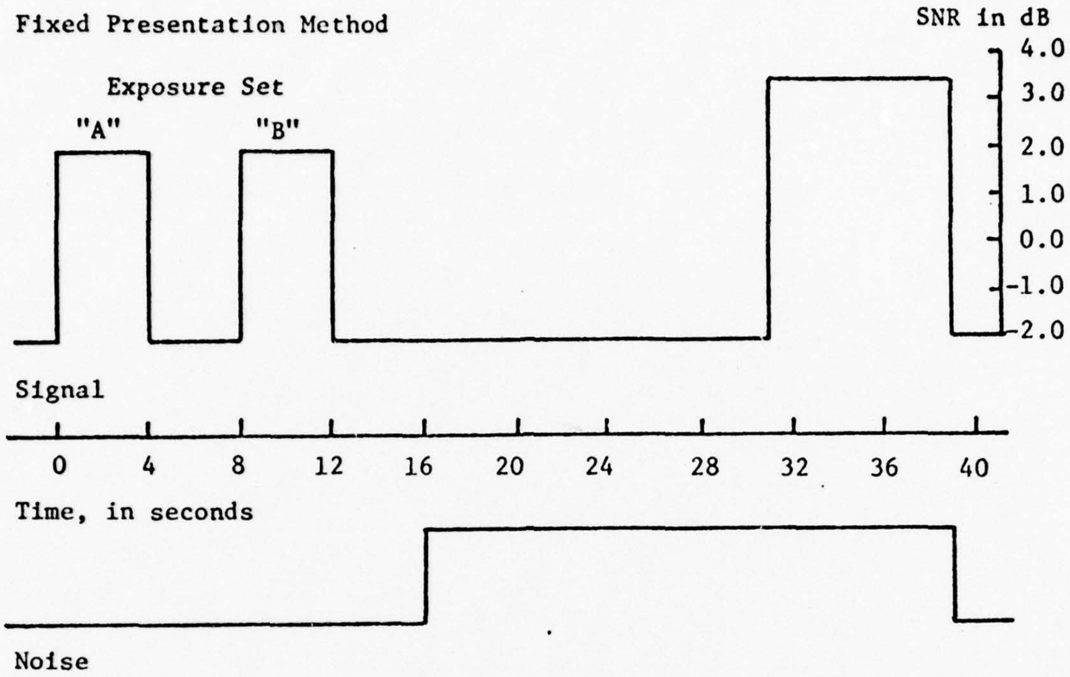
Both methods are schematically diagrammed in Figure 1. The ordering of stimulus presentation can be seen here as well as the incremental character of the SNR, in dB, of the MTFR method.

The final intensity of the signal is the same in both methods. In addition, the durations of the signal at the peak intensity are identical. This duration, eight seconds, is sufficiently long to give the subject an opportunity to attend to the signal.

A total of seven different signal intensities are used with each method. This enables comparisons to be made between the two procedures under easy and difficult detection situations. Five SNR levels are crossed with each method, yielding a 2 × 5 repeated measures design. The two remaining SNR values are not fully crossed, but nested in each method. Janota's results indicated that some difference in detection may be attributable to the order of signal presentation in the exposure set. To control for this effect, the signals of the pair are presented equally as "A" and "B" across sessions. This is added as a third variable to yield a 2 × 5 × 2 design. (Analysis of the results indicated that there was little difference in performance as a function of stimulus order, so this factor was deleted.)

Only one pair of signals is used throughout the entire experiment. Both signals contain an octave band of noise centered at 4 KHz. In addition, one signal contains another octave band of noise centered at 500 Hz. This second octave band is the only difference in the two signals. The task is then one of detecting the dichotomous band of noise. Hence, the experiment can be described as one of detection and not recognition.

Fixed Presentation Method



Modified Threshold-Forced Response Method

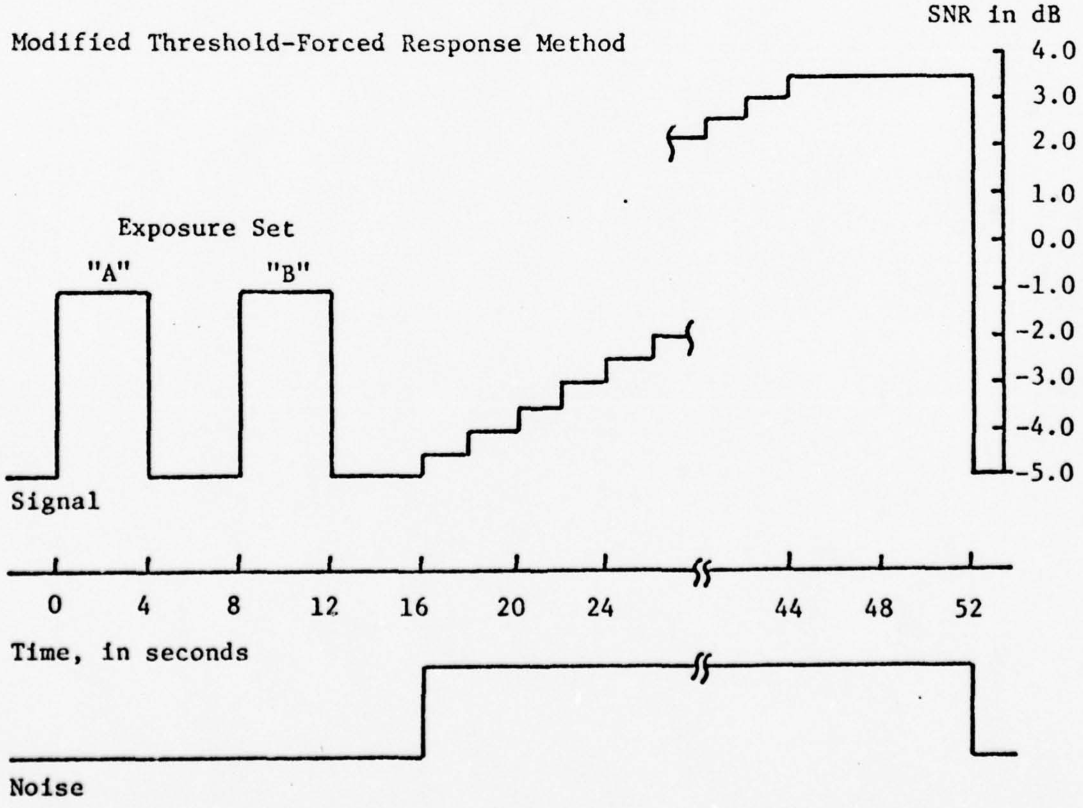


Figure 1. Fixed Presentation and Modified Threshold-Forced Response Paradigms

A six-point confidence rating scale is used to record subject's responses. This scale is symmetric, and the values are described as follows:

+3 - sure signal A, odds 6 to 1 in favor of A
+2 - reasonably certain A, odds 5 to 2 in favor of A
+1 - maybe signal A, odds 4 to 3 in favor of A
-1 - maybe signal B, odds 4 to 3 in favor of B
-2 - reasonably certain B, odds 5 to 2 in favor of B
-3 - sure signal B, odds 6 to 1 in favor of B

The subject is required to respond with one of these values on every trial. No feedback is given to the subjects due to the lack of evidence in support of its use.

CHAPTER VIII

HYPOTHESES

The greatest determinant of performance in a psychoacoustical experiment is the method by which the stimulus is presented to the subject. This is best shown by the 5.4 dB difference between Janota's modified threshold technique and the results obtained by Green with a 2AFC procedure. Many factors contribute to this large difference. The purpose of this experiment is to establish the importance of some of these factors by comparing two techniques of stimulus presentation.

In the two techniques examined, only one signal pair is used. This not only reduces signal uncertainty, but reduces the memory load on the subject as well. This is facilitated by specifying each signal, by means of a cue, before each trial. Furthermore, in both methods, the onset time of the signal is less variable. Conjointly, these two procedural changes reduce the amount of uncertainty tremendously. Hence, it is hypothesized that the FP and MTFR methods will result in better performance than the modified threshold technique.

When comparing the MTFR technique with the FP, it is expected that the latter will yield better performance for two reasons. First, there is less time uncertainty in the FP than the MTFR. Second, subjects do not have to attend to the noise stimulus presented prior to the final SNR because no signal is present at this time. This reduces the amount of noise introduced into the "system." These two characteristics of the FP method combine to make detectability better in this procedure.

The use of a rating response scheme will account for a reduction in performance, relative to designs employing binary responses. However, it will not be possible to quantitatively determine the size of this reduction from this experiment. The FP and MTFR techniques are essentially SI tasks, and this accounts for an additional decrease in performance when compared to 2AFC tasks. Using the equation to predict performance, it is expected that Green's (1960) data will show the best performance, followed by the FP, MTFR, and the modified threshold results.

Janota's results were fairly close to that predicted by Stallard and Leslie. The FP and MTFR procedures will be used to further evaluate Stallard and Leslie's hypotheses. In both methods, signal uncertainty is minimal, and its effect will be excluded from the predicted results. With the FP method, time uncertainty is also less, and will be omitted. Predictions will be in terms of the expected shift, in dB, of the psychometric function relating $10 \log d'_{opt}$ to P(C). The quantitative estimates can be found in Section 10.7.

One potential difficulty found with the modified threshold technique is the observer's inconsistency. Two reasons for this could be the desire of the subjects to avoid no response trials and the possibility of their using the passage of time as a criterion. Both of these influences are eliminated in the FP and MTFR methods. This should increase the subject's reliability. Consistency is further enhanced by having definitions of the six criteria readily available during every session. This should reduce inter- as well as intrasession variability. Estimations of consistency are made by correlating responses to identical sets of trials collected at different points in time. Comparisons with correlations obtained with the modified threshold procedure are expected to indicate higher correlations with both of these methods.

By sampling several signal intensities, a more accurate assessment can be made of the two methods. As was stated earlier, the FP method is expected to yield better performance. This will be indicated by higher P(C)'s and confidences at the lower intensities. Even at the higher intensities, where little difference will be noticed between either method in terms of P(C), FP responses should reflect higher confidences.

An additional effect to be examined is that found by Shipley. She found that the variance of the distribution was not monotonically related to intensity. This is examined in the present experiment by comparing the slopes of the ROC's obtained with different intensities. The ratio of the variances are expected to decrease with increasing intensity because, at higher intensities, detection is easier.

Through this investigation, some of the factors that reduce detection performance in the modified threshold technique will be evaluated. Based on these results, it is hoped that a better understanding can be acquired of the factors that influence detection in a signal-plus-noise environment.

CHAPTER IX

METHODS

9.1 Subjects

Four male graduate students in acoustics served as subjects. All the subjects had participated in acoustical experiments at The Pennsylvania State University earlier in the academic year. The previous work lasted four months, during which time each subject listened to approximately 260 trials using the modified threshold technique. Twelve signal pairs were used in this prior research, one of which was used in this experiment.

The subjects were paid for each session, but received no bonuses or rewards for superior performance.

9.2 Apparatus

Signal Characteristics. The pair of signals used in this experiment was used by Janota in earlier work. The pair comprised "Treatment 5," and the two signals were labeled w_2 and w_7. This particular pair was chosen because it was desirable to have a signal pair of medium difficulty. Previous experimentation with these subjects using the modified threshold technique resulted in a probability correct of 0.84. The average SNR at response was 11.06 dB, with a standard deviation of 3.24.

Both signals contain an octave band of noise centered at 4 KHz. The band-edges of this band were 3.10 and 5.17 KHz. A spectrum analysis of w_2 and w_7 is shown in Appendix A. One of the signals, w_7, contained an additional octave band of noise centered at 500 Hz with band-edges of 387 and 646 Hz. This band of noise was the only difference between the two signals and comprised the dichotomous feature.

The signals were created by passing the output of a General Radio Company Type 1390-B Random Noise Generator through two SKL Variable Electronic Filters, Model 302. One of these devices low pass filtered the noise at 5.17 KHz and high pass filtered it at 3.10 KHz. Two outputs were taken from this filter and one was used as signal w_2. The other was mixed with the noise band specified by the other filter to form w_7. This second filter, using the same noise source, low pass filtered it at 646 Hz and high pass filtered it at 387 Hz.

The background noise was bandlimited at 70 and 10000 Hz. This was necessitated by the limitations of the recording instrument and the automatic gain control used in the system. The ambient noise was generated by taking another 1390-B noise source and passing it through a third SKL filter.

Both signals were adjusted so that the power in the 4 KHz centered bands was equal. A typical spectrum analysis of the two signals and the noise source is shown in Appendix A.

The intensity of the signal was determined by the largest SNR in the 4 KHz centered band. This level yielded an effective SNR of the dichotomous feature, which is later used in the calculation of d'_{opt}. An example of how these values are calculated is given in Appendix A. The signal intensities that were used are given in Table 1.

TABLE 1

SIGNAL INTENSITY AND d'_{opt} CALCULATIONS

SNR Calibrated*	SNR Effective**	d'_{opt}	10 log d'_{opt}
-0.5	-2.5	4.05	6.07
0.5	-1.5	4.89	6.89
1.0	-0.5	5.84	7.66
2.5	0.5	6.89	8.38
3.5	2.0	8.57	9.33
4.5	3.0	9.70	9.87
6.5	5.5	12.96	11.13

$$d'_{opt} = (WT)^{\frac{1}{2}} \frac{\sigma_s^2}{\sigma_n^2} \left[\tfrac{1}{2}(\sigma_s^2/\sigma_n^2)^2 + (\sigma_s^2/\sigma_n^2) + 1\right]^{-\frac{1}{2}} ,$$

where W = effective bandwidth of the noise; 355.5 for f_o = 500 Hz,

T = integration time, duration; 400 msec,

σ_s^2 = measured variance of signal distribution

and

σ_n^2 = measured variance of noise distribution.

*Nondichotomous feature, f_o = 4.0 KHz

**Dichotomous feature, f_o = 500 Hz

Tape Construction. The tapes were recorded using the facilities at the Applied Research Laboratory at The Pennsylvania State University. An apparatus designed by Janota was used to record and sequence each trial. Once the signals and noise were generated, they were input into this system. This apparatus enabled the experimenter to record the order and length of each part of a trial automatically. A balanced mixer combined the signal and noise to the desired SNR and maintained a constant overall loudness of the stimulus.

All the trials were recorded on one-inch Soundcraft Instrumentation Tape with an Ampex FR1200 Fourteen-Channel Recorder. These tapes, 17 in number, were designated primary tapes and contained all the trials of a particular cell. The audio tapes, which contained a complete experimental session, were recorded from cuts of the primary tapes on one-quarter-inch Ampex 756 and 736 Series Tape. By selecting different cuts, different audio tapes could be generated. All audio tape recording was accomplished by connecting the output of the Ampex FR1200 Recorder to the input of a Crown 700 Recorder. The verbal instructions (see Appendix D) were added by using a Superscope CS-200 Cassette Player which was also connected to the Crown Recorder. The instructions for each presentation technique remained the same throughout the experiment, and were pre-recorded on the Superscope using a Sony C-90 Cassette Tape. The controls of the Crown Recorder were set to record at a constant loudness of 65 phons (GD) (ISO R532). The schematic for this set-up is illustrated in Appendix B.

The tapes were made such that, across all sessions, both signals, w_2 anf w_7, were presented equally as "A" and "B." This was counterbalanced with respect to SNR and technique. The order of the signals on each tape was random, with each signal being equally likely on each trial.

Testing Environment. Testing took place in an enclosed audiometric booth. The booth was located in an acoustical laboratory in the Applied Research Laboratory at The Pennsylvania State University. The size of the booth was 1.90 × 1.00 × 1.20 m, and accommodated one subject at a time. A 75-watt light bulb provided illumination. The booth contained a fiberglass chair, a ledge to write on and a 60 × 40 cm window. The Crown 700 Recorder was used to play back the trials to the subjects, and was located outside the booth. The trials were presented over TDH 30 headphones, which were calibrated to insure accurate playback.

At his convenience, the subject would go to the laboratory and obtain his file from a cabinet adjacent to the booth. This file contained the tape assignments for each individual. The subject would then mount his assigned audio tape on the Crown Recorder, procure a response sheet, start the tape, and enter the booth.

Each response sheet listed the verbal description of each confidence rating on the top, middle, and bottom of the page. The criteria were also defined in terms of odds at the top of the page. This information was contained on every response sheet in an effort to increase the subject's consistency from session to session. A sample response sheet is presented in Appendix C.

9.3 Training

The subjects had previous experience with the modified threshold technique using a criterion of "reasonably certain." To train them in the use of a multiple criterion method of responding, six tapes were made. The first tape consisted primarily of comments and instructions regarding the new techniques and procedures. Each subsequent tape contained fewer verbal comments and more trials. Appendix D contains the six sets of comments and the number of trials that were contained on the training tapes.

Previous results with the modified threshold method suggested that SNR's of 13.0, 10.5, and 8.5 would adequately differentiate the two presentation methods. As training progressed, however, it became evident that much lower SNR's were needed to attain P(C)'s of less than 1.00. In each successive training tape, therefore, the signal was presented at a lower SNR. By the end of training, subjects had listened to signals presented at SNR's as low as 3.5 dB.

The six tapes were listened to in a span of two weeks, and each tape lasted between 40 and 50 minutes. A total of 74 trials with the MTFR technique and 81 with the FP technique were gathered during training. Individual discussions with the subjects indicated that this was an adequate introduction to the new procedures.

9.4 Procedure

A complete experimental session was recorded on each audio tape. This allowed the subjects to participate at their leisure, without having to notify the experimenter. Each subject had a list of the order in which he was to listen to the tapes. The order was determined randomly, so that a subject might listen to any combination of intensity and technique at any time during the 20 weeks of testing.

Once a session was started, it ran continuously without any breaks for approximately 40 minutes. A set of instructions for the session was given first. These instructions described the presentation technique and the confidence rating procedure. The set of instructions given the subjects depended on the presentation method. Both instructional sets can be found in Appendix D.

After the instructions, a double exposure of the two signals was given in the order "A," then "B." The durations of the first exposure were eight seconds, with a silent interval of four seconds between them. The second exposure, and all subsequent exposures, was four seconds long with a silent interval of four seconds in between. The signals were presented without any background noise.

The stimulus was presented next. In the FP method, this was noise alone. After a period of 15 seconds, the signal was added to the noise for eight seconds, after which time both were gated. Trials lasted approximately one minute with this procedure, and a total of 30 were given per session. In the MTFR method, the stimulus presented to the subject contained the signal, but at a low SNR. The SNR of the signal increased until the desired intensity was reached. It remained at this level for eight seconds, after which the stimulus was terminated. These trials lasted approximately one and one-half minutes and there were a total of 24 collected per session. In both techniques, only one signal intensity was used per session.

A period of 15 seconds occurred between each trial, and subjects recorded their confidence ratings at this time. This interval was silent except for the announcement of the next trial. This procedure of exposure, trial, and silence was used throughout each session. Diagrams of the procedures are given in Appendix E.

Audio tapes were numbered such that, if the number was less than 200, the FP technique was used. If the number was 200 or greater, then the MTFR method was used. Tapes with the last digit equal to zero presented w_7 as signal "A." Tapes with the last digit equal to five presented w_2 as signal "A." (The design was counterbalanced so that each signal was presented as "A" the same number of times it was presented at "B.") This was the only a priori knowledge the subjects had of the session that they were to participate in. They were never told at what SNR the signals were to be presented, nor were they given any feedback regarding their performance. Subjects were encouraged to write down any comments concerning their strategies and criterion.

Subjects were allowed to listen to no more than two tapes a day. They were requested to listen to five tapes a week, but this depended on the individual's schedule. Several of the tapes were listened to twice. There were a total of 52 sessions recorded per subject.

CHAPTER X

RESULTS

10.1 Order of Presentation

Some of the data collected with Janota's modified threshold technique indicated that the order of stimulus presentation may have had an effect on performance. The counterbalancing used in this study was done to control for this possible factor.

In the 2 × 5 × 2 factorial design, two sessions were recorded for each cell, which amount to 48 (four additional sessions at a high SNR value were added for practice, making a total of 52 sessions). P(C) for each cell was calculated by averaging the two sessions. To examine order effects, the averaged P(C)'s were compared. In Table 2 are given the differences in P(C) obtained by subtracting the P(C) found with w_7 as "A" from that with w_2 as "A." This table is broken down by subject, by SNR, and by method. In general, the difference is small, but occasionally, a difference as great as 0.21 is noted. No consistent relationship is shown and, therefore, it seems justified to delete order as a factor in the analysis.

The design now becomes 2 × 5, method and intensity being the two variables; this makes four sessions per cell per subject. Thus, each cell is based on 120 trials in the FP method and 96 in the MTFR method.

TABLE 2

DIFFERENCES IN PERFORMANCE AS RELATED TO ORDER*

	Fixed Presentation					Modified Threshold-Forced Response				
SNR	1.71	1.72	1.73	1.77	Mean	1.71	1.72	1.73	1.77	Mean
-0.5	0.19	0.03	-0.09	-0.17	-0.010	--	--	--	--	--
0.5	0.03	0.00	0.09	-0.05	0.035	0.00	-0.21	0.02	-0.13	-0.080
1.0	0.01	0.21	0.03	0.05	0.073	0.11	0.03	0.02	-0.02	0.035
2.5	0.00	0.01	0.02	0.05	0.020	0.15	0.11	0.05	0.03	0.085
3.5	0.03	0.00	-0.02	0.00	0.003	-0.06	-0.02	-0.14	-0.04	-0.065
4.5	0.00	-0.05	0.00	0.02	-0.008	-0.11	-0.11	0.10	-0.16	-0.070
6.5	--	--	--	--	--	0.02	0.09	-0.02	-0.06	-0.008
Mean	0.043	0.033	0.005	-0.017	0.017	0.018	-0.018	0.005	-0.063	-0.015

---

*P(C) obtained with w_7 as "A," w_2 as "B" minus P(C) obtained with w_2 as "A," w_7 as "B."

10.2 Within-Subject Variability

To get an idea of the amount of within-subject variability, subjects were required to listen to 15 tapes twice. The confidence ratings were then rank ordered and correlated. If a high correlation was obtained, this would indicate that the subject was consistent in his responses. This procedure was done using both methods and at a variety of SNR's.

The results of this analysis are contained in Table 3. For each subject, the difference in P(C), the correlation, the number of days between sessions, and the signal intensity are given. This information is separated by method.

Generally, the correlations obtained with the FP method are higher than those obtained with the MTFR. The highest correlations achieved by the MTFR technique are usually no greater than the lowest ones found with the FP method. Both methods, however, show a trend of increasing within subject reliability as the SNR of the signal increases.

The number of days occurring between the two sessions does not appear to have much of an effect, indicating that the subjects were consistent over a long period of time. Both the correlation and P(C) are independent of this possible factor. A correlation between intervening days and difference in P(C) yielded a value of 0.068 for the FP method and -0.147 with the MTFR. The correlation between intervening days and the within-subject variability resulted in a value of 0.013 for the FP, and -0.045 for the MTFR. This is fortunate in that, in several instances, the number of days between sessions is large, and to analyze these results separately would be cumbersome.

TABLE 3

CORRELATIONS OF RESPONSES TO IDENTICAL TAPES

Subject 1.71

	Intervening Days		ΔP(C)*		Correlation	
SNR	FP	MTFR	FP	MTFR	FP	MTFR
-0.5	22	--	0.10	--	0.73	--
0.5	13	25	-0.04	-0.33	0.94	-0.27
1.0	--	113	--	0.00	--	-0.06
2.5	51	36	0.00	-0.08	0.98	0.34
2.5	--	42	--	-0.08	--	0.12
3.5	81	37	0.00	0.05	1.00	0.71
3.5	--	41	--	0.00	--	0.64
4.5	18	--	0.00	--	1.00	--
4.5	18	--	0.00	--	1.00	--
6.5	28	21	0.00	0.16	1.00	0.60
6.5	--	53	--	0.04	--	0.21

Subject 1.72

	Intervening Days		ΔP(C)*		Correlation	
SNR	FP	MTFR	FP	MTFR	FP	MTFR
-0.5	23	--	0.00	--	0.94	--
0.5	49	52	0.07	-0.08	0.57	-0.07
1.0	--	54	--	0.13	--	0.07
2.5	49	66	0.03	-0.12	0.89	0.20
2.5	--	20	--	-0.25	--	-0.04
3.5	73	26	0.00	0.08	0.97	0.29
3.5	--	125	--	-0.21	--	0.05
4.5	94	--	0.10	--	0.84	--
4.5	47	--	0.00	--	0.97	--
6.5	54	52	0.00	-0.09	0.90	0.48
6.5	--	49	--	-0.08	--	0.48

---

*ΔP(C) = P(C) second run - P(C) first run

TABLE 3 (Continued)

Subject 1.73

	Intervening Days		ΔP(C)*		Correlation	
SNR	FP	MTFR	FP	MTFR	FP	MTFR
-0.5	82	--	0.03	--	0.53	--
0.5	31	33	0.03	0.00	0.74	0.17
1.0	--	122	--	0.00	--	0.57
2.5	14	36	0.00	0.04	0.95	0.42
2.5	--	36	--	-0.08	--	0.32
3.5	75	32	0.03	0.29	0.92	0.53
3.5	--	41	--	0.08	--	0.15
4.5	62	--	0.00	--	1.00	--
4.5	41	--	0.00	--	0.98	--
6.5	31	41	0.07	0.17	0.85	0.66
6.5	--	54	--	-0.05	--	0.49

Subject 1.77

	Intervening Days		ΔP(C)*		Correlation	
SNR	FP	MTFR	FP	MTFR	FP	MTFR
-0.5	10	--	-0.21	--	0.14	--
0.5	34	12	-0.10	0.21	0.63	0.13
1.0	--	78	--	-0.16	--	0.05
2.5	5	21	0.00	-0.09	0.94	-0.02
2.5	--	10	--	-0.04	--	-0.11
3.5	89	5	-0.04	-0.12	0.76	0.32
3.5	--	91	--	0.04	--	0.13
4.5	36	--	0.00	--	0.97	--
4.5	7	--	0.00	--	0.96	--
6.5	50	5	0.00	0.01	0.93	0.32
6.5	--	39	--	0.29	--	0.28

---

*ΔP(C) = P(C) second run - P(C) first run

This also shows that the subjects were sufficiently trained prior to the experiment. If a high correlation had existed between intervening days and the difference in P(C), then the adequacy of the training would be suspect.

The differences in P(C) obtained with the FP method are small, showing little effect of practice. With one exception, all the values fall within a range of -0.10 to 0.10. This difference in P(C) is generally larger with the MTFR method, as would be expected as a result of the subject's greater variability with this technique. Overall, P(C) differences are small in light of the fact that they are based on small amounts of data. These values were obtained by subtracting P(C) from one session from that of another. Three more correct in one session would result in a difference in P(C) of at least 0.10. This fluctuation in P(C) is best explained as within-subject variability rather than a practice effect.

It should be recalled that the correlations are of the ratings given by subjects, not of their performance in the two sessions. The higher the correlation, the smaller the difference in P(C). The converse of this is not true, however. A subject can obtain identical P(C)'s for two sessions, but respond at different levels of confidence. Thus, ratings were chosen to be correlated due to their greater sensitivity to the subject's criterion variability.

In Table 4 are given the correlations of the SNR at responses for the subjects using the modified threshold technique. These results were obtained from earlier experimentation. Unlike confidence ratings, the SNR at response is not an arbitrary value; hence, these are not

TABLE 4

CORRELATIONS OF SNR AT RESPONSE USING
THE MODIFIED THRESHOLD TECHNIQUE

Subject	Tape 42	Tape 39	Tape 37	Mean
1.71	0.92	0.87	0.52	0.82
1.72	0.89	0.74	0.54	0.76
1.73	0.70	0.76	0.51	0.67
1.77	0.53	0.71	0.69	0.65
Mean	0.81	0.78	0.57	0.74

rank correlation coefficients. The correlations are between separately recorded sessions of each subject listening to the same tape.

The correlations ranged from 0.51 to 0.92. These values were averaged using Fisher's z-transformation which yielded a mean of 0.74. The average correlations for each subject ranged from 0.65 to 0.82. This indicates a fairly consistent criterion of "reasonably certain" was maintained by the subjects using the modified threshold method. It is interesting to note that Tape 42 was listened to near the end of the experiment, and resulted in the highest correlations. In contrast to Tape 37, the sessions of which were collected early in the experiment, it appears that the more experienced the subjects were with the procedure, the greater their consistency.

10.3 Between-Subject Variability

In psychophysical research, the results are usually analyzed on a per-subject basis. Individual differences are commonly large and necessitate this action. To determine individual differences in this study, an analysis similar to that employed in the previous section was used.

The rank ordered responses of each subject were correlated with those of each other subject. The tapes used in this analysis are the same as those used before and, thus, each subject has two sets of responses, one for each session with the 15 tapes. This results in an 8 × 8 correlation matrix. These matrices can be found in Appendix F. They contain the within- as well as the betwee-subject correlations with the selected tapes. The results are summarized in Table 5.

TABLE 5

AVERAGE WITHIN- AND BETWEEN-SUBJECT CORRELATIONS TO IDENTICAL TAPES, USING FISHER'S z-TRANSFORMATION

	FP			MTFR		
SNR	Average Within	Average Between	Diff.*	Average Within	Average Between	Diff.*
-0.5	0.681	0.405	0.286	--	--	--
0.5	0.766	0.680	0.086	-0.011	0.178	-0.167
1.0	--	--	--	0.168	0.110	0.058
2.5	0.946	0.894	0.052	0.245	0.139	0.106
2.5	--	--	--	0.098	0.119	-0.021
3.5	0.993	0.972	0.021	0.478	0.446	0.032
3.5	--	--	--	0.264	0.226	0.038
4.5	0.999	0.998	0.001	--	--	--
4.5	0.999	0.998	0.000	--	--	--
6.5	0.991	0.926	0.065	0.544	0.485	0.059
6.5	--	--	--	0.380	0.405	-0.025

---

*Average within- minus average between-correlation

In Table 5 can be found the average, using Fisher's z-transformation, within- and between-correlations, and their difference. These numbers are based on four correlations for within, and 24 for between. The difference between these means becomes smaller with increasing SNR. With both methods of presentation, the difference is no greater than 0.065 at SNR's 3.5 dB or greater. This indicates that the average between-subject variability is only slightly larger than the average within.

On this basis, it appears justified to pool the results across subjects, ignoring individual differences. Although the legitimacy of this action may be debatable, the data shown in Table 5 provides adequate support for the pooling and creation, as it were, of a theoretical subject. This operation makes all subsequent analyses simpler, reducing the number of figures and tables necessary to adequately present the data. Another advantage is that the obtained ROC's are more accurate. On account of the pooling, each curve with the FP method is based on 480 trials, and with the MTFR method, 384 trials. All subsequent results are discussed in terms of this theoretical subject, a product of the pooling across subjects.

10.4 Performance

Using the d'_{opt} values listed in Table 1, P(C) is plotted as a function of signal intensity and graphically illustrated in Figure 2. Each point represents a single session of 30 or 24 trials, depending on the method. Data are also plotted from Green's (1960) results with a 2AFC. The d'_{opt} values are transformed to $10 \log d'_{opt}$ so that they may be compared to Green's results, which are also transformed in this

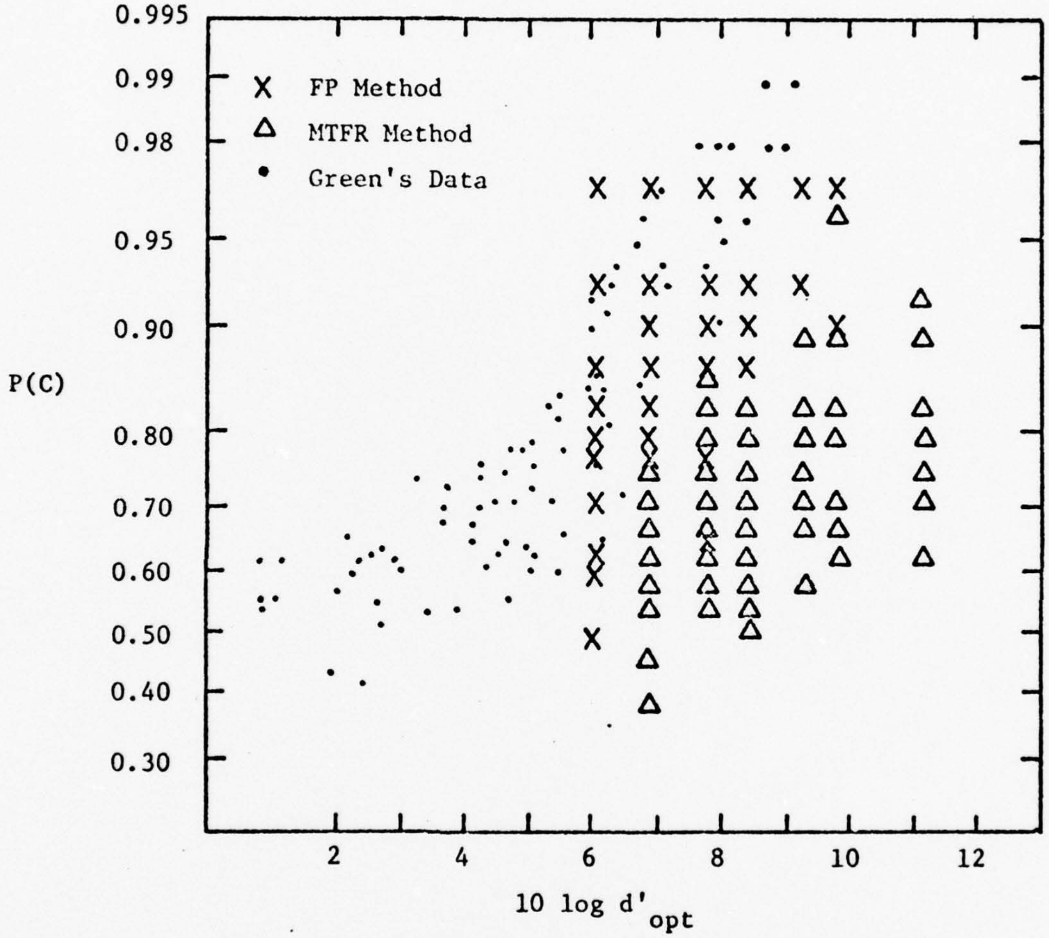


Figure 2. Performance in P(C) as It Relates to 10 log d'_{opt}

manner. This transformation allows computation of the difference in performance in terms of power units.

In several sessions, a P(C) of 1.00 was obtained. Due to the fact that this value cannot be represented on probability paper, it is excluded from Figure 2. In several sessions with the FP technique, this level of performance was attained, especially at the higher signal intensities. Only once was a P(C) of 1.00 achieved with the MTFR.

These three sets of data were used as input to a Biomed (BMDO3R) polynomial regression program. In all cases, a third-degree polynomial was used. The results of this analysis are shown in Figure 3. All the sessions were used in this computation, including those with P(C)'s of 1.00.

The FP and 2AFC functions have similar slopes, but the slope of the MTFR is considerably different. The data of the FP are shifted to the right, indicating poorer performance. The data of the MTFR are shifted even further to the right, showing that it resulted in the poorest performance. The FP function is shifted about one-half dB, while the shift of the MTFR varies from 4.0 to 6.0 dB. These shifts are relative to the results of the 2 AFC.

Also shown in Figure 3 is the obtained result using the modified threshold technique with the same subjects in an earlier experiment. Included here are the 90% confidence intervals for both P(C) and 10 log d'_{opt}. To obtain this datum, d'_{opt} was calculated using the SNR at response. This point is based on a total of 45 trials with the modified threshold technique. The modified threshold method differs from Green's by 6.0 dB, from the FP by 5.0, and from the MTFR by 0.8 dB.

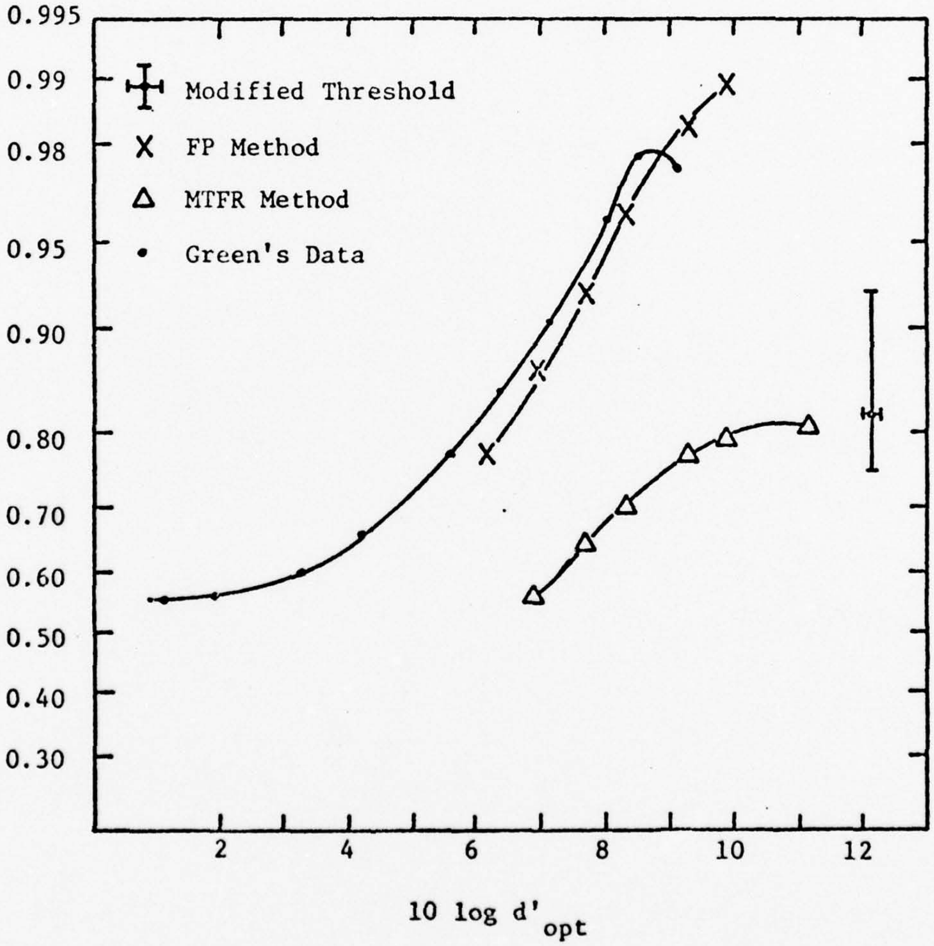


Figure 3. Third-Degree Polynomial Fit of Psychometric Functions

Table 6 gives the averaged P(C) obtained with each SNR for each method. This value is calculated from the P(C)'s obtained from the four subjects participating in four sessions with each method × SNR cell. This amounts to 16 sessions per cell. It can be seen from Table 6 that, with only two exceptions, P(C) increases steadily with signal intensity.

10.5 ROC Analysis

In this study, subjects had to determine whether or not the dichotomous feature, the 500 Hz centered octave band of noise, was present in the stimulus. As Janota points out, this is similar to testing the null hypothesis. In the ensuing discussions, the null hypothesis H_0 will be that the dichotomous feature is absent. The alternative H_1 is feature presence. Therefore, to make the calculations for a ROC, one must determine the probability of responding H_1 given H_1 was presented, and the probability of responding H_1 given H_0 was presented.

Since no difference in performance was found as a function of signal labeling, or presentation, the data were manipulated such that the positive side of the rating scale corresponds to the subject believing H_1 was presented. A correct response for a presentation of H_0 would thus be a negative value.

In Figure 4 are shown the ROC's which resulted from the FP method. Performance with this method was so high at SNR's 3.5 dB and higher that only SNR's up to 2.5 dB are shown. A shift upwards on the negative diagonal is evident with each increase in signal intensity.

TABLE 6

COMPARISONS OF DIFFERENT MEASURES OF DETECTABILITY

SNR	P(C)	d'	d(s)	d(gm)	Slope	P(C)	d'	d(s)	d(gm)	Slope	SNR
-0.5	0.76	1.28	1.44	1.58	0.76	--	--	--	--	--	--
0.5	0.87	1.72	2.06	2.26	0.70	0.56	0.38	0.38	0.38	1.05	0.5
1.0	0.91	2.56	2.58	2.50	0.99	0.67	0.70	0.60	0.56	1.16	1.0
2.5	0.97	3.56	3.56	3.36	1.06	0.68	0.56	0.60	0.64	0.87	2.5
3.5	0.98	4.38	4.38	3.40	1.26	0.76	1.41	1.32	1.18	1.16	3.5
4.5	0.99	--	--	--	--	0.81	1.49	1.40	1.26	1.14	4.5
6.5	0.98	--	--	--	--	0.79	1.42	1.44	1.42	0.94	6.5

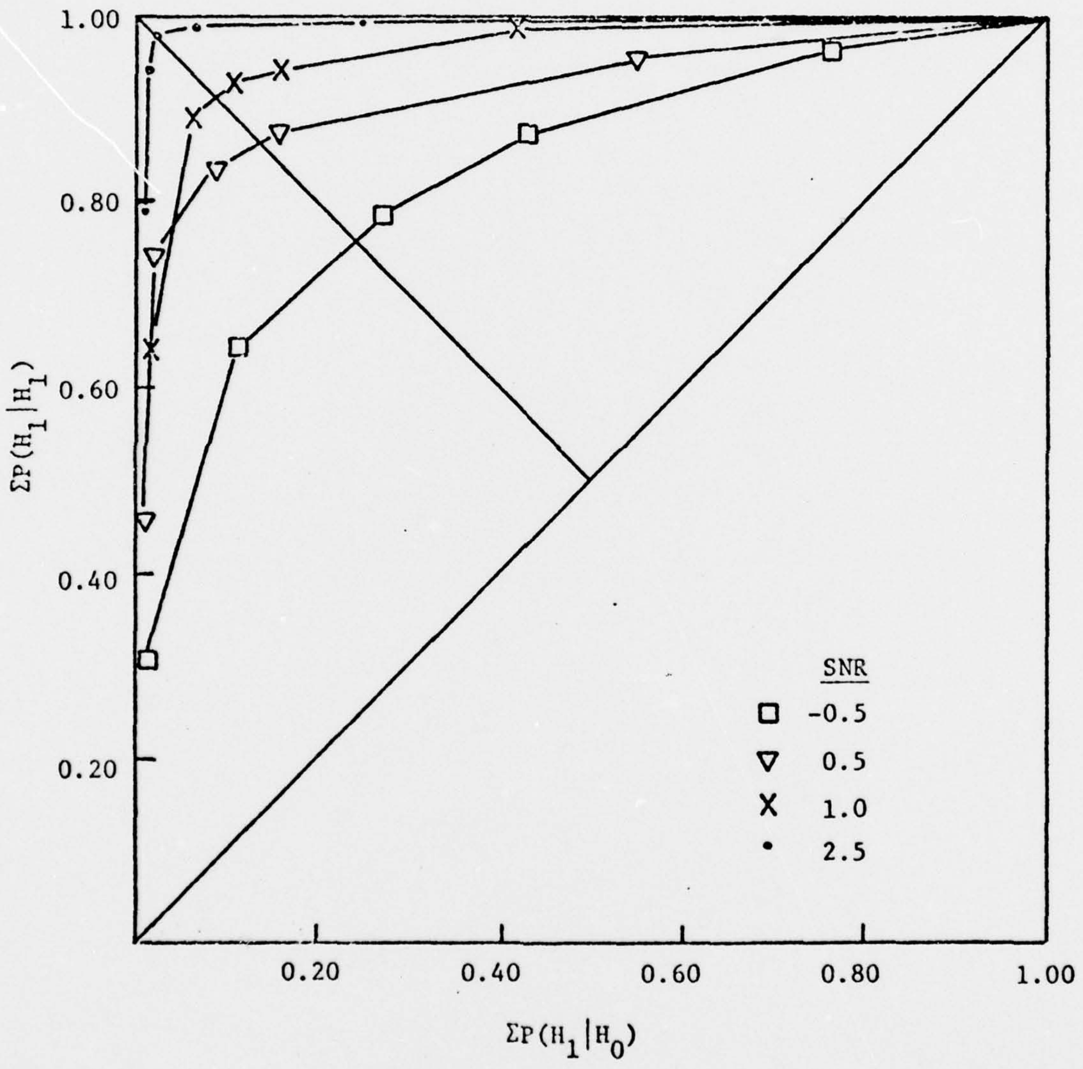


Figure 4. ROC Using the FP Method

Although performance does increase with intensity using the MTFR method, it is not nearly as marked. This is shown in the ROC curves of Figure 5. As was also shown in Table 6 and Figure 3, the highest SNR level, 6.5, resulted in poorer performance than the next highest, 4.5. The best performance achieved with the MTFR method is comparable to the second lowest level found with the FP method. Clearly, the FP method results in superior performance.

Further ROC analysis in Figures 6 and 7 show the curves on normal-normal coordinates. An SR 51-II calculator was used to perform a least squares linear regression of the normalized data. These lines are shown as well as the raw data for five signal intensities with the FP method in Figure 6 (only five signal intensities were used in the FP ROC due to the near perfect performance of the subjects at the SNR value of 4.5 dB or higher). Six ROC's are shown in Figure 7 for the MTFR method. Increases in intensity show less of an effect with this method. The slopes for all the lines are given in Table 6.

10.6 Measures of Detectability

Using Figures 6 and 7 and the least squares linear regression program in the SR 51-II, three different measures of detectability were computed for each SNR level for both methods. The d' measure was obtained by computing the intercept of the ROC with the y-axis. The coordinates of the intercept of the ROC with the negative diagonal were added to obtain d_s. Perpendicular lines were drawn from the ROC to the point (0,0) to obtain d_{gm}. Both of these latter two values are subject to error due to the difficulty in determining the precise coordinates. The error is small, however, and is estimated to be no greater than

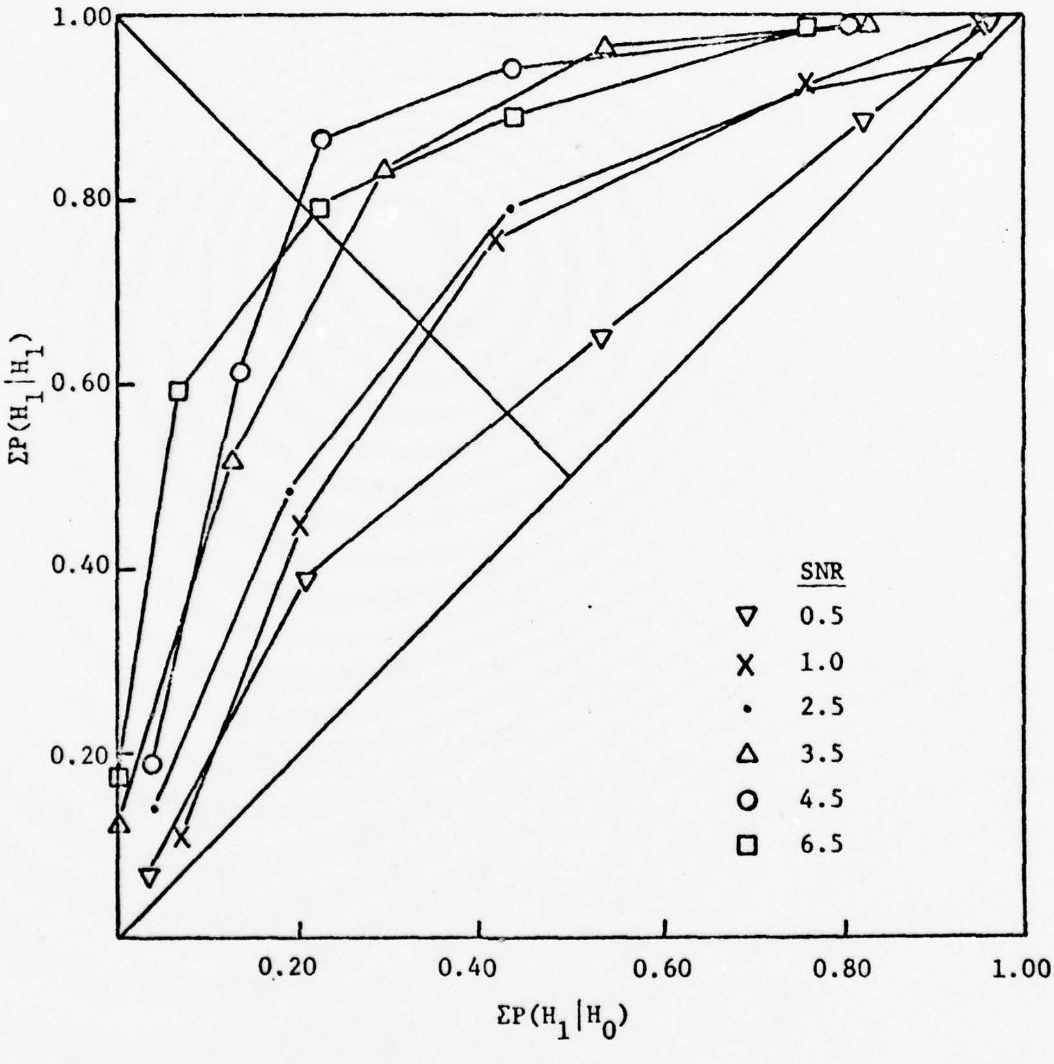


Figure 5. ROC Using the MTFR Method

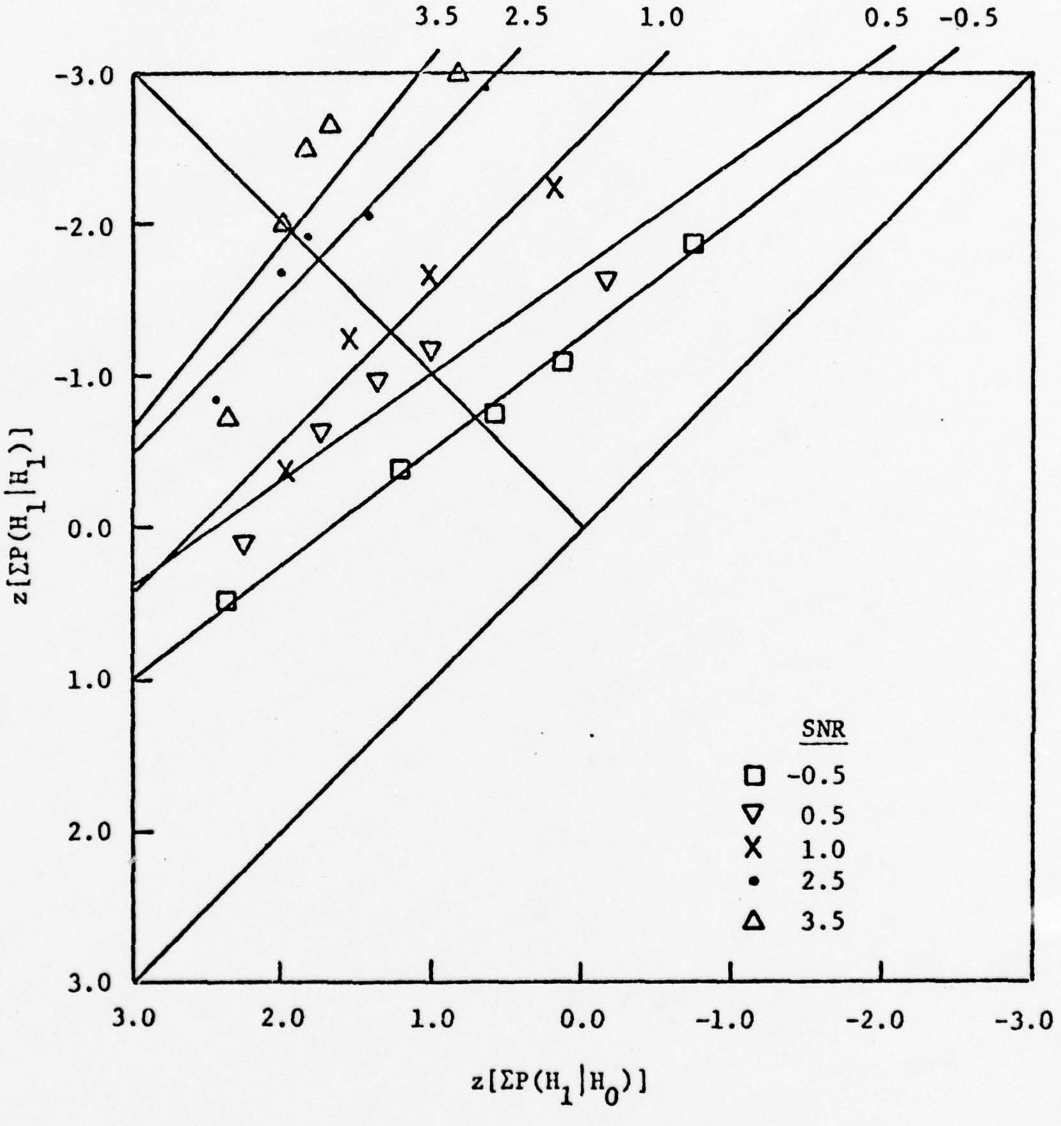


Figure 6. ROC of FP Method Plotted on Normal-Normal Coordinates

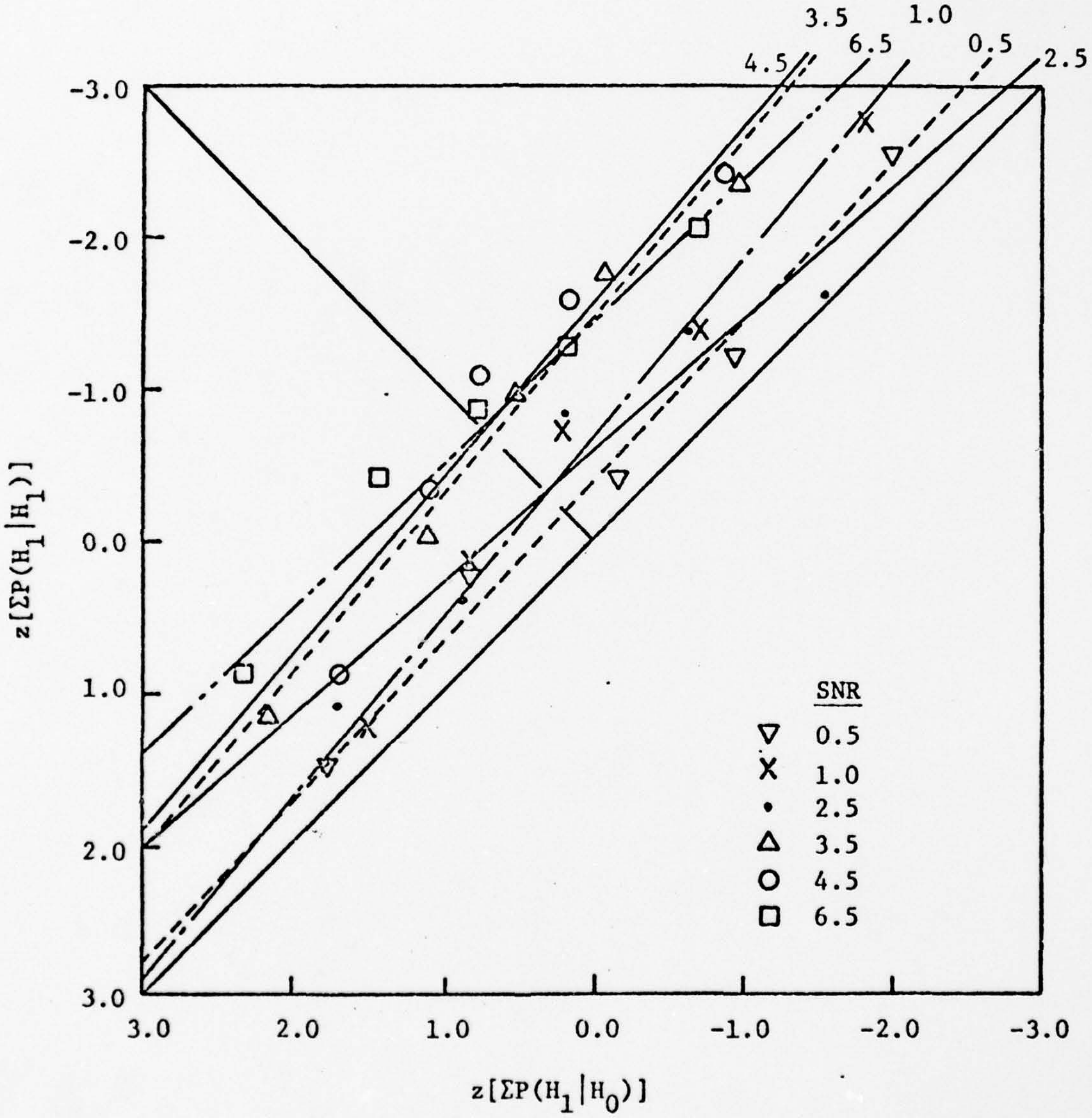


Figure 7. ROC of the MTFR Method Plotted on Normal-Normal Coordinates

0.10. These values are presented in Table 6 along with the slope of the ROC and the P(C) of each level for each method.

The similarities between the measures is dependent on the slope of the ROC. When the slope is less than one, d' is usually less than or equal to d_s and d_{gm}. The average difference between d' and d_s is -0.12, and -0.17 between d' and d_{gm}. When the slope is greater than one, d' is always greater than or equal to d_s and d_{gm}. The average differences are 0.15 with d_s and 0.31 with d_{gm}.

As with d', the difference between d_s and d_{gm} depends on the slope. With slopes less than one, d_{gm} is usually larger, the difference averaging out to 0.06. With slopes greater than one, this relationship reverses itself, with d_s averaging 0.20 larger than d_{gm}. With all these measures, the closer the slope is to unity, the smaller the difference between them.

10.7 Green's Model and d'

Stallard and Leslie cited three causes for the reduced detectability in passive sonar environments as compared to 2AFC experiments. These were time and signal uncertainty, and the presence of an SI situation, rather than a multiple interval one. Using their calculations, predictions can be made of performances in the FP and MTFR methods. The FP differs from a 2AFC in that it is an SI paradigm. Detectability, d', in the 2AFC should thus be the square root of two times larger than that of the FP:

$$d'_{2AFC} = \sqrt{2} \times d'_{FP} \quad .$$

The MTFR is also an SI task. In addition, however, some time uncertainty exists in this procedure. Stallard and Leslie point out that this halves the obtained detectability. These two factors combined would make the detectability in the 2AFC 2.8 times better than the MTFR, as shown in the following:

$$d'_{2AFC} = \sqrt{2} \times 2 \times d'_{MTFR} \quad .$$

If these differences are transformed into 10 log d' units, as Green and Stallard and Leslie did, then one can express the differences as a shift in dB:

$$10 \log d'_{2AFC} = 10 \log (1.4 \times d'_{FP})$$

$$= 1.5 + 10 \log d'_{FP} \quad .$$

The observed detectabilities of the FP should thus be 1.5 dB lower than that of the 2AFC. Similarly, the MTFR should be 4.5 dB lower:

$$10 \log d'_{2AFC} = 10 \log (2.8 \times d'_{MTFR})$$

$$= 4.5 + 10 \log d'_{MTFR} \quad .$$

The FP and MTFR should differ from each other by 3.0 dB:

$$1.5 + 10 \log d'_{FP} = 4.5 + 10 \log d'_{MTFR} \quad ,$$

thus,

$$10 \log d'_{FP} = 3.0 + 10 \log d'_{MTFR} \quad .$$

This is due to the absence or reduction in time uncertainty in the FP method.

Figure 8 illustrates the observed d' as a function of 10 log d'_{opt}. The curves of Green's 2AFC and the passive sonar environment are also illustrated (for d' greater than 2.5, these curves were extrapolated from those given in Stallard and Leslie).

The following equation was used in conjunction with the SR 51-II linear regression routine to determine 90% confidence intervals for the d' of the various signal intensities in Figure 8:

$$Se = \text{sqrt}\ \{[\Sigma(Y_i - Y_i')^2] \ / \ (n - 2)\} \quad ,$$

where Se is the standard error, and $(Y_i - Y_i')$ is the difference between the predicted and obtained values of the ROC in Figures 6 and 7.

The range of the shift for the FP method, relative to the 2AFC, is 0.3 to 0.9 dB. For the MTFR, the range is 3.2 to 5.1 dB. The difference between the FP and MTFR ranges from 3.0 to 4.7. (These values are suspect, however, due to the small degree of overlap of the curves; that is, only for a small range of d' are there values for both the FP and MTFR.)

The observed performance was better than predicted for the FP. The difference in dB with the MTFR is variable, due to the nature of the slope of that method. The predicted difference of 4.5 falls within the range of observed differences, however.

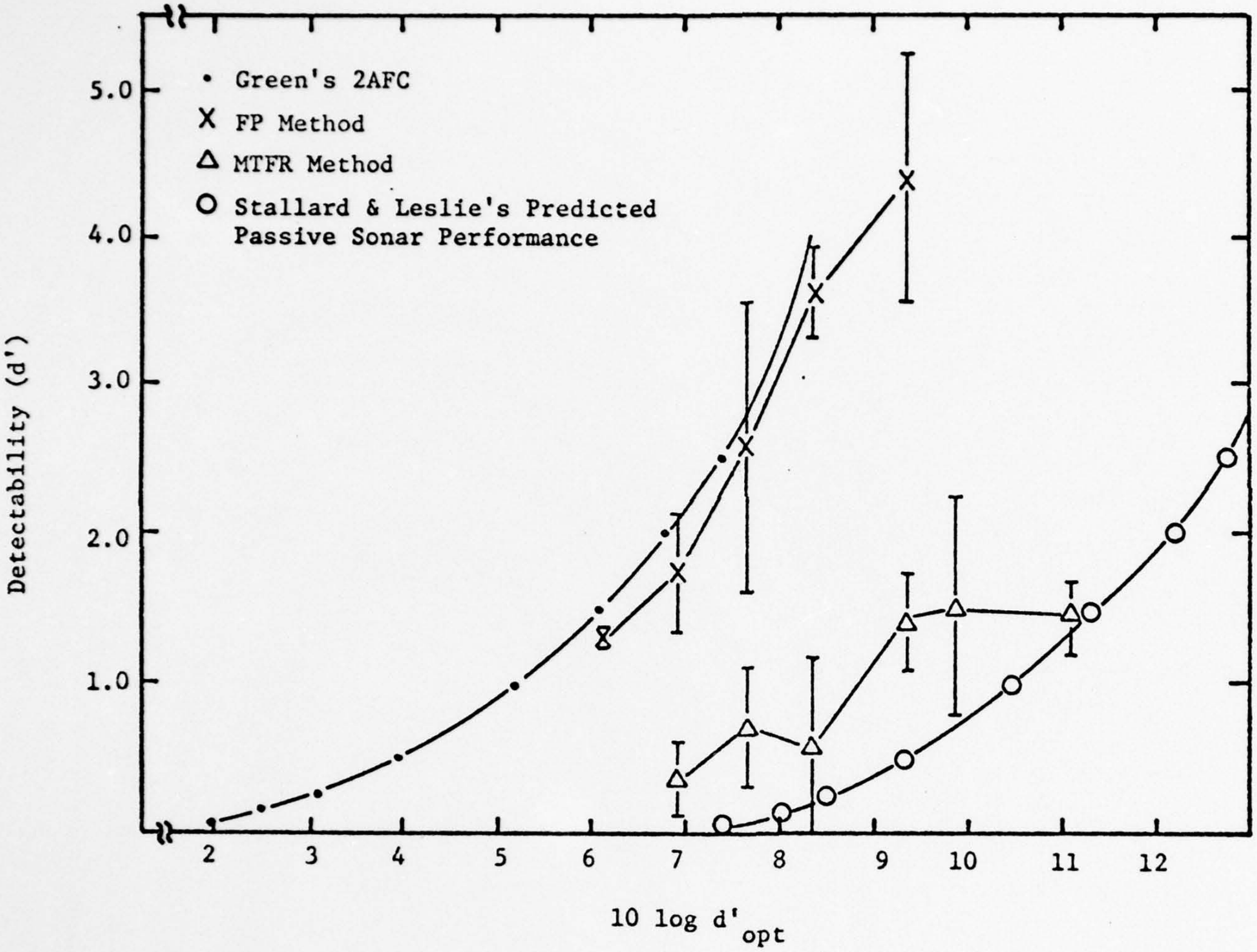


Figure 8. Detectability (d') as a Function of 10 log d'_{opt} for Four Stimulus Presentation Environments, 90% Confidence Intervals Included

CHAPTER XI

DISCUSSION

11.1 Consistency

A better understanding of the consistency of the subjects in this study might be acquired by comparing them to some earlier results. Green's (1964) results, in which responses to identical tapes were studied, appear to show a lower level of reliability than the FP, but a higher level than the MTFR. Green determined an agreement score in percent, whereas in this study, correlations of the response criterion were calculated. Green used a 2AFC and the subjects had to determine in which interval the signal occurred, a binary decision. Shipley points out that there is a considerably larger margin for variability when, as in this experiment, the subjects have to respond on a multiple criterion scale. Hence, Green's results are not directly comparable.

A more similar measure was that used by Bell and Nixon. They obtained correlations of responses to identical tasks on a rating scale. The coefficient ranged from 0.33 to 0.81. These are generally lower than those of the FP and generally higher than those obtained with the MTFR. Bell and Nixon did not, however, present their subjects with identical tapes, just identical stimuli arranged in random orders. If some sequential effect exists, it would influence these correlations, reducing them to some degree. Therefore, these correlations are not directly comparable either.

One further complication exists in comparing these studies. As indicated in Table 5, the correlation coefficients increase with signal intensity. The intensities used by Green, and Bell and Nixon may have been lower or higher than those used here. No comparative statement regarding these reliabilities can be made until the relationships between the intensities can be specified.

The higher reliabilities found with the FP technique are due primarily to the reduced time uncertainty in that task. The subjects were aware that the signal would not be presented until some time after the onset of the stimulus. Although its onset and offset were not coincident with a light or other stimulus, it always occurred after a specific length of time. A noticeable change in the stimuli occurred with the addition of the signal to the noise. This alerted the subject as to the presence of the signal. Time uncertainty did exist, but it was less than that found with the MTFR. Due to the ramping characteristic, the subject was not as aware of the signal. Also, the offset was variable with the MTFR, requiring the subject to attend to the majority of the stimulus. As a result, more internal noise was introduced into the "system." Internal noise is assumed to be a random variable, and its effect would be to reduce reliability. This is reflected in lower correlations with the MTFR.

The results shown in Table 4 do not support this hypothesis. The modified threshold technique, in which time and signal uncertainty are greater, resulted in rather high correlations of the SNR at response. One would presume that internal noise would show a marked influence here, reducing the correlations considerably. A possible explanation is that subjects used a time criterion which could be maintained more reliably. That is, they may have responded after a period of time was exceeded, not when a criterion threshold of "reasonably certain" was exceeded. Regardless of the randomization of the starting SNR, subjects might wait a specified length of time into the trial, and if they had not reached a criterion threshold, they responded in terms of a time threshold. However, this criterion should also have exhibited some fluctuation, decreasing the consistency. As with the other studies, these consistency measures are not directly comparable to those of Table 5. Inferences on the differential reliability of subjects under different experimental procedures are untenable at this time.

Internal noise may also be at fault for the lower reliabilities obtained with small SNR's. At low intensities, the internal noise plays a prominent role in the responding because there is more overlap in the signal-plus-noise to noise distributions. Internal noise is assumed random; this reduces the consistency of the observer. As the intensity increases, the internal noise does not decrease, but its effect becomes less pronounced as the difference in the distributions increases. This is indicated by higher correlations.

Another explanation for the high reliabilities is the subjects themselves. Being graduate students in acoustics may have helped in that they were aware of the kinds of noises they were listening for. Their past experiences would have allowed them to set up their filters, as it were, for the dichotomous noise band and possibly to eliminate some of the extraneous noise. Their higher level of intelligence might also have been an influence. Perhaps they were able to develop better and more consistent strategies of responding. They may have been more capable of using the scale descriptors provided on the response sheet. Furthermore, all of them had had prior experimental experience in psychoacoustics. As is indicated in Table 4, these subjects showed increased consistency with experience in the modified threshold technique.

These ideas are supported by the between-subject variabilities shown in Table 5. The subjects had similar backgrounds and were alike in their approach to the problem. As a group, they exhibited similar patterns of variability as they did individually.

More study of within-subject variability needs to be carried out before a judgment can be passed on those found here. Although it would appear that little more could be expected with the FP method, much improvement could be made with the MTFR. Overall, the results are encouraging. They suggest that fewer sessions are necessary, and the ability of each subject can be determined in less time. This would depend on the procedure used, however. A recommendation for future study is that within-subject variability be assessed. Such results have major implications on the inferences drawn from the results, and also on the effort necessary to obtain accurate findings.

11.2 Performance and Procedural Differences

A marked difference in performance was found between the two procedures used in this study. This difference is shown in the polynomial fit of Figure 3, the ROC's of Figures 4 and 5, the normalized ROC's of Figures 6 and 7, and the observed detectabilities of Figure 8.

Figures 4 and 5 emphasize the differences in the confidence ratings with each method. Each point of the ROC represents a value on the rating scale. Even the lowest ROC curve of the FP, an SNR of -0.5, matches or is higher than those of the MTFR. Figures 6 and 7 show how responsive performance is in the FP method as compared to the MTFR. With each increase in SNR, there is a clear improvement in the detectability of the signal. The absence of this relationship with the MTFR suggests that some factor, not present in the FP, is inhibiting improvement with this technique.

The outstanding characteristic that distinguishes these two procedures is the ramping of the signal in the MTFR (this was cited earlier as a difference in time uncertainty). This source of uncertainty is surely the major source of the reduced detectability with the MTFR.

Unfortunately, another contributing factor exists. The MTFR is a modified, modified threshold procedure. The initial SNR of the signal was therefore variable, which made the overall length of the trial variable. On a whole, the length of these trials was longer than that of the FP. This causes some confounding with memory, the subjects having to remember the signal pair for a longer period of time. The extent of this confound is felt to be small, primarily because one signal pair was used throughout the experiment. If this were not the case, the increased length of the trial would be influential, reducing the performance in the MTFR method relative to the FP.

Both of these methods result in better performance than the modified threshold. Recall that a P(C) of 0.84 was found with an SNR at response of 11.06 dB. At SNR's four dB and lower, P(C)'s were found to be as high or higher. One reason for this is the reduced signal uncertainty. Sixteen different signal pairs were used in the modified threshold research, with no more than six trials of any pair presented during a session. Subjects in this study showed a rapid improvement in performance during training when just one signal pair was used. The SNR's in this study were originally based on the SNR at response found with the modified threshold procedure. Performance was near perfect at these levels, necessitating the use of much lower SNR's. Stallard and Leslie claim the effect of signal uncertainty is to halve the detectability. Further research with the modified threshold method using just one signal throughout a session would test this hypothesis. The results of this experiment could be compared to those where several signals are used during a session. This would help determine the quantitative aspects of signal uncertainty.

A second reason for the difference in the modified threshold procedure is the free response situation. More responsibility is placed on the subject, causing him to attend more closely to the stimulus. As discussed earlier, this introduces more noise into the system. In the FP and MTFR, the subject can be a little lazier. It is possible for them to even let the stimulus terminate and attend to the preperceptual auditory image that remains. It would seem that reduced attention to the stimulus would reduce performance as well. In these procedures, however, this is not necessarily so.

The FP was predicted to result in poorer performance than the 2AFC because it was an SI task. Time uncertainty is present to some extent in the FP, and this should also reduce detectability. A multiple criterion response procedure was used and should further diminish performance. Of these three, only the first was used in the prediction. The prediction overestimated the decrement by one dB, suggesting that these three factors do not affect performance significantly.

Experimental evidence suggests that SI tasks may not result in poorer performance in all situations (Swets, 1963; Emmerich, 1968b). Such might be the case here as well. Lower performance resulting from using multiple criteria is another hypothesis that has not received universal support (Egan, Greenberg, and Schulman, 1959; Binford and Loeb, 1966; Markowitz and Swets, 1967; Emmerich, 1968a). If these two factors were of little consequence, perhaps the observed difference shown in Figure 3 was caused by time uncertainty alone; and since time uncertainty was not maximal, it would not halve the effective detectability as Stallard and Leslie suggest. No hypothesis can be stated at this time concerning the differential contributions each factor made to the reduction.

The predicted performance of the MTFR was close to that observed, except that the slope of the function in Figures 3 and 8 is different than that expected. A slope similar to the 2AFC is allowable within the 90% confidence interval, however (see Figure 8). This indicates that some factors may be operating in one procedure that are not in another. Indeed, since the FP and 2AFC slopes are similar, it may be the increased time uncertainty in the MTFR that changes the shape of the function. Performance in the passive sonar setting is only predicted by Stallard and Leslie, not observed. Experimental simulation of this environment may not show the slope to be like that of the 2AFC, but more like that observed with the MTFR.

The data obtained with the modified threshold method appears to be close to the function predicted by Stallard and Leslie. It was not possible to calculate a d' measure for these data due to the small number of trials available. To obtain d', the hit, correctly identifying H_1 as H_1, and false alarm rates, incorrectly identifying H_0 as H_1, are necessary. This automatically excludes much of the data, i.e., misses and correct rejections. A d' based on as few as 45 trials is dubious, using considerably less than this number is not justified. Thus, the modified threshold method is not represented in Figure 8. It is shown in Figure 3, however. A shift of 6.1 dB to the right of the 2AFC is evident. At either extreme of the confidence interval, a shift of 4.5 and 6.5 dB is observed. Stallard and Leslie's prediction (5.4 dB) falls within this interval and appears to be a good estimate.

Stallard and Leslie's assessment of reduced detectability is accurate only to a point, in light of the fact that the SI nature of the situation appears to be of little consequence, as shown by the FP method. Time uncertainty does contribute significantly to the reduction in performance. The effect of this factor is variable, however, and does not in all cases reduce the detectability by one-half. It may reduce the detectability by more than one-half, compensating for the absence of an SI effect. A better estimation of the effects of different quantities of time uncertainty is necessary before accurate predictions can be made. Though Stallard and Leslie do not discuss it, the effect of uncertainty may not just shift the function to the right, but interact with signal intensity or other factors present in the environment. This would result in a changing of the slope of the ROC, as is observed in Figures 3 and 8. If the factors combined in the manner they suggest, the psychometric functions should be parallel, displaced to the left or right in accordance with the procedure used. Such is not the case. More experimentation is needed to quantify the effects of time and signal uncertainty in these tasks, and their interaction with other variables before Stallard and Leslie's hypothesis can be accepted.

11.3 Green's Model

As discussed earlier, the FP and 2AFC methods show similar psychometric functions. This lends good support to Green's model of the human as an energy detector. The addition of a factor, perhaps some time uncertainty, has reduced the effectiveness of the detector, adding a noise source and shifting the function to the right.

Such compatibility is not found with the MTFR method. Green and Sewall improved the similarity of the predicted to the observed by specifying the onset of the signal more clearly. A similar procedure employed here may do the same.

In the computation of d'_{opt} in Table 1, a value of 400 msec was used for "T," the integration time. Stallard and Leslie state the effective integration time of the human ear is 480 msec. Janota uses a value of 500 msec in his computations, but this is said to include response time as well. Green found good fit to his model with integration times up to 300 msec. The value of 400 msec was chosen as a median of these. If 300 msec was used, the FP and MTFR functions would be shifted to the left about 0.62 dB. (Recall that d'_{opt} is a function of the WT product; the larger the "T," the larger the d'_{opt}.) The shape of the functions would not change. Using 300 msec would result in essentially zero difference between the FP and 2AFC tasks. The implications of the results would change somewhat if this lower "T" were used; the FP and 2AFC would be identical and the MTFR would be 2.6 to 4.5 dB from the 2AFC. These shifts are much less than those predicted. More research delineating the integration time of the ear is necessary before the validity of assuming "T" equal to 400 msec can be assessed.

11.4 Measures of Detectability

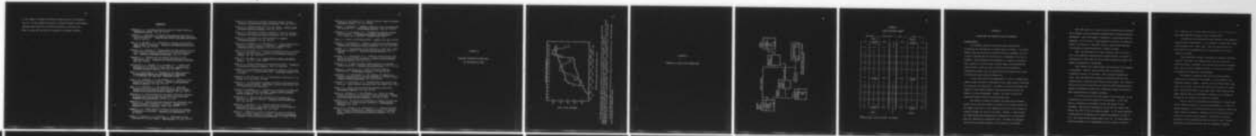
Egan, Greenberg, and Schulman, and Egan and Clarke show that the inverse of the slope of the normalized ROC is the ratio of the variance of the signal-plus-noise distribution to the variance of the noise distribution. The assumption of the d' measure is that this ratio is unity and, hence, that the slope is one. As Table 6 shows, this is not the case. A consistent relationship between d', d_s, and d_{gm} holds, however. As the variance of the signal-plus-noise becomes larger, relative to the noise variance, the slope approaches zero, d_{gm} becomes largest, and d' smallest. As the variance of the noise gets larger than that of the signal-plus-noise, the slope gets larger, d' becomes the biggest of the three measures, and d_{gm} the smallest.

Support is provided for the d_s of Egan et al. They hypothesize that the point at which the ROC crosses the negative diagonal is the pivot of the ROC. As the ratio changes, the ROC rotates about this pivot. The data indicate that d_s fluctuates the least with changes in the ratio. It therefore appears to be the best of the three measures tested.

Examination of Table 6 shows that little difference between the measures is found for slopes of 1.16 to 0.87 (ratios of 0.86 to 1.15).

AD-A053 756 PENNSYLVANIA STATE UNIV UNIVERSITY PARK APPLIED RESE--ETC F/G 6/16
DETECTION OF AN OCTAVE BAND OF NOISE AS A FUNCTION OF STIMULUS --ETC(U)
NOV 77 P T CORNELL N00017-73-C-1418
UNCLASSIFIED ARL/PSU/TM-77-311 NL

2 OF 2
AD
A053756



END
DATE
FILMED
6-78
DDC

In this range, the largest difference between any two of the indices is 0.23. As the confidence intervals in Figure 8 suggest, each measure probably falls within the confidence intervals of the other two. Thus, it seems that the choice of measures is somewhat arbitrary.

REFERENCES

Atkinson, R. C. A variable sensitivity theory of signal detection. Psychological Review, 1963, 70, 91-106.

Auerbach, C. Improvement of frequency discrimination with practice: An attentional model. Organizational Behavior and Human Performance, 1971, 6, 316-335.

Bell, D. W. and Nixon, J. C. Reliability of ratings in an auditory signal-detection experiment. Journal of the Acoustical Society of America, 1971, 49, 435-439.

Binford, J. R. and Loeb, M. Changes within and over repeated sessions in criterion and effective sensitivity in an auditory vigilance task. Journal of Experimental Psychology, 1966, 72, 339-345.

Campbell, R. A. Feedback and noise-signal detection at three performance levels. Journal of the Acoustical Society of America, 1965, 37, 434-438.

Carterette, E. C., Friedman, M. P. and Wyman, M. J. Feedback and psychophysical variables in signal detection. Journal of the Acoustical Society of America, 1966, 39, 1051-1055.

Egan, J. P. and Clarke, F. R. Psychophysics and signal detection. In Experimental Methods and Instrumentation in Psychology. (J. B. Sidowski, Ed.) New York: McGraw-Hill, 1966.

Egan, J. P., Greenberg, G. Z. and Schulman, A. I. Interval of time uncertainty in auditory detection. Journal of the Acoustical Society of America, 1961, 33, 771-778.

Egan, J. P., Schulman, A. I. and Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. Journal of the Acoustical Society of America, 1959, 31, 768-775.

Emmerich, D. S. ROCs obtained with two signal intensities presented in random order, and a comparison between YN and rating ROCs. Perception and Psychophysics, 1968(a), 3, 35-40.

Emmerich, D. S. Receiver-operating characteristics determined under several interaural conditions of listening. Journal of the Acoustical Society of America, 1968(b), 43, 298-307.

Emmerich, D. S. Cueing signals in auditory detection and frequency discrimination experiments. Perception and Psychophysics, 1971, 9, 129-134.

Fidell, S., Pearsons, K. S. and Bennett, R. Prediction of aural detectability of noise signals. Human Factors, 1974, 16, 373-383.

Green, D. M. Detection of multiple component signals in noise. Journal of the Acoustical Society of America, 1958, 30, 904-911.

Green, D. M. Auditory detection of a noise signal. Journal of the Acoustical Society of America, 1960, 32, 121-131.

Green, D. M. Detection of auditory sinusoids of uncertain frequency. Journal of the Acoustical Society of America, 1961, 33, 897-903.

Green, D. M. Consistency of auditory detection judgments. Psychological Review, 1964, 71, 392-407.

Green, D. M., Birdsall, T. G. and Tanner, W. P. Signal detection as a function of signal intensity and duration. Journal of the Acoustical Society of America, 1957, 29, 523-531.

Green, D. M. and Sewall, S. T. Effects of background noise on auditory detection of noise bursts. Journal of the Acoustical Society of America, 1962, 34, 1207-1216.

Green, D. M. and Swets, J. A. Signal Detection Theory and Psychophysics. New York: Wiley, 1966.

Gundy, R. F. Auditory detection of an unspecified signal. Journal of the Acoustical Society of America, 1961, 33, 1008-1012.

Hammerton, M. An investigation into changes in decision criteria and other details of a decision-making task. Psychonomic Science, 1970, 2, 23-204.

Hammerton, M. and Altham, P. M. E. A non-parametric alternative to d'. Nature, 1971, 234 (5330), 487-488.

Janota, C. P. An experimental treatment of auditory discrimination of complex noise-like sounds. Unpublished doctoral dissertation, 1977, The Pennsylvania State University.

Jesteadt, W. and Bilger, R. C. Intensity and frequency discrimination in one- and two-interval paradigms. Journal of the Acoustical Society of America, 1974, 55, 1266-1276.

MacMillan, N. A. Detection and recognition of increments and decrements in auditory intensity. Perception and Psychophysics, 1971, 10, 233-240.

Markowitz, J. and Swets, J. A. Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. Perception and Psychophysics, 1967, 2, 91-100.

McNicol, D. Feedback as a source of information and as a source of noise in absolute judgments of loudness. Journal of Experimental Psychology: Human Perception and Performance, 1975, 104, 175-182.

Pastore, R. E. and Sorkin, R. D. Adaptive auditory signal processing. Psychonomic Science, 1971, 23, 259-260.

Pollack, I. and Hsiah, R. Sampling variability of the area under the ROC curve and of d'. Psychological Bulletin, 1969, 71, 161-173.

Robinson, E. R. and Watson, C. S. Psychophysical methods in modern psychoacoustics. In Foundations of Modern Auditory Theory (J. V. Tobias, Ed.), New York: Academic Press, 1972.

Sakitt, G. Indices of discriminiability. Nature, 1973, 241, 133-134.

Sandusky, A. and Ahumada, A. Contrast in detection with gated noise. Journal of the Acoustical Society of America, 1971, 49, 1790-1794.

Schulman, A. I. Detectability of the deletion of a tone from a tone-plus-noise background. Perception and Psychophysics, 1971, 9, 496-498.

Schulman, A. I. and Mitchell, R. R. Operating characteristics from yes-no and forced-choice procedures. Journal of the Acoustical Society of America, 1966, 40, 473-477.

Shipley, E. F. A signal detection theory analysis of a category judgment experiment. Perception and Psychophysics, 1970, 7, 38-42.

Simpson, A. J. and Fitter, M. J. What is the best index of detectability? Psychological Bulletin, 1973, 80, 481-488.

Snelbecker, G. G. and Fullard, W. The influence of feedback and stimulus characteristics on absolute judgments of auditory stimuli. Journal of General Psychology, 1972, 87, 245-250.

Stallard, J. M. and Leslie, C. B. Psychoacoustics and passive sonar detection. Naval Ordnance Laboratory Tech. Rep. NOLTR 74-27, 1974.

Swets, J. A. Central factors in auditory frequency selectivity. Psychological Bulletin, 1963, 60, 429-440.

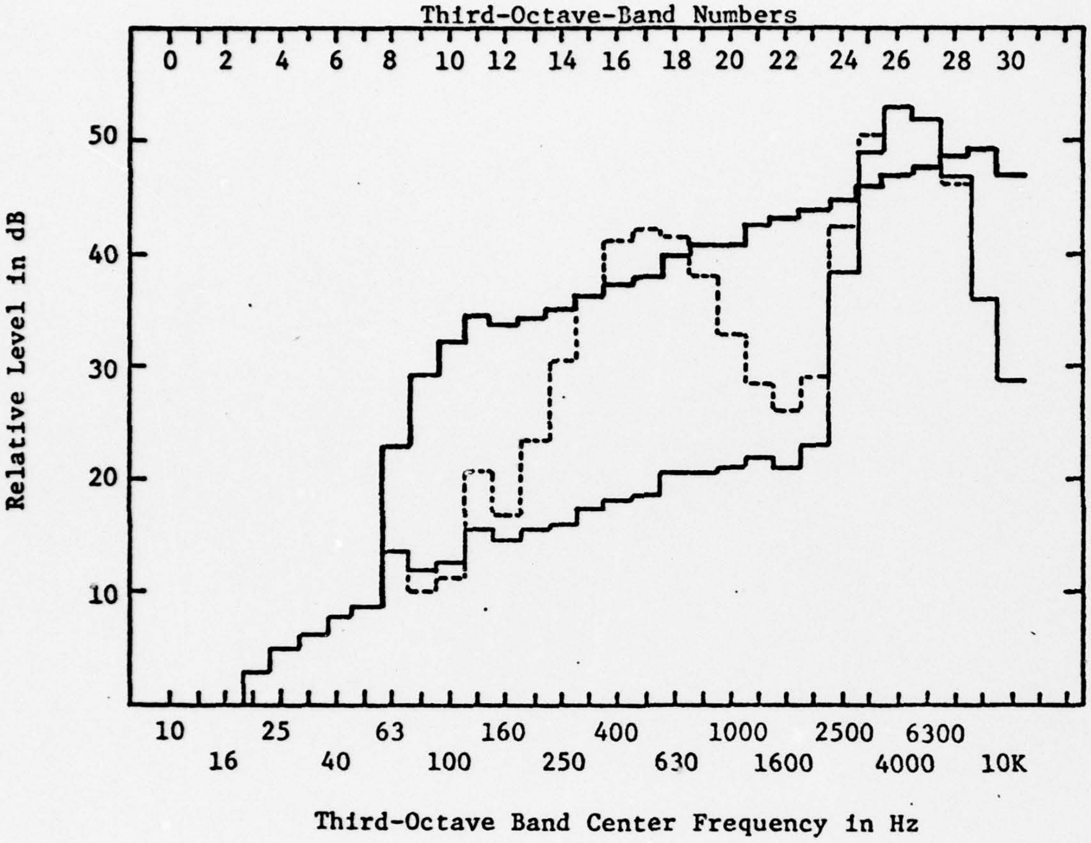
Tucker, A., Williams, P. I. and Jeffress, L. A. Effect of signal duration on detection for gated and for continuous noise. Journal of the Acoustical Society of America, 1968, 44, 813-816.

Wagenaar, W. A. The effect of fluctuations of response criterion and sensitivity in a signal detection experiment. Psychologische Forschung, 1973, 27-37.

Watson, C. S., Rilling, M. W. and Bourbon, W. T. Receiver-operating characteristics determined by a mechanical analog to the rating scale. Journal of the Acoustical Society of America, 1964, 36, 283-288.

APPENDIX A

FREQUENCY SPECTRUM OF SIGNAL PAIR AND CALCULATION OF SNR

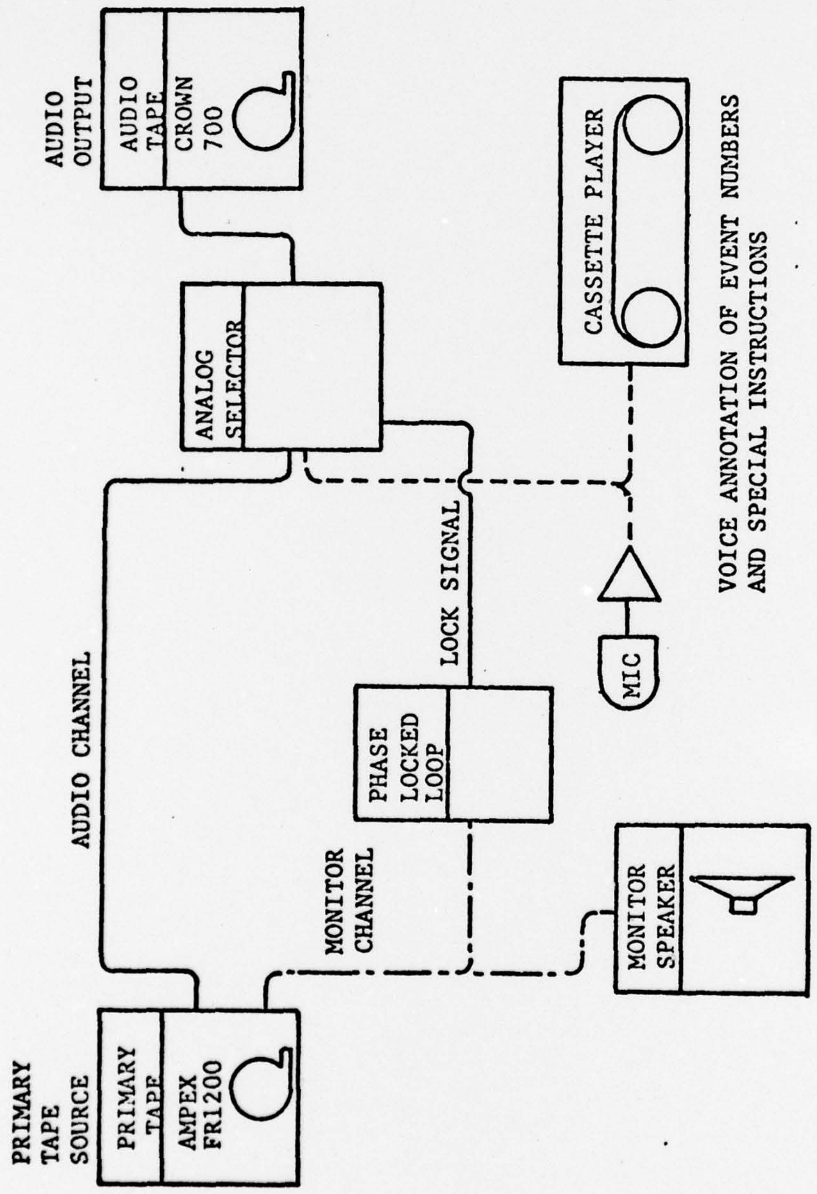


The SNR is calculated by taking the difference, in dB, between the peak level of the signal and the noise in the third-octave-band which contains the center frequency of the feature. Example:
Nondichotomous feature, band number 26, calibrated SNR. (signal level) - (noise level) = 53.0 - 47.5 = 5.5 dB.
Dichotomous feature, band number 17, effective SNR. (signal level) - (noise level) = 42.0 - 38.0 = 4.0 dB.

In the spectrum given above, the dotted line is signal w_7, and the solid line is w_2. The thick solid line represents the background noise.

APPENDIX B

SCHEMATIC OF AUDIO TAPE CONSTRUCTION


PRIMARY TAPE SOURCE
PRIMARY TAPE
AMPEX FR1200
AUDIO CHANNEL
MONITOR CHANNEL
PHASE LOCKED LOOP
LOCK SIGNAL
MONITOR SPEAKER
MIC
ANALOG SELECTOR
CASSETTE PLAYER
VOICE ANNOTATION OF EVENT NUMBERS AND SPECIAL INSTRUCTIONS
AUDIO OUTPUT
AUDIO TAPE
CROWN 700

APPENDIX C

SAMPLE RESPONSE SHEETS*

Tape Number: ______ Subject Number: _____ Date: __________ Method: ________

SIGNAL A				SIGNAL B		
sure	reasonably certain	maybe	EVENT	maybe	reasonably certain	sure
6 to 1	5 to 1	4 to 3		4 to 3	5 to 2	6 to 1
+3	+2	+1	1	-1	-2	-3
+3	+2	+1	2	-1	-2	-3
+3	+2	+1	3	-1	-2	-3
+3	+2	+1	4	-1	-2	-3
+3	+2	+1	5	-1	-2	-3
+3	+2	+1	6	-1	-2	-3
+3	+2	+1	7	-1	-2	-3
+3	+2	+1	8	-1	-2	-3
sure	reasonably certain	maybe		maybe	reasonably certain	sure
+3	+2	+1	9	-1	-2	-3
+3	+2	+1	10	-1	-2	-3
+3	+2	+1	11	-1	-2	-3
+3	+2	+1	12	-1	-2	-3
+3	+2	+1	13	-1	-2	-3
+3	+2	+1	14	-1	-2	-3
+3	+2	+1	15	-1	-2	-3
+3	+2	+1	16	-1	-2	-3
sure	reasonably certain	maybe	EVENT	maybe	reasonably certain	sure
SIGNAL A				SIGNAL B		

---

*Reduced size, actual size 8½ × 11 inches.

APPENDIX D

INSTRUCTIONS FOR TRAINING TAPES AND SESSIONS

Training Tape 1

The research conducted to date has been concerned with determining the detectability of several pairs of signals. The method employed thus far has been the modified threshold technique, in which a subject responds with his decision once a threshold or criterion is exceeded. This threshold can be described as a degree of confidence that a subject has in his choice of which signal is being presented with the noise. Until now, the criterion has been labeled as "reasonably certain," and each subject has been allowed to set in his own mind where this criterion threshold is.

In this next phase of experimentation, I plan to systematically examine several criteria or confidence levels, under several conditions, but using just one signal pair. As before, you will be presented with samples of each signal followed by the presentation of one of the signals in a noise background. However, the means by which the signal-plus-noise stimulus is presented will vary.

One method is termed the modified threshold-forced response method, and as the name implies, is similar to the method used before. The signal starts out at a low signal-to-noise ratio and increases in strength through the course of the event. However, in this method, the signal-plus-noise presentation is terminated at a predetermined SNR, and therefore not terminated by you. You make your response, which will be discussed shortly, at the offset of the stimulus.

The other method to be used is the fixed presentation technique, and is similar to the traditional techniques used in signal detection experiments. After each signal is presented, a noise-alone period is presented. After a period of approximately 15 seconds, the signal is added to the noise at a fixed SNR. After an eight-second interval, the stimulus is then terminated and a response is made.

Thus, the difference in the two techniques comes in the fact that in the modified threshold-forced response method, the signal is incremented to a final SNR, and in the fixed presentation method, the signal is delivered at a single SNR.

This research will further differ from prior work in the method of responding. Earlier, you were instructed to respond when "reasonably certain" of a decision. Your response terminated the event. Now, the signal will terminate automatically and you will respond during a 15-second post-stimulus offset period. Your response will be one of six alternatives, the choice of which depending on the degree of confidence you have that a particular signal was presented. These alternatives are arranged as follows:

A "+3" means you are sure that signal "A" was presented, or the chances are six to one that signal "A" was presented. A "+2" will denote a confidence of reasonably certain that signal "A" was presented. This would be analogous to saying that the chances were five to two that "A" was the presented signal. A "+1" would represent a small amount of confidence, but a decision nonetheless in favor of "A." This would be like saying, maybe it was "A." To attach odds, it would be somewhere in the neighborhood of four to three in favor of

"A." Likewise, for the other three responses, "-3," "-2," and "-1," only these represent ratings in favor of signal "B."

The odds that are attached to these confidence levels imply an interval scale of measurement. It is ludicrous to assume that anyone could establish such a rigid scale. They are used here only to facilitate your conceptualization of the placement of one value relative to another.

Using confidence ratings is analogous to using six different criteria. The higher the confidence that one has in favor of a particular signal, the higher the criterion that has been surpassed. Thus, when I speak of confidence levels, or ratings, and criteria, I'm speaking of roughly the same phenomena.

The degree of confidence that will be attained depends primarily on the strength of the signal. It also depends on the individual himself, however. Factors, such as internal noise, vary from trial to trial, and thus will affect the confidence one might have. Small fluctuations in the signal and/or noise distributions might likewise affect an individual's confidence, even though the strength of the signal on two trials is identical.

This is one of the purposes of my experiment, to examine what factors affect confidence or criterion establishment. In order to achieve this goal, it is necessary that you, as subjects, utilize the entire number of responses. Therefore, try to avoid being overly cautious or careless in your ratings. The merits of a rating procedure are only realized when the full scale is used. Another difficulty in ratings can arise when the established criteria fluctuate. The more stable the criteria, the better. With practice, it is hoped that you will be able to achieve some stability with these confidence levels and also make full use of the scale. You are encouraged to record your strategies so as to aid you in this endeavor. Also, I will give you feedback during training which will proceed for the next few weeks. This will occur via my examining your responses, and then either talking them over with you or writing them down for you.

From the attained confidence ratings, receiver operating characteristic curves will be generated. From these curves, detectability measures of the signal pair will be attained for each method of stimulus presentation. The two techniques will then be compared to determine the differnces in detectability, if any, that one precedure yeilds relative to the other.

Each experimental session will consist of 24 events in the modified threshold-forced response method, or 30 events in the fixed presentation method. Each session will be approximately 30 to 35 minutes long and you are requested to listen to no more than two sessions in any one day. It would be appreciated if you would attempt to listen to five sessions a week. Since extra time is required in set-up, etc., you will be paid three hours work per five sessions that you participate in. As before, you will be assigned tapes to listen to and an order in which to listen to them. No cassettes will be necessary, as you will record all your responses on a response sheet which will be provided. (In the earlier research subjects participated in, cassette tapes were used in recording their responses with the modified threshold procedure.)

As stated earlier, it is requested that you record all your comments regarding strategies, distractions, and/or opinions on the reverse side of the response sheets. Such information will be helpful in drawing inferences from the results. It would also aid in developing further study.

The beginning of each session will consist of a brief summary of the instructions and confidence levels, and a statement as to which method of signal presentation is to be used. The first event will contain two exposures of each signal prior to the event itself. All subsequent events will contain just one exposure of each signal. As would be expected, the signals are presented in a random order, with the a priori probability of either signal being 0.50.

The following are some practice trials to acquaint you with the methods of signal presentation and the rating technique. First will be five trials of the modified threshold-forced response method. Please mark off the appropriate circle on the response sheet.

. . . .

(Five trials with the MTFR are given.)

. . . .

Now, five trials using the fixed presentation technique will be presented.

. . . .

(Five trials with the FP are given.)

. . . .

This concludes the first training tape. Please record your comments, if any, on the reverse side of the response sheet. If there are any particular questions regarding any of the procedures, write them down and I will get back to you. If you understand everything, which would be a pleasant surprise for me, then proceed to the next training tape at your earliest convenience.

Training Tape 2

As stated earlier, one of the purposes of this experiment is to discover if any differences in detectability exist when signals are presented via the two methods. The performance measure will be determined by the percent correct and the resultant confidence ratings.

Of additional interest in this study is the effects of different signal-to-noise ratios on obtained confidence. Obviously, one would expect that these would be directly proportional, with increases in the SNR increasing confidence.

To test this hypothesis, three different levels of SNR will be used. These values, though somewhat arbitrary, are based on the results obtained from the work you did last fall. (All the subjects participated in work with the modified threshold technique.) Since these levels differ in a systematic way, it should be possible to make accurate conclusions regarding the effect of the final SNR on the ratings.

Since the methods of stimulus presentation differ, the SNR variable will be incorporated differently in each method. In the modified threshold-forced response, the signal will start out at a low SNR and then increase to the final SNR. It will remain at this final level for eight seconds before the stimulus is terminated. With the fixed presentation method, the final SNR is the only SNR at which the signal is presented. It also is presented for an interval of eight seconds.

In each session, only one SNR level will be sampled with only one method of presentation. Since there are three SNR levels and two methods, there is a total of six different kinds of tapes that you will listen to. (This had to be changed when it became evident that the subjects were performing at near perfect levels.)

These training tapes will differ from subsequent tapes in that each SNR level will be presented. In this tape, for instance, there are three groups of nine events, one group per final SNR, presented via the fixed presentation procedure. In Training Tape 3, there will be three groups of eight events presented via the modified threshold-forced response technique.

We will now begin with the events of Training Tape 2. Please make sure you have the proper response sheet and that all information is filled out at the top of the page.

Training Tape 3

Welcome to Training Tape 3. You'll be happy to know that I do not have any additional instructions for you. By this time, I hope that you are becoming fairly familiar with the confidence rating technique and are finding it easy to use. Let me reiterate that if you have any questions, do not hesitate to ask.

As with the other training tapes, you should have a special response sheet entitled "Training Tape 3." Clever, huh?

You will be presented with three groups of eight events, each group at a different SNR. The modified threshold-forced response will be used with each event.

Training Tape 4

Contrary to what was said in an earlier tape, there are going to be four SNR levels at which signals will be presented. Attribute the incorrect information to lack of organization, lack of time, or lack of intelligence.

Just in case you have not noticed, the signals are not labelled the same from one tape to the next. With some tapes, a particular signal will be "A" and, on others, it will be "B." On each specific tape, however, signal "A" will remain signal "A" for all the events.

Other than that, there is nothing new to say. This tape will contain four groups of six events, one group per final SNR. They will be presented via the modified threshold-forced response method.

Training Tape 5

This is the last of the training tapes that you will listen to. As in Tape 4, it will contain four groups of events, each group being a different SNR level. There will be a total of 28 events, and they will be presented with the fixed presentation technique.

From here on out, each one of you will have a different schedule of tapes to listen to. In an attempt to control for any sequential effects, all of your schedules will be randomly determined. Therefore, it is important that you fill out the information at the top of the response sheet before each session.

By the way, I will put the response sheets in a folder in the drawer. When you come in, just grab one. When done, place them in your folder. Also be sure to sign in on the Master (a sign-up sheet located outside the audiometric booth). This allows me to check what data are and are not available.

To insure that you are getting paid for the proper amount, record the amount of time plus about five minutes for set-up on the time sheets.

I trust that by now, if you have had any questions, we have talked about them. If not, then see me.

Training Tape 6

The results from the first five training tapes indicate that you all are doing quite well; much better, in fact, than during Fall term. This is evident in that very few, if any, errors have been made. If any of you are familiar with ROC analysis, you know it is impossible to construct these types of curves unless errors are made. In light of this, therefore, it has become necessary to create new tapes which present the signals at lower SNRs. There will be no procedural change, however, and you are to continue to use the ratings as they have been set up.

Even though some of the sessions will be difficult, please attempt to use the full scale during each session. The confidence ratings, or criteria, are session relevant. Thus, the confidence you have should be weighted in terms of the level at which the signals are presented.

For instance, in a session where the SNR is low, a confidence of "+2" would not be the same as a confidence of "+2" with a high SNR event. The confidences are not arranged on some absolute psychological continuum, but are arranged differently with each condition.

All I'm trying to say is: attempt to use all the values during each session. If this still is not clear, which it probably is not, come see me. (No one came to see me concerning this.)

By the way, I am pleased with the results so far, even if it does mean some extra work. Some interesting hypotheses are suggested to account for the discrepancies between the earlier and present research.

This tape will consist of 32 events, or four groups of eight. The first group will be presented with the fixed presentation method, the second two by the modified threshold-forced response method, and the last by the fixed presentation again. Two different SNR levels will be sampled with each technique of stimulus presentation.

One last comment, if you can think of any reason why last fall's and this winter's results are so different, I'd like to hear about it. I will be seeing each one of you shortly anyway, so maybe we can discuss it then.

Instructions for Sessions with the FP Technique

The following events will be presented via the fixed presentation technique. After an exposure set of each signal, a noise-alone stimulus will be presented. After an interval of approximately 15 seconds, one of the signals will be presented at a fixed SNR for an interval of eight seconds, after which the entire stimulus will be gated. During the period which follows, you are to indicate the degree of confidence you have in your decision regarding which signal was presented in the noise. Do this by circling the appropriate value on the response sheet. A definition of each confidence rating is given on the response sheet.

The session will begin with a double exposure of each signal presented in the order "A"-"B"-"A"-"B." These signals may not be in the same order as on previous tapes. Following this is the first event. All subsequent events are preceded by a single exposure of each signal in the order "A," then "B."

There is a total of 30 events on this tape. The signal presented in the noise is determined randomly. Both signals have equal probability of occurrence on every event.

Please record any comments regarding strategies, complaints, fatigue, etc., on the back of the response sheet.

The session will now begin with the fixed presentation method.

Instructions for Sessions with the MTFR Technique

The following events will be presented via the modified threshold-forced response technique. With this method, the signal is initially presented with a noise background at a very low SNR. This initial value will vary from event to event. As the event proceeds, the signal strength is incremented to a final SNR. It remains at this final value for eight seconds, at which time the entire stimulus is gated. During the silent interval which follows, you are to indicate the amount of confidence you have regarding which signal was presented with the noise. Do this by circling the appropriate value on the response sheet. A definition of each confidence rating is provided at the top of the response sheet.

The session will begin with a double exposure of each signal presented in the order "A"-"B"-"A"-"B." These signals may not be in the same order as on previous tapes. Following this is the first event. All subsequent events are preceded by a single exposure of each signal in the order "A," then "B."

There is a total of 24 events on this tape. The signal presented in the noise is determined randomly. Both signals have equal probability of occurrence on every event.

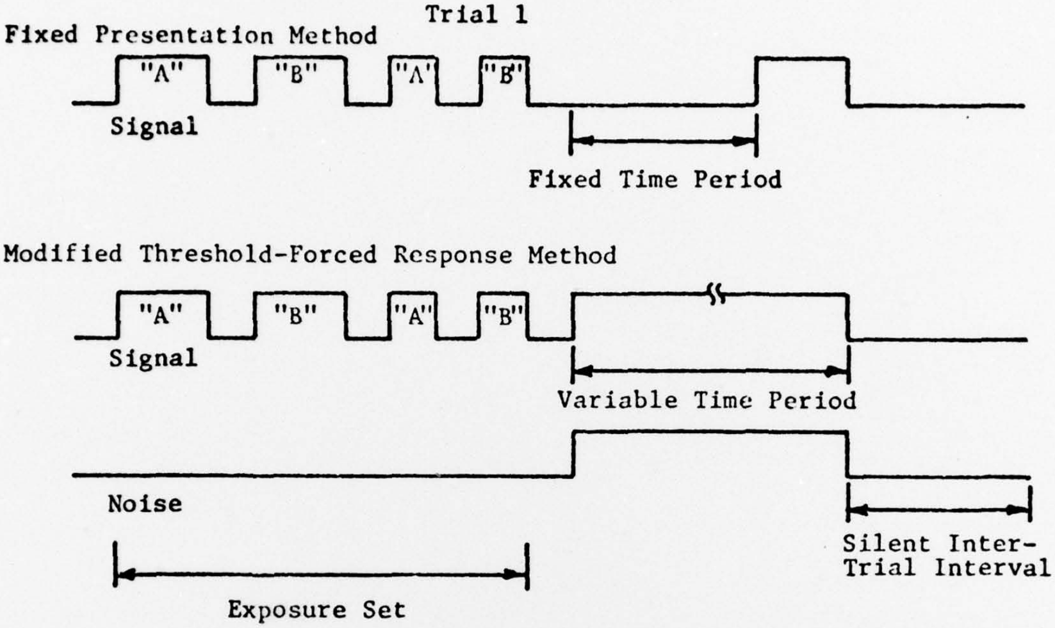
Please record any comments regarding strategies, complaints, fatigue, etc., on the back of the response sheet.

The session will now begin with the modified threshold-forced response method.

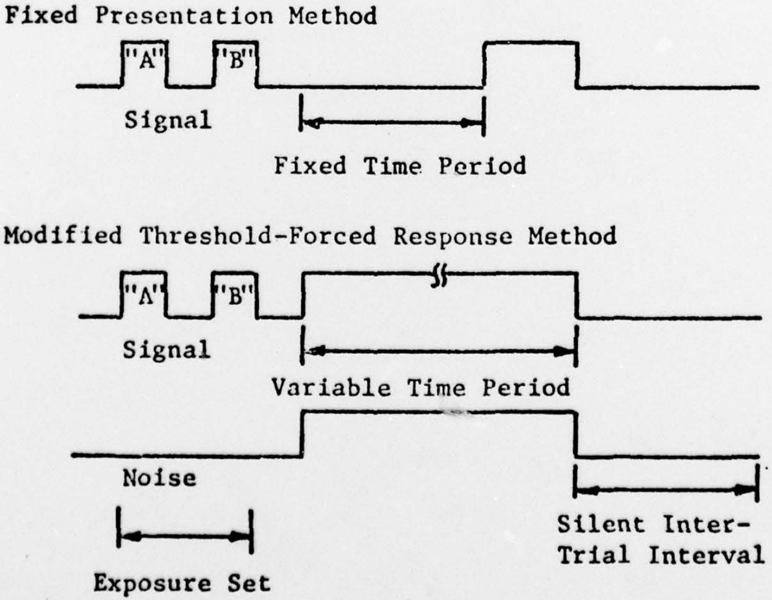
APPENDIX E

SCHEMATIC OF SESSION PROCEDURES

Trial 1



Remaining Trials



APPENDIX F

BETWEEN- AND WITHIN-SUBJECT CORRELATIONS OF RESPONSES TO IDENTICAL TAPES

FIXED PRESENTATION METHOD

Tape 170, SNR -0.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.728	0.568	0.596	0.458	0.545	0.495	0.298
1.71	2		1.000	0.569	0.606	0.561	0.711	0.386	0.178
1.72	1			1.000	0.944	0.491	0.370	0.255	0.017
1.72	2				1.000	0.561	0.418	0.315	0.005
1.73	1					1.000	0.529	0.319	0.117
1.73	2						1.000	0.415	0.098
1.77	1							1.000	0.140
1.77	2								1.000

Tape 120, SNR 0.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.936	0.496	0.730	0.820	0.861	0.788	0.526
1.71	2		1.000	0.424	0.672	0.798	0.822	0.701	0.501
1.72	1			1.000	0.569	0.562	0.484	0.594	0.928
1.72	2				1.000	0.641	0.655	0.519	0.587
1.73	1					1.000	0.741	0.704	0.641
1.73	2						1.000	0.690	0.535
1.77	1							1.000	0.627
1.77	2								1.000

FIXED PRESENTATION METHOD

Tape 190, SNR 2.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.978	0.831	0.862	0.978	0.953	0.932	0.913
1.71	2		1.000	0.830	0.897	0.977	0.972	0.934	0.886
1.72	1			1.000	0.888	0.819	0.781	0.721	0.714
1.72	2				1.000	0.858	0.864	0.812	0.755
1.73	1					1.000	0.946	0.924	0.907
1.73	2						1.000	0.886	0.867
1.77	1							1.000	0.935
1.77	2								1.000

Tape 160, SNR 3.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	1.000	0.988	0.988	0.924	1.000	0.923	0.840
1.71	2		1.000	0.988	0.988	0.924	1.000	0.923	0.840
1.72	1			1.000	0.973	0.909	0.988	0.905	0.824
1.72	2				1.000	0.940	0.988	0.905	0.824
1.73	1					1.000	0.924	0.855	0.728
1.73	2						1.000	0.923	0.840
1.77	1							1.000	0.761
1.77	2								1.000

FIXED PRESENTATION METHOD

Tape 90, SNR 4.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	1.000	0.842	1.000	1.000	1.000	0.985	0.990
1.71	2		1.000	0.842	1.000	1.000	1.000	0.985	0.990
1.72	1			1.000	0.842	0.842	0.842	0.833	0.864
1.72	2				1.000	1.000	1.000	0.985	0.990
1.73	1					1.000	1.000	0.985	0.990
1.73	2						1.000	0.985	0.990
1.77	1							1.000	0.968
1.77	2								1.000

Tape 145, SNR 4.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	1.000	0.983	0.990	0.990	0.986	0.987	0.970
1.71	2		1.000	0.983	0.990	0.990	0.986	0.987	0.969
1.72	1			1.000	0.973	0.973	0.987	0.966	0.947
1.72	2				1.000	0.979	0.975	0.975	0.969
1.73	1					1.000	0.975	0.975	0.956
1.73	2						1.000	0.969	0.948
1.77	1							1.000	0.963
1.77	2								1.000

FIXED PRESENTATION METHOD

Tape 115, SNR 6.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	1.000	0.959	0.930	0.866	0.990	0.937	0.951
1.71	2		1.000	0.959	0.930	0.866	0.990	0.937	0.951
1.72	1			1.000	0.895	0.818	0.948	0.922	0.907
1.72	2				1.000	0.806	0.915	0.861	0.904
1.73	1					1.000	0.854	0.771	0.792
1.73	2						1.000	0.921	0.947
1.77	1							1.000	0.931
1.77	2								1.000

MODIFIED THRESHOLD-FORCED RESPONSE METHOD

Tape 305, SNR 0.5

Subject	Order	1.71 1	1.71 2	1.72 1	1.72 2	1.73 1	1.73 2	1.77 1	1.77 2
1.71	1	1.000	-0.272	0.339	-0.153	-0.159	-0.139	-0.311	0.180
1.71	2		1.000	0.301	0.158	0.121	0.193	-0.296	0.317
1.72	1			1.000	-0.070	-0.213	-0.10	0.101	0.784
1.72	2				1.000	0.247	0.858	0.221	0.118
1.73	1					1.000	0.168	0.090	-0.218
1.73	2						1.000	0.109	0.101
1.77	1							1.000	0.130
1.77	2								1.000

Tape 285, SNR 1.0

Subject	Order	1.71 1	1.71 2	1.72 1	1.72 2	1.73 1	1.73 2	1.77 1	1.77 2
1.71	1	1.000	-0.055	0.007	0.204	0.115	0.002	0.046	0.210
1.71	2		1.000	0.163	0.111	-0.061	0.299	0.025	0.100
1.72	1			1.000	0.068	-0.104	-0.106	-0.132	0.407
1.72	2				1.000	0.058	0.236	-0.125	0.335
1.73	1					1.000	0.568	0.303	-0.018
1.73	2						1.000	0.364	0.118
1.77	1							1.000	0.048
1.77	2								1.000

MODIFIED THRESHOLD-FORCED RESPONSE METHOD

Tape 220, SNR 2.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.339	0.341	-0.031	0.175	0.389	0.299	-0.042
1.71	2		1.000	0.315	0.076	0.332	0.300	0.017	0.144
1.72	1			1.000	0.201	0.392	0.096	0.184	-0.122
1.72	2				1.000	-0.004	0.128	-0.121	0.279
1.73	1					1.000	0.418	-0.046	-0.069
1.73	2						1.000	0.089	0.094
1.77	1							1.000	-0.015
1.77	2								1.000

Tape 205, SNR 2.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.118	0.805	-0.093	0.396	0.185	-0.066	0.219
1.71	2		1.000	-0.101	0.121	0.181	0.231	0.038	0.233
1.72	1			1.000	-0.036	0.382	0.167	0.148	0.195
1.72	2				1.000	-0.127	0.074	-0.079	-0.179
1.73	1					1.000	0.324	-0.203	-0.051
1.73	2						1.000	-0.142	-0.051
1.77	1							1.000	-0.113
1.77	2								1.000

MODIFIED THRESHOLD-FORCED RESPONSE METHOD

Tape 295, SNR 3.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.705	0.261	0.578	0.494	0.583	0.482	0.320
1.71	2		1.000	0.481	0.588	0.534	0.560	0.605	0.351
1.72	1			1.000	0.289	-0.022	0.283	0.290	0.464
1.72	2				1.000	0.568	0.395	0.505	0.297
1.73	1					1.000	0.527	0.682	0.211
1.73	2						1.000	0.518	0.449
1.77	1							1.000	0.316
1.77	2								1.000

Tape 270, SNR 3.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.635	0.274	0.427	0.106	0.357	0.278	0.275
1.71	2		1.000	0.446	0.038	0.112	0.463	0.232	0.297
1.72	1			1.000	0.053	0.271	0.202	0.160	0.202
1.72	2				1.000	0.211	0.051	0.088	-0.099
1.73	1					1.000	0.150	0.343	0.172
1.73	2						1.000	0.472	-0.060
1.77	1							1.000	0.127
1.77	2								1.000

MODIFIED THRESHOLD-FORCED RESPONSE METHOD

Tape 240, SNR 6.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.597	0.375	0.464	0.702	0.737	0.262	0.601
1.71	2		1.000	0.651	0.606	0.516	0.581	0.086	0.481
1.72	1			1.000	0.479	0.411	0.551	0.103	0.411
1.72	2				1.000	0.241	0.445	0.174	0.473
1.73	1					1.000	0.660	0.200	0.746
1.73	2						1.000	0.503	0.723
1.77	1							1.000	0.324
1.77	2								1.000

Tape 235, SNR 6.5

Subject		1.71	1.71	1.72	1.72	1.73	1.73	1.77	1.77
	Order	1	2	1	2	1	2	1	2
1.71	1	1.000	0.212	0.391	0.389	0.304	0.369	0.114	0.377
1.71	2		1.000	0.579	0.586	0.476	0.361	0.412	0.546
1.72	1			1.000	0.485	0.279	0.491	0.201	0.602
1.72	2				1.000	0.281	0.399	0.430	0.436
1.73	1					1.000	0.493	0.180	0.457
1.73	2						1.000	0.098	0.669
1.77	1							1.000	0.284
1.77	2								1.000

DISTRIBUTION

Commander (NSEA 09G32)
Naval Sea Systems Command
Department of the Navy
Washington, D. C. 20362 Copies 1 and 2

Commander (NSEA 0342)
Naval Sea Systems Command
Department of the Navy
Washington, D. C. 20362 Copies 3 and 4

Defense Documentation Center
5010 Duke Street
Cameron Station
Alexandria, VA 22314 Copies 5 through 16