THE UNIVERSITY OF MICHIGAN
ENGINEERING RESEARCH INSTITUTE
ANN ARBOR

DETECTION OF COMPLEX AUDITORY SIGNALS IN NOISE, AND THE CRITICAL BAND CONCEPT

Technical Report No. 82

Electronic Defense Group
Department of Electrical Engineering

By: David M. Green

Approved by: A. B. Macnee

W. P. Tanner, Jr.

AFCRC-TR-58-51
ASTIA Document
No. AD 152 566

Contract No. AF19(604)-2277

Operational Applications Laboratory
Air Force Cambridge Research Center
Air Research and Development Command

April 1958

ACKNOWLEDGEMENTS

I would like to express my appreciation to the people who aided me in preparing this thesis. The aid of the group under the direction of Dr. J. Karlin at the Bell Telephone Laboratories enabled me to conduct part of the second experiment. In particular, Miss N. Bowles provided very able assistance with this experiment. Many persons at the Electronic Defense Group provided assistance. Mr. T. G. Birdsall patiently tutored many of my first attempts to express my thoughts mathematically.

The members of my committee were both helpfully critical and understanding. Dr. G. Peterson and Dr. J. C. R. Licklider patiently tried to help me express my thoughts more clearly and easily. For the many virtues required of a thesis chairman let me thank my chairman, Dr. J. D. Birch.

Finally, let me thank Mr. W. P. Tanner, Jr., who first encouraged my interest in auditory theory, and who has continued to encourage and stimulate much of the work reported in this investigation.

The National Science Foundation provided two pre-doctoral fellowships which made much of this work possible. This work was sponsored by the Operational Applications Laboratory, Air Force Cambridge Research Center, and the U. S. Army Signal Corps.

TABLE OF CONTENTS

LIST OF FIGURES

ABSTRACT

In the last two decades, considerable information has become available concerning the mechanism of auditory frequency analysis as it is inferred from behavioral data. In the first section of this thesis a review of many of these papers is presented. These include experiments from the area of masking, frequency discrimination, and loudness. All of the studies are analyzed in terms of the critical band concept proposed by Fletcher. While many of the results are consistent with the deductions obtained from this simple concept, certain relationships still remain obscure. For example, the estimates of critical bandwidth obtained from masking data differ by as much as an order of magnitude from the estimates obtained from frequency discrimination data.

Two experiments concerning the detection of complex auditory signals in noise are presented. The first experiment employed as a complex signal two sinusoidal stimuli. Both the duration and the frequency separation of the signal serve as variable parameters in the experiment. It is concluded that the complex signal is more detectable than either of the two stimuli which comprise the complex. A mathematical model, which is a simple extension of the critical band concept, is presented. The model is based on the assumption that the outputs of several critical bands may be combined in detecting these stimuli. Two other models, which do not predict the data as well as the first model, are compared with the experimental data.

The second experiment employed bandlimited white Gaussian noise as the signal to be detected. A statistical model which incorporates the extension of the critical band concept as it was used in the first experiment is presented. This new model provides excellent predictions for the results

obtained in the second experiment. Both bandwidth and signal duration of these noise signals may be accounted for by this statistical model. The basic equation for this model is:

$$d'_{opt} = \sqrt{W T}\, \frac{S_o}{N_o}$$

where d'_{opt} is a measure of the detectability of the signal, W is the bandwidth of the noise signal, T is the signal duration, and S_o/N_o is the signal power to noise power ratio in a one cps band.

As a result of this experimental work, a general model of the receiving mechanism is proposed stating that the ear may be likened to a series of bandpass filters. The output of these filters may be linearly combined, with weighting constants, so that an effectively larger bandpass can be obtained. This adjustable bandpass model is contrasted with the fixed critical band concept which Fletcher suggested.

CHAPTER I

INTRODUCTION

An attempt will be made to use behavioral data as a basis for a mechanistic and mathematical description of the auditory frequency analysis process. The critical band theory, as proposed by Fletcher, is used as an interpretive device in discussing the research in this area. While many of the results obtained from such research are consistent and can be integrated, still several areas need further clarification before complete unification is possible. Because of this lack of complete integration, no speculations as to the physiological bases of the frequency analysis process will be attempted.

A review of the psychological data pertaining to auditory frequency analysis will provide some definite and quantitative information on how the mechanism of attention operates in at least this modality. Because attention is usually considered some process or system within the perceptual act, its description should be in terms of the mechanisms or equations for the process. Unfortunately most existing descriptions of this perceptual process are vague, general, and almost entirely verbal. One of the common analogies used to describe the process of attention is that of some sort of filtering process. This filtering process somehow allows for the selection of certain informational aspects of the stimulus and for the rejection of others. Such a description is very similar to a theory often used in the area of hearing to explain many of the results obtained from psychophysical experiments. In this area the concept of a filter is used in the strict sense to refer to a mathematical operation with respect to the variable, frequency.

The thesis may be divided into three main sections. The first section (Chapter II) contains a review of the research concerned with the problem of auditory frequency analysis. This review is constructed from papers drawn from four general areas of research on hearing. The first subsection is a summary of masking experiments, i.e., experiments where a stimulus is to be detected in the presence of some sort of auditory interference, usually white noise. The second subsection includes data concerned with the ability to detect changes in the frequency of a sine wave stimulus. In the last two subsections some papers are briefly summarized which have to do with certain data obtained in loudness experiments and data from masking experiments where the signals are unknown in frequency. All of the research is viewed in terms of a single simple model, the critical band theory.

In the second section (Chapter III) of the thesis two new experimental studies are reported in some detail. The first of these has to do with the detection of multiple component signals in noise. A model extending the critical band theory, as well as two other models, is considered in analyzing the results of this first experiment. The second study deals with an extension of the ideas developed in the first experiment by applying these ideas to a situation where the signal is a sample of noise, instead of a sinusoid. An equation developed from theory is suggested for these noise signals. Two parameters of the noise signal are investigated: (1) the bandwidth of the signal; and (2) the duration of the signal.

The third section (Chapter IV) discusses what the results of these two experiments imply about the auditory frequency analysis mechanism.

CHAPTER II

REVIEW OF STUDIES CONCERNED WITH THE

AUDITORY FREQUENCY ANALYSIS MECHANISM

2.1 Introduction: The Critical Band Concept

In 1940 Fletcher (1) introduced the critical band concept. According to this hypothesis certain masking data may be explained by assuming that the hearing mechanism acts as a narrow band filter when detecting sine wave signals in noise. Data from such diverse areas as masking experiments, frequency discrimination experiments, and even cochlear anatomy have been analysed according to this theory.

The assumption of some sort of filtering or resonance process in the hearing mechanism is hardly new. Ohm's acoustic law is consistent with such an assumption. Helmholtz speculated about the physiological and anatomical structures which might provide a means for the analysis. For many years after Helmholtz the means or method whereby the ear accomplished its frequency analysis occupies a sizeable part of auditory theory. Theorists who advanced rival views of how the frequency analysis was accomplished marshalled physiological and anatomical data to support their positions, while psychophysical data were not used as crucial evidence. Frequency discrimination data, for example, were used only to indicate a lower bound on the number of resonators. Helmholtz (8) could not find as many "resonators" as discriminable pitches. He, therefore, claimed that if two adjacent "resonators" were activated the perceived pitch would fall somewhere between the pitches of the two active resonators. Wever (28) points out that had Helmholtz identified the resonators with the basilar membrane fibers instead of the rods of Corti, as he later did, this difficulty would not have arisen.

Masking experiments are largely responsible for an increasing concern regarding the way in which the auditory filters affect psychophysical data. Wegel and Lane (27) were the first to demonstrate clearly that the amount of masking one tone produces on another is related to the frequency separation of the tones. More than a decade passed before another masking experiment, this time using random noise as the interfering stimulus, led Fletcher to postulate the critical band concept. This concept specified in some detail equations for this frequency analysis, or filtering process, as it operates in hearing.

Since the following sections will use the filter concept as an interpretive device it seems wise at this point to define the concept. Probably the most general manner of defining a filter is to consider it as a mathematical operator. Given any linear system or mechanism with which one can define an input and an output, it is possible to define a filter characteristic of the system. Let the input to the system be of the form

$$X = A \sin 2\pi ft$$

where f is the frequency of the input and t is time. Then the output, after steady state conditions have been obtained, can be described in the form

$$Y = G(f)A \sin \left[2\pi ft - T(f)\right] .$$

G(f) and T(f) will be defined as the filter characteristics of the device. G(f) is the gain characteristic of the filter. T(f) is, of course, the phase characteristic of the device. For the most part, phase effects are not of crucial importance in the studies summarized in the review; hence, the function T(f) will be ignored in most discussions. The function G(f) is the important transformation which characterizes the filter. Two types of filter characteristics are referred to frequently in the review which follows. One is the rectangular filter. For such a filter;

$$G(f) = 0 \qquad |f - f_o| > \omega$$
$$G(f) = \alpha \qquad |f - f_o| < \omega \ ,$$

where α and ω are constants. This characteristic is impossible to build in an analog manner, although good approximations are possible. The second filter characteristic is the single tuned or universal resonance filter. This filter is extremely common in simple electrical or mechanical systems and it is described in detail by Terman (25). For this filter G(f) has a single maximum at f_o, called the center frequency or peak, and is monotonic decreasing in either direction from f_o. This definition of a filter is adequate for the purposes of this paper. While this steady state definition is simple, it is hardly the most powerful means of considering the concept. Many important aspects of the problem, for example, the transient responses of a filter, are ignored. For a more detailed discussion of the concept, E. A. Guillemin's "Introductory Circuit Theory" is recommended.

The concept of a filter will be an extremely useful model in discussing the data which follow. The claim is not made that this model represents the only, or, indeed the best, model to apply to the subject matter. The model is simple, which is a virtue, and it does provide an economical and fairly accurate method of analysis. In any psychophysical experiment the only statement that can be made, from the data *per se*, is that the subject behaves as though something were true. No stronger statement is ever intended.

2.2 Masking Experiments

This section presents a review of the masking experiments interpreted in terms of critical bands. Fletcher's original paper is reviewed, and his formulation of the hypothesis is discussed. Later work has led to a modification of the original hypothesis and this appears to provide a better model for the data.

Fletcher (1) used a random noise source with approximately flat frequency characteristics. He had the observer adjust the amplitude of a sine wave signal so that it was "just detectable" in the presence of the masking noise. Once this determination had been made the noise was attenuated on each side of the sine wave signal by a bandpass filter. Once again the amplitude of the signal necessary to be just detectable was determined. Fletcher, using a variety of bandwidths to filter the masking noise, repeated this procedure several times. He found that as long as the noise bandwidth was wider than some critical value the amplitude necessary to detect the tone was constant. Once the noise bandwidth was smaller than this critical value the amplitude of the tone could be decreased in magnitude and still remain detectable. This basic experiment was repeated using several different frequencies for the signal. Fletcher summarizes the results of the experiment by the following equation,

$$C = \frac{I_m}{WI_f} \qquad (1)$$

where I_f is the intensity per cycle of the noise, I_m is the intensity of a tone of frequency, f, which is just perceptible in noise, and W is the width of the critical band in cps. This equation is applicable as long as the noise bandwidth is large compared with W, or, equivalently, as long as

I_f is a constant in the region of interest. If the noise bandwidth is smaller than W, I_f will be reduced so that if a square bandpass could be used, the intensity of the signal, I_m, should be proportional to the noise bandwidth in order for the tone to remain just detectable.

This statement of the equation was a fair first approximation. The following simple example provides a rationale for the equations. Suppose the auditory filter is rectangular in shape with a certain width, W. Assume further that it is possible to produce masking noise with a rectangular spectrum. Let the width of this noise be α_n. Now, for all conditions where the noise spectrum is wider than W, the tone to be heard must be set at a constant power level, I_m. If the width of the noise is less than W, then the tone may be reduced by the fraction $\frac{\alpha_n}{W}$, so that if the noise is one-half the width of the auditory filter, then I_m may be reduced by that fraction and still be detectable.

Such an explanation of the equation follows the assumption made by Fletcher that the auditory filter has a rectangular shape. Without such an assumption there is no simple rationale for the equation he presents. Such an equation leads one to expect that the data should fall along a function consisting of two straight lines. Figure 1 shows the data Fletcher actually obtained in his experiment. The lines display the two relations that will obtain, depending upon whether the noise bandwidth is wider or more narrow than a critical band. The data certainly do not support the abrupt breaks implied by the lines, but the general trend of the data indicates relations similar to those suggested by Fletcher. Later work (20, 24) has given better approximations to the actual characteristic of the auditory filter, but the rough square bandpass assumption made by Fletcher has served to elucidate the major result.

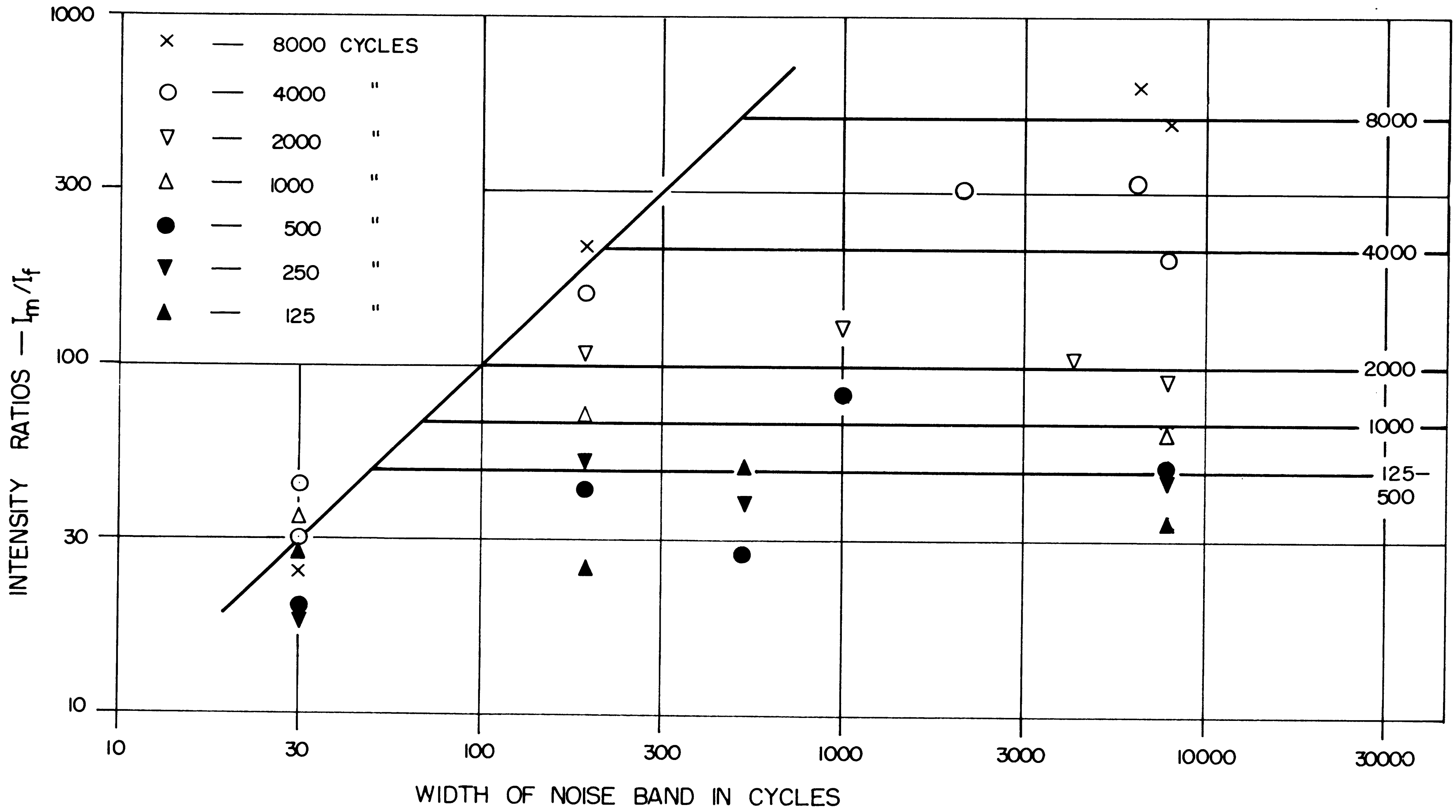


FIG. I THRESHOLD SIGNAL—to—NOISE RATIO *vs.* WIDTH OF NOISE BAND (AFTER FLETCHER)

The points represent the signal power to noise power per unit bandwidth necessary for the tone to be just detectable. The width of the band of noise used to mask the signal is given by the abscissa.

From Fletcher's study it is seen that the bandwidth of the auditory filter is strongly dependent upon the center frequency of the filter. This relationship is extremely important, for it has become the major argument in extending the concept to other areas (1, 13). The bandwidth versus center frequency relation is displayed in Figure 2 along with other data which yield similar functions.

It was not until 1952 that Webster, Miller, Thompson, and Davenport (26) reported data based on the natural inverse of the method employed by Fletcher in his original work. Fletcher demonstrated that reducing the noise power some distance in frequency from the signal does not influence its detectability. Webster, et al., demonstrated that lowering the noise power in the immediate vicinity of the signal increases its detectability. Specifically, they employed masking noise which was filtered so as to produce a gap in the noise spectrum. By presenting pure tones in the region of this gap they were able to demonstrate that the signal energy necessary for detection was substantially reduced in the neighborhood of the gap. The data almost completely supported the contention of Fletcher. The only departures from the data involved the exact form that should be assumed for the auditory filters. Their findings supported the position of Schafer, Gales, Shewmaker, and Thompson (20) who claim the auditory filter is better approximated by a simple tuned filter. This problem will be discussed more fully later.

While the research reported above will be used to define the concept of critical bandwidth, there is at least one other manner of measuring the width of the critical band which Fletcher has used. This second method of determining the critical bandwidth will be called the critical ratio method, following

2659 A-G-17 AB 12-26-57

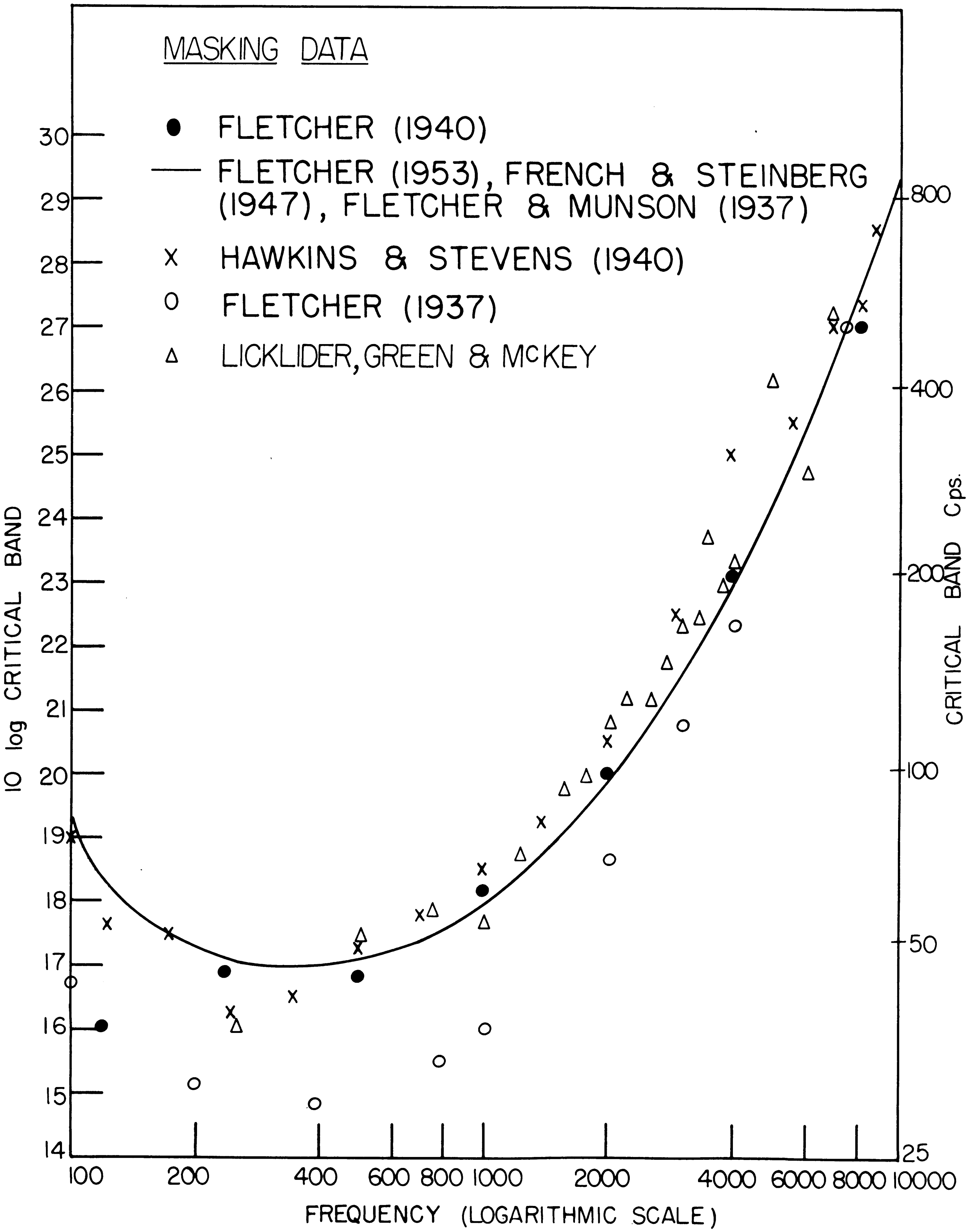


FIG. 2 THE CRITICAL BANDWIDTH AS A FUNCTION OF CENTER FREQUENCY AS INFERRED FROM MASKING DATA

The width of a critical band in cps is given by the ordinate (right side). The center frequency in cps is the abscissa.

Zwicker, Flottorp, and Stevens (29). It is called a critical ratio method since it uses the ratio of Equation 1 above to determine W. This method is most easily explained by rewriting Equation 1 in a slightly different manner. Rewriting the Equation 1 in logarithmic form:

$$\log C = \log I_m - \log I_f - K;\ K = \log W . \qquad (2)$$

Now for values of I_m that are at least an order of magnitude above absolute threshold Fletcher claims C is nearly unity. Hence, by changing the level of the noise and determining the value of I_m, the number K may be determined.

Fletcher presents data which support such a conclusion (2). Once again the bandwidth estimates show dependence upon the frequency of the tone. Figure 2 displays this relation by the solid curve. This solid curve apparently represents an accumulation of data obtained at the Bell Telephone Laboratories, for earlier Fletcher had presented different estimates of bandwidth, which are displayed as circles in Figure 2. While no particulars of the experiment are reported in this reference, the theoretical argument is similar to that presented above.

Hawkins and Stevens (7) have determined the critical bandwidth by the critical ratio method. They explored the masked threshold of a pure tone for a variety of frequency and noise levels. Their noise levels ranged from zero to a spectrum level 60 db above 0.0002 dyne/cm^2. Once the noise reaches a certain level (about 10 db above absolute threshold) the signal power necessary to just hear the tone is roughly proportional to the noise level. This relation is, of course, consistent with Equation 1. For lower levels of noise the data are consistent with a different equation, one which adds some constant to the external noise. Fletcher has suggested this other relation in his discussion of the problem (2, p. 167). The manner in which critical bandwidth and center frequency vary for the higher noise levels is shown in Figure 2.

Licklider, Green, and McKey (14) also determined the masked threshold of a gated sine wave signal in noise about 50 db re 0.0002 dyne/cm^2. The important difference between this and other studies was that in the two previous studies the signal was left on continuously and the observer adjusted its amplitude until it was detectable in the noise. In the study of Licklider, Green, and McKey the signal was turned on for 0.1 seconds and a two-alternative forced-choice method was used to measure detectability. Equation 1 must therefore be corrected to include the variable, signal duration. Just how this will affect the equation is not known.

In general, for finite duration signals, equations which multiply signal power and some function of signal duration best fit the major part of the data (5). Thus, Equation 1 should be corrected by multiplying I_m by some function of signal duration. Since both W and duration act as multiplying constants in Equation (1), no estimates of bandwidth alone can be obtained. However, the form of the data when plotted against center frequency may be compared with other experiments. Assuming duration acts only as a multiplying factor, it is possible to obtain estimates which are directly proportional to the bandwidth. Choosing the constant for signal duration to be 12, the estimates of bandwidth depicted in Figure 2 are obtained.

With the exception of the paper by Webster, et al., the preceding papers have not questioned the assumption that the auditory filter or critical band is rectangular. That is, the function G(f) is a rectangle with a certain height and width which represent the critical band. This assumption was introduced by Fletcher as a first approximation and has had considerable usefulness. Schafer, Gales, Shewmaker, and Thompson (20) as well as Tanner, Swets, and Green (24) have attempted to provide a better approximation to the actual shape of the auditory filter. One of the difficulties involved in measuring the exact form of the auditory filter lies with certain physical limitations imposed on electronic filters. Ideally, according to Fletcher's assumption, he should use a perfectly rectangular filter for the masking noise. While fairly good approximations are possible a very good electronic filter may show an amplitude decline off center frequency of 60 db/octave. If such is the case the response of the filter will change by only an order of magnitude in power after a change of about 150 cycles near 1000 cps. Since the critical band is estimated to be about 60 cycles per second at this frequency the limitation of electronic filtering is obvious.

Schafer, Gales, Shewmaker, and Thompson tried to avoid this difficulty by constructing their masking stimulus from a sum of sinusoids rather than by trying to filter noise. First they determined the spacing of a collection of sine waves in random phase that would produce effects similar to that of noise. They then repeated Fletcher's original experiment, but this time, in order to change the masking stimulus, they simply turned off the outermost frequencies in the masking "noise". The model which appears to fit the data best is one which assumed the auditory filter is of the single tuned filter or universal resonance filter.

They obtained bandwidth estimates at three frequencies; i.e., 200, 800 and 2000 cps. Their estimates of bandwidth (in cps) are 21.5 for 200, 21.6 for 800, and 82.1 for 2000. Some difficulty is involved in comparing these bandwidth estimates with those obtained in Fletcher's original experiment or those using the critical ratio method. The usual method of specifying the bandwidth of the universal type of resonance curve is to use the frequency difference between the half-power points of the filter. The half-power points are those frequencies where $G(f)^2$ is at one-half the maximum value. This measure of bandwidth, call it W, is smaller by a factor of $\pi/2$ than the equivalent rectangular filter width.+ Hence, the equivalent square filter width (in cps) obtained in the study are 33.8 at 200, 34.0 at 800, and 129 at 2000. [In terms of 10 log Δf, 10 log 33.8 = 15.3, 10 log 34 = 15.3, 10 log 129 = 21.1.]

Tanner, Swets, and Green (24) have also attempted to determine the characteristic of the auditory filter. Rather than trying to construct a filter with a rectangular characteristic, or constructing a psuedo noise, they employed the mathematically simple single tuned filter to change the noise. Essentially, they repeated Fletcher's procedure using a single tuned filter to attentuate the noise components on each side of the signal. They then considered several possible shapes for the auditory filter. The problem is treated mathematically as two filters in series; one representing the external single tuned filter, the other representing some form of internal or auditory filter. They attempted to select that auditory filter characteristic which yielded the most consistent estimates of bandwidth for the various conditions of the experiment. The results indicate that the single tuned filter characteristic is the best choice for the auditory

---

+ If the rectangular filter has a characteristic $G^2(f) = 1$ when $|f-f_o| < \frac{W}{2}$, and $G^2(f) = 0$ when $|f-f_o| > \frac{W}{2}$. Let the universal resonance curve be of the form $F(f)$ with center frequency f_o, and $(f_o \pm \frac{\omega}{2})$ the half power points, $(F(f_o) = 1)$, then $\int_{-\infty}^{\infty} G^2(f)\,df = \int_{-\infty}^{\infty} F^2(f)\,df$, when $\frac{\pi}{2}\omega = W$.

filter. Only one center frequency (1000 cps) was used and the auditory bandwidth was estimated to be about 41 cps. This yielded an equivalent square bandwidth of 64 cps (10 log 64 = 18.06). This value is, of course, extremely close to those obtained in the previous experiments (See Figure 2).

This exceptional agreement may be spurious, since Figure 2 contains experiments where monaural observation was used. Tanner, Swets, and Green used binaural observations. French and Steinberg (4) claim there is a difference between those estimates obtained with monaural and binaural conditions. In general, the differences between the estimates of the 10 log critical bandwidth are about 1.5 for the low frequencies with larger differences occurring above 6000 cps. The critical bandwidths are always smaller in the binaural than in the monaural case. This result is cited in the French and Steinberg article (4). They refer to Fletcher and Munson (3) as the source of the data. Fletcher and Munson do not make any claim concerning the difference between monaural and binaural critical bands. They do provide evidence that for absolute threshold measurements there is a difference between monaural and binaural listening. In general, they find binaural listening yields lower thresholds than monaural listening. Munson (17) has stated that French and Steinberg had assumed that it was logical to extrapolate the absolute threshold data to experiments where noise is used. Thus, using the critical ratio assumptions of Equation 1, there would be different functions for the monaural and binaural critical bands. No direct evidence on this point is available. However, indirect, contrary evidence is available since there is a great deal of similarity between the estimates of critical bandwidth in the study of Hawkins and Stevens (7) who used monaural observation, and Fletcher's (1) study where binaural observation was used.

This concludes the review of masking experiments, done in this country, which deal directly with the critical band hypothesis. There have been several recent experiments conducted in Germany which deal with this problem. The original papers are not available to the author, but are summarized in Zwicker, Flottorp, and Stevens (29). Some of the data from these papers will be discussed in the next section.

This section has reviewed two main experimental approaches to the problem of determining the shape and width of the critical band. The first approach involves changing the frequency characteristic of the masking stimulus in some way and determining how the tone must be changed in intensity to remain detectable. The second approach is called the critical ratio method. This method assumes a certain equation e.g., Equation 1 to describe the masked threshold. The data are then used to infer a parameter of the equation, which is the critical bandwidth. The two methods yield similar estimates of the bandwidth and a high degree of consistency in the data from different laboratories is evident. Both methods, and all the experiments reported, involve tasks of detecting sinusoidal signals in some sort of masking noise.

2.3 Frequency Discrimination Experiments

If the concept of a critical band is shown to be useful in an area of auditory research other than masking, then the degree of confidence in the concept will be increased. Both Fletcher (1) and Schafer et al. (20) have suggested such an extension to the area of frequency discrimination The argument runs as follows. If the auditory filters exist, then a change in frequency causes a change in amplitude at the output of the various filters. This change in amplitude pattern may be directly related to the apparent change in pitch. It appears reasonable that the change in amplitude pattern should be related to the size of the bandwidth of the auditory filter. Thus, one way to test these assumptions is to determine how the change in frequency which is just detectable, Δf, changes with frequency f. If more detailed assumptions are made in the preceding argument, for example concerning the shape of the filter, one can claim Δf is proportional to the critical bandwidth. Thus, a plot of critical bandwidth or of Δf against frequency should be similar, except for the constant of proportionality, to the data from masking experiments which related bandwidth to center frequency. This demonstration that both masking data and frequency discrimination data yield similar functions has been used to show the importance of the critical band function (1, 2 and 13).

Fletcher has used such an argument when he likens the critical band to "patches" on the basilar membrane. The change in pitch is assumed to be a constant change along a linear, or in a later interpretation, an areal extent, of the basilar membrane (1, 2). The pitch discrimination data of Shower and Biddulph (21) were used by Fletcher. Shower and Biddulph's data and that of the critical bandwidth versus frequency, is the same except for a constant multiplicative factor. Schafer, Gales, Shewmaker, and Thompson (20) have also presented an argument to link the width of the critical band and the "just-detectable" increment in frequency.

While the theoretical arguments for linking bandwidth and the "just-detectable" change in frequency are somewhat appealing, there is considerable difficulty in determining exactly what pitch discrimination data to select. Unfortunately, the theoretical arguments are not of sufficient precision to dictate how frequency discrimination should be measured. The difficulty is that there are many ways of determining the "just-detectable" change in frequency, for to produce a signal of some frequency one must select some intensity for that frequency. Now it is well known that the size of the "just-detectable" change in frequency is dependent on the intensity of the frequency. Therefore, the question arises as to what intensity to use for each frequency region. Usually each frequency is presented at a constant number of decibels above absolute threshold. Whether or not such a technique is proper depends upon other assumptions made concerning the hearing mechanism. For the most part, these assumptions have not been made explicit.

A second major problem encountered in experiments in frequency discrimination is how the frequency of the stimulus should be changed. Two earlier studies (11, 21) presented a signal of a certain frequency, changed to a second frequency, and then returned to the original frequency. Shower and Biddulph made the change in a nearly sinusodial fashion. Knudsen made the changes abruptly. These techniques correspond to a frequency modulation of the signal. This frequency modulation can be expressed by a Fourier analysis. For the sinusodial modulation the stimulus can be expressed as a series of components; i.e., one component is at a frequency which is the average of the original frequency and the changed frequency, and the others are at this average frequency plus or minus the rate of modulation. The exact appearance of these spectra are, therefore, heavily determined by the rate at which the change occurs as well as by the magnitude of the change. Using some other nonsinusodial method of change will, of course, develop quite different Fourier spectra. Koch (12) has, in fact, claimed that the data could be explained on the basis of the Fourier analysis without appealing to the frequency analysis of the ear+.

In light of the two considerations stated above the use of frequency discrimination data as strong evidence about the critical bands might be viewed somewhat skeptically.

This skepticism will probably be increased as the results of several studies from this area are reviewed. Knudsen's study was the earliest one to try to control the intensity level of the stimuli. He found the absolute threshold for the various frequencies and determined the just detectable change in frequency at a level 40 db above the threshold value. Knudsen abruptly switched the stimulus between two frequencies, using the amount of change in frequency which the observers could reliably detect to measure Δf. Due to the abrupt changes the observers heard clicks with frequencies greater than 3000 cps so that no measurements were taken above this value.

Shower and Biddulph have provided the most complete investigation of the problem in terms of the number of frequencies, intensities, and observers used. They varied the stimulus sinusoidally at a rate of 2 cps. The subjects were again asked to state when they heard two frequencies.

+ Koch's paper is misleading. The opening part of the paper contains a discussion of Heisenberg's uncertainty principle. This principle in one form expresses an order of uncertainty which will be encountered in determining the frequency for finite duration signals. Koch then developes the equations for the Fourier analysis of the frequency modulation used in Shower and Biddulph's experiment. By assuming some arbitrary form of the spectra which Koch claims "would be heard as two separate tones" he shows the data may be explained on this basis. This arbitrary form assumed by Koch is claimed to be the frequency change just detectable by a "perfect receiver". Why this is a perfect receiver in any sense is not explained in the paper. Heisenberg's principle certainly is not used to define this perfect receiver. Thus, while the paper presents some interesting speculation about how the data might be explained, it does not in any sense deny that the frequency analysis ability of the ear is in fact being measured in these experiments.

Harris, (6) in a more recent study using quite different procedures, has provided some extremely interesting data in this area, based on a two-alternative forced-choice technique. The subject listened to pairs of tones and was required to judge whether the second (variable) tone was "higher" or "lower" than the first (standard) tone. The variable stimulus was chosen at random from any of 8-12 frequencies, half of which are above and half of which are below the standard. At a sensation level of 30 db, groups ranging from 40 to 90 subjects were tested at various frequencies. The data from the three experiments are displayed in Figure 3.

The differences in the form of the data obtained by the various techniques are obvious. The abrupt continuous change from variable to standard frequencies used by Knudsen produces the smallest value for Δf for low frequencies. Shower and Biddulph's data in this region yield the larger estimates of Δf. At high frequencies, the studies of Harris and Knudsen yield larger estimates of Δf than do those of Shower and Biddulph. Harris presented the stimulus for 1.4 second duration with an interval of silence of this same duration between the pairs. No satisfactory explanation of why these differences should arise has been suggested. Harris suggests that Shower and Biddulph are probably measuring a different function from that assessed by Harris's technique. Whatever the solution, the use of frequency discrimination data for estimating critical bandwidth is a rather hazardous undertaking.

This section has presented an argument, first used by Fletcher, to link the critical bandwidth and the size of the "just-detectable" increment in frequency, Δf. Theoretical considerations of this suggestion have cast several doubts upon its validity. Moreover, the basic data on frequency

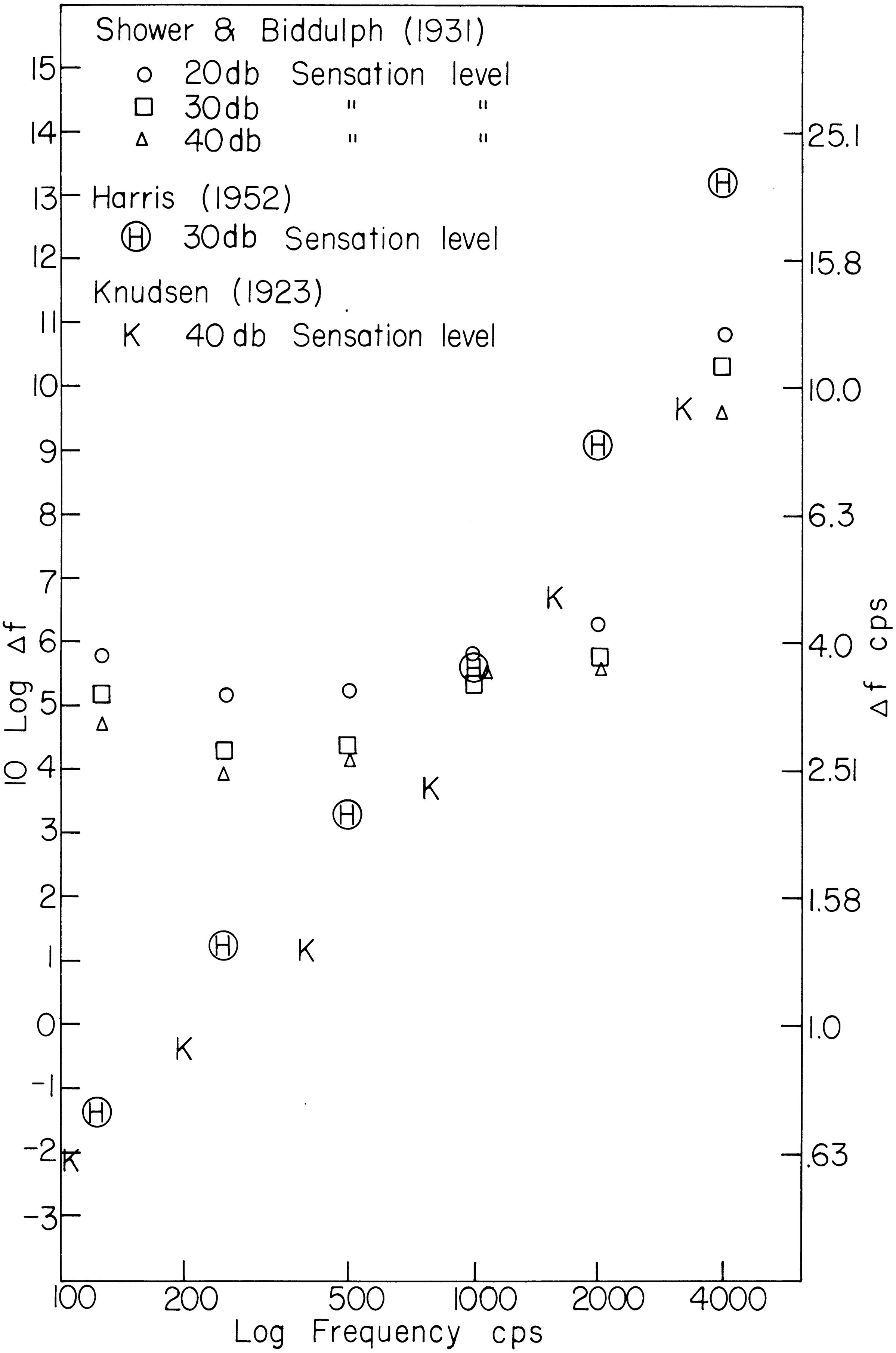


FIG. 3 THE JUST DETECTABLE CHANGE IN FREQUENCY vs. FREQUENCY

The points represent the change in frequency in cps which is reliably detected. The frequency from which this change is noted is the abscissa. Sensation level is the intensity of the tone in db above absolute threshold.

discrimination are different from experiment to experiment. The use of the critical band argument to explain the frequency discrimination data is certainly appealing. However, until the theoretical arguments develop to the extent that they can be used to predict the type of experiment which should yield data consistent with critical band functions, the hypothesis remains interesting but untested.

2.4 Loudness Experiments

This subsection and the following one will present still further data which have some bearing on the critical bandwidth of the auditory system. The data from loudness experiments have been recently obtained by German investigators who called the bandwidth a Frequenzgruppe. Zwicker, Flottorp, and Stevens (29) have reviewed many of the pertinent papers and have presented data collected at the Psycho-Acoustic Laboratory which demonstrate the main phenomena. So far as the author is able to discern, there is no logical or theoretical argument which explains why the data from loudness experiments behave as they do. That is to say, the critical bandwidth hypothesis, as expressed in the earlier part of this paper, was not used to deduce the effects found in the area of loudness. Nevertheless, the data obtained certainly demonstrate a critical region in frequency, and the width of this region plotted against center frequency is reminiscent of the data obtained from masking or some frequency discrimination experiments.

In the loudness experiment, reported by Zwicker, Flottorp, and Stevens, the situation is as follows; four tones of equal spacing are selected centered around some center frequency. The subject's task is either to adjust a standard tone to equal the loudness of the complex or vice versa. The width of the spacing in the complex is varied. The results are usually of the following form; as long as the overall spacing (that is, the frequency difference between the highest and lowest tone) is less than some value the loudness of the complex is constant. Once this critical frequency difference is exceeded the loudness of the complex grows with the frequency difference in a manner which appears to be dependent on both the center frequency of the complex and the sound pressure level of the tones. For 1000 cps center frequency, and a complex where all the components are within the critical bandwidth, the loudness of the complex is the same as the loudness of the 1000 cps standard having the same sound pressure level.

A similar experiment may be performed using band limited noise in place of the collection of sine wave signals. The loudness balance is obtained by using another band of noise as a comparison stimulus while increasing the bandwidth of one noise signal. The results are similar to those obtained with pure tones. As long as the band of noise has a width less than the critical value the loudness is constant. When the band of noise is increased in width beyond this value the loudness increases. The point at which this increase occurs is independent of level, to a first approximation, and is consistent with the value obtained using sine wave stimuli. Zwicker, Flottorp, and Stevens report other experiments. Some are concerned with masking and some with changes in phase. The results of all these experiments (Figure 4) lead them to present a relation between bandwidth and center frequency. A summary of this same relationship, as it is obtained from the masking experiments and frequency discrimination experiments, is presented along with this relation. One line represents an average of Shower and Biddulph's data, and the other represents Harris's work.

The form of the data obtained from these different experiments is obviously somewhat similar. They are different largely in terms of a

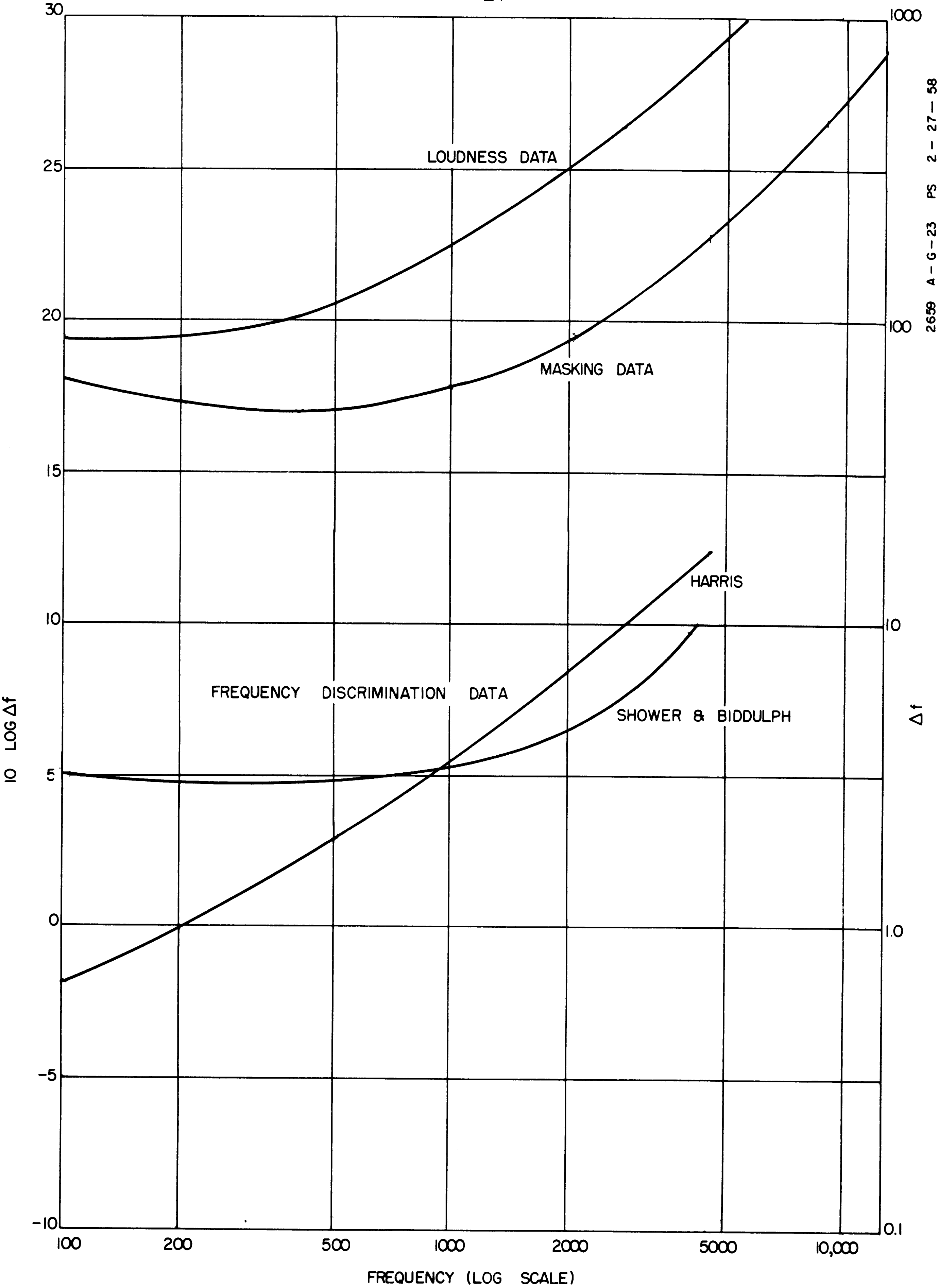


FIG.4 SUMMARY OF THE DATA ON BANDWIDTH vs. CENTER FREQUENCY

For the Loudness and Masking Data Δf is the width of the critical band is cps. For the frequency Discrimination Data Δf is the just detectable change in frequency.

constant shift in the logarithmic axis, which corresponds to a constant multiplier. As yet, no theory explains why these different techniques yield these different estimates of critical bandwidth.

2.5 Detection with Unknown Frequency

The three experiments reviewed in this section result from deductions based on an auditory filter hypothesis. They pertain most directly to the problem of attention and how the critical band concept may play an important role in the more perceptual, as opposed to sensory, type of experimentation.

The first phenomenon to be noted was reported at the Psychophysical Laboratory of the University of Michigan. A group of subjects had been listening for some time to a 1000 cps gated signal imbedded in noise. Unknown to the subjects, the frequency of the signal was changed to about 1300 cps. The subject's task was to state in which of four temporal intervals the signal occurred. The subjects in responding to this unexpected signal, performed near the chance level. That is, they behaved as if the signal had been turned off. Later, when they were told to listen for the 1300 cps signal, they obtained about 90 percent correct detections, with all physical parameters the same. Since this is by nature a difficult experiment to repeat, while maintaining the subject's trust, no more elaborate investigation was undertaken.

Karoly and Isaacson (10) have repeated this basic finding. The experiment was conducted in an experimental laboratory section of a psychology class. The speaker was placed in front of the room and the class (subjects) sat at varying distances. Frequencies of 500, 1000, and 1500 cps were used in a counterbalanced design. 1000 cps was the expected signal while 500 or 1500 was unexpected. Various levels were employed for all signals. To obtain some constant detectability in a two-alternative forced-choice test, a difference of about 10 to 15 db between the expected and unexpected signals was required.

The second experiment (24) along this same line demonstrated that if the subject was uncertain as to which of two signals is presented a decrease in performance would result.

This experiment consists of three conditions. In the first condition some frequency for the signal is selected; call it f_1. The signal is presented in one of four time intervals and the subject is asked to detect in which interval it was presented. The second condition is identical to the first except some other frequency is selected, f_2. Once both frequencies have been adjusted in amplitude so that they are about equally detectable the third condition of the experiment is begun. This last condition consists in randomly presenting either f_1 or f_2 in one of four time intervals. Once more the subject is asked in which interval the signal occurred. The subject knows that one and only one stimulus will be presented in any given four intervals, and is uncertain only as to which frequency will be used. The results indicate that for signal durations of about one tenth second and frequency separation greater than 200 cps the subject obtains less percent correct detections for the third condition than in either the first or the second conditions. In fact, the data are consistent with a model which assumes the observer can "listen" with only one critical band.

If the model is true, then, assuming the observer has no information as to which of two frequencies will be presented, he should be "listening" at the correct frequency one-half of the time. His percent correct in the experimental condition will then be one-half of the average percent correct in the control conditions plus one-half of the chance level (25 percent). While no great amount of data was collected, for the conditions where an extreme frequency separation was used, this model appeared to fit the data reasonably well. Both of these experiments demonstrate the importance of some central determinants in explaining the behavior. The critical band hypothesis suggests a mechanism which is completely consistent with the result.

2.6 Summary of Chapter II

A review of many of the pertinent investigations concerned with the critical band hypothesis has been presented. The main source of evidence has come from the area of masking, where the hypothesis was formed. However, both frequency discrimination data and experiments in the area of loudness have added considerable support to the general concepts. There is, as yet, no detailed theory of how the data from the several areas can be completely integrated.

While the critical bandwidth versus center frequency functions are somewhat similar in shape when plotted on logarithmic axes, no theory explains in detail why the multiplicative factors have different values. Until such a theory is forthcoming the parallel functions must remain interesting and intriguing findings which need explanation.

Perhaps the data discussed above may be better unified under some other type of hypothesis. The critical band hypothesis is, of course, easily interpreted by assuming a band pass filter characteristic in the auditory mechanism. Such a simple assumption is most easily employed to explain the data gathered in masking experiments. Its extension to frequency discrimination data is made fairly easily. The main assumption needed in such an extension is that frequency information, or, rather, a change in frequency, is reflected in a change in amplitude. Such a property is consistent with a filter model. Thus, the filter model asserts that intensity is the main variable while frequency information is obtained by determining which filter is active. This model is, therefore, almost exactly a place theory of hearing.

The loudness data are based mostly on empirical findings which demonstrate a critical frequency region. This critical region acts as a parameter in the equations for the data. If the complex sound is within this region, one function represents the data. If the complex sound falls outside this region, another function applies. While the critical band hypothesis does not present an immediate interpretation of the data, this may be due in part to the nature of loudness judgements. As Howes, (9) and others (29) have pointed out, the judgement of loudness depends upon the attitudes of the subject. Two prevalent attitudes are the "analytic" and the "integrative." The former attitude leads to a loudness judgement that is independent of the number of components. The latter leads to a loudness judgement that grows as the number of components increases. Hence, the loudness data may be more completely explained in terms of some central mechanism which merely reflects the frequency analysis mechanism to some minor degree. In short, the main problem to be explained is how the outputs of the several critical bands are employed by the central mechanism, not how the critical bands *per se* affect the judgement.

This latter question, how the outputs of the critical bands are employed, is the topic for the experimental data presented in the next section.

CHAPTER III

THE DETECTION OF COMPLEX SIGNALS IN NOISE

3.1 Introduction

Two experiments pertaining to the critical band concept are presented in this section. These experiments do not test directly the critical band concept; rather, both are extensions of this concept to new areas. By introducing new assumptions it will, perhaps, be possible to reconcile the inconsistencies discussed in Chapter II. One very important area needing clarification is the manner in which the critical bands affect the detection of stimuli which do not have a simple sinusoidal form. It is the purpose of the two studies reported in this section to suggest some assumptions which will permit the critical band concept to be applied to these complex signals. The experiments involve the detection of multiple component signals and signals with randomly varying components, such as band limited noise.

3.2 Experiment I: Detection of Multiple Component Auditory Signals In Noise

Most of the evidence for the critical band concept arises from situations where the observer is asked to detect a sinusoidal signal in the presence of some sort of interference. The problem considered in the present experiment is slightly different. Consider a complex signal generated by adding two sinusoidal signals. The aim of this investigation is to provide a model that will predict, from the detectability of the component sinusoids, the performance on complex signals. This model must entail certain assumptions concerning the nature of the auditory mechanism. The test of these assumptions should therefore provide important information concerning how the ear operates.

In order to provide a method for predicting the detectability of the complex signals, and to provide a method of interpreting the data, three models will be considered. The first model, the no-summation model, may be considered the null hypothesis. This model asserts the detection of a complex signal is no better than the detection of the most easily detectable component of the complex. The second model, the multiple independent-thresholds model, claims that an increase in detectability of the complex will result because only one of several thresholds need be exceeded by the complex signal. The third model, the statistical summation model, asserts that the detection of the complex signal may be predicted by assuming that the outputs of several critical bands are linearly combined.

In order to clarify the conditions under which the models apply, the experimental procedure and the results will be presented first. The three models and their predictions will then be explained. A comparison of the obtained results and these predictions will be used to test the models.

3.2.1 History of the Problem. Fletcher's critical band concept (1) answers the question of what are the relevant variables for predicting the detectability of a single sinusoid in noise. The concept asserts that one can neglect all the frequency components of the interference with the exception of a narrow band of frequencies centered about the signal to be detected. Thus, for sine wave signals, the auditory system can be likened to a narrow bandpass filter. This model has been tested and appears to be adequate.

How this model should be applied when the signal consists of two or more sinusoids widely separated in frequency is not stated. One possibility is that the observer can listen with only a single critical band. When a composite signal is presented the observer's behavior can be explained by assuming he is listening to only one narrow frequency region where the signal-to-noise ratio is most favorable for detecting the signal.

Such a hypothesis has received support from the research of Marill (15). Marill used four frequencies: 500, 540, 1100 and 1060 cps. When presenting the pairs with small frequency differences (i.e., 500 and 540, or 1100 and 1060 cps) his observers behaved, roughly, as if a single member of the pair had been presented at twice the energy. That is to say, a power summation model will predict the data. Using the frequency pair 500-1100 cps, Marill found generally the detectability of the pair to be no greater than the detectability of the most easily detectable single component. Certainly the power summation model predicts much too high for this pair of widely separated frequencies.

Tanner, Swets, and Green (24) have also reported research to support the conclusion that for widely separated frequencies the observer apparently listens to only one frequency region. Their experimental situation required the subject to detect one of two possible signals presented in one of four time intervals. The observer was simply asked to detect the signal's presence, never to state which frequency it was. For signals of sufficiently short durations, and of sufficiently wide frequency separation, the observers behaved as if they were listening to only one or the other of the two possible frequency regions where the signal might be presented.

Schafer and Gales (19) attempted to determine the detectability of multiple component signals. For two tones separated by a critical band or more, they conclude, the energy of each component can be reduced 0 to 2 db and the composite signal will remain as detectable as when the signals are presented individually. For 4 to 8 tones the value is 0 to 3 db. If one accepts the 0 db figure in each case, then Marill's conclusion is supported. In any event, there is some loss of efficiency due to splitting the power into two frequency regions. If there were perfect power summation, two tones should be 3 db, 4 tones 6 db, and 8 tones 9 db lower to maintain a constant performance. Thus, Schafer and Gales find, on the average, that for widely separated frequencies neither perfect power summation nor complete lack of summation fits the data.

It is unfortunate that all the results do not agree; for the answer to the question of multiple tone detection is necessary if one is ever going to explain the detection of anything but the simplest of auditory signals. The research reported below is both a repeat of the previous studies and an investigation of two more parameters of the process which may be a source of interaction. These two parameters are: (1) the duration of the signal and, (2) the frequency relation of the components.

3.2.2 Experimental Procedure. The experiment was conducted at the Psychophysical Laboratory, Electronic Defense Group, University of Michigan. The basic apparatus has been described elsewhere (5). A four-alternative forced-choice method was used to measure the detectability of a signal. Two changes were made from the procedure described in (5). Each sequence of four test intervals was preceded by an interval in which the noise dropped 10 db in level and the signal was presented. This served to remind the observers of the frequency characteristics and duration of the signal. Also, the signal was simply gated for the specified duration without regard to phase.

Four frequencies were employed: 500, 1000, 1823, and 2000 cps. The complex signal was generated by adding the voltages of two of these signals. The frequencies were generated by two independent oscillators. All possible pairs were employed. For each signal duration, each frequency was used as a signal and was adjusted in amplitude so that about 75 percent correct detections were obtained. One of the six possible pairs of frequencies, using the amplitudes previously determined, was then used as the complex signal. The percentage of correct detections for this complex signal was determined. There are therefore ten signals used in this experiment: four single frequency signals, and six complex signals. Four blocks, each of one hundred trials, were used to estimate the probability of a correct detection for each signal. Each signal received one block of trials before a second block was conducted for any other signal. The choice of conditions was determined by a random sequence. Signal durations of 50, 200 and 1000 milliseconds were employed. Four hundred observations were completed for each signal before another duration was selected.

Measurement of amplitude has been described in Reference (5). Measurements of frequency and duration of the signals were made with a Hewlett-Packard Model 521C Frequency Meter. Frequency drift throughout an experimental session was less than about 3 cps. Permoflux PDR-8 headphones were used. The observers listened binaurally. A noise level of about 55 db re 0.0002 dyne/cm^2. was used throughout all of the tests.

3.2.3 Results. The data are summarized in graphical form in Figure 5.[+] Each entry is based upon 400 observations. For each single component signal the percentage of correct detections is represented by a vertical line. For

[+] The physical parameters used in the experiment are listed in the table below. E is the signal energy, or time integral of the signal power. N_o is the noise power per unit bandwidth (dimensions of energy).

E/N_o	Duration 50	200	1000
500 cps	8.7	6.2	13.2
1000 cps	10.7	8.8	19.8
1823 cps	15.6	15.4	30.8
2000 cps	15.8	15.8	33.0

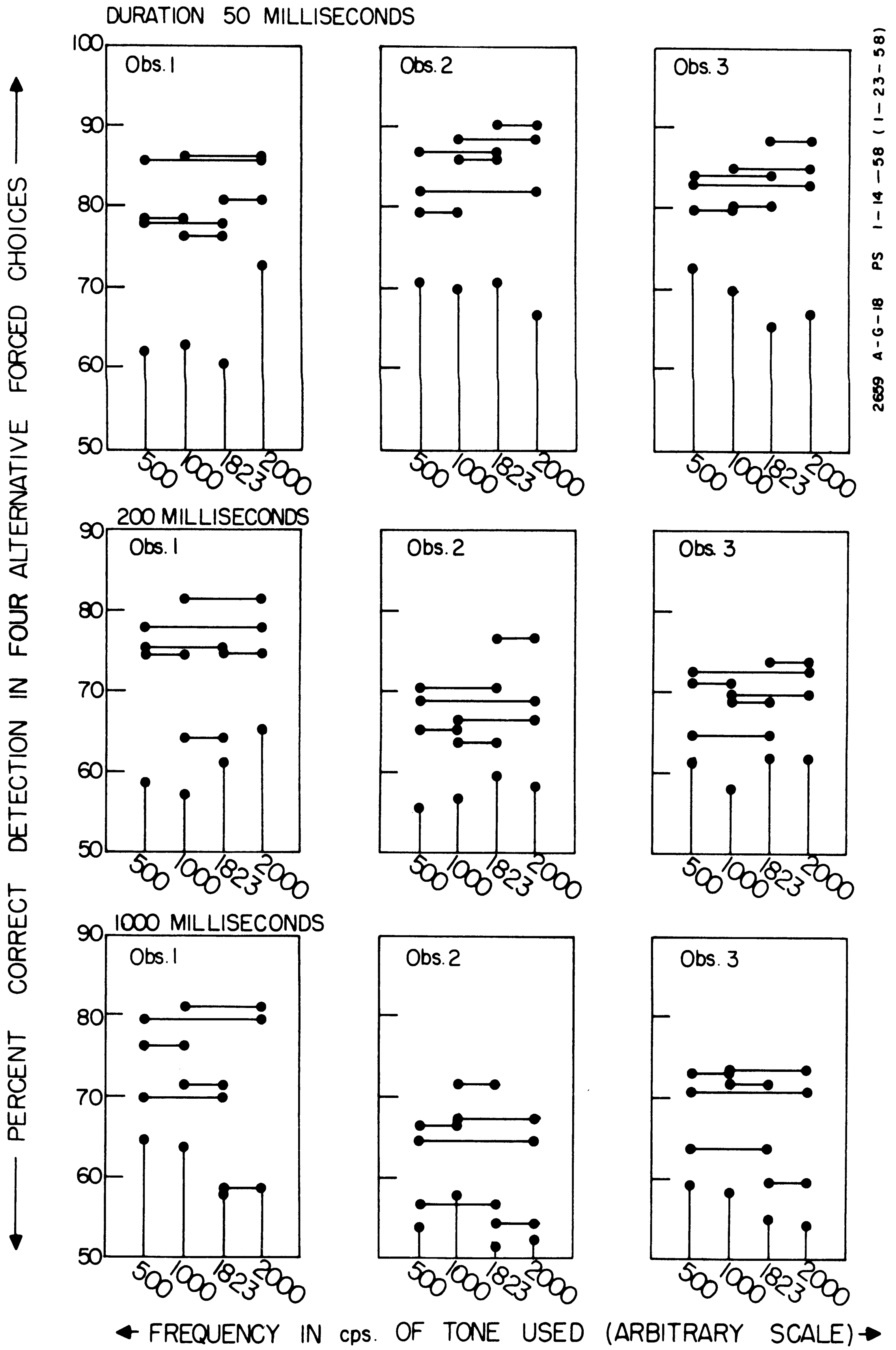


FIG. 5 RESULTS OF THE DETECTION OF TWO COMPONENT SIGNALS IN NOISE

The points attached to horizontal lines represents the percent correct detections obtained for the single sine wave stimulus. The frequency of this signal is given by the abscissa. The points connecting the horizontal lines is the percent correct detections obtained on the complex stimulus. The two end points represent the frequency of the pair of sine waves used to generate the complex stimulus.

example, with the 50 millisecond, 500 cps signal, Observer I obtained 62 percent correct detections. The percentage of correct detections for the complex signal is represented by a horizontal line connecting the two frequencies which were used to generate the complex signal. For the same conditions, and the same observer, the complex signal of 500 and 1000 cps yielded 79 percent correct detections. The complex signal is generated by using the same physical parameters for each component of the complex as were used when an individual component was presented. Thus, if the horizontal line representing the complex is above both of the vertical lines representing the single component signal, the complex is easier to detect than either single component. These graphs show the individual data obtained from three observers at three durations. About the only point that can be made with the data in this form is that the detectability of the complex signal is somewhat better than the detectability of either single component of the complex. For a finer analysis of the data it is obvious that some procedure must be adopted to normalize thedata so that the amount of increase in detectability for the complex is independent of the value of detectability for the single component stimuli. There is no accepted method of accomplishing this result. In the next section of the paper, however, three models will be considered. These will allow a form of normalizing the data so that the detectability of the complex signals may be compared.

3.2.4 Models. Three models will be used to analyze the data. All of the models make predictions which are independent of the duration or frequency separation of the components. The predictions of two models will then be compared with the data obtained in the various conditions of the experiment; such a comparison will allow for the evaluation of the models and the determination of whether the amount of improvement is indeed independent of these other factors.

3.2.4.1 Model 1: - No-Summation Model. This model predicts that the percentage correct obtained on the complex signal will be no greater than the percentage correct obtained on the most easily detectable component of the pair. This statement of the prediction assumes the conditions which prevailed in this experiment; that is, the level of the individual components is the same when the complex is tested as when the individual stimuli are tested. The rationale for such a prediction might be that the observer can listen with a critical band to only one frequency region at any given time. When faced with detecting a complex signal, where the frequencies involved are more than a critical-band apart, he listens at that region in frequency where the signal is easiest to detect.

The data collected in the experiment indicate this hypothesis is incorrect. Of the 54 complex signals tested, the detectability of the complex was greater than the detectability of the most detectable component in 53 cases. Thus, for the condition employed in this experiment, the no-summation model is definitely rejected.

3.2.4.2 Model 2: The Two-Independent-Thresholds Model. Schafer and Gales (19) suggested that such a model predicted fairly accurately the data collected in their experiment. This model asserts that in order to detect a single component signal, some hypothetical variable must exceed some critical value. This value is called a threshold. Assuming that the variable rarely exceeds this critical value when no signal is presented, the probability of a correct detection, P(C), in a forced choice test employing n temporal alternatives is given by:

$$P(C) = p + \frac{1}{n}(1 - p), \tag{3}$$

where p is the probability of the signal causing the threshold to be exceeded.

The second term in the equation is the probability of not obtaining a supra-threshold value, (1-p), times the probability of correctly guessing the correct alternative with no information, (1/n).

Now, consider the case where two or more signals are presented simultaneously and the hypothetical variables associated with each signal are independent. Four cases can arise: (1) the first signal can cause a supra-threshold value and the second not, or (2) the second signal can lead to a supra-threshold value and the first not, or (3 and 4) both signals can lead to either supra-threshold or sub-threshold values for the hypothetical variable.

Several combination rules could be used, but the one investigated is that a correct response will result if either or both signals lead to a supra-threshold value of the variable. The probability of either or both signals exceeding threshold is, therefore, given by Equation 4.

$$p_m = 1 - \prod_{i=1}^{m} (1 - p_i), \qquad (4)$$

where p_i is the probability that the ith signal will be supra-threshold, and m is the number of signals.

Hence, the probability of a correct choice $P_m(C)$ among n temporal alternatives with m signals, can be written by substituting Equation 4 into 3.

$$P_m(C) = 1 - \prod_{i=1}^{m} (1 - p_i) + \frac{1}{n} \prod_{i=1}^{m} (1 - p_i) . \qquad (5)$$

For the temporal four-alternative forced-choice method (n = 4) used in these studies, and for complex signals using two components (m = 2), this formula reduces to:

$$P_2(C) = 1 - \frac{4}{3} \left[1 - P_1(C)\right] \left[1 - P_2(C)\right] . \qquad (6)$$

This model allows us to predict the percentage of correct detections obtained on the complex from the percentage correct obtained on the single component signals.

If the hypothetical variables of this model are not independent these predictions will not be correct. If the variables are perfectly correlated, only two cases could arise. Both variables would either exceed the threshold or, they would not. The predictions obtained from this model would then be identical to those obtained with the first model.

Other changes could be made in the model. We have assumed that the observer guesses only when both components fail to reach the threshold. If more than two components are used for the complex signal it is possible that the signal would be detected only when a majority of the signals cause a supra-threshold value of the variable. Such a model becomes considerably more complicated. Yet another alternative is to assume that, even without a signal, the threshold is exceeded some constant percentage of the time. Depending, then, upon the value of this constant, the model would predict quite different values for the detectability of the complex.

3.2.4.3 Model 3: A Statistical Summation Model. This model will be explained in some detail because it appears to predict the data obtained in this experiment better than either of the other two models. This model is a logical extension of the statistical decision theory proposed by Tanner and Swets (23), and especially of the ideas expressed by Tanner (22). The critical band hypothesis is used as an interpretive device in explaining several assumptions of the model. These are, strictly considered, assumptions, and hence do not depend upon the critical band hypothesis being correct. However, the assumptions are consistent with some sort of auditory filtering process, and the interpretations will be used so as to make the model less abstract.

Before explaining how the model is extended to encompass the problem of detecting multi-component signals in noise, a brief review of how the model explains the detection of a single component stimulus will be given. Presumably, the energy located in a narrow frequency region (critical band) is transformed to a single variable so that the value of this variable will be greater when a signal is present than when noise alone is present. Let this variable be denoted $X(f_1)$. As a convenience for a later part of the discussion, this notation will make obvious the dependency of X on the frequency region f_1. Now, since noise is a random variable, $X(f_1)$ will have some distribution of values. Actually it will have two distributions of values, one when the signal (f_1) is present in the noise and another when there is no signal. When no signal is present the condition will be termed noise alone. It is assumed that in the noise alone condition the variable $X(f_1)$ will be normally distributed. When signal plus noise is present, $X(f_1)$ will also be normal with a greater average value but the same variance.

The parameter of importance is (d'), the difference in the means of the two distributions divided by the standard deviation. In order to avoid unnecessary notation, the noise alone distribution may be normalized so that the mean is zero and the standard deviation unity. Then the parameter d' becomes simply the mean of the signal plus noise distribution. This parameter will be denoted d'_{f_1}.

In a temporal four-alternative forced-choice test the observer listens to four temporal intervals. In one and only one of these intervals the signal is added to the noise. The observer is instructed to choose the interval in which the signal occurred. The model assumes that in each interval the variable $X(f_1)$ is measured. It is assumed that the observer operates with a decision rule, which is equivalent to choosing the interval in which the largest value of the variable $X(f_1)$ occurred. Hence, the probability that the largest value of $X(f_1)$ will occur when the signal (f_1) plus noise is present is simply the probability that a sample taken from a normal deviate with mean d' and variance unity will be larger than any of the other three samples from a normal deviate with mean zero and the same variance. A table relating the probability of a correct decision in a four-alternative forced-choice test and the parameter d' has been presented elsewhere (5).

In extending the model to a situation involving more than a single sine wave stimulus the problem is essentially that of determining the distribution associated with the complex signal plus noise and the distribution associated with noise alone. Particularly, the parameter which characterizes the difference divided by the standard deviation of the distributions must be determined. This parameter is essentially the parameter of importance for the complex signal.

The extension of the model will be discussed for two component complex signals. This limitation will greatly simplify the mathematics, and the extension of this result to signals involving many component signals will be obvious once this derivation is complete.

If two components are widely separated in frequency the output of the critical band associated with each will be independent. This statement is simply a way of interpreting the assumption that $X(f_1)$ and $X(f_2)$ are independent. When noise alone is presented both variables are normally distributed with some mean and variance. When signal $(f_1 + f_2)$ plus noise is presented the mean values of both variables, $X(f_1)$ and $X(f_2)$, will increase, compared with the noise alone conditions, and will have the same variance. The correlation of the variables when the complex signal plus noise is presented will still be zero.

One reasonable method of combination is simply to add each variable, weighted by a constant. It is presumed that each variable is normalized so that the noise alone distribution has zero mean and unit variance before the summation. Actually, such normalization could be included in the constants of the addition, but it greatly simplifies the following derivation to assume such a normalization at this point in the argument. The two variables, therefore, have the following properties:

Noise Alone:

$X(f_1)$ is normally distributed with zero mean and unit variance;

$X(f_2)$ is normally distributed with zero mean and unit variance.

Signal Plus Noise:

$X(f_1)$ is normally distributed with mean d'_{f_1} and unit variance;

$X(f_2)$ is normally distributed with mean d'_{f_2} and unit variance.

Consider the linear combination;

$Z = aX(f_1) + bX(f_2)$, where a and b are constants.

With Noise Alone:

Z is normally distributed with mean zero and unit variance:

$a^2 + b^2 + 2r\left[X(f_1)X(f_2)\right]ab$, where $r\left[X(f_1)X(f_2)\right]$ is the correlation of the variables $X(f_1)$ and $X(f_2)$. Since $r\left[X(f_1)X(f_2)\right] = 0$, the variance of Z is $(a^2 + b^2)$.

With Signal $(f_1 + f_2)$ plus Noise:

Z is normally distributed with mean $ad'_{f_1} + bd'_{f_2}$ and variance $a^2 + b^2$. Hence, in the variable Z the difference between the means of the two distributions divided by the standard deviation or (d'_Z) is given by:

$$d'_Z = \frac{ad'_{f_1} + bd'_{f_2}}{\sqrt{a^2 + b^2}} \qquad (7)$$

The d' for the combination is now determined except for the constants a and b. In order to predict anything about the detectability of the complex from the detectability of the individual components, a and b will be chosen so as to maximize d'_Z. This is equivalent to assuming that the probability of a correct detection is maximized since d' and this probability are monotonically related.

To accomplish this maximization it is convenient to write a and b as a function of a single variable. Consider a and b as two sides of a right triangle.

$$\frac{a}{\sqrt{a^2 + b^2}} = \sin \gamma \tag{8}$$

$$\frac{b}{\sqrt{a^2 + b^2}} = \cos \gamma \tag{9}$$

The maximum occurs when

$$\cos \gamma \, d'_{f_1} = \sin \gamma \, d'_{f_2} \tag{10}$$

$$\frac{d'_{f_1}}{d'_{f_2}} = \tan \gamma = \frac{a}{b} \tag{11}$$

One solution is to set $a = d'_{f_1}$, $b = d'_{f_2}$, (12)

then,

$$d'_Z = \frac{(d'_{f_1})^2 + (d'_{f_2})^2}{\left[(d'_{f_1})^2 + (d'_{f_2})^2\right]^{1/2}} = \left[(d'_{f_1})^2 + (d'_{f_2})^2\right]^{1/2} . \tag{13}$$

Thus an expression for the detectability of the complex signal has been determined in terms of the detectability of each component of the complex. The geometric interpretation of this result is very simple. Consider $X(f_1)$ and

$X(f_2)$ as the coordinates of a plane. If the two values associated with each frequency region are plotted in this plane, one finds the noise alone distribution centered about the point (0,0). When the complex signal plus noise is present the values form a new distribution centered about the point (d'_{f_1}, d'_{f_2}). Now suppose that each observation taken by the observer yields a value for $X(f_1)$ and $X(f_2)$. The maximum d' for the combination is obtained by projecting such points in the space on a line running from the point (0,0) to (d'_{f_1}, d'_{f_2}). The distribution functions for such projections is normal when either noise alone or signal plus noise is present. The difference in the means of these two distributions divided by the standard deviation has the value d'_Z.

If this line of argument is extended to a situation where many components are used for the signal, and assuming each component affects independent processes, then the maximum d' for the linear combination is simply the square root of the sum of the squares of d's for the individual components.

Since the frequency separations between components used in the experiments reported in this paper are wide compared with the width of a critical band, $r\left[X(f_1), X(f_2)\right]$ will be assumed to equal zero. The model may be extended to handle cases where the correlation is not zero. The following gives a brief discussion of the manner in which the model handles such cases. Presumably such cases arise when the components are located near each other in frequency so that the outputs of the critical bands associated with each component are affected by energy which is common to both bands. In other words, the filter bands overlap. One result of such a condition is to make the correlation coefficient $r\left[X(f_1), X(f_2)\right]$ greater than zero. The second result is that the signal energy associated with one frequency, say f_1, influences the variable $X(f_2)$, which is associated with the other band.

Since the frequencies f_1 and f_2 are very close together it may be assumed that this influence is symmetric; that is, f_2 affects $X(f_1)$ in the same way f_1 affects $X(f_2)$.

When the signals were widely separated in frequency the mean of the signal plus noise distribution was d'_{f_1}. For the case now being considered it will be assumed that the mean of the signal plus noise distribution for $X(f_1)$ will be $d'_{f_1} + kd'_{f_2}$. The positive constant k has a value somewhere between zero and unity. It reflects the amount of overlap between the critical bands. Both $r\left[X(f_1), X(f_2)\right]$ and k should be monotonic decreasing function of the distance in frequency between the two components, f_1 and f_2. If the difference in the frequency of the two components is very small both r and k should be near unity.

In order to simplify the following derivation it will be assumed that $d'_{f_1} = d'_{f_2} = d'$. Proceeding as before one finds the difference in means divided by the standard deviation, d'_Z.

$$d'_Z = \frac{d'(a + b)(1 + k)}{\sqrt{a^2 + b^2 + 2rab}} \tag{14}$$

Using the same method to maximize d'_Z, the maximum occurs when a = b, letting a = b = 1.

$$d'_Z = \frac{d'\ 2(1 + k)}{\sqrt{2 + 2r}} = \sqrt{2}\, d' \frac{1 + k}{\sqrt{1 + r}} \tag{15}$$

If both signals are nearly the same in frequency; $k \approx 1$ and $r \approx 1$, then $d'_Z = 2d'$. The result agrees with Marill's finding for the frequency pair 500 and 540 cps. If the frequency pair is very far apart; $k \approx 0$, $r \approx 0$, and $d'_Z = \sqrt{2}\, d'$, which agrees with the previous equation 13.

This model, therefore, gives a plausible derivation of the detectability of a complex signal using any two frequencies. No quantitative predictions are made when the frequency pair are close together in frequency since the relation between k, r, and frequency separation is not known. This relation depends theoretically only upon the shape of the critical band.

In summary, the model makes it possible to predict the detectability of a complex signal from the detectability of each component of the complex. The application of the model to those cases where the noise affects both correlated and uncorrelated processes is explained. For components separated by several critical bands one may assume k and r are zero. Once this separation has been reached, the predictions of the model are independent of any further separation. These are the types of separations used for the complex signals discussed in the experimental section of this paper. Thus, the predictions cited in the following sections assume both k and r equal to zero, as in Equation 13.

3.2.5 Evaluation of the Models. Only the two independent thresholds and the statistical decision model will be compared with the obtained data. The "no-summation" model in this experiment could be rejected as being inadequate in 53 out of 54 cases as discussed previously.

For each of the three observers and three durations there are six possible pairs of tones used as the complex signal. For both models the detectability for the single component stimuli was used to generate a prediction concerning the detectability of the complex tone. The difference between the predicted percent correct detections and those actually obtained is displayed in Figure 6. Two points appear for each complex stimulus and condition, one for each of the models used to generate predicted values. At the extreme right of each condition the average error in prediction is plotted. The average error over all conditions and subjects is about 1.5 percent for the

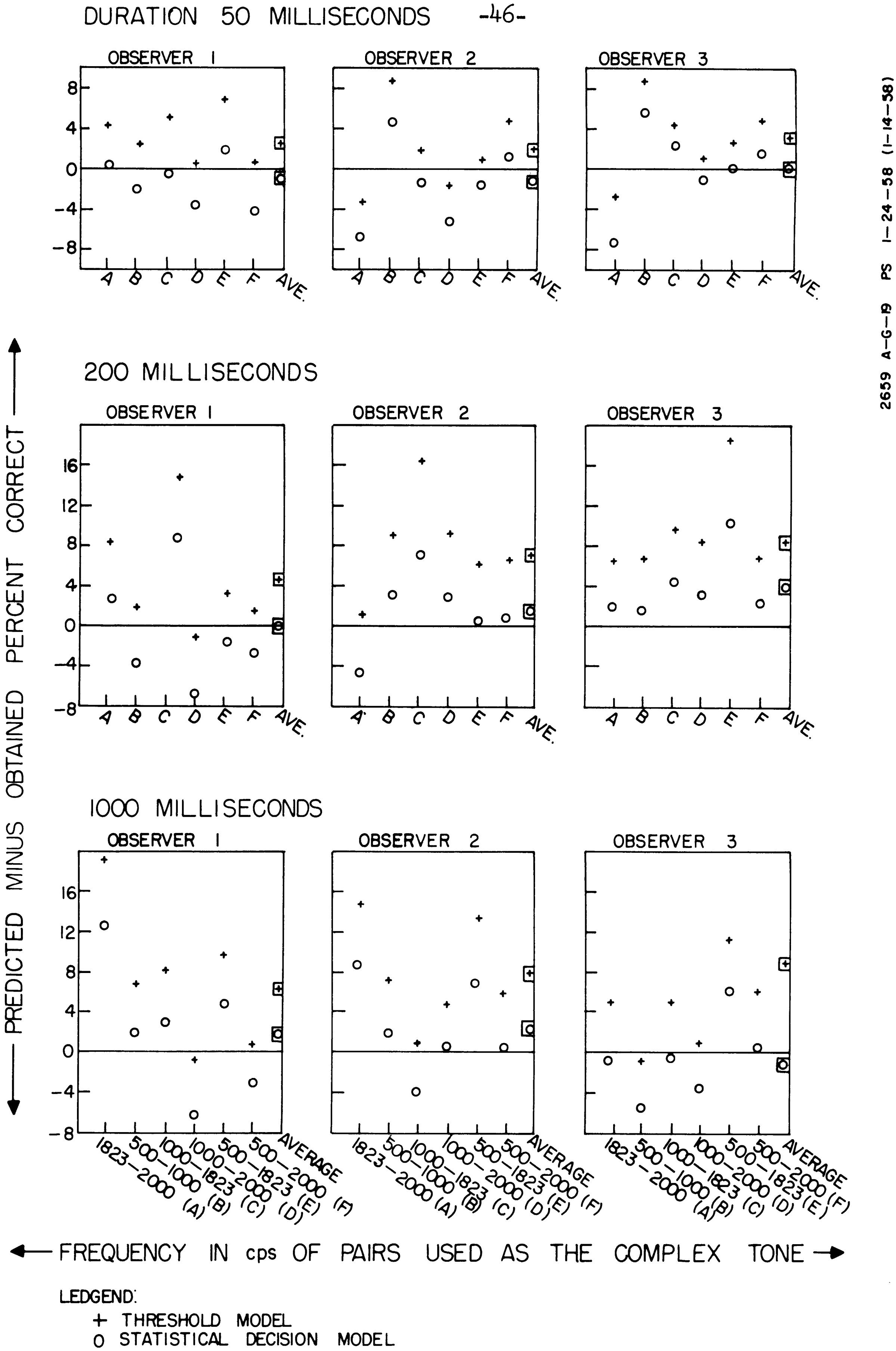


FIG.6 A COMPARISON OF THE RESULTS OF THE FIRST EXPERIMENT WITH TWO MODELS

The ordinate is the difference between the predicted and the obtained percent correct detections. Two points are plotted for each complex signal; one point when the Threshold Model is used another when the Statistical Decision Model is used. The abscissa is frequency pair used for the complex signal.

statistical decision model and about 5 percent for threshold model. A "t" test was calculated for the entire set of data. The value of the parameter t was about 1.25 for the threshold model and about .25 for the statistical decision model. The threshold model can thus only be rejected at the 20 percent confidence limit while the statistical decision model can only be rejected at the 80 percent confidence limit. In terms of the likelihood ratio the statistical decision model is to be preferred about 4 to 1 over the threshold model. Obviously an experiment involving more than two components will provide a stronger test of the differences between the theories. With only two components used for the complex signal the difference between the predictions of the two theories never exceeds 10 percent and averages much less than that.

The predictions made by both models are independent of the frequency separation and of the duration of the complex stimuli. Such a position can be checked to some extent by using the graphs of Figure 7. The statistical decision model is employed since it best predicts the available data and conveys essentially the same picture that one would obtain with the other model. The percent error as plotted in the previous figures has simply been averaged across observers and plotted as the ordinate of Figure 7. The combination signals are plotted as the abscissa in the order of their frequency differences. The scale is arbitrary. As one can observe, there is no dependence upon frequency separation or signal duration.

3.2.6 Discussion. The statistical decision model is tentatively accepted as the most adequate model for predicting the detectability of a complex signal in noise. This model is consistent with the detection theory previously proposed for single component stimuli. It is also consistent with a critical band model which asserts one can linearly combine the outputs of several crit-

2659 A-G-20 PS 1-22-57

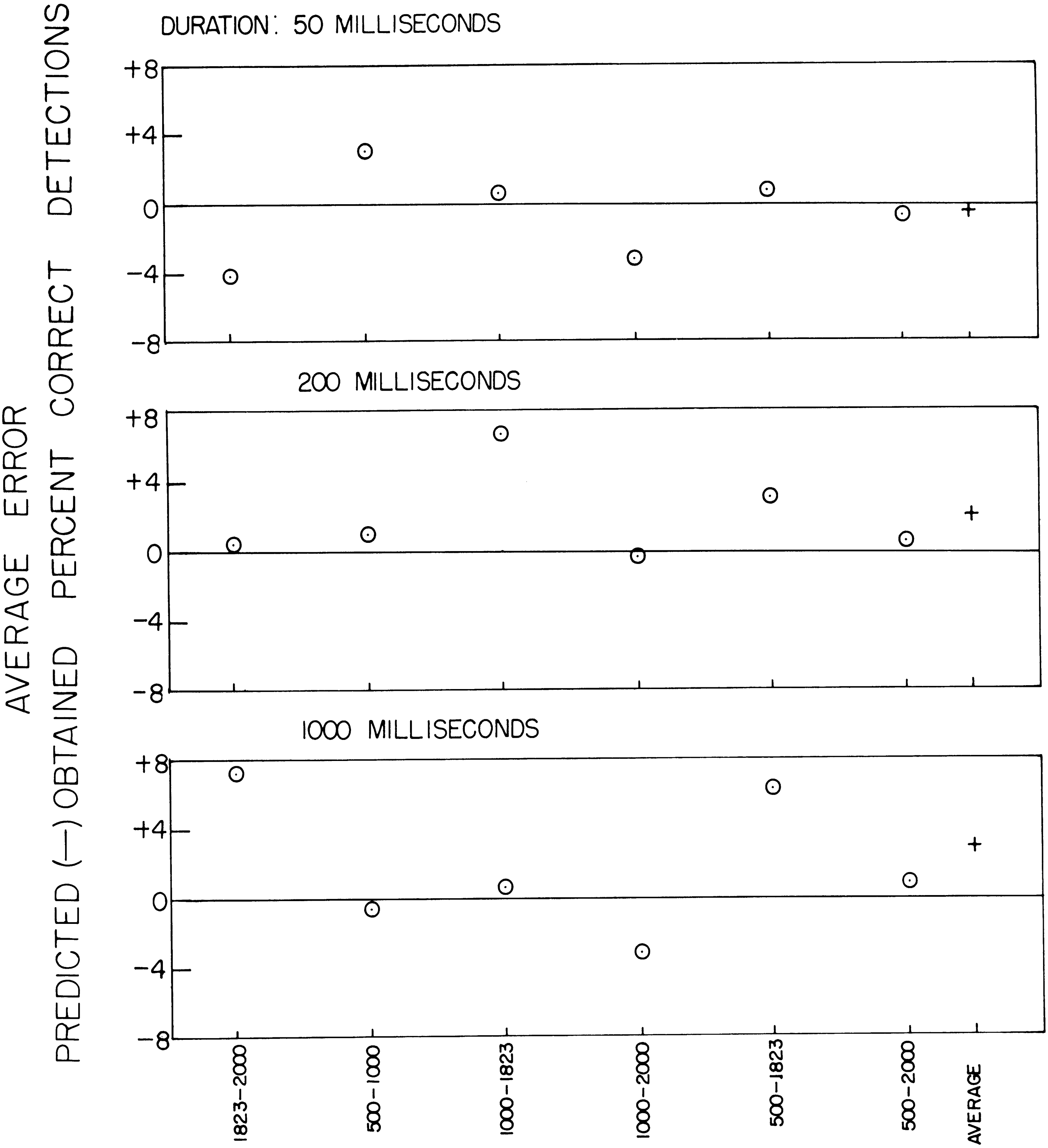


FIG. 7 THE AVERAGE ERROR AS A FUNCTION OF DURATION AND FREQUENCY

The ordinate is the average of the errors obtained for each subject. The abscissa is the frequency pair used for the complex stimulus. Only the statistical decision model is used to obtain the error values.

ical bands. The model has the advantage of incorporating all the variables which affect the detectability of a single component, since the detectability of each signal component is used in determining the detectability of the complex. The manner of combination is theoretically linear and no interaction term is used.

Since d' is roughly proportional to signal energy, the level of two equally-detectable components can be lowered by 1.5 db in a threshold experiment and the same detectability can be maintained for the complex signal. This result agrees with the average values cited by Schafer and Gales (19), but is inconsistent with the results obtained in Marill's work (15).

The statistical decision model has been applied to an experiment where the observer is asked to detect one of two possible signals. If the two variables $X(f_1)$ and $X(f_2)$ are combined the d' for the combination will be decreased, as compared with the d' associated with either component, since only one signal is presented on any series of trails. Thus, the observer listens to twice as much noise power when only one of the two signals is present. The predictions generated by such considerations are similar to those made when one assumes the observer listens to only a single frequency region at any instant in time. Thus, the present model is consistent with most of the data obtained by Tanner, Swets, and Green (24).

The use of only two components as a complex signal does not provide the best condition for attempting to choose between models. Rather, it provides a simple condition for formulating a theory which will work in the more general case. The discrepancy between the results obtained in this study and those obtained by Marill (15) provides an interesting challenge. Assuming both results are correct, it is obvious that some factor neglected by both Marill and the author plays an important role in this situation.

3.3 Experiment II: The Detection of An Auditory Noise Signal in Noise

3.3.1 Introduction. In these experiments the signal consisted of a burst of noise with certain frequency characteristics and of specified duration. A continuous noise of constant power was used as a masking stimulus. The observer's task was to detect the presence of the noise signal. The word "signal" will be used to refer to the noise signal. The masking stimulus will be called "noise". When both signal and noise have idential frequency characteristics the task of the subject is to detect a change in the power of the noise. In this special case the experiment is exactly equivalent to the determination of the observer's differential intensity sensitivity to noise. The advantages of defining the terms signal and noise, as described above, will become obvious when a theoretical model is presented. The model can be used in the more general case when the signal power spectrum is not the same as the noise power spectrum. The impetus for these studies grew out of a consideration of the critical band concept and how this concept might apply when the signal was a sample of noise.

The critical band concept was developed in experiments where the signal was a sine wave and the masking stimulus was wide-band noise. The results of these experiments suggested that the observer's behavior could be explained by assuming the auditory system resembled in some ways a narrow band-pass filter. Such a narrow band-pass filter is clearly an effective way of excluding a great deal of interfering energy. If the signal to be detected is not a narrow band signal, such as a sinusoid, then employing a narrow band analysis device is not an efficient scheme for detecting such signals. Since the auditory system is like a narrow band analyzer, it is possible that observers will do rather poorly when attempting to detect a wide-band signal. A noise signal is ideal for investigating this problem, because a noise signal which has a wide power spectrum can be used.

Two theoretical approaches could be used in considering this problem. One approach could maintain that the auditory system is constructed in such a way that only one critical band can be used at any instant in time to detect signals. Such an approach is consistent with a model presented by Tanner, Swets, and Green (24) for an experiment where the signal frequency was unknown. It is also consistent with the results reported by Marill (15) in an experiment concerned with detecting two widely separated pure tones imbedded in noise. It is unfair to attribute such an approach to any of these authors, since their work was confined to sine wave stimuli, and none of it dealt explicitly with the problem of a noise signal.

The second approach assumes that the critical bandwidth, as inferred from pure tone data, simply represents a minimum bandwidth of the auditory system. It has been suggested, in the earlier sections, that for other types of signals several of the bands may be linearly combined. Thus, if enough continguous bands are combined the auditory system may become, effectively, a wide band system.

In order to investigate these two approaches, two experiments were performed. The first is most germane to the issue just discussed. The second was an attempt to determine whether the model developed for the first experiment could be extended to encompass another parameter, the duration of the signal. Before presenting the experimental results a theoretical model will be introduced. This model will set forth explicitly how signal bandwidth, duration and power affect the detection of noise signals in noise. This model is an optimum detection scheme that is extremely useful in contrasting the two theoretical approaches just discussed.

3.3.2 <u>The Statistical Decision Model</u>. The statistical decision model was first developed by Peterson and Birdsall in "The Theory of Signal Detectability"(18). The present development borrows many of the ideas and conclusions of that paper. The exact form of the development given here is different and is a more convenient formulation for the research that is discussed.

The problem of detecting random signals in the presence of random interference obviously requires a statistical treatment. In particular, the detection of a Gaussian noise signal in Gaussian noise may be interpreted, mathematically, in the following way. During a finite interval in time, called the observation interval, the "receiver" obtains a sample wave form which is some single-valued function of time. There are two possible hypothesis concerning the origin of the sample. The sample either was drawn from a noise distribution or from a distribution of signal plus noise. The frequency characteristics and the power levels of both signal and noise are known. The duration of the signal is equal to the observation interval and is also known.

According to Peterson and Birdsall the sampled waveform may be represented as a point in a space of n dimensions. ($n = 2WT$), where W is the bandwidth of the signal and T is the duration of the signal. One method of obtaining the values for the n coordinates is to measure the magnitude of the input waveform at n discrete units in time. By sampling in this way the magnitude of the waveform at each sample point is statistically independent of the magnitude at any other sample point. If this finite sampling procedure is followed, and the likelihood ratio test is employed, the optimum decision axis may be shown to be linear with the power, or energy, in the sample waveform. Thus, if the distribution of the power in the sample under each hypothesis can be obtained, it will be possible to state exactly how the probability of a correct detection will depend on the various physical parameters. The following arguments will derive the distributions.

Consider the distribution of the values for the power in the sample waveform when noise alone is presented. Let the magnitude of the waveform at the ith sample point in time be X_i. Since Gaussian noise is used, X_i will be normally distributed with a zero mean and variance N.* The quantity N refers only to that noise power which is in the signal band. In deriving the distribution of power in the sample waveform it will be more convenient to work with a variable P_N which is linearly related to this power:

$$P_N = \sum_{i=1}^{2WT} X_i^2$$

P_N is a random variable, since it is a transformation of the random variables X_i^2. Since the summation

$$\sum_{i=1}^{2WT} \left(\frac{X_i}{\sqrt{N}} \right)^2$$

---

* To be consistent, if N is the average power of the noise, the average total energy in the sample must be:

$$E\left[\int_o^T X(t)^2 dt \right] = NT$$

where T is the duration and X(t) is the voltage waveform.

A Riemann sum must approximate this integral, thus:

$$E\left[\sum_{i=1}^{n} X_i^2 \Delta t \right] \approx NT$$

where X_i is the discrete approximation to X(t), and n = 2WT. Since $\Delta t = \frac{1}{2W}$,

$$E\left[\sum_{i=1}^{n} X_i^2 \Delta t \right] = n\overline{X^2} \frac{1}{2W} = T\,\overline{X^2}, \text{ where } \overline{X^2} = E(X_i^2),$$

but then $\overline{X^2}$ must equal N, which is the variance of X_i.

is distributed as chi square with 2WT degrees of freedom, the expected value of the distribution is 2WT, or,

$$E(P_N) = E(\sum_{i=1}^{2WT} X_i^2) = 2WTN. \tag{16}$$

Since the variance of a chi-squared distribution is simply twice the number of degrees of freedom,

$$E\left[\text{variance of } P_N\right] = 4WTN^2 \tag{17}$$

For 2WT greater than 30, a chi-squared distribution closely approximates a normal distribution; thus, P_N is normally distributed. Consider the distribution of values for the power in the waveform when signal plus noise is presented. Let the magnitude of the waveform at the ith sample point be y_i. Since both signal and noise are Gaussian, y_i will be normally distributed with mean zero and variance S + N.*

The quantity S + N refers only to the signal and noise power which is in the signal band. In the case of the signal plus noise, consider the following random variable.

$$P_{S+N} = \sum_{i=1}^{2WT} y_i^2$$

Since the summation

$$\sum_{i=1}^{2WT} \left(\frac{y_i}{\sqrt{S+N}}\right)^2$$

is distributed as chi-squared with 2WT degrees of freedom the expected value of this distribution is 2WT, or,

---

* An argument similar to that given in the first footnote says $E(y_i^2) = S + N$.

$$E(P_{S+N}) = E\left(\sum_{i=1}^{2WT} y_i^2\right) = 2WT\,(S + N), \qquad (18)$$

and the expected value of the variance

$$E\left[\text{variance of } P_{S+N}\right] = 4WT\,(S + N)^2. \qquad (19)$$

As long as 2WT is greater than 30, P_{S+N} will also be normally distributed.

Figure 8 shows the results of these derivations. The distribution of power in the sampled waveform, under each hypothesis, is displayed.

In the experiments presented in this paper a two-alternative forced-choice technique was used. Two time intervals were marked off for the observers, and, in one of these intervals, the signal was added to the noise. The ideal detector would measure the power in the two waveforms presented during each interval. The optimum decision rule would be to say that the sample having the largest energy measure was the one that contained the signal. This decision rule maximizes the probability of a correct decision. Thus the probability that this decision is correct is simply the probability that a random sample from the signal plus noise distribution will be larger than the drawing from the distribution of noise alone.

To determine this probability it is convenient to consider the difference between the distributions of signal plus noise and noise alone. When the drawing of signal plus noise is greater than the drawing from noise alone, the difference between the signal plus noise sample minus the noise alone sample will be greater than zero. Thus, the problem is to determine the distribution of the quantity, $(P_{S+N}) - (P_N)$.

Now since both P_{S+N} and P_N have normal distributions, their difference also has a normal distribution. The mean of the differences is the difference of the means, and since the samples are taken at two different intervals in

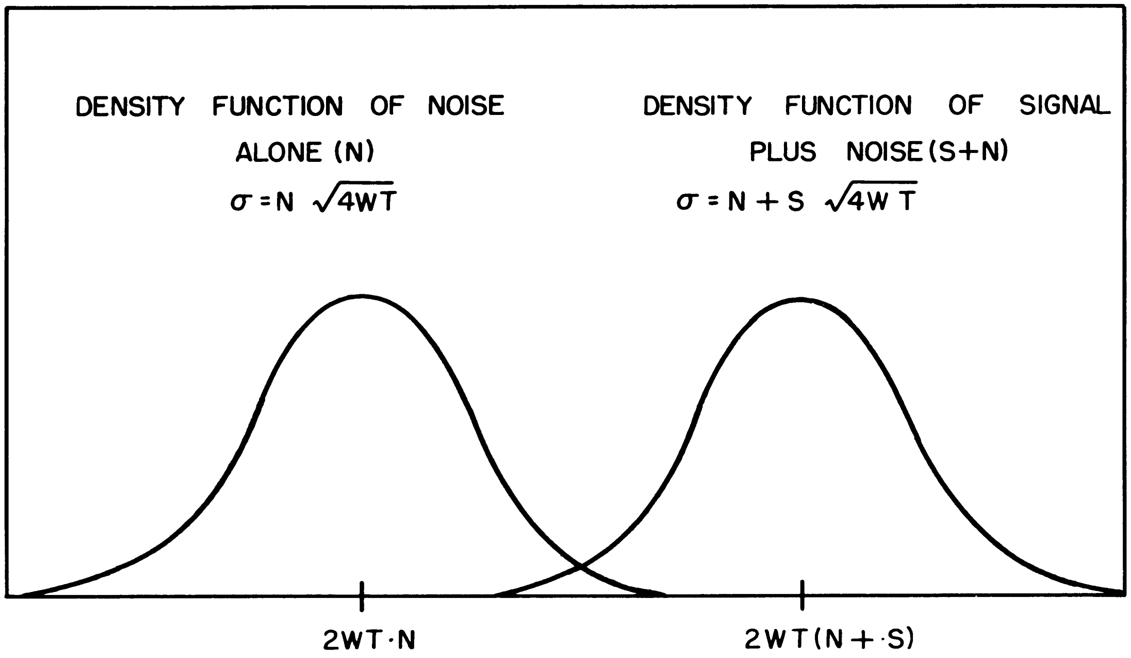


FIG.8 THE DISTRIBUTION OF POWER IN A SAMPLED WAVEFORM FOR A SIGNAL WHICH IS A SAMPLE OF NOISE

time they are uncorrelated; therefore, the variance of the difference is the sum of the individual variances. The probability of a correct decision is then the probability that a drawing from this normal deviate will be greater than zero. It is convenient to normalize this difference distribution so that the variance is unity. The mean of this normalized distribution is then:

$$M = \sqrt{\frac{WT}{2}}\ \frac{S}{N} \frac{1}{\sqrt{\frac{1}{2}\left(\frac{S}{N}\right)^2 + \frac{S}{N} + 1}} \ . \tag{20}$$

Both the signal power (S) and the noise power (N) are measured in the same bandwidth. Therefore, the noise power density, or spectrum level of the noises, may be used in Equation (20). Since $N_o W = N$, and $S_o W = S$, the equation may be rewritten as

$$M = \sqrt{\frac{WT}{2}}\ \frac{S_o}{N_o} \frac{1}{\sqrt{\frac{1}{2}\left(\frac{S_o}{N_o}\right)^2 + \frac{S_o}{N_o} + 1}} \ . \tag{21}$$

This is the form that is used in analyzing the data.

Let the quantity $d'_{opt.}$ (d'optimum) be defined in the following way.

$$d'_{opt} = \sqrt{2}M \tag{22}$$

The quantity d' has been introduced in other papers on signal detection (5, 23, 24). Previously this quantity was used only when the distribution of signal plus noise and noise alone both had the same variance. For the types of signals considered in the present paper this condition may not be true. However, when the ratio of signal power to noise power per unit bandwidth becomes small (i.e., ($S_o/N_o << 1$), the radical in Equation 21 becomes unity. Then d'_{opt} has the following value.

$$d'_{opt} = \sqrt{WT}\ \frac{S_o}{N_o}\ , \quad \frac{S_o}{N_o} << 1 \qquad (23)$$

In this case the variance of both the signal plus noise and noise alone distributions is nearly the same (see Figure 8). This is a convenient definition, since the quantity d'_{opt} is consistent with the previous definitions as S_o/N_o approaches zero.

Using either the quantity M or d'_{opt}, a table of the normal probability integral can be used to determine the expected percentage of correct detections for the ideal receiver. For example, at the value of ($d'_{opt} = 1$) the expected percentage correct is 75 percent.

While the mathematics have become somewhat imposing, the rational for these equations is really quite simple. Consider Equation 23: from this equation it is seen that the quantity d'_{opt} is monotonically related to the signal-to-noise ratio. Also, given a fixed signal-to-noise ratio, d'_{opt} increases as the square root of the number of sampling points. This corresponds to the familiar statistical result that the standard deviation of the sample mean decreases as the square root of the number of samples. In the types of signals considered in this paper more sample values can be obtained by extending either the duration or the bandwidth of the signal. This result can be anticipated, since symmerty between the time and the frequency domain exists in an ideal receiver.

While the preceding derivations are strictly true only for the ideal receiver, the general results are also applicable to a variety of receivers that are less than optimum. Any receiver which first filters the incoming waveform, and passes it through a non-linear device that provides an output monotonic with power, should behave roughly the same as the ideal receiver.

In particular, if this non-optimum receiver increases its bandwidth to be more nearly equal to the signal bandwidth, the detectability of the signal should increase. Obviously, if the receiver bandwidth is increased beyond the bandwidth of the signal, a decrease in performance can be expected, since the receiver is only listening to more noise which does not contain signal energy.

While it is hardly expected that the human observer will perform as well as the ideal receiver, it is not unreasonable to expect that the human will perform like the ideal receiver with respect to certain parameters, such as signal bandwidth and duration. At the very least, such a hypothesis seems to be worth investigating. The first experiment was designed to determine if the signal bandwidth, or more precisely, if the observer's critical band, influenced the detection of these noise signals.

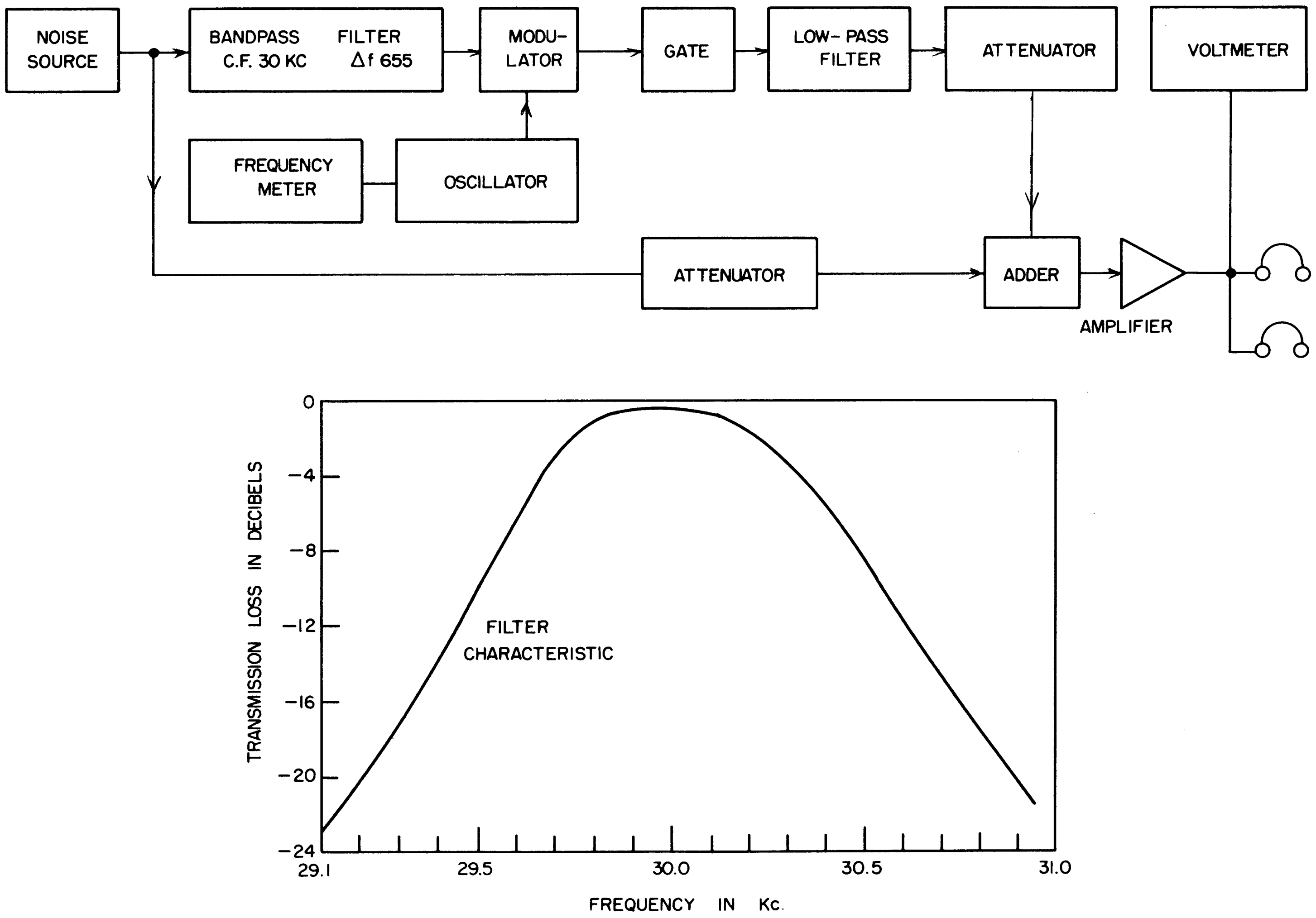
3.3.3 Experiment II-A. In this first experiment the detectability of a noise signal with a bandwidth of about 600 cps was determined as a function of the center frequency of the signal. It is well known that the detectability of a sine wave signal masked by white noise decreases as the frequency of the tone is raised. This is usually attributed to the fact that the critical bands are wider at the higher frequencies. For these pure tone signals, according to the critical band theory, the observer is forced to listen to more noise at the higher frequencies. The signal-to-noise ratio is therefore reduced, and higher frequency signals are harder to hear. When applying this model to the detection of noise signals masked by noise, there are two theoretical approaches which may be employed as discussed earlier.

The first approach maintains that only one critical band may be used to detect signals. All of the estimates of critical bandwidth from masking experiments claim that the critical band will always be smaller than the bandwidth of this 600 cps signal. However, the critical bandwidth will be wider at the higher frequencies. Thus, according to the mathematical model presented earlier, more samples should be obtained at the high frequencies than at the lower frequencies, and the bands of noise at the higher frequencies will be easier to hear than the bands of noise centered at the lower frequencies. According to Fletcher's estimates of critical bandwidth, and also according to the variation suggested by Equation 21, the signal noise centered at 400 cps will be about 8 db harder to hear than a band centered at 7500 cps.

The second approach maintains that the critical bandwidth obtained from the pure tone data are simply minimum bandwidths. By employing several bands in combination, the effective bandwidth of the auditory system can be adjusted to any width greater than this minimum value. Therefore, for this 600 cps noise signal, the auditory bandwidth should be matched to the bandwidth of the signal, and, following the statistical model, the detectability of the signal will then be independent of center frequency. Furthermore, if a wider band signal is used, one which covers the entire region, the detectability of this signal may be explained by an equation like that presented in the earlier discussion (Equation 21).

The first experiment provided a test of these views. This experiment was conducted by the author at the Bell Telephone Laboratories, Murray Hill, New Jersey, in 1957.

3.3.3.1 <u>Equipment.</u> The audio circuit is shown in Figure 9. The noise source was a General Radio Model 1380-A; this provided a Gaussian output with an approximately flat power output up to 200 kc. In order to change the center frequency of the filtered signal without changing the linear frequency characteristics, a modulation scheme was used. The noise was filtered



2659 A-G-22 PS 2-27-58

FIG.9 BLOCK DIAGRAM OF EQUIPMENT USED IN THE FIRST PART OF THE NOISE IN NOISE EXPERIMENT at a center frequency of 30 kc with an equivalent square bandwidth of 655 cps. The filter characteristic is displayed in Figure 9. By setting the oscillator at a frequency x cps above 30 kc, the audio output at the headphones would be a band-passed noise signal with a center frequency of x cps and a bandwidth which is, on a linear frequency scale, independent of center frequency. The total power of the signal is, therefore, independent of the center frequency of the band. Frequency measurement of the carrier oscillator was accomplished by employing a Berkeley Model 5510 Universal Counter and Timer. A true rms meter was used to establish the noise power per unit cycle for signal and noise. The background masking noise power per cycle was equivalent to a sine wave at 1000 cps set at about 40 db re 0.0002 dyne/cm^2.

3.3.3.2 Procedure and Recording Apparatus. The stimuli were presented binaurally through Permoflux PDR-8 headphones. Throughout the test block, the subjects heard wide-band noise (the masking noise). Each subject was seated in front of a small box. Four lights flashed on the box in sequence. The first light warned the subject when a test cycle was to begin. During this warning light, the signal was presented in exactly the same fashion as it would occur in one of the two following intervals. This interval, therefore may be interpreted as a preview of the signal to be detected. The second and third lights then flashed in succession. During one of these intervals the signal was presented. The selection of the interval in which the signal was presented in each cycle was determined from a table of random numbers which was programmed from a tape reader. The fourth light notified the subject of the answer interval. At this time, each subject was to indicate, via pushbottons, the interval in which he believed the signal occurred. After the answer interval, the subjects were notified in which interval the signal had actually been presented. The cycle was then repeated. 150 such cycles constituted a trial block. During each trial block, the signal and noise parameters were fixed. At each center frequency, five signal levels were employed, and two trial blocks were run at each level. All trial blocks were run at a single center frequency before the center frequency was changed. Before each trial block, the subjects listened to several practice trials to acquaint them with the stimulus parameters during the test session. Random sequences determined the order of signal levels as well as the order of center frequencies. Each cycle required about four seconds, permitting a trial block to be completed in ten minutes. The observers, who were both female employees of Bell Telephone Laboratories, Inc., were unfamiliar with the nature of the problem. They worked one hour in the morning and one hour in the afternoon. About 1000 observations were taken per day. A rest period of at least five minutes was allowed between sessions. The data were recorded on counters which tabulated the number correct in each interval; thus a complete matrix of the subjects' responses and the stimuli presented was available.

3.3.3.3 Measured Quantities. The results will be presented in terms of the following quantities. N_o is the noise power per cycle (spectrum level) of the masking noise. S_o is the noise power per cycle of the signal noise. W_s is the bandwidth of the signal. The bandwidth will be specified in terms of the equivalent square bandwidth. The equivalent square bandpass is a mathematical convenience. It is a mathematical filter which has a rectangular frequency characteristic. This mathematical filter's height is equal to the maximum gain of the physical filter, and its width is such that total power output from each filter is the same for a white Gaussian noise input. T is the duration of the signal. N_o and S_o have the dimensions of energy. The ratio of these quantities is, of course, dimensionless. $(S_o \cdot W_s)$

is the total power of the signal. The total energy of the signal is (S_oW_sT).

3.3.3.4 Results. The results of this experiment are presented in Figures 10 and 11. The data for all the conditions of the experiment are presented for each observer. The ordinate is the percentage of correct decisions in the two-alternative forced-choice tests. Three hundred observations were used to define each point. The abscissa is the logarithm of d'_{opt}. This scale is used because it permits one to plot the results of both the narrow-band and wide-band noise signal on the same graph. The solid line represents the performance one would expect from the ideal receiver. The observer's performance departs from that which would be obtained with an ideal receiver in two ways. The observer's performance is about 5 to 6 db less than optimum at the 75 percent correct detection point. Also, the observers appear to have a slightly steeper psychophysical function than that which would be obtained with the ideal receiver. This finding is reminiscent of the data obtained with pure tones masked by noise. In the experiments using pure tones as the signal the observers' performance approaches the ideal performance more closely for the more intense signals (5).

Despite these differences between the ideal performance and the observer's performance, the parameter bandwidth appears to influence both in the same manner. The change in bandwidth from about 650 cps to 5100 cps can be accounted for by the theory.

Figure 12 is a summary of the data just discussed. In this figure the ratio of signal-to-noise power in a one cycle band necessary for 75 percent correct detections is plotted against the center frequency of the narrow-band noise signal. The 75 percent point was chosen since, for the ideal receiver, it corresponds to a d'_{opt} of unity. The values were obtained by plotting the psychophysical function appropriate for each condition. A smooth line was

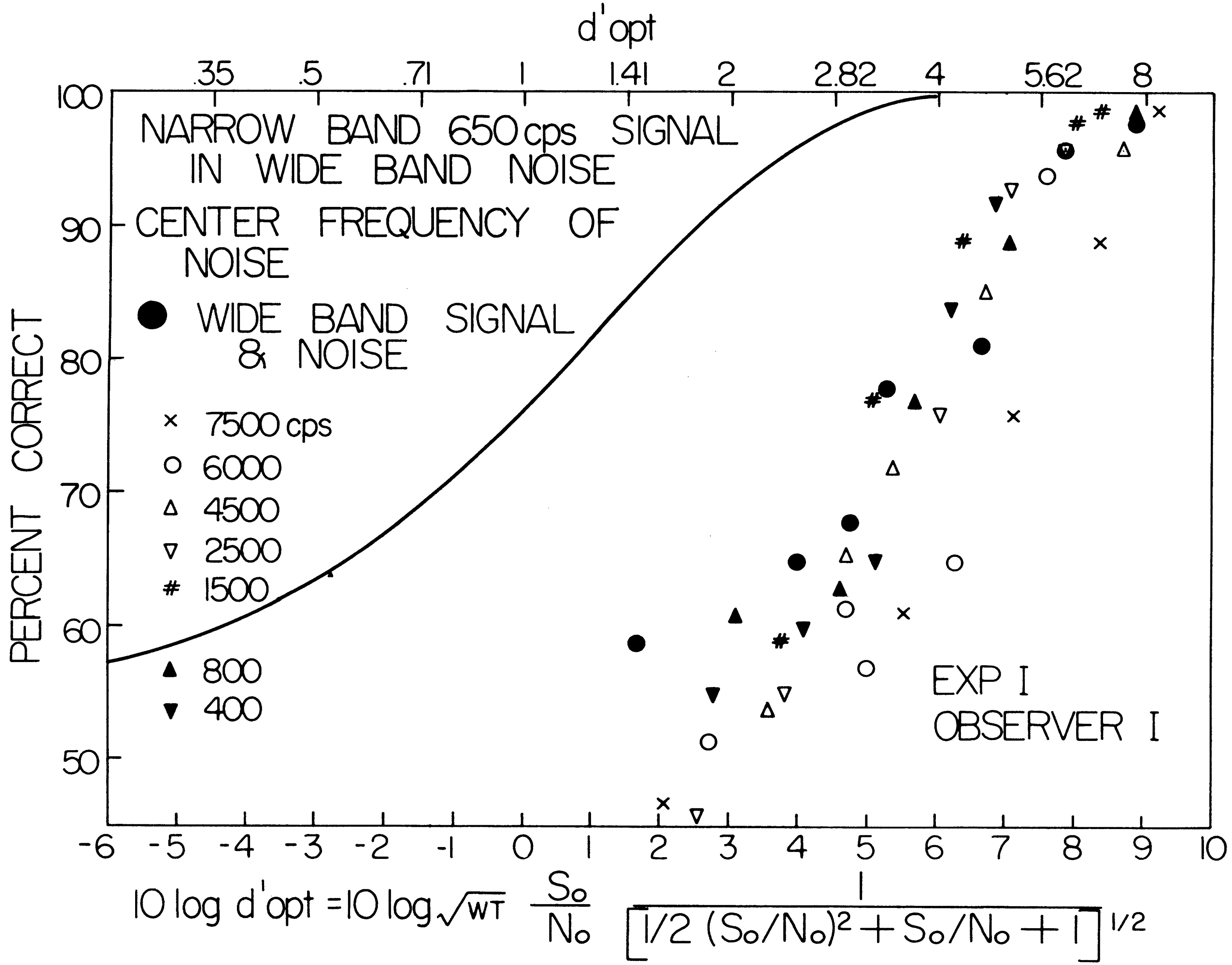
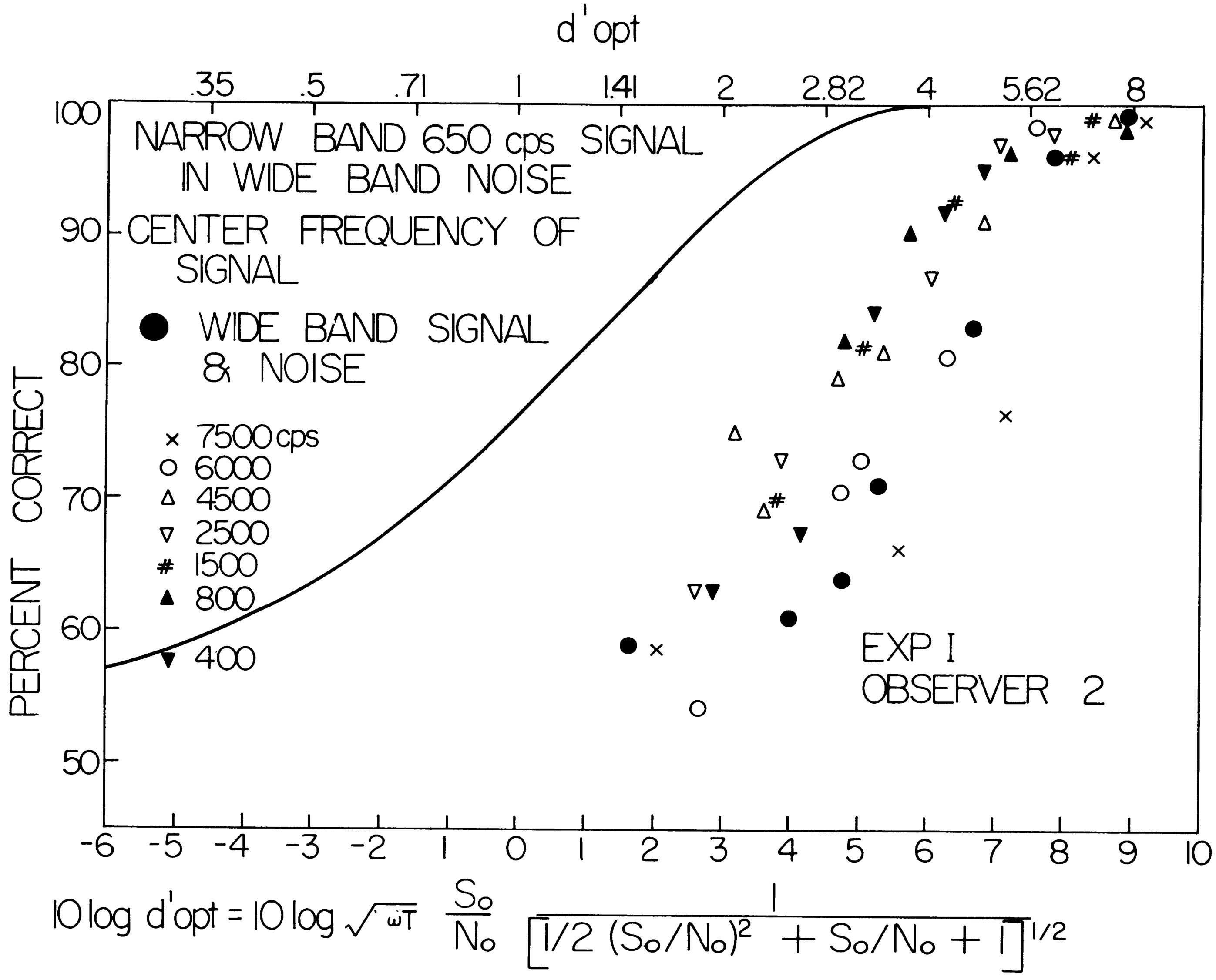


FIG. 10 THE RESULTS OF THE FIRST PART OF THE NOISE IN NOISE EXPERIMENT — OBSERVER I

The ordinate is the percent of correct obtained in a two choice forced choice test. The abscissa represents a transformation of the physical parameters used in the test. The solid curve represents the expected behavior of the ideal detector.

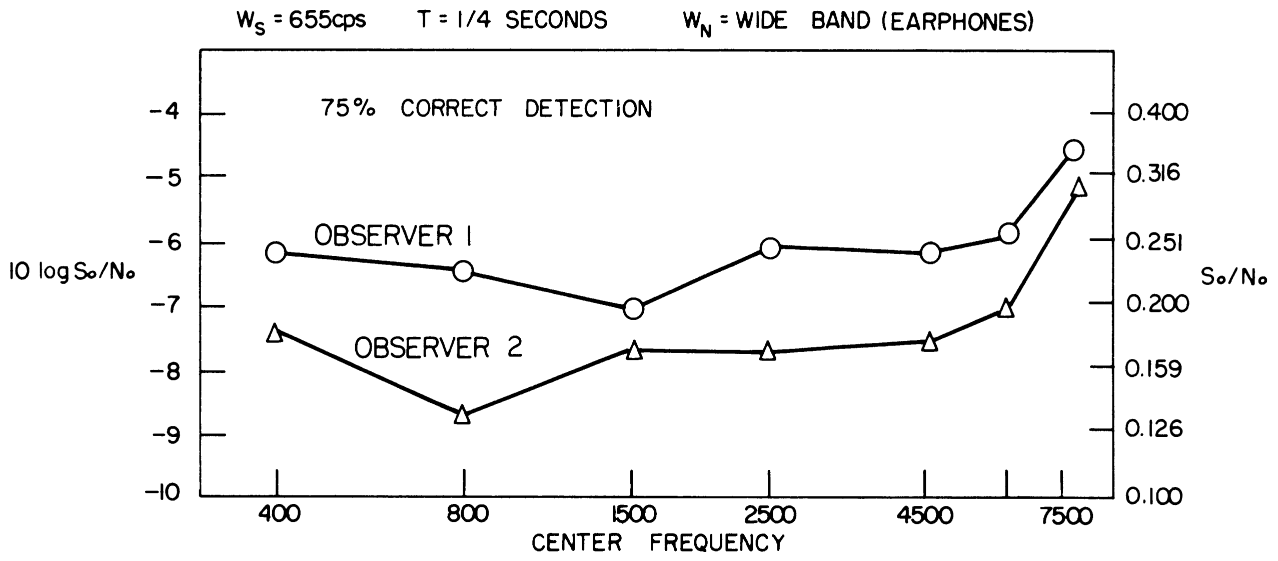


2659 A-G-15 PS 2-10-58

FIG. 11 THE RESULTS OF THE FIRST PART OF THE NOISE IN NOISE EXPERIMENT — OBSERVER 2

The ordinate is the percent of correct obtained in a two choice forced choice test. The abscissa represents a transformation of the physical parameters used in the test. The solid curve represents the expected behavior of the ideal detector.

NARROW BAND SIGNAL IN WIDE BAND NOISE



WIDE BAND SIGNAL IN WIDE BAND NOISE , $W_S = W_N$ = 5143cps

T = 1/4 SECONDS

	OBSERVER 1		OBSERVER 2	
	10 log So/No	So/No	10 log So/No	So/No
FOR 75% CORRECT	−10.0	.100	− 9.6	.1096

FIG. 12 SUMMARY OF THE DATA OBTAINED IN THE FIRST PART OF THE NOISE IN NOISE EXPERIMENT

The points represent the Signal to Noise power per unit bandwidth necessary to obtain 75% correct detections. The abscissa is the center frequency of the band of noise used as the signal. The table below is the summary of the data when a wider bandwidth is used for the signal.

fitted to the data by eye, and the 75 percent point was determined. Clearly, for the narrow-band signal, the center frequency of the signal had little to do with the signal's detectability. This result is consistent with the second theoretical approach discussed earlier; namely, the observer either widens the critical band to match the signal bandwidth or, equivalently, linearly combines the outputs of several critical bands to obtain the same result. Only the data obtained at a center frequency of 7500 cps seems to disagree with this conclusion. Here the frequency response characteristics of the headphones certainly contribute heavily to the results. Anomolous data at such frequencies have been obtained in two other studies (7, 14). No complete explanation of such a result can be suggested at this time.

The table beneath Figure 12 expresses the same data for the wide-band signal. These data were obtained by passing both signal and noise through the same filter; hence these data represent the differential intensity sensitivity for a noise stimulus. S_o/N_o is thus equivalent to the Weber function $\Delta E/E$ for sound energy. In order to avoid confusion it should be pointed out that the measure S_o/N_o is independent of signal bandwidth. This ratio is about .251 for Observer I with the 655 cps bandwidth signals, and about .10 for the 5134 cps bandwidth signal. In terms of the total energy of the signal ($W_s S_o T$), the wide band signal had a total energy of about 3.3 times the total energy of the narrow band signal.

3.3.4 <u>Experiment II-B</u>. This experiment was undertaken at the Psychophysical Laboratory* with different observers than those employed in the previous experiment. The previous experiment indicated that the observer behaved consistently with the equation for the ideal observer, at least with

---

* Electronic Defense Group, University of Michigan, Ann Arbor, Michigan.

respect to the signal bandwidth. The parameter investigated in the present experiment is signal duration. In this experiment, both signal and noise were passed through the same filter. The equivalent square bandpass of this filter was about 3800 cps. Six signal durations were used. Signal duration and power were the independent variables. At each duration, a function relating signal power to the percentage of correct detections was obtained. A two-alternative forced-choice test was used. The experimental procedure and the equipment used have been reported in Ref. 5. Only one change was made in implementing the present test; instead of using an oscillator to generate the signal, a noise source was employed. The signal and the noise were filtered at the input to the power amplifier. The masking noise level was about 50 db re 0.0002 dyne/cm^2 in a one cycle per second band.

Figures 13, 14 and 15 show the raw data obtained in this experiment. As in Figures 10 and 11, the percentage of correct detections is related to the physical measure, d'_{opt}. About 300 observations were used to define each point. In some of the conditions some of the observers were absent so that less than this number of observations was obtained. These data are not presented on the graphs. Except for an increase in the variability of the data, which the author is unable to explain, the results are similar to those obtained in the previous experiment. A summary of the data is presented in Figure 16. The points represent the signal-to-noise ratio in a one cycle band necessary to obtain 75 percent correct detection. The solid curve represents the same data which should be obtained by the ideal receiver. It is seen that the difference between the ideal receiver and the observers' performance is about 5 to 6 db. At the 1 sec. duration, Miller (16) has obtained similar data. He found the ratio of S_o/N_o to be .102 for one subject, and .107 for another, at this noise level.

2659 A-G-17 PS 2-10-58

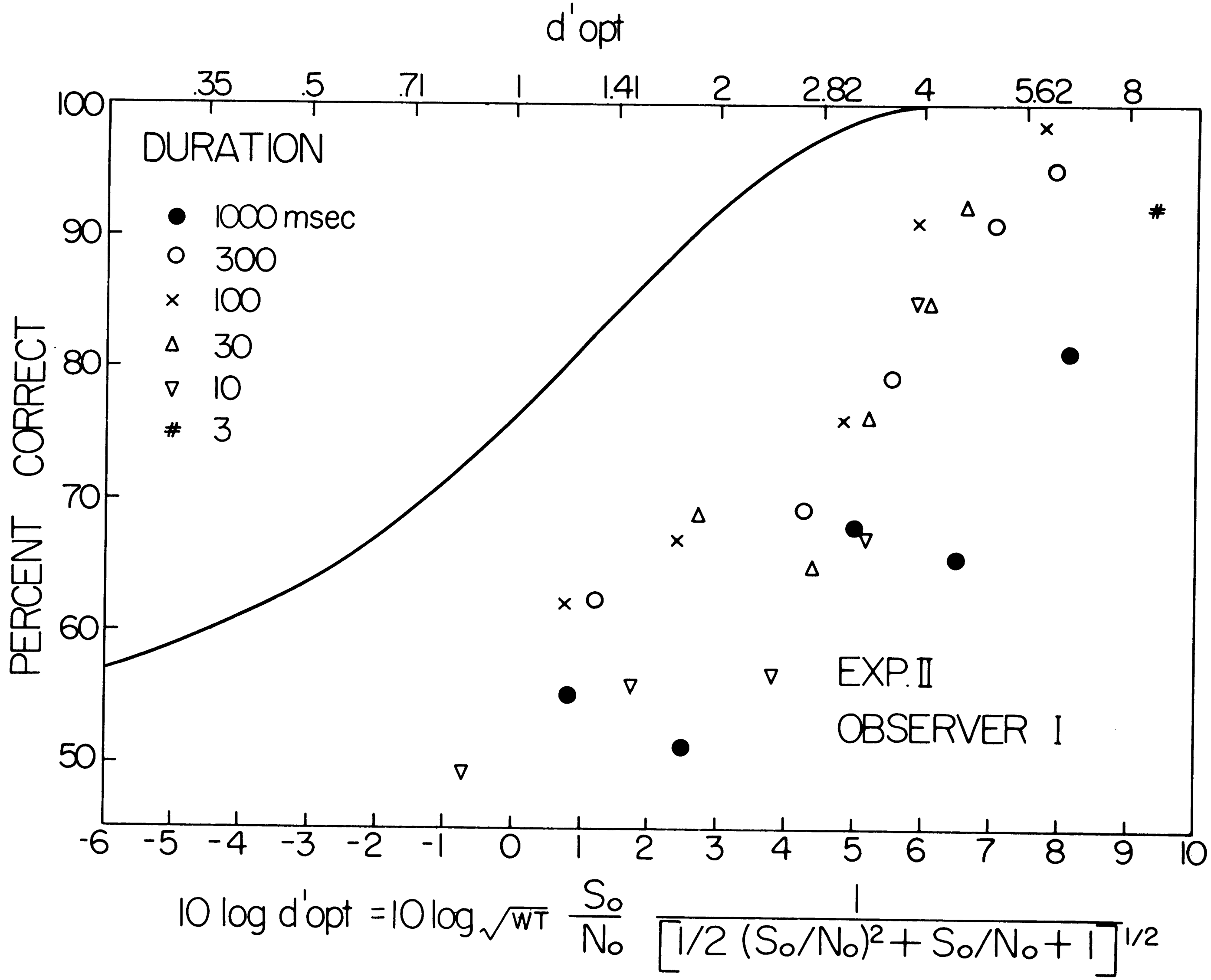


FIG. 13 THE RESULTS OF THE SECOND PART OF THE NOISE IN NOISE EXPERIMENT — OBSERVER I

The ordinate is the percent of correct obtained in a two choice forced choice test. The abscissa represents a transformation of the physical parameters used in the test. The solid curve represents the expected behavior of the ideal detector.

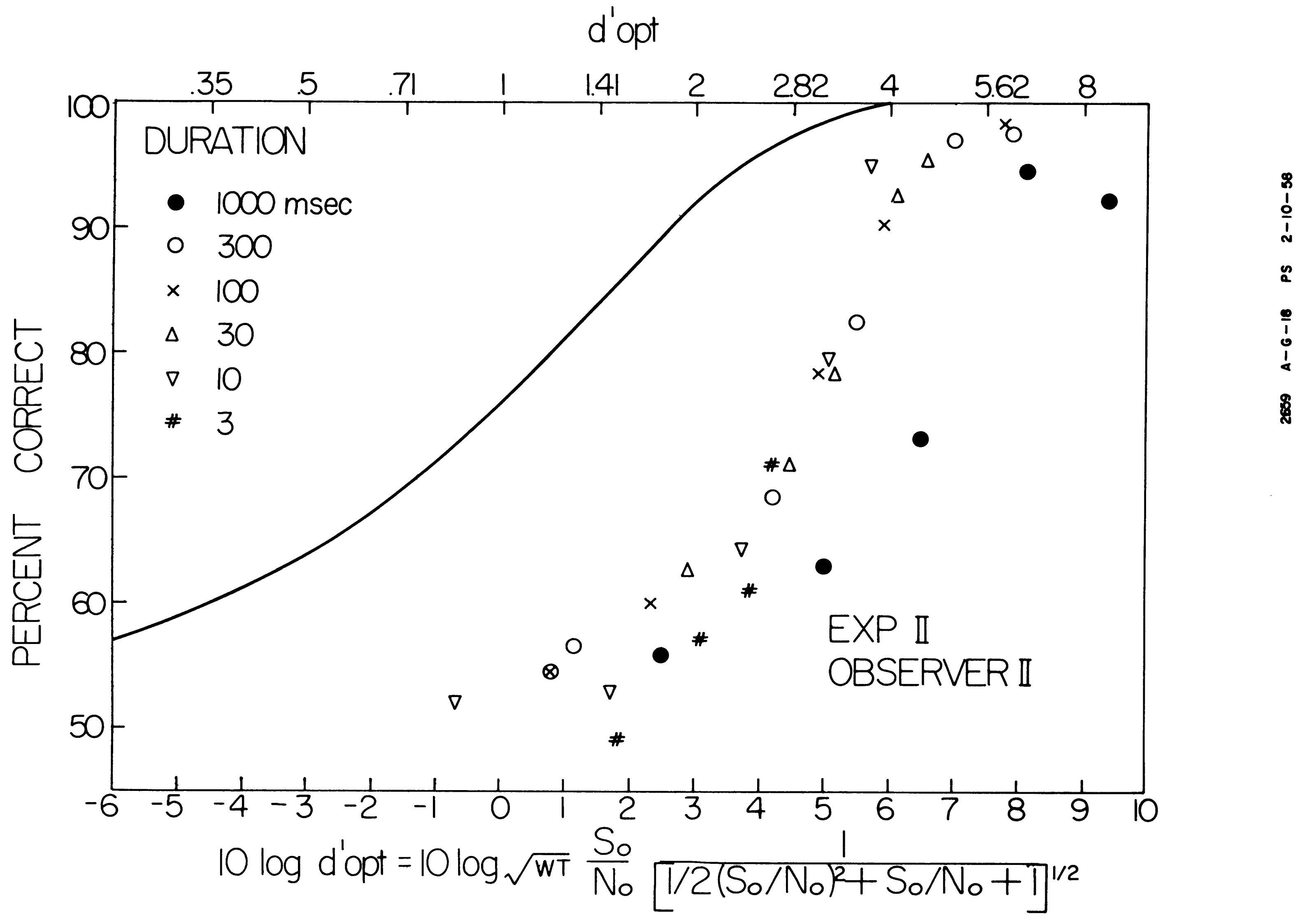


FIG. 14 RESULTS OF THE SECOND PART OF THE NOISE IN NOISE EXPERIMENT - OBSERVER 2

The ordinate is the percent of correct obtained in a two choice forced choice test. The abscissa represents a transformation of the physical parameters used in the test. The solid curve represents the expected behavior of the ideal detector.

2659 A-G-19 PS 2-10-58

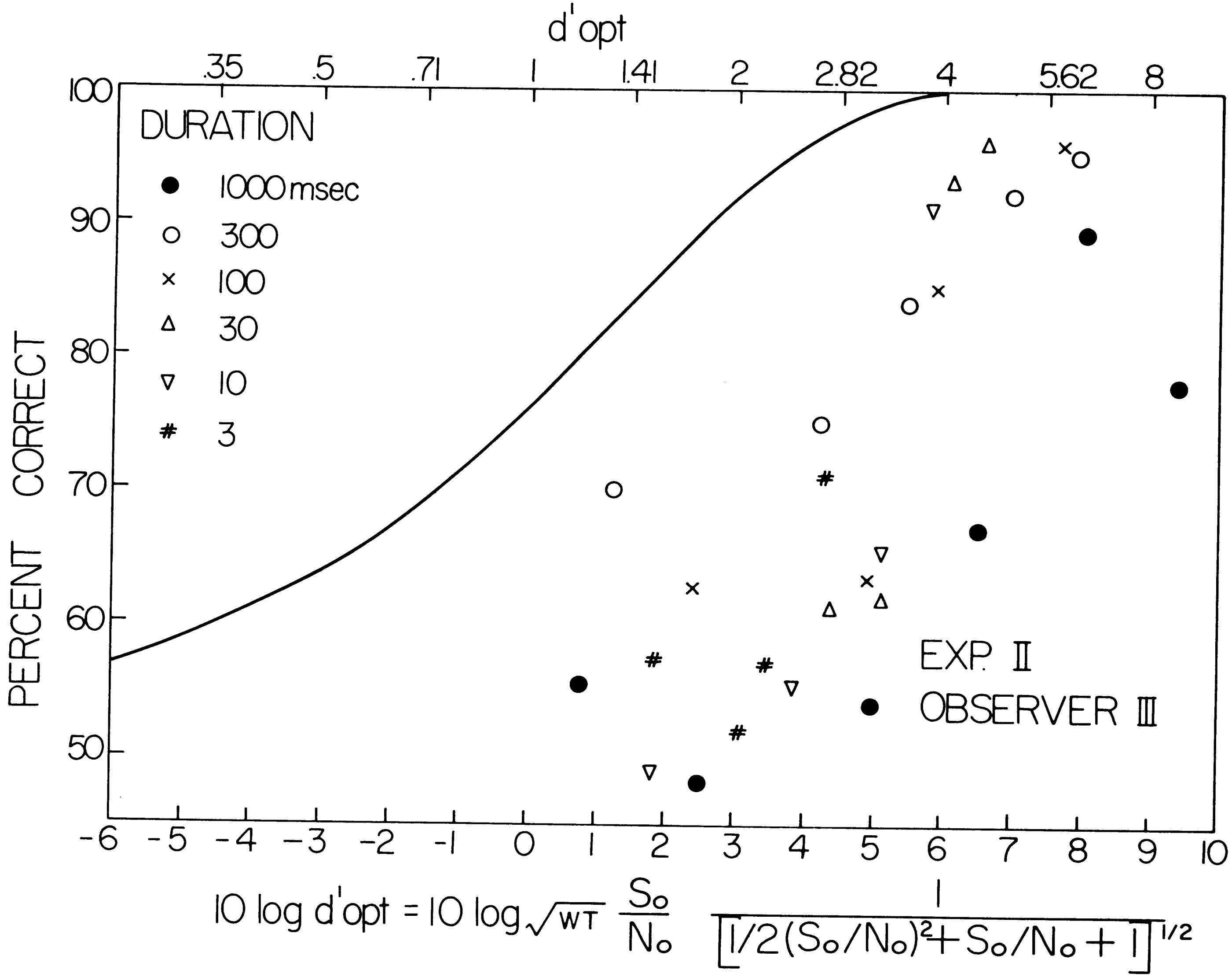


FIG. 15 RESULTS OF THE SECOND PART OF THE NOISE IN NOISE EXPERIMENT - OBSERVER 3

The ordinate is the percent of correct obtained in a two choice forced choice test. The abscissa represents a transformation of the physical parameters used in the test. The solid curve represents the expected behavior of the ideal detector.

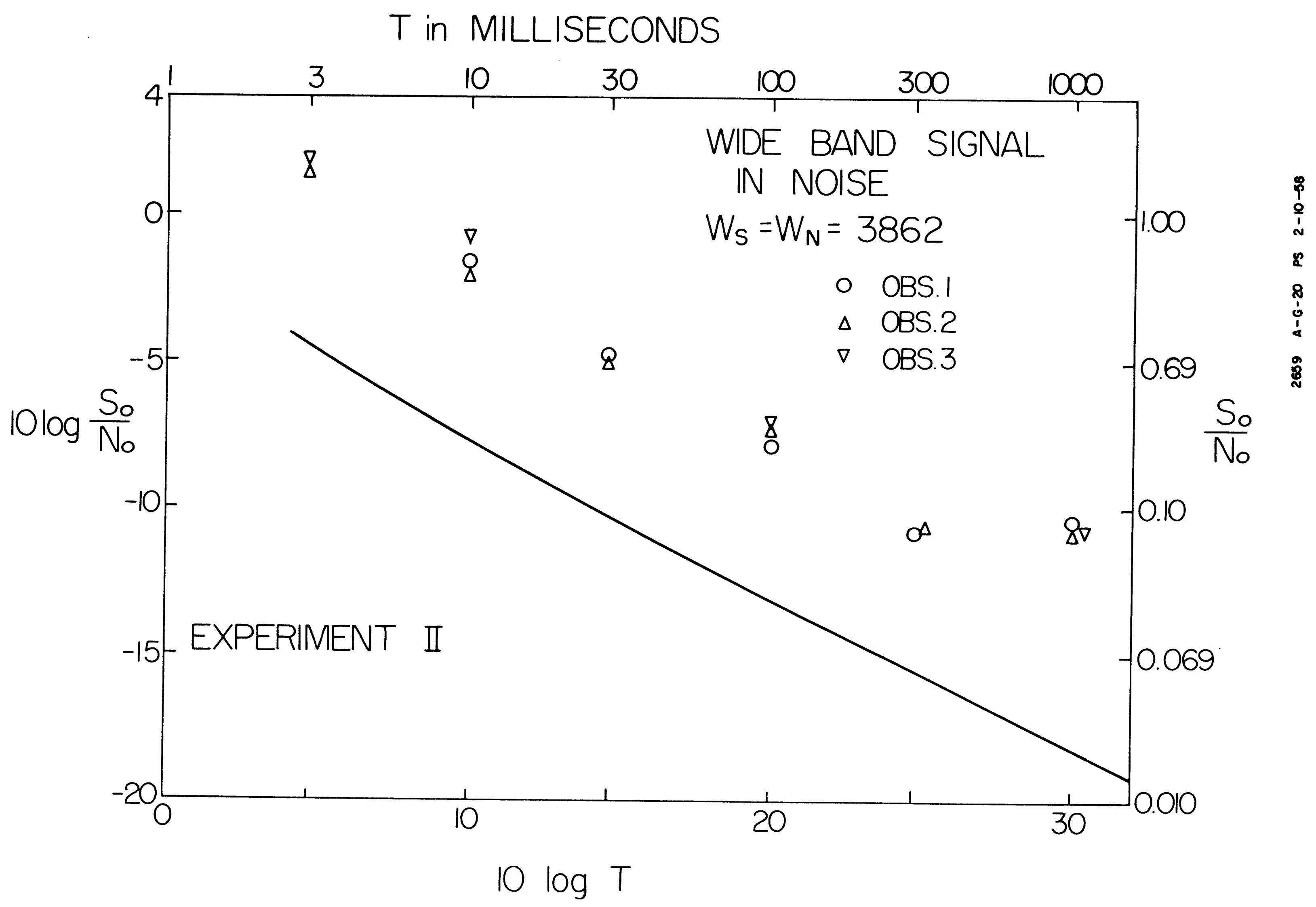


FIG. 16 SUMMARY OF THE DATA OBTAINED IN THE SECOND PART OF THE NOISE IN NOISE EXPERIMENT

The points represent the ratio of signal to noise power per unit bandwidth necessary for 75% correct detections. The abscissa is the duration of the signal. The solid curve represents the expected behavior of the ideal detector.

The statistical decision model appears to account for the variation in the observers' performance as a function of the signal duration. One would expect that if even longer durations had been used, this relationship might fail for then the memory capacity of the observer would be severely taxed. The data from the 1 second signal duration suggest such a result.

3.4 Summary and Conclusions

For those situations where some constant detectability is determined (e.g., threshold), the optimum receiver and the observer differ by some constant number of decibels which depends only upon the definition of the threshold point. Using the 75 percent point, all of the observers are about 6 db below optimum for all of the conditions encountered in both experiments. The change in the percentage of correct detections as a function of signal power density is not the same for the ideal receiver and for the human observer. There are, of course, many factors which could lead to this discrepancy. Until further experiments are undertaken the author will refrain from speculating on this point.

The results are completely consistent with the assumption that the bandwidth of the auditory system can be adjusted over some range so as to match the bandwidth of the signal to be detected. This is, in the author's opinion, the most important result of the study. In many cases the auditory system is assumed to be a fixed physical system. One then attempts to extrapolate certain parameters of the system, which, since it is fixed, should generalize to many experiments. Often such attempts fail. The model presented in this paper assumes that the auditory bandwidth is a variable, rather than a fixed parameter. In order to obtain some generality it was necessary to assume that the parameter is changed so as to maximize the percentage of correct detections. This experiment demonstrates the success of such an approach.

CHAPTER IV

THE IMPLICATIONS OF THE RESEARCH

Two types of signals were investigated: one was a combination of two sine waves, and the other was a band of white Gaussian noise. Both of these signals may be characterized by a frequency spectrum which is wider than the nominal width of the critical band. An adjustable bandpass model appears to provide good predictions of the detectability of these signals. At the very least, this model is clearly superior to a single, fixed, critical band model.

The adjustable bandpass model claims that two or more critical bands may be used simultaneously to detect these complex signals. No restriction is imposed as to how these multiple bands may be selected; only two bands which are widely separated in frequency might be employed. The hearing mechanism is, therefore, likened to a collection of simple, tuned filters whose outputs may be added together. By employing appropriate weighting constants, this linear sum provides a single combined output.

The choice of the combination to be used and the constants to be selected is determined by some higher-order system. The model assumes that the combination is determined so as to maximize the probability of a correct detection in a particular experimental situation. Such a scheme is quite different from the previous notion of the critical band concept, especially that advanced by Fletcher. In his work there appears the definite suggestion that the critical bands are simple reflections of the geography of the basilar membrane. Such a fixed system is not consistent with the model presented by the present author, or the results obtained in the two experiments reported in Sections 3.2 and 3.3.

The adjustable bandpass model is, in essence, a model which claims that parameters of the hearing mechanism are adjustable. Such a view could, therefore, provide the flexibility necessary to explain some of the inconsistencies which were pointed out in the review of the literature of the critical band concept (Chapter II). While the adjustable bandpass model has not been extended to cover these inconsistencies, it is apparent that, at least, such a model has the potential for doing so. One method of reconciling divergent results is to claim that the mechanism somehow changes from one experiment to the next. The model presented provides a rational and rigorous means of approaching such a problem.

This model is, of course, more complicated than the simple, fixed-bandpass model. While this is unfortunate, it seems that the data have already forced at least this level of complication upon us. The maximization assumption somewhat offsets this disadvantage, since, unlike some more complicated models, this one does not cost the experimenter any more degrees of freedom. Indeed, for the noise signals, only one free parameter is employed.

Finally, the use of a model such as the adjustable bandpass model provides a natural division of what may be called the sensory and perceptual categories. The maximization assumption serves to link the two. Such a theoretical division of the two concepts, with some method of linkage, even if not correct, is certainly a virtue of this approach.

BIBLIOGRAPHY

1. Fletcher, H., "Auditory Patterns," Rev. Mod. Phys., Vol. 12, 47-65 (1940).

2. Fletcher, H., Speech and Hearing, D. VanNostrand Company, Inc., New York (1953).

3. Fletcher, H. and Munson, W. A., "Relation Between Loudness and Masking," Jour. Acoust. Soc. Am., Vol. 9, 1-10, (1937).

4. French, N. R., and Steinberg, J. C., "Factors Governing the Intelligibility of Speech Sounds", Jour. Acoust. Soc. Am., Vol. 19, 90-119 (1947).

5. Green, D. M., Birdsall, T. G., Tanner, W. P., Jr., "Signal Detection As a Function of Signal Intensity and Duration," Jour. Acoust. Soc. Am., Vol. 29, 523-531, 1957.

6. Harris, J. D., "Pitch Discrimination," Jour. Acoust. Soc. Am., Vol. 24, 750-55, 1952.

7. Hawkins and Stevens, "The Masking of Pure Tones and of Speech by White Noise", Jour. Acoust. Soc. Am., Vol. 22, 6-13, (1950).

8. Helmholtz, H. L. F., On the Sensations of Tones, Dover Publications, Inc., New York, 1954.

9. Howes, D. H., "The Loudness of Multicomponent Tones", Amer. Jour. of Psych., Vol. 63 (1950).

10. Karoly, A. J., and Isaacson, R. L., "Scanning Mechanisms in Audition", Paper presented at Michigan Academy of Science, Ann Arbor, Michigan, 1956.

11. Knudson, V. O., "The Sensibility of the Ear to Small Difference of Intensity and Frequency", Phys. Rev., Vol. 21, 84-102, (1923).

12. Koch, W. E., "A New Interpretation of the Results of Experiments on the Differential Pitch Sensitivity of the Ear", Jour. Acoust. Soc. Am., Vol. 9, 129-134, (1937).

13. Licklider, J. C. R., "Basic Correlates of the Auditory Stimulus", in Handbook of Experimental Psychology, S. S. Stevens Editor, Wiley and Sons, Inc., New York, 1951.

14. Licklider, J. C. R., Green, D. M., McKey, M., "The Detection of Masked Tones as a Function of Frequency", Mass. Inst. of Tech. (in preparation).

15. Marill, T. "Psychophysics and Signal Detection Theory", PhD. Thesis, Massachusetts Institute of Technology, 1956.

16. Miller, G. A., "Sensitivity to Changes in the Intensity of White Noise and Its Relation to Masking and Loudness", Jour. Acoust. Soc. Am., Vol. 19, 609-619, July, 1947.

17. Munson, W. A., Personal Communications.

18. Peterson, W. W., and Birdsall, T. G., "Theory of Signal Detectability", Part I and II, Technical Report No. 13, Electronic Defense Group, University of Michigan, 1953. Also, available in 1954 Symposium on Information Theory, Transactions of the IRE, September 1954.

19. Schafer, T. H., and Gales, R. S., "Auditory Masking of Multiple Tones by Random Noise", Jour. Acoust. Soc. Am., Vol. 21, 392-398, 1949.

20. Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O., "The Frequency Selectivity of the Ear as Determined by Masking Experiments", Jour. Acoust. Soc. Am., Vol. 22, 490-496, 1950.

21. Shower and Biddulph, "Differential Pitch Sensitivity of the Ear", Jour. Acoust. Soc. Am., Vol. 3, 275-287, 1931.

22. Tanner, W. P., Jr., "Theory of Recognition", Jour. Acoust. Soc. Am., Vol. 28, 882-888, September 1956.

23. Tanner, W. P., Jr., and Swets, J. A., "A Decision-Making Theory of Visual Detection", 401-09, Psychol. Rev., Vol. 61, 1954.

24. Tanner, W. P., Swets, J. A., Green, D. M., "Some General Properties of the Hearing Mechanism", Technical Report No. 30, Electronic Defense Group, Engineering Research Institute, University of Michigan, Ann Arbor, Michigan.

25. Terman, F. E., Radio Engineering, McGraw-Hill Book Company, Inc., New York, Third Edition.

26. Webster, J. C., Miller, P. H., Thompson, P. O., Davenport, E. W., "The Masking and Pitch Shifts of Pure Tones Near Abrupt Changes in Thermal Noise", Jour. Acoust. Soc. Am., Vol. 24, 147-152, 1952.

27. Wegel and Lane, "Auditory Masking and Dynamics of the Inner Ear," Phys. Rev., Vol. 23, 266-285, 1924.

28. Wever, E. G., Theory of Hearing, John Wiley and Sons, Inc., New York, 1949.

29. Zwicker, E., Flotlorp, B., and Stevens, S. S., "Critical Bandwidth in Loudness Sum ation", Jour. Acoust. Soc. Am., Vol. 29, 1957, 548-557.

DISTRIBUTION LIST

1 Copy	Document Room Willow Run Laboratories University of Michigan Willow Run, Michigan
12 Copies	Electronic Defense Group Project File University of Michigan Ann Arbor, Michigan
1 Copy	Engineering Research Institute Project File University of Michigan Ann Arbor, Michigan
250 Copies	Operational Applications Laboratory Air Force Cambridge Research Center Air Research and Development Command Bolling Air Force Base 25, D. C. Contract No. AF19(604)-2277
1 Copy	H. H. Goode Engineering Research Institute University of Michigan Ann Arbor, Michigan