DEFENSE RESEARCH LABORATORY

PRAESIDIUM DISCIPLINA CIVITATIS

THE UNIVERSITY OF TEXAS · AUSTIN 12, TEXAS

DRL-A-245 14 June 1965

MASKING AND BINAURAL PHENOMENA

by

Lloyd A. Jeffress

Copy No. 154

DRL Acoustical Report No. 245

Resulting from research done under U. S. Navy
Bureau of Ships Research and Development Contract
NObsr-72627, NE 051247-6, NE 051456-4, Problem 1-U,
and under U. S. Navy Office of Naval Research
Contract Nonr-3579(04), NR 142-190, on National
Aeronautics and Space Administration Fund Transfer R-129

DEFENSE RESEARCH LABORATORY
THE UNIVERSITY OF TEXAS
AUSTIN, TEXAS 78712

ACKNOWLEDGMENTS

The two chapters comprised in this report were written for publication in a book, Modern Foundations of Auditory Theory, to be edited by J. V. Tobias and E. D. Schubert, and published by The Academic Press. Work on the chapters was supported by the Bureau of Ships through Contracts NObsr 72627 and 93124 and by the National Aeronautics and Space Administration (Grant R-129) through Office of Naval Research Contract Nonr 3579(04).

I wish to acknowledge this support and also to express my thanks to a number of people who read earlier versions of the manuscripts and made many helpful suggestions: to Charles S. Watson, for continued assistance throughout; to Wilson P. Tanner, Jr., for help with TSD; to Bertram Scharf, for preventing my "critical band" widths from becoming too narrow; to George Moushegian, for keeping me up-to-date on the goings-on in the accessory olivary nucleus; to David Green and John Swets, for letting me see chapters of their forthcoming book and for much enlightening discussion; to many of my present and past graduate students, who have done most of the work; to my secretary, Mrs. Willie Herron, whose presence here made it possible to say "yes"; and finally to Jerry Tobias and Earl Schubert, who offered the opportunity.

CHAPTER 13

MASKING

Lloyd A. Jeffress

Department of Psychology and

Defense Research Laboratory

The University of Texas

TABLE OF CONTENTS

TABLE OF CONTENTS (Contd)

CHAPTER 13

MASKING

Lloyd A. Jeffress

Department of Psychology and

Defense Research Laboratory

The University of Texas

I. INTRODUCTION

A. Historical

Masking is the obscuring of one sound by another. As Licklider (1951) pointed out, it is the opposite of analysis; when we fail to hear the signal in the noise, it is because the analysis has been inadequate, or because we were not listening. Analysis implies some sort of filter system and most of our theories of hearing have been filter theories from the time of Helmholtz. We should, therefore, expect that a study of the phenomena of masking would bring us closer to an understanding of the basic problem of how we hear.

Much of the early work was conducted by the Bell Telephone Laboratories because of the close relation of masking to the problems of telephonic communication, and is summarized in Fletcher's book, Speech and Hearing (1929). Much earlier (1876), Mayer had found that a tone could be rendered inaudible by another tone of lower frequency, but not readily by one of higher frequency. Figure 1, which summarizes a series of experiments done at the Bell Telephone Laboratories by Wegel and Lane (1924), support Mayer's observation that frequencies below that of the signal are more effective in masking it than frequencies above. The figure also shows that frequencies

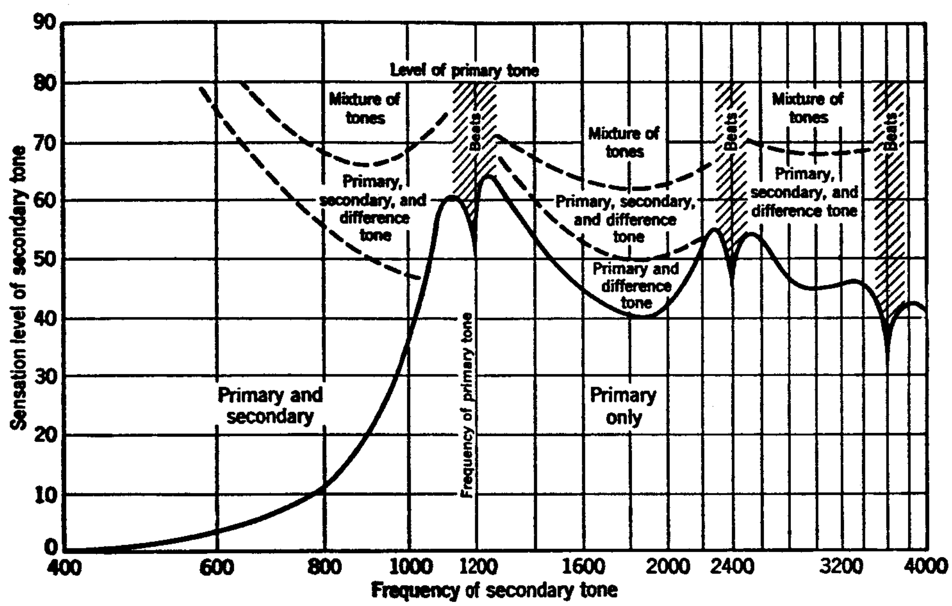

Fig. 1. The various sensations produced by a two-component tone. The primary component is a sinusoid of 1200 cps, 80 dB above threshold. The secondary component is a sinusoid of the frequency and sensation level indicated by the coordinates. When the secondary component falls below the solid curve it is masked. When the secondary component is above its masked threshold, however, the auditory sensation may be quite complex, as indicated by the descriptions in the several regions of the graph. (From S. S. Stevens, (Ed.), Handbook of Experimental Psychology, New York: John Wiley and Sons, 1951, after Fletcher, 1929, from Wegel and Lane, 1924, by permission.)

near the signal, whether above it or below, are more effective than frequencies farther removed, especially if they are higher.

The curves of Wegel and Lane are characterized by notches occuring at frequencies near that of the signal. These are the results of beats between the masker and the signal--fluctuations of level which render the signal more conspicuous and easier to detect. When steps are taken to avoid these beats by employing a narrow band of noise rather than a tone as the masker, or by using a signal duration too short to permit a full cycle of beating to occur, the notches are eliminated and the curves show a peak rather than a notch at the signal frequency. Figure 2, taken from Egan and Hake (1950) shows the masking effect of a narrow band of noise, and exhibits a peak rather than a notch at the signal frequency.

B. White Noise

Probably the most commonly used masking stimulus is white noise, noise which has a uniform power spectrum from one extreme of its frequency range to the other. The power is usually measured for a bandwidth of one cycle per second, and when expressed in decibels relative to 0.0002 microbars, is called the spectral level of the noise. A white noise therefore has the same spectral level at all frequencies within its frequency range. If we were to make a series of measurements of the instantaneous voltage associated with a band of white noise, whether wide or narrow, we would find that the mean of these voltage measurements was zero, and that the distribution around it was a normal distribution. Because of this normal or Gaussian distribution, the noise is often referred to as Gaussian, and because it often arises from thermal agitation as in a resistor, it is also referred to as thermal noise.

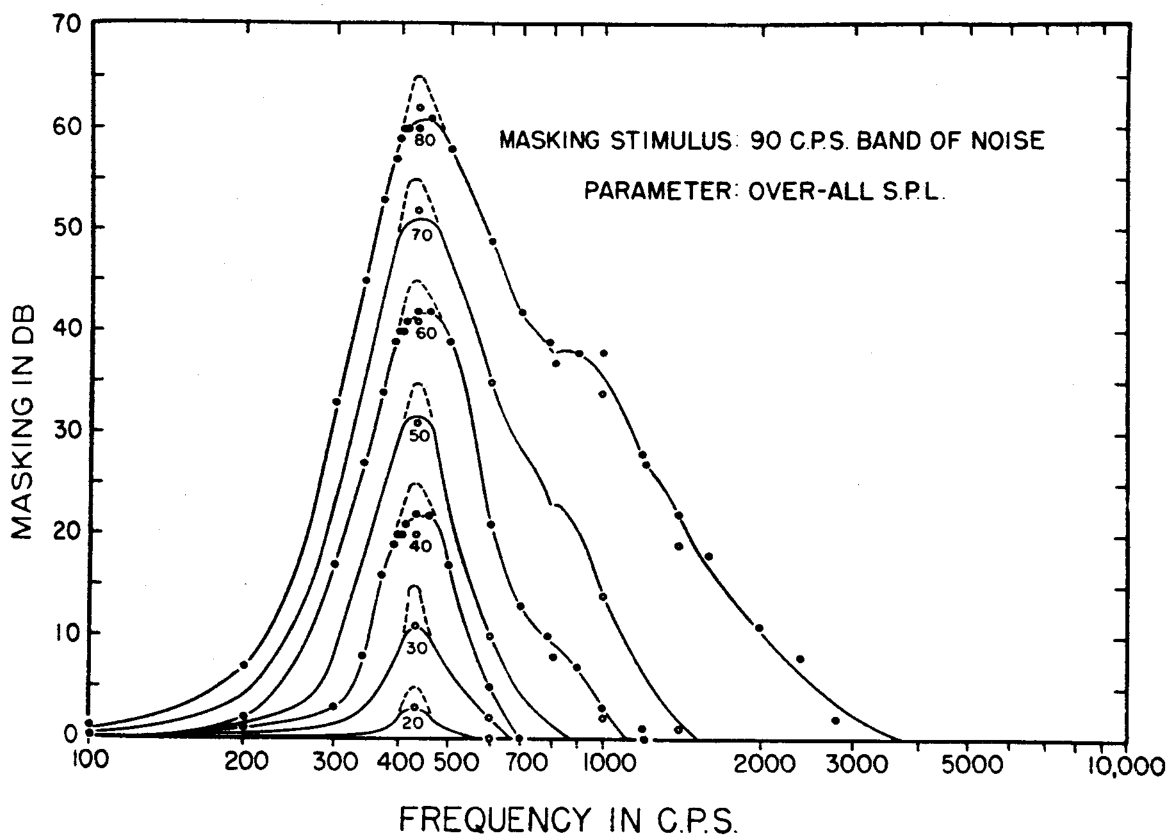

Fig. 2. Masking audiograms of a narrow band of noise (90 cps wide, centered at 410 cps) presented at various over-all sound-pressure levels (decibels re 0.0002 dyne/cm^2). The pressure spectrum level of the noise may be obtained by subtracting 19.5 dB from the corresponding number under the curve. The peak of each masking curve is extended by 4.2 dB in order to represent better the amount of excitation near the frequency of the masking stimulus. (From Egan and Hake, 1950, by permission.)

C. Critical Bands

Fletcher (1940) proposed the critical band concept to account for many of the phenomena of masking. He suggested that the basilar membrane provides a filtering action, with different frequencies producing their maximal effects at different locations along the membrane, and that each filter band is responsive to a limited range of frequency. The range of frequency to which a particular filter responds is its critical band. Masking occurs, according to Fletcher, when the noise pre-empts a filter (or its output channels) that would otherwise respond to the frequency of the signal, and only those frequencies of the noise which fall within the bandwidth of filter will be effective in masking the particular signal. The signal will be just detectible, according to Fletcher, when its energy equals the energy of that part of the noise which is affecting the filter. Fletcher says, "When the ear is stimulated by a sound, particular nerve fibers terminating in the basilar membrane are caused to discharge their unit loads. Such nerve fibers then can no longer be used to carry any other message to the brain by being stimulated by any other source of sound. Masking experiments appropriately chosen, then, should enable us to determine what portions of the membrane are being stimulated by an external sound." (1929, p. 167).

D. Noise-Level and Masking

Hawkins and Stevens (1950) studied the masking of tones of different frequencies by a wide band of white noise. Their results are presented in Fig. 3. The abscissa is noise level, and the ordinate is the amount of masking, expressed as the increase in signal level required for detection over that required in the absence of external noise. Had Hawkins and Stevens

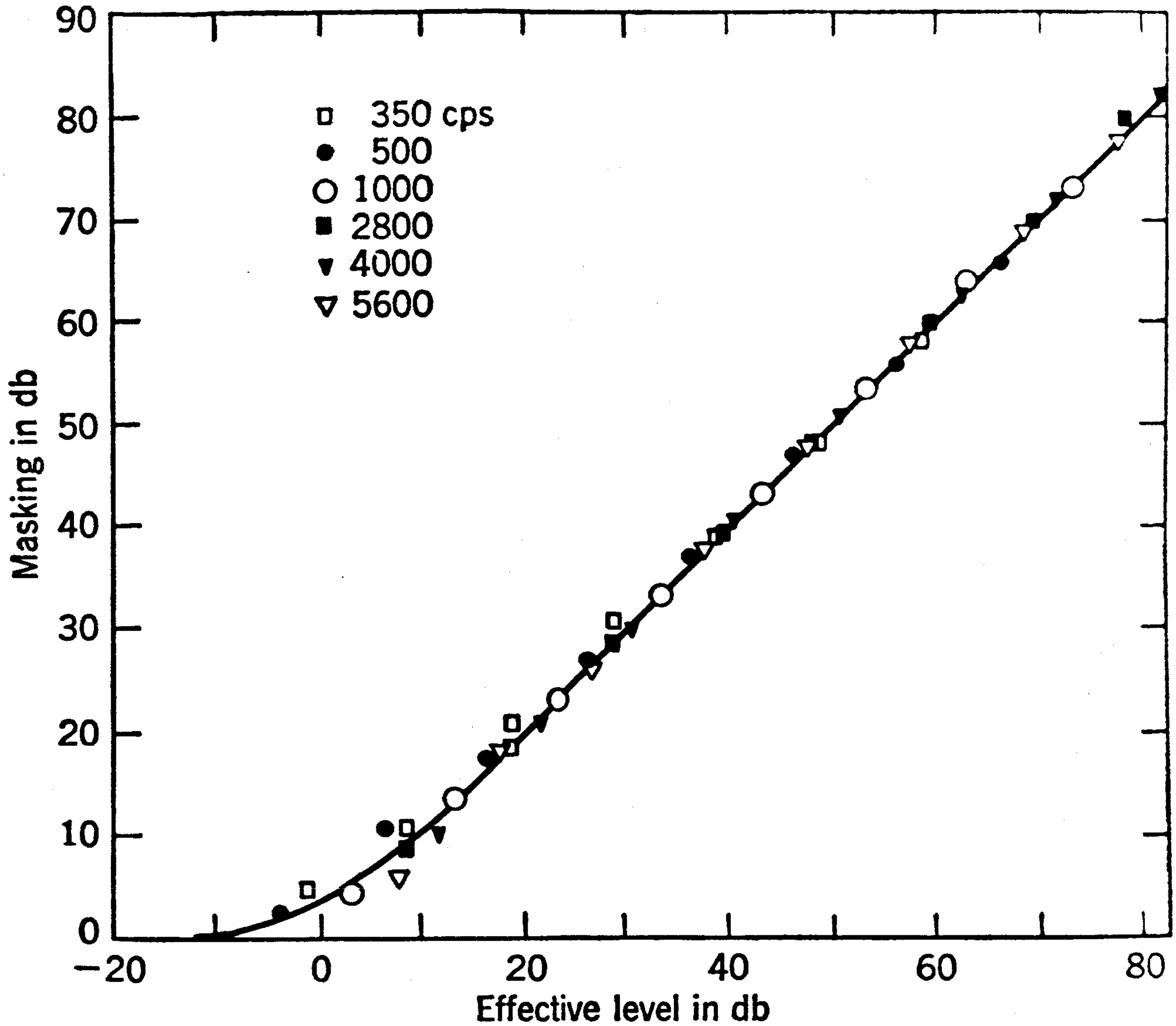

Fig. 3. The relation between the masking produced by a white noise and the effective level of the noise. The effective level is the amount of noise power in a narrow frequency band, the "critical band" (see text), centered about the frequency of the masked sinusoid. It is expressed in decibels relative to the absolute threshold (in power units) at that frequency. When the intensity is given in terms of effective level, the function shown in the graph is essentially independent of the frequency of the masked sinusoid. (From S. S. Stevens, (Ed.), Handbook of Experimental Psychology, New York: John Wiley and Sons, 1951, after Hawkins and Stevens, 1950, by permission.)

employed the spectral level of the masking noise as their abscissa, they would have obtained six parallel lines instead of their single line. There would be a line for each frequency used. Instead, they employed as abscissa the "effective level" of the noise--the overall level within a specified band around the signal frequency. The effective level is numerically equal to the spectral level plus 10 log W, where W is the bandwidth in cycles per second. For each frequency, they chose W so as to make the line for that frequency pass through _a_ point where the masking in dB equalled the effective level in dB. The fact that the line was straight over most of its course and also passed through other points where the amount of masking and the effective level are equal indicates a linear relation between masking and noise level. Watson (1963) obtained a similar function for masking in cats, Fig. 4, by the choice of an appropriate value of W. The value of the bandwidth, W, chosen in this way has been referred to by some experimenters as the width of the "critical band," others prefer to call it the "critical ratio" and reserve the term "critical band" to denote the bandwidth arrived at by band narrowing experiments. (See Chapter 22 for a discussion of this topic).

II. VARIETIES OF MASKING

A. Masking, Difference Limens, and Absolute Thresholds

G. A. Miller (1947), in an article about the difference limen for noise intensity, pointed out that the difference limen and the masked threshold are essentially the same thing. When we discover the size of the increment needed to produce a just noticeable difference in loudness, we may express this as Δp where the original stimulus was _p_, or we may speak of the masked threshold and express it in decibels. If, for example, $p = 0.002$ microbars, and for this intensity $\Delta p/p = 0.15$, we can equally well say that the SPL of

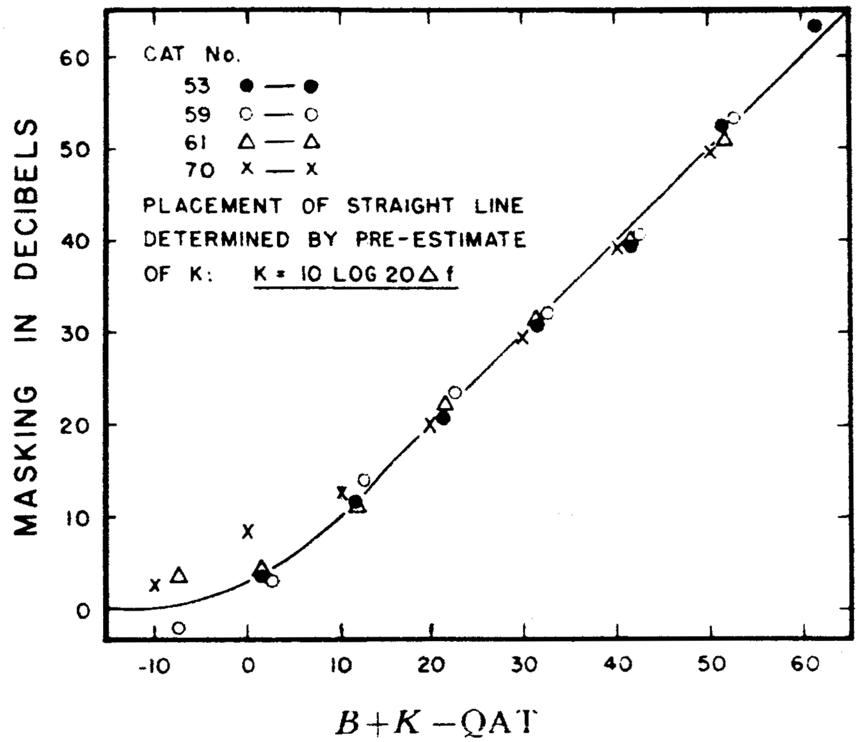

Fig. 4. A function relating masking M to effective level, Z = B + K - QAT, where B is the spectrum level of a band of noise broader than the estimated critical bandwidth K, and QAT is the quiet absolute threshold. The function is estimated from the relation K = 20 Δf; the points represent data from each of four cats. (From Watson, 1963, by permission.)

the noise is 20 dB and the masked threshold is 21.2 dB, or that there is 1.2 dB of masking.

Diercks and Jeffress (1962) went a step further and argued that the absolute threshold is itself really a masked threshold. By comparing thresholds for an antiphasic tone with thresholds for diotic stimulation, they obtained results from which they concluded that there was noise present in the cochlea and that the noise at one ear was partially correlated with the noise at the other. They pointed out that the amount of masking was commensurate with the amount of noise measured physically in the external meatus by Shaw and Piercy (1962).

B. Remote Masking

In addition to the masking by noise frequencies lying within the same critical band as the signal, there can be masking by frequencies which appear to lie well outside of the band, even above it. This phenomenon was discovered by Bilger and Hirsh (1956) and was called by them, "remote masking." It occurs only at high noise levels, 60 to 80 dB spectral level, and exhibits itself as an elevation of threshold for frequencies below those of the band of noise. Thus a band of noise having frequencies from 2450 to 3120 cps at a spectral level of about 70 dB will elevate the threshold for tones from 100 to 1000 cps by about 20 dB. Deatherage, Davis, and Eldredge (1957) were able to demonstrate a similar phenomenon in the guinea pig. The action potentials and cochlear microphonics recorded from the third turn of the cochlea in response to 500 cps tone bursts were masked when an intense high frequency noise was introduced. At the same time a random, low-frequency, cochlear microphonic potential appeared. The authors explain their finding as being the result of non-linear distortion that generates out of the high-frequency noise a low-frequency disturbance, fluctuating in frequency and amplitude. A study by Hirsh and Burgeat (1958) appears to confirm this hypothesis. They found that the binaural masking level differences associated with reversing the phase of a low-frequency tone, masked remotely by a high-frequency noise, were similar to those obtained with low-frequency noise; suggesting that the masking was the result of low frequencies actually existing in the cochlea. Cox (1958) showed that low frequencies can indeed be generated from a high-frequency band of noise by limiting (clipping).

C. Backward and Forward Masking

Backward masking is the masking of a signal by a noise which occurs later; forward masking is the reverse, the noise being terminated before the signal is begun. Both phenomena have other names; backward masking has been called precedent masking, and forward masking has been called residual masking, poststimulatory threshold shift, and adaptation. Lüscher and Zwislocki (1949) review earlier experiments and present data on the spread of adaptation (forward masking) as a fucntion of level, time, and frequency. Masking as a function of frequency is similar to the function for simultaneous masking; a tone is masked by an earlier tone when the frequencies are close together or when the earlier tone is lower in frequency. There is little forward masking when the masker has a higher frequency than the signal. The masking effect of an earlier sound falls off rapidly with the size of the interval between, and the slope of this drop is a function of the level of the masking sound. For an 80-dB, 400 msec, 3000-cps masker, the masking is about 40 dB, when the signal follows in 20 msec. This drops to zero when the interval is increased to 200 msec. The drop, when expressed in decibels of masking, is approximately linear with time.

Pickett (1959) and Elliott (1962a) summarize earlier work on backward masking, much of it done in the Soviet Union, and both present the results of several experiments in which the masker was a burst of white noise and the signal, a 1000-cps tone of short duration. The dependence of masking on the level of the masker appears to be linear, but the linear relation between interval and masking, found by Lüscher and by Zwislocki for forward masking, apparently does not hold for backward masking. The masking decreases much more rapidly as the interval between signal and masker is increased. Elliott, in one case for example, found about 60 dB of masking when the noise began 1 msec after the termination of the signal, and this dropped to about 20 dB when the interval was increased to 10 msec. Virtually no masking was found for intervals longer than about 25 msec.

Both forward and backward masking are the result of time-dependent properties of the neural mechanism of hearing. Forward masking suggests that cells which have recently been stimulated are not as sensitive as rested cells, a not very surprising fact. Backward masking, however is the interference of the later noise with some process initiated by the signal but not completed by the time of the onset of the noise. The times appear to be too long to be explained in terms of energy integration at the cochlea, and suggest instead some kind of interaction at higher centers, where the later activity produced by the more intense stimulus can overtake and obscure the effects of the earlier stimulus. The effect is large; a short tone terminating 1 msec before the onset of the noise may experience 60 dB of masking, where the same tonal pulse, starting 1 msec after the termination of the noise, experiences only about 30 dB.

D. Masking of Speech

While speech is probably our most important signal, it is an awkward one to use, and has not been much employed in masking experiments.

Where it has been used, the purpose has often been to study the nature of speech itself, or the nature of the masker--room reverberation, street noise, etc. Licklider (1948) used speech in a study of binaural phenomena, and Pollack and Pickett (1958) in their study of the masking of speech by speech were primarily concerned with binaural effects. Swets (1964) devotes four chapters to the masking of speech, but there the major concern was with speech as signal in TSD. Some aspects of the masking of speech are discussed in Chapter 3, on applications of TSD, and some in Chapter 25, on the perception of speech.

III. MASKING AND THE THEORY OF SIGNAL DETECTABILITY

A. Noise and Noise-Plus-Signal Distributions

The Theory of Signal Detectability (TSD) is a very general theory covering the detection of a great variety of signals. It is discussed in this general sense in Chapter 5. We are concerned with it here in a much more restricted sense as it applies to the detection of a (usually) tonal signal in a background of (usually) Gaussian noise. The two probability density curves (for noise and for noise plus signal) of TSD are usually exhibited along an unspecified abscissa, representing whatever it is about the stimulus that the subject is responding to. In the present section an attempt will be made to specify the abscissa, to determine just what aspect of the stimuli it is that causes the subject to vote more frequenctly for the interval containing the signal.

1. Narrow-Band Noise

The concept of critical bands is one of the important ideas to grow out of the experimental work on masking. The picture of the cochlea as a series of narrow-band filters helps in understanding not only many phenomena of masking, but also a number of other functions of the ear.

Figure 5, from Licklider (1951), shows estimates of critical bandwidths based on data from studies of masking, of pitch discrimination, of pitch scaling, and of speech intelligibility. The data for pitch discrimination, pitch scaling and speech intelligibility did not yield critical bandwidths directly. They were adjusted along the ordinate to conform with the masking data at one frequency to show the form of the function, not its magnitude. It can be seen that the functions agree surprisingly well considering the diversity of the sources of data. A fuller discussion of critical bands is to be found in Chapter 22.

Figure 5 tells us that the bandwidth associated with a frequency of, say, 500 cps is about 50 cps. This means that in a masking experiment, where the noise level is not excessively high, only the noise-frequencies from about 475 to 525 cps play an important role in the masking of a tonal signal of 500 cps.

It must be realized that the idea of the "critical band" as a filter, resembling an electrical filter, is only an analogy, and that the analogy is imperfect in many respects. The bandwidth of an electrical filter is the same over a wide range of measured levels, the ear's is not. Also when we speak of the bandwidth of a filter we refer either to its width at the half-power points (3-dB down) or to its equivalent rectangular width and both of these measures lose some of their meaning (especially their predictive value) when the filter response is unsymmetrical as we know the ear's response to be. The skirt of the ear's filter is considerably higher on the low frequency side than on the high, since low frequencies mask higher frequencies more effectively than the contrary. Most of the masking studies show the shape in reverse, since the masker is kept at a constant frequency and the signal probe-tone is varied, thus making each measurement in a different "critical band."

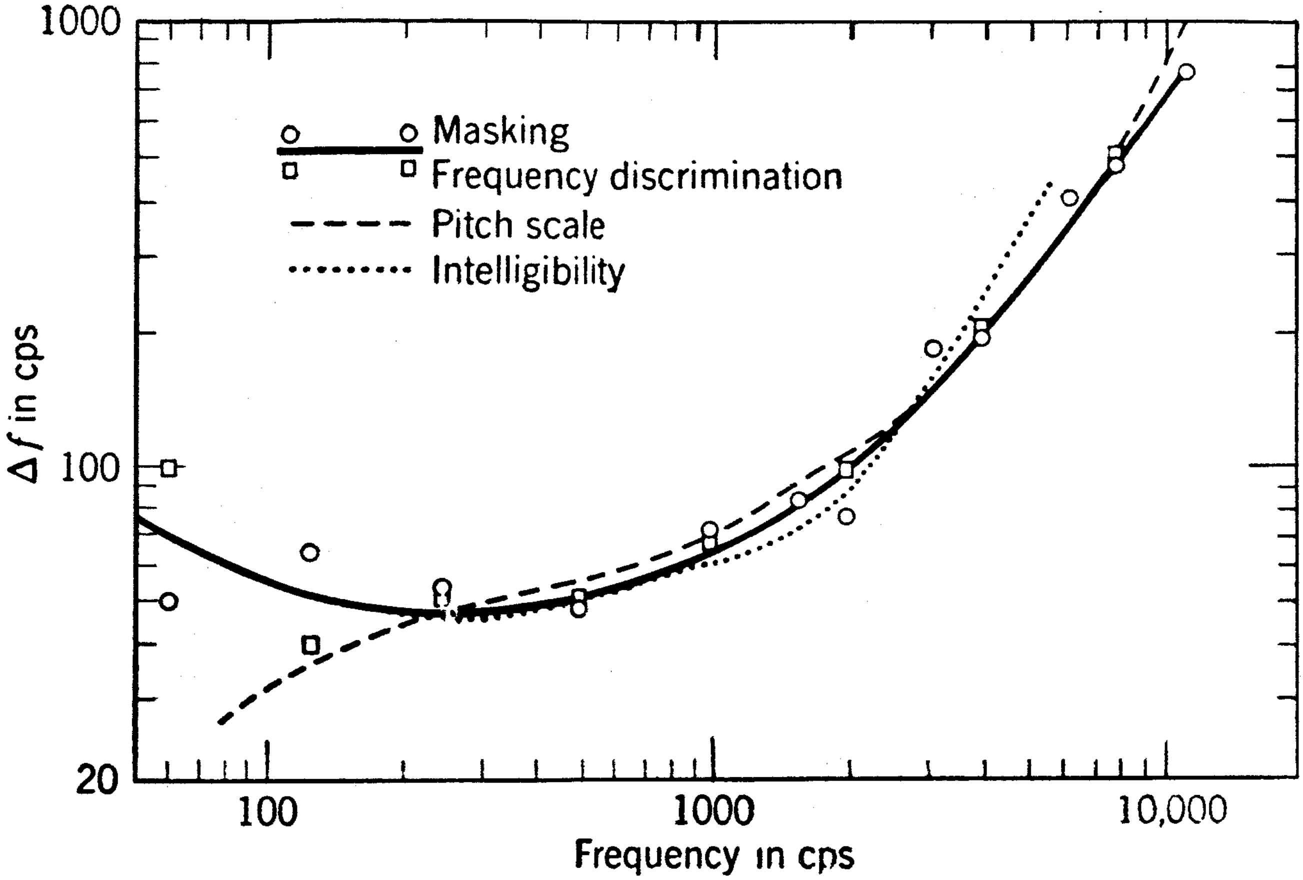

Fig. 5. Four functions relating Δf to f. The critical band function (circles and solid curve) shows the width of the band of noise that contributes to the masking of a sinusoid at the center of the band. In the frequency-discrimination curve (squares and solid curve), Δf is 20 times the jnd. The curve based on the pitch scale gives the width in frequency of intervals that are 50 mels wide in pitch. The curve based on intelligibility data shows the widths of frequency bands that contribute equally--2 per cent of the total--to the intelligibility of speech. The similarity of the curves suggests that they have a common basis in the auditory mechanism. (From Licklider in Stevens Handbook of Experimental Psychology, New York: John Wiley and Sons, 1951, by permission.)

Figure 6 shows what a narrow band (50 cps) of noise, centered at 500 cps, looks like. The upper picture is an oscilloscope photograph of the noise, and the lower is a photograph of noise to which a 500 cps tone has been added. The pictures were taken simultaneously and show the same stretch of noise with and without the signal. The signal has the same rms voltage as the noise. We see from the photographs that both functions closely resemble sine waves that are slowly fluctuating both in amplitude and in frequency or phase; the axis crossings are not quite evenly spaced. We may think of these functions as sinusoids having a basic frequency of 500 cps, and randomly modulated in both amplitude and phase. The fluctuations of phase are equivalent to fluctuations in frequency; 360° per second being 1 cps. Because of the narrowness of the filter, the rates of modulation, both of amplitude and phase, are slow compared with the frequency of the sinusoid.

The instantaneous displacement (voltage or pressure) for the narrow band of noise can be written

$$y(t) = a(t) \sin[2\pi ft + \varphi(t)], \qquad \text{Eq. (1)}$$

where _f_ is the center frequency (500 cps in the example), $a(t)$ is the amplitude, frequently called the envelope, and $\varphi(t)$ is the phase angle between the noise and some reference zero. Both $a(t)$ and $\varphi(t)$ are slowly varying, random functions of time. The amplitude, or envelope, $a(t)$, is always positive in sign and varies from zero upward. If the noise is Gaussian, the instantaneous voltage, $y(t)$, will be normally distributed around zero as its mean, and the standard deviation of the distribution of instantaneous voltages, σ, will be identical with the rms voltage of the band of noise (the square-root of the mean of the squared voltages).

If, for a Gaussian noise, the voltages are normally distributed, it is apparent that the amplitude cannot be. Its distribution must be skewed, since it is bounded on one side by zero and can reach toward infinity on the

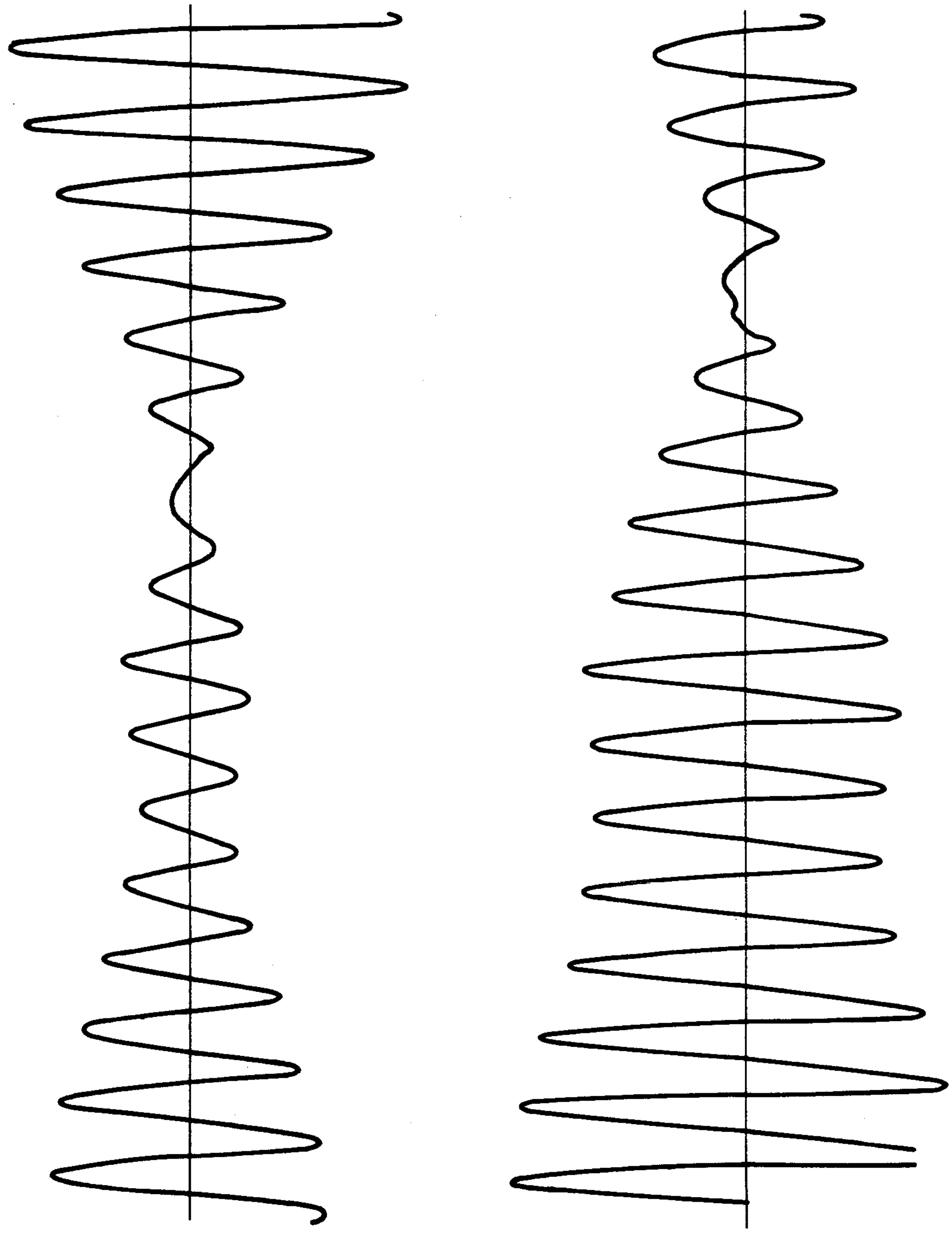

Fig. 6. Narrow-band noise (bandwidth 50 cps, center frequency 500 cps) (upper trace). Narrow-band noise plus signal (lower trace).

other. The literature of both acoustics and physics is often confusing here through the widespread, careless uses of the work "amplitude" when voltage or displacement is meant. The amplitude of a pure sinusoid is a constant, it is the voltage or displacement or pressure that varies sinusoidally. In the case of a narrow band of noise the amplitude, a(t), fluctuates but it remains positive in sign, it is the voltage that shows a Gaussian distribution.

2. Rayleigh's Distribution

Rayleigh (1894, p. 35-42) discussed the distribution function associated with the amplitudes of Eq. (1). He considered a narrow band of noise, obtained by combining _n_ sinusoids of equal amplitude and of random phases, and derived the expression for the distribution function when _n_ is allowed to become infinite. He found the function to be

$$f(a) = a/\sigma^2 \operatorname{Exp}[-a^2/2\sigma^2], \qquad \text{Eq. (2)}$$

where σ is the standard deviation of y(t) of Eq. (1), and hence is the rms voltage of the band of noise. The probability that _a_ in this expression will exceed some magnitude, a_i, is given by

$$P(a > a_i) = \int_{a_i}^{\infty} f(a)\, da = \operatorname{Exp}[-a_i^2/2\sigma^2]. \qquad \text{Eq. (3)}$$

A graph of the probability-density function of Eq. (2) is shown as the left-hand curve of Fig. 7.

The density function of Eq. (2) can be obtained from a bivariate normal distribution function of x and y, where the means of x and y are equal to zero, and σ_x and $\sigma_y = \sigma$, and where the correlation between x and y is zero. The radius, from the center (origin) to any point (x, y), is the quantity, a.[1]

3. Narrow-Band Noise Plus Signal

When we add a signal of frequency, f, to the noise of Eq. (1), we obtain the function pictured in the lower part of Fig. 6. It too may be thought of as a sinusoid, modulated in amplitude and phase, and can be written

$$y(t) = r(t) \sin[2\pi ft + \theta(t)], \qquad \text{Eq. (4)}$$

where r(t) is the new amplitude--the vector sum of the random variable, a(t) and the constant signal amplitude, A. The new phase angle, $\theta(t)$, is the angle between the resultant and the reference zero.

4. Distribution for Noise Plus Signal

Knowing the distribution function for narrow-band noise, let us attempt to determine the function for noise plus signal. The function for the amplitude of the noise is the Rayleigh, or circular-normal distribution. The function for phase is rectangular; all angles are equally likely. Let us select a number, say ten, equally likely values of the noise amplitude. These would be the mid-decile values, i.e., the values for P = 0.05, 0.15, 0.25, etc. The corresponding values of a, computed from Eq. (3), are 0.32, 0.57, 0.76, etc. Let us select also a number of equally likely phase angles--every 15° will serve for this. By using all

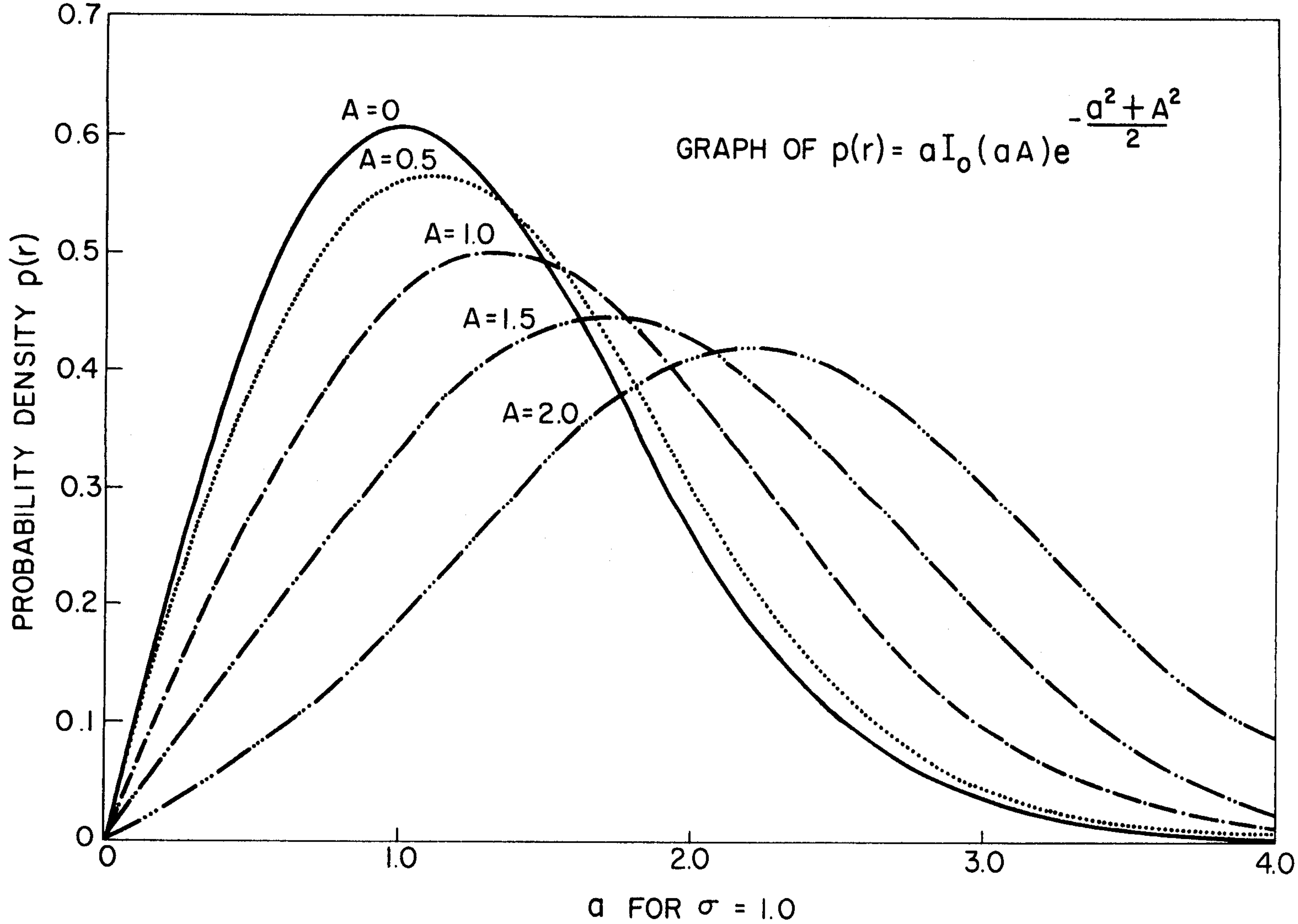

Fig. 7. Distribution curves for narrow-band noise and noise plus signal. The parameter A is the amplitude of the signal relative to the rms noise voltage. The curve for A = 0.0 is for noise alone--the Rayleigh distribution. (From Jeffress, 1964, by permission.)

combinations of amplitude and phase we arrive at 240 values of a(t), all equally likely. These values are shown as dots in Fig. 8.

Now let us add a signal of amplitude, A, and for convenience, in phase with the reference zero. Figure 8 shows the result. The signal vector is shown for a value of A equal to 4 times the standard deviation of the noise (the rms noise voltage), and pointing along the x axis (0° phase). One of the 240 values of the noise vectors is shown, and the resultant drawn in. The resultant, the SN vector, has a length, r, and makes an angle, θ, with the x axis.[2]

By repeating this process for each of the dots of Fig. 8, and measuring the length of the SN vector for each triangle, we can obtain a set of data from which to construct a frequency polygon. The polygon will approximate the distribution function for noise plus signal for the case where $A = 4\sigma$.[3] We could now combine this distribution with the one for noise alone and from the two, obtain an ROC curve for $A = 4\sigma$.

Fortunately, we are spared the necessity for solving vector triangles graphically. Rice (1954, pp. 236-241) has derived the expression for the probability density corresponding to our frequency polygon. In our notation, with a for noise amplitude, A for signal amplitude, and σ for the rms noise voltage, the function is

$$f(a, A) = (a/\sigma)\, I_o(aA/\sigma^2)\, \mathrm{Exp}[-(a^2 + A^2)/2\sigma^2], \qquad \text{Eq. (5)}$$

where I_o is a Bessel function for which tables are readily available.[4] When A = 0.00, the Bessel function, I_o, is unity, and Eq. (5) reduces to Eq. (2), the expression for noise alone. The right-hand curves of Fig. 7 represent the function f(a, A) for various values of the signal amplitude A, for σ = 1.00. The curves for noise, and for small values of A, are

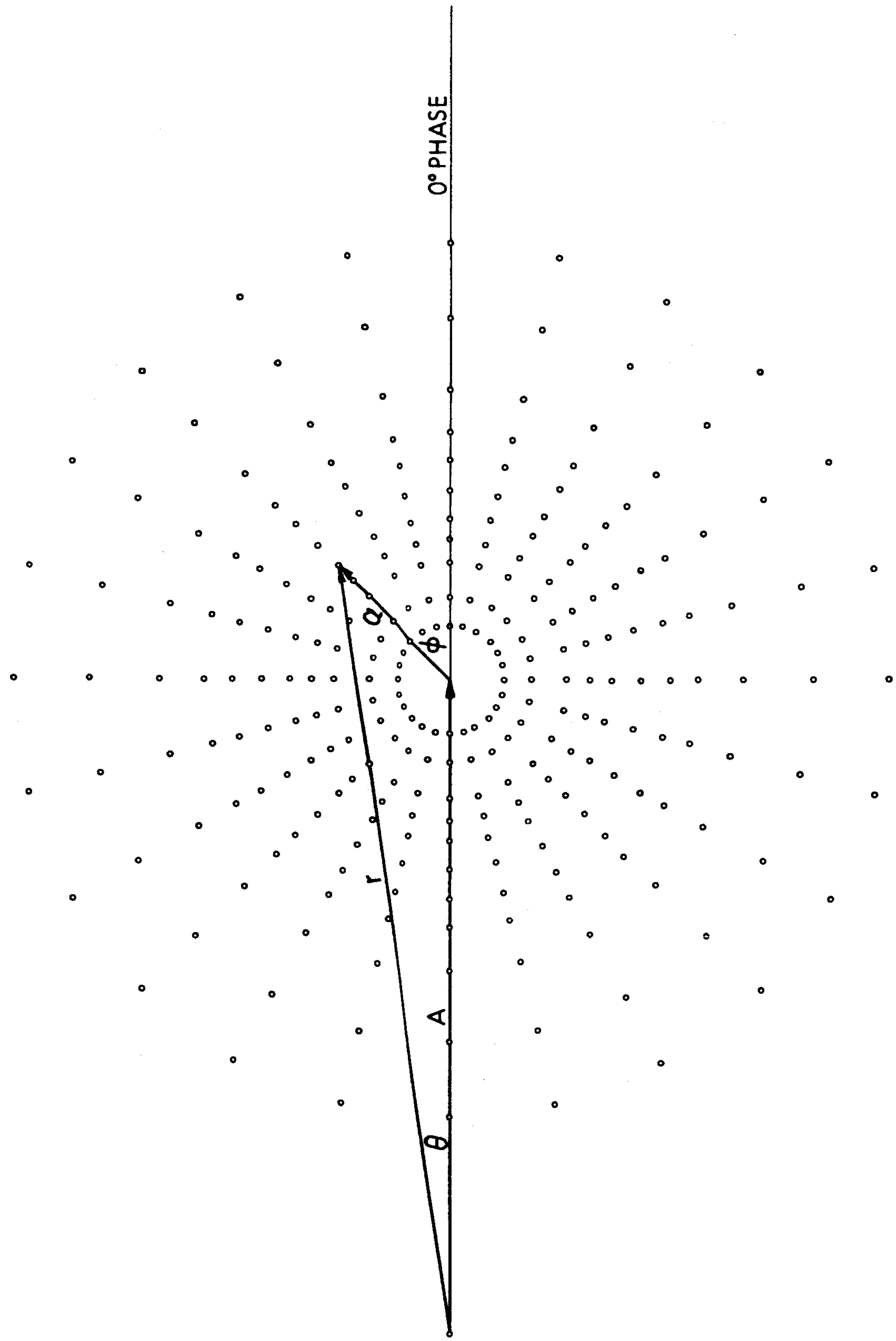

Fig. 8. Adding a signal to noise. The vector A represents the signal and the vector a represents one possible value for the noise. The momentary phase angle between the signal and the narrow-band noise is φ, and between the signal and the signal-plus-noise vector, r, is θ. The dots indicate equi-probable locations of the terminus of the noise vector.

decidedly skewed. The curves for large values of A approach the normal distribution, with a standard deviation equal to σ, the rms noise voltage.

In order to obtain ROC curves from the probability density functions of Fig. 7, it is necessary to accumulate the probabilities under the curves. The probabilities for noise are given by Eq. (3) and can be readily determined from tables of the exponential. The corresponding expression for SN is

$$P(r > a_i) = \int_{a_i}^{\infty} f(a, A)\, da. \qquad \text{Eq. (6)}$$

The expression in Eq. (6) is not integrable, but it can be evaluated numerically. Marcum (1950) has done so, and has prepared a set of tables of the integral (with σ taken as unity) for values of a_i ranging in steps of 0.1 units from 0.1 to 20.0, and for A, in steps of 0.05 units, from 0.00 (the Rayleigh distribution) to 24.90. The probabilities are given to six decimal places. Using Marcum's table instead of our graphically derived frequency polygon, we can determine $P(a > a_i)$ and $P(r > a_i)$ for various values of the criterion, a_i, and of the signal amplitude, A. These probabilities correspond to $P(y|n)$ and $P(y|sn)$, and can be used in plotting a family of ROC curves. The result is shown in Fig. 9.[5]

5. ROC Curves for Amplitude Distribution

The curves of Fig. 9 are slightly different in shape from the familiar ones of TSD, but they are to be found in the TSD literature. Peterson, Birdsall, and Fox (1954, p. 193) present the family, and show that they represent the behavior of the ideal detector for the case where the signal is completely specified except for phase. The more familiar curves are derived from overlapping normal curves of equal variance, and represent the behavior of the ideal detector for the case where the signal is completely specified, including phase.

Marill (1956) showed that ogives derived from Eq. (6) (Marcum's table), where P(c) is plotted against signal level, fitted his subjects' data better than ogives derived from normal curves. Similarly, Jeffress (1964) showed that the ROC curves of Fig. 9 fitted rating-scale data better than ROC curves derived from normal distributions. It had been known for some time (see for example, Egan, Schulman, and Greenberg, 1959) that rating-scale data plotted on normal-normal probability paper yield lines having slopes almost always less than unity, and therefore violate the equal-variance assumption. Several <u>ad hoc</u> explanations of the unequal variance have been offered, none very satisfactory.

Watson, Rilling, and Bourbon (1964) employed a rating-scale device with which the subject indicated his assurance that a signal was present in the stimulus interval. With it they were able to obtain 36 points on an ROC curve. Their data, when plotted on normal-normal probability paper could be fitted fairly well by straight lines having slopes less than unity, but the lines were not quite straight and the authors commented, "It could be further conjectured that the functions that generate these ROC curves are somewhat more peaked than normal distributions . . . ." Jeffress later found that the ROC curves of Fig. 9 fitted the curves of the study by Watson <u>et al</u> better than the curves used by the authors and derived from normal distributions.

From the foregoing results we may conclude that the subject, in detecting a tonal signal in a background of Gaussian noise, responds to the <u>amplitude</u> of the stimulus, and does not utilize phase information. The

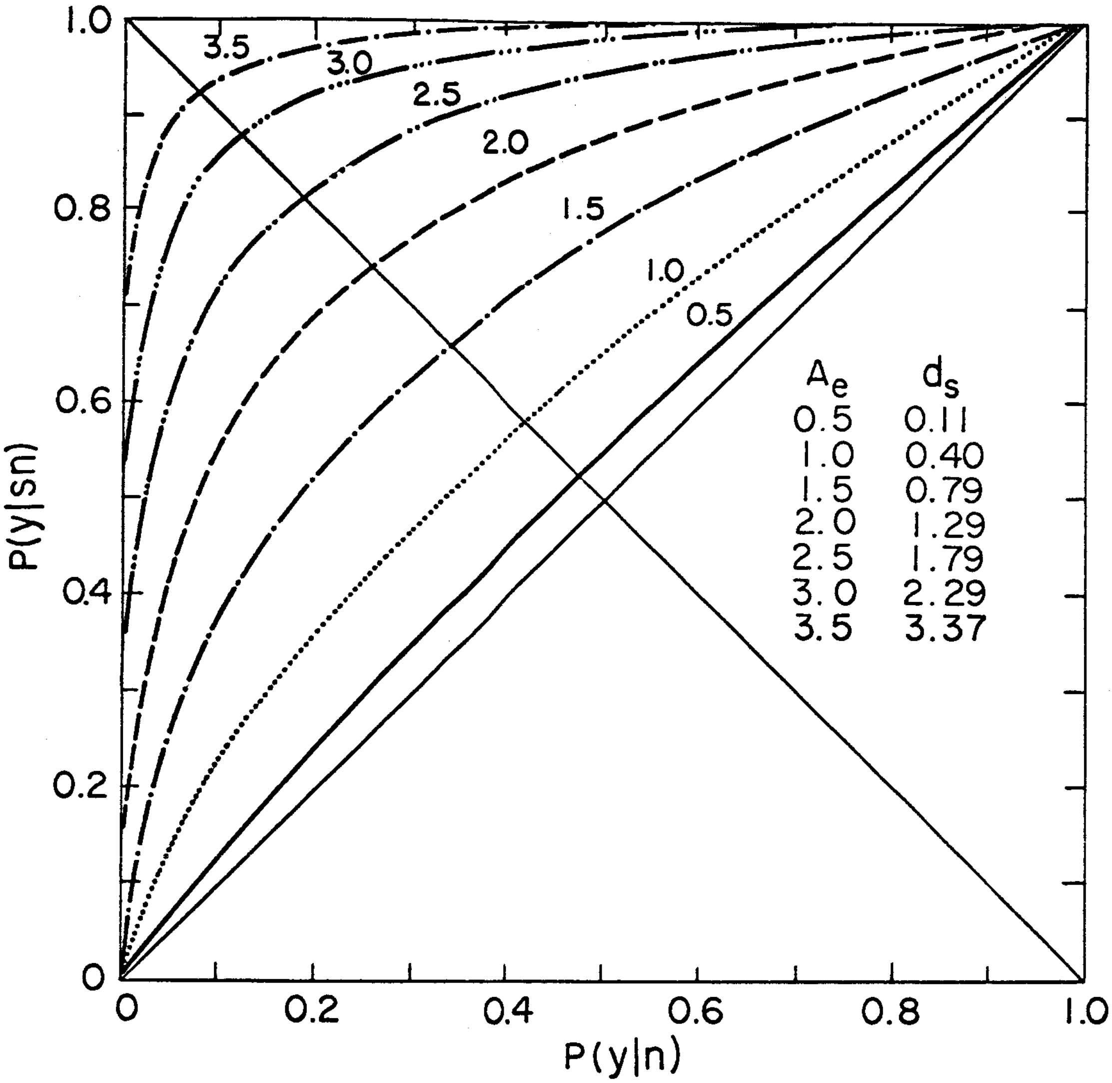

Fig. 9. ROC curves derived from the distribution functions for noise and noise plus signal. The parameter is the same as that of Fig. 7. (From Jeffress, 1964, by permission.)

abscissa of our family of probability-density curves is the ratio of voltages (SN amplitude divided by rms noise voltage), and supports the observation by Tanner and Birdsall (1958) that d' is, at least for audition, a voltage-like quantity.

6. "Pedestal" Experiments

Gaston (1964a) pointed out that, since the curves of Fig. 9 become more nearly normal as A is increased, using one of the curves (say, for A = 2.0) for "noise" and another for SN should yield ROC curves like those for the normal, equal-variance assumption. An ROC plot on normal-normal probability paper for A = 2.0 (N) and A = 3.0 (SN), yields a line which is substantially straight and has a slope almost equal to unity.

The use of a noise mixed with a sinusoid of the signal frequency for "noise" constitutes a "pedestal" experiment. The "signal" is achieved by increasing the level of the sinusoid, and can be thought of as standing on the "pedestal." The experiment suggested by Gaston is therefore a pedestal experiment. At the present time no one has determined an ROC curve using a pedestal, but there have been several experiments in which the shape of the psychophysical function relating signal level and detection percentage has been determined. Green (1960) points out that for a two-alternative, forced-choice experiment without a pedestal, the psychophysical function does not fit the function for the ideal observer (phase known), when a pedestal is used, the fit is good. (Data from Tanner, Bigelow, and Green, see Green, 1960.)

The fact that the psychophysical function for pedestal experiments fits the function for the ideal detector (phase known) better than data taken without a pedestal, can be (and has been) interpreted as indicating that the human observer can use phase information when it is made available.

Pfafflin and Mathews (1962), in discussing a series of pedestal experiments, point out that to account for the use of phase by the observer requires the assumption of some sort of correlation detector, whereas the assumption of a simple envelope detector will predict successfully the outcome of experiments both with and without a pedestal, and even with a pedestal to which the signal is added in quadrature phase. For convenience their mathematical treatment is based on an energy (square-law) detector, but they point out that a simple amplitude detector (rectifier) would yield almost the same functions and be more realistic neurophysiologically. Their approach to the detection problem is considerably different from ours, but the functions at which they arrive are substantially the same.[6]

7. Meaning of d' for Rayleigh-Type Distributions

Tables of d' (P. B. Elliott, 1959) yield the values of d' associated with various combinations of $P(y|n)$ and $P(y|sn)$. If we look up the value of d' associated with the probabilities taken from one of the ROC curves of Fig. 9, we discover that each point on the curve gives us a different value. To circumvent this difficulty, Clarke, Birdsall, and Tanner (1959), and Egan (1961), employ the value of d' for the point where the ROC curve passes through the negative diagonal as the measure of detection. Egan calls this value d_s. Green (1964) has shown that for two-alternative, forced-choice data, P(c) is equal to the area under the ROC curve, no matter what shape the underlying probability-density functions may take. He therefore suggests using P(c) as the measure of detection. This would still leave us in difficulty with data from a yes-no experiment. Unless we know the shape of the ROC curve, we do not know what curve the data point belongs to, and cannot discover the appropriate P(c), nor for that matter, the appropriate d_s. It appears quite possible that some of the variability of yes-no data stems from this difficulty, that points which really fall on the same curve yield different d's because of the shape of the curve. The moral appears to be that the first step in dealing with a new stimulus situation should be to determine an ROC curve for it; or else to employ the two-alternative, forced-choice procedure, and use P(c).

For the Rayleigh-type distributions P(c) can be obtained from Marill's (1956) expression, for P(c) for two-alternative, forced-choice data: $P(c) = 1 - \frac{1}{2}\exp[-E/2N_o]$. If we find the z-score corresponding to this P(c), and plot it as a function of $\sqrt{2E/N_o}$, we obtain a (very nearly) straight line having a slope of unity. The line bends to the origin for small values of $\sqrt{2E/N_o}$. The straight portion appears to intersect the ordinate at -0.707 (see Jeffress, 1964). This quantity, the z-score, appears to be numerically equal, or very nearly equal, to the d_s associated with the negative diagonal for these distributions.

B. Energy, Bandwidth, and Duration

1. Energy vs Amplitude as Stimulus

The literature of TSD has sometimes confused the reader about whether subjects respond to the energy of the signal or to its voltage. All of the evidence of the present section (Section III) points to the latter, to the envelope of the waveform as the aspect of the stimulus to which the subject responds. The shape of ROC curves, the shape of the psychophysical function relating detection to signal level, both with and without a pedestal, and the parameter of the family of distribution functions involved, all indicate that the ear (and brain) is acting like an envelope (amplitude) detector. That our methods are sensitive enough to discover other bases for detection when they are operative is shown in Chapter 15, where interaural time difference proves to be the aspect of the stimulus employed under some binaural conditions. The sensitivity of the methods is also shown by the fact that the data of the present section deny the use of phase information by the subject, when phase information is not accessible to him.

2. Duration and Bandwidth

The parameter employed by Peterson, Birdsall, and Fox (1954) in their derivation of the probability density functions of Fig. 7 is $\sqrt{2E/N_o}$. It is employed in the same way in their treatment as A/σ was in ours. For the same curve, the two quantities are equal: $A/\sigma = \sqrt{2E/N_o}$. Let us see what is implied by this relationship. Our quantity, A, is the signal amplitude, and is therefore equal to $\sqrt{2}$ times the rms signal voltage, S. Our σ is the rms noise voltage of the band of noise. If for arithmetical convenience we make the conventional assumption of a unit resistive load, the noise power is σ^2 and this is equal to the per-cycle noise power, N_o times the bandwidth, W. The signal power is S^2, and the signal energy E is S^2 times the duration, T. By squaring both sides of our original equality, $A/\sigma = \sqrt{2E/N_o}$, we obtain $A^2/\sigma^2 = 2E/N_o$, and substituting for A, σ, and E, we have $2S^2/WN_o = 2S^2T/N_o$. Cancelling leaves us the relationship $1/W = T$. Since the curves of Peterson *et al.* were derived for the ideal detector for the case where signal phase is unknown, our relationship tells us that the ideal detector (phase unknown) is an envelope detector with a filter having a bandwidth equal to the reciprocal of the signal duration. This is the conclusion drawn by Peterson *et al.* by way of a somewhat different line of reasoning. In the light of this fact, let us examine the relation between duration and masking.

3. Signal Duration and Masking

There have been several studies of the effect of signal duration on masking. Probably the earliest was an experiment by Garner and Miller (1947) who found, for four different frequencies of the signal, a linear relationship between the masked threshold expressed in dBs of masking, and the logarithm of signal duration, for durations between 12.5 and 200 msec. Increasing the duration by a factor of 2 increased detection by 3 dB. Other investigators, Hamilton (1957), Blodgett, Jeffress, and Taylor (1958) have found similar relationships, but have found changes in the slope of the function at different durations. All have found that the slope increases for short signals where $\underline{T}$ is less than the reciprocal of the critical bandwidth. Durations longer than about 200 to 500 msec are apparently not so efficiently employed as shorter durations and the line relating masking and duration tends to level off. Green, Birdsall, and Tanner (1957), in an experiment where the signal energy was kept constant by increasing the power in proportion to the decrease in duration, found that the detection index, d', remained constant through a range of durations that varied from subject to subject. All had constant values of d' from 20 to 150 msec, but some subjects extended this range to about 10 msec at one end, and to nearly 300 msec at the other.

The results from studies of the effect of duration suggest that the ear integrates energy. The durations involved, however, are too long to make it reasonable that there is actual accumulation of energy before stimulation occurs. It seems more likely that the integration is neural, and that rather than accumulating energy, the mechanism accumulates neural events associated with envelope (amplitude) peaks. There is some evidence that this is so. Elliott (1962b) has found that subjects having substantial high-frequency hearing loss, presumably neural, when tested at frequencies where the loss is serious, show less temporal integration than normal subjects. They require nearly as strong a signal at long durations as they do at short. It is difficult to see how this fact could be explained in terms of energy integration.

4. Bandwidth and Masking

Since the original experiments by Fletcher (1940), there have been several experiments attempting to discover the bandwidth and shape of the ear's filter system. Schafer, Gales, Shewmaker, and Thompson (1950) investigated the masking effects of narrow bands of noise obtained *a la* Rayleigh by combining a number of sinusoids of equal amplitudes and random phase. They determined the equivalent-rectangular bandwidths at three frequencies--200, 800, and 3200 cps--to be 65, 65, and 240 cps respectively. If the ear's filters are taken as single-tuned circuits, the corresponding Qs are 9.3, 37, and 39 respectively. The shapes of the masking function obtained by Schafer *et al*. resembled curves for single-tuned circuits.

Swets, Green, and Tanner (1962) also studied the relation between masking and the bandwidth of the masking noise, using filters ranging from much wider, to much narrower than the values usually assumed for the critical band. They made estimates of the bandwidth for various assumptions about the response characteristics: single tuned, 41 cps; rectangular, 95 cps; Gaussian (3 dB points), 79 cps; Gaussian (one sigma points), 95 cps. They conclude, "We would suggest a consideration . . . of the possibility that the parameters of the mechanism of frequency selectivity vary from one task to another under intelligent control. If they do, then, of course, we cannot speak of, or measure, the critical band." The validity of the foregoing statement is supported by many experiments employing other than tonal signals, and yielding bandwidths considerably wider then those mentioned above. The significance of a number of such experiments is discussed in Chapter 22.

5. Duration, Bandwidth, and Masking

The reciprocal relation between bandwidth and duration indicated by the equivalence of A/σ and $\sqrt{2E/N_o}$ as parameters, has interesting implications for hearing. The most obvious is that the subject employs a bandwidth appropriate to the duration (or expected duration) of the signal. Something of the sort appears to be implied in the statement by Swets, Green, and Tanner quoted earlier. It seems also to be implied in Békésy's (1959) concept of neural funneling, and in Marill's reference to the bandwidths employed by his subjects. Jeffress (1964) specifically considers the possibility of a bandwidth that can be narrowed for long signals and widened for short, and examines Hamilton's (1957) data with this in mind. Hamilton varied both the duration of the signal and the bandwidth of the noise. The interactions of bandwidth and duration were in the direction predicted, but were not clearcut enough to be completely convincing. Hamilton's estimates of critical bandwidth and those of Greenwood (1961) are considerably wider than most of those discussed in the present chapter. (See Chapter 22 for a further consideration of this problem.)

C. Signal Uncertainty

1. Frequency Uncertainty and Masking

If we were physically looking for a signal of known frequency and duration (but unknown phase) in Gaussian noise, we would employ a filter having a bandwidth equal to the reciprocal of the duration, and a center frequency equal to that of the signal. If the signal might be either of two frequencies, we would employ two appropriate filters and combine their outputs by means of an OR gate, unless the frequencies were close enough together so that one filter would suffice. An alternative to employing *n* filters to look for *n* different frequencies would be to employ a filter wide enough to encompass the range of frequencies involved. This would be a less efficient method, since the wider filter would not exclude the short, signal-like bursts of noise that would be rejected by the optimal filters.

The question of interest to us here is, does the ear (and brain) do anything of the sort. Several experiments have been designed to answer the question. Marill (1956) found that when two signals are employed simultaneously, if their frequencies lie close together (within one filter) they are detected more readily than either alone, but if they are well separated, they are detected no better than one alone. This result is in agreement with the multifilter hypothesis. If the signals affect the same filter, they are occurring in a single band of noise and should increase the signal-to-noise ratio; if they occupy different filters, each will have its own band of noise and no improvement should occur.

The problem becomes more complicated when the signals are employed singly and the subject does not know which of two (or of several) to expect. A still different situation arises when the subject expects one frequency and gets another. Both situations have been studied. Greenberg (1964) found that when a subject is expecting a signal of 1100 cps, and is given one of some other frequency, without knowing that this can occur, his detection drops sharply. The curve of detection-vs-signal frequency resembles the masking curves of Schafer et al. This result suggests that the subject is employing a single filter and disregarding signals that lie outside of its band. Now if we change the instructions so that the subject knows that there are two possible signals equally likely to occur, what happens? One possibility suggested by Tanner, Swets, and Green (1956) is that the subject scans the filter back and forth between the frequencies involved, frequently missing the signal because he is looking in the other place. Green (1958) examined this idea and rejected it because it predicts too low a detection score. He suggested instead a multifilter model in which the outputs of the separate filters are added. Creelman (1960) made the further suggestion that the detector decides on the basis of the maximal output of the filters taken separately. Some of the experimental data appear to agree with one hypothesis and some with another. Creelman found that some of his data even suggested that the subject widened his filter band to encompass the range of frequencies involved.

There seems to be no doubt, as Swets (1963) points out, that the subject is somehow selecting what to listen for; that there is some sort of control of the peripheral apparatus by the central nervous system. The question appears to be, what kind. Possibly at this point we should invoke the Huggins-Licklider (1951) pinciple of diversity, which says in effect that if there are two ways of doing something, the nervous system will employ both.

Let us examine the possibility suggested earlier, that the outputs of the filters are combined by way of an OR gate. In a yes-no experiment, if the criterion were maintained at a constant level, $P(y|sn)$ would remain unaltered, but $P(y|n)$ would be doubled; a signal-like noise occurring in either filter would appear in the output of the gate, and receive a "yes" vote. (The gate responds to either A or B or both.) If we assume that the ROC curves of Fig. 9 are appropriate here (and they should be), we may use them to predict the outcome of a two-frequency experiment. We very quickly discover that we can predict almost any degree of drop in detection, depending upon our choice of criterion and of signal level. Let us take, for example, a d_s of 2.0 for a single frequency, and assume that $P(y|sn) = 0.57$. The corresponding value of $P(y|n)$ will be 0.05. Doubling this moves us to a new ROC curve, the one for $d_s = 1.7$. If instead of 0.57, we assume the initial detection to be $P(y|sn) = 0.80$, $P(y|n) = 0.12$. Doubling this moves us to the curve for $d_s = 1.55$, a considerably greater drop in detection efficiency. The former corresponds to about a 1 dB change of level, the latter to about 1.5 dB (see Jeffress, 1964, p. 771). For a low signal level, and hence a low value of d_s, the effect of doubling $P(y|n)$ is much greater and may amount to 3 or 4 dB. This increase of the effect of frequency uncertainty at low signal levels has been noted by both Creelman (1960) and Swets (1963).

When a two-alternative, forced-choice procedure is used, the subject is forced to remain near the negative diagonal of the ROC curve, and the OR gate becomes the equivalent of Creelman's model; the subject responds to the interval containing the larger stimulus. He may therefore respond correctly for the wrong reason: because the interval containing the signal also contains a strong burst of noise from the other filter. Data taken with uncertain signal frequency are very erratic; the foregoing discussion may serve to show why this is the case.

2. Other Uncertainties

In addition to the deliberately-introduced frequency uncertainty, another form of uncertainty develops when the signal level is so low that the subject seldom hears the signal clearly. Here the subject may become uncertain about the frequency, the duration, and the time of onset of the signal. The frequency uncertainty probably would not be as great as in the two-frequency experiments, and we can imagine that the subject might, if he had the machinery for it, widen his filter band slightly to take care of the uncertainties.

Marill (1956) in his work at very low signal levels, avoided this "forgetting" of the frequency and duration of the signal by presenting a sample of the signal without noise in advance of each stimulus trial. By rewriting his expression for P(c) in a two-alternative, forced-choice experiment in terms of signal voltage and bandwidth instead of E/N_o, he obtained $P(c) = 1 - \frac{1}{2}\,\mathrm{Exp}[-S^2/WN_o]$, where W is bandwidth. He found that his subjects maintained the same bandwidth at all signal levels.

Gaston (1964b) in an experiment devised to determine the relation between d_s and signal voltage at low levels, found that his subjects responded as though they were employing a wider band at low levels than at high. The experimenter did not employ a cuing signal in the initial study, but in a replication of it, Gaston and Jeffress (1964) did use a cuing signal sufficient to be heard clearly above the noise, and found some improvement of detection at low levels. Their findings, however, did not quite reach the constancy of bandwidth reported by Marill. Greenberg (1962) has studied the effect of a variety of cuing signals. He found that a cuing signal is most effective when it precedes the stimulus intervals by about one-half second. He employed a cuing signal having the same level as the signal to be detected and did not get as large an effect as Gaston and Jeffress obtained with a larger cuing signal.

It is of course, equally possible to describe the foregoing results in terms of the efficiency measure, η, of TSD. Instead of thinking of the response to signal uncertainty as an increase of bandwidth, we may think of a decrease of efficiency. The two quantities η and *W* are reciprocally related.

REFERENCES

Békésy, G. von. Vibration of the head in a sound field and its role in hearing by bone conduction. J. Acoust. Soc. Amer., 1948, 20, 749-760.

Békésy, G. von. Neural funneling along the skin and between the inner and outer hair cells of the cochlea. J. Acoust. Soc. Amer., 1959, 31, 1236-1249.

Bilger, R. C. and Hirsh, I. J. Masking of tones by bands of noise. J. Acoust. Soc. Amer., 1956, 28, 623-630.

Blodgett, H. C., Jeffress, L. A., and Taylor, R. W. Relation of masked threshold to signal duration for various interaural phase combination. Amer. J. Psychol., 1958, 71, 283-290.

Clarke, F. R., Birdsall, T. G., and Tanner, W. P., Jr. Two types of ROC curves and definitions of parameters. J. Acoust. Soc. Amer., 1959, 31, 629-630.

Cox, J. R., Jr. Spectrum of noise passed through a symmetrical limiter (with applications to remote masking). J. Acoust. Soc. Amer., 1958, 30, 696 (A).

Creelman, C. D. Detection of signals of uncertain frequency. J. Acoust. Soc. Amer., 1960, 32, 805-810.

Deatherage, B. H., Davis, H. and Eldredge, D. H. Physiological evidence for the masking of low frequencies by high. J. Acoust. Soc. Amer., 1957, 29, 132-137.

Diercks, K. J. and Jeffress, L. A. Interaural phase and the absolute threshold for tone. J. Acoust. Soc. Amer., 1962, 34, 981-984.

Egan, J. P. Method of free response in signal detection: a simplified technique. J. Acoust. Soc. Amer., 1961, 33, 993-1007.

Egan, J. P. and Hake, H. W. On the masking pattern of a simple auditory stimulus. J. Acoust. Soc. Amer., 1950, 22, 622-630.

Egan, J. P., Schulman, A.I., and Greenberg, G. Z. Operating characteristics determined by binary decisions and by ratings. J. Acoust. Soc. Amer., 1959, 31, 768-773.

Elliott, Lois L. Backward masking: monotic and dichotic conditions. J. Acoust. Soc. Amer., 1962, 34, 1108-1115. (a)

Elliott, Lois L. Comments on temporal integration of the ear. J. Acoust. Soc. Amer., 1962, 34, 1973 (A). (b)

Elliott, P. B. Table of d'. University of Michigan: Electronics Defense Group, 1958. (Technical Report No. 97) (Also available as appendix in Swets (1964)).

Fletcher, Harvey. Speech and hearing. New York: D. Van Nostrand, 1929.

Fletcher, Harvey. Auditory patterns. Rev. Mod. Physics, 1940, 12, 47-65.

French, N. R. and Steinberg, J. C. Factors governing the intelligibility of speech sounds. J. Acoust. Soc. Amer., 1947, 19, 90-119.

Garner, W. A. and Miller, G. A. The masked threshold of pure tones as a function of duration. J. Exper. Psychol., 1947, 37, 293-303.

Gaston, A. D. (Personal communication), 1964. (a)

Gaston, A. D. Effect of signal level on masking. Unpublished M. A. Thesis. The University of Texas, 1964. (b)

Gaston, A. D. and Jeffress, L. A. (Unpublished experiment) 1964.

Green, D. M. Detection of multiple component signals in noise. J. Acoust. Soc. Amer., 1958, 30, 904-911.

Green, D. M. Auditory detection of a noise signal. J. Acoust. Soc. Amer., 1960, 32, 121-131.

Green, D. M. General prediction relating yes-no and forced-choice results. J. Acoust. Soc. Amer., 1964, 36, (A) 1042.

Green, D. M. (Personal communication) 1965.

Green, D. M., Birdsall, T. G., and Tanner, W. P., Jr. Signal detection as a function of signal intensity and duration. J. Acoust. Soc. Amer., 1957, 29, 523-531.

Greenberg, G. Z. Cueing signals and frequency uncertainty in auditory detection. Indiana University: Hearing and Communication Laboratory, Tech. Rept. AF 19(628)-266 ESD-TDR 62-38, 1962.

Greenberg, G. Z. (Personal communication) 1964.

Greenwood, D. D. Auditory masking and the critical band. J. Acoust. Soc. Amer., 1961, 33, 484-502.

Hamilton, P. M. Noise masked thresholds as a function of tonal duration and masking noise band width. J. Acoust. Soc. Amer., 1957, 29, 506-511.

Hawkins, J. E., Jr. and Stevens, S. S. Masking of pure tones and of speech by white noise. J. Acoust. Soc. Amer., 1950, 22, 6-13.

Hirsch, I. J. and Burgeat, M. Binaural effects in remote masking. J. Acoust. Soc. Amer., 1958, 30, 827-832.

Huggins, W. H. and Licklider, J.C.R. Place Mechanisms of auditory frequency analysis. J. Acoust. Soc. Amer., 1951, 23, 290-299.

Jeffress, L. A. Stimulus-oriented approach to detection. J. Acoust. Soc. Amer., 1964, 36, 766-774.

Licklider, J.C.R. Influence of interaural phase relations upon the masking of speech by white noise. J. Acoust. Soc. Amer., 1948, 20, 150-159.

Licklider, J.C.R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York: John Wiley and Sons, 1951.

Lüscher, E. and Zwislocki, J. Adaptation of the ear to sound stimuli. J. Acoust. Soc. Amer., 1949, 21, 135-139.

Marcum, J. I. Tables of Q functions. Santa Monica, Calif. Rand Corporation, RM-339, 1950. Also available as ASTIA Document Number AD 116551.

Marrill, T. Detection theory and psychophysics. Mass. Inst. Technol. Res. Lab. Electron., Rept. 319, 30 Oct. 1956.

Mayer, A. M. Researches in acoustics. LX. Phil Mag. (Ser. 5) 1894, 37, 259-288.

Miller, G. A. Sensitivity to changes in the intensity of white noise and its relation to masking and loudness. J. Acoust. Soc. Amer., 1947, 19, 609-619.

Peterson, W. W., Birdsall, T. G., and Fox, W. C. The theory of signal detectability. Trans. IRE Professional Group on Information Theory, 1954, PGIT-4, 171-212.

Pfafflin, S.M. and Mathews, M. V. Energy-detection model for monaural auditory detection. J. Acoust. Soc. Amer., 1962, 34, 1842-1853.

Pickett, J. M. Backward masking. J. Acoust. Soc. Amer., 1959, 31, 1613-1615.

Pollack, I., and Pickett, J. M. Stereophonic listening and speech intelligibility against voice babble. J. Acoust. Soc. Amer., 1958, 30, 131-133.

Rayleigh, Lord (J. W. Strutt.) The theory of sound. (2nd ed. 1894) Dover Publications, Inc., New York, 1945.

Rice, S. O. Mathematical analysis of random noise. In N. Wax. Selected papers on noise and stochastic processes. New York: Dover Publications Inc. 1954.

Schafer, T. H., Gales, R. S., Shewmaker, C. A., and Thompson, P. O. Frequency selectivity of the ear as determined by masking experiments. J. Acoust. Soc. Amer., 1950, 22, 490-496.

Shaw, E.A.G. and Piercy, J. E. Physiological noise in relation to audiometry. J. Acoust. Soc. Amer., 1962, 34, 745 (A).

Shower, E. G., and Biddulph, R. Differential pitch sensitivity of the ear. J. Acoust. Soc. Amer., 1931, 3, 275-287.

Stevens, S. S. and Volkmann, J. The relation of pitch to frequency. Amer. J. Psychol., 1940, 53, 329-353.

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

Swets, J. A. (Ed.) Signal detection and recognition by human observers. New York: John Wiley and Sons. 1964.

Swets, J. A., Green, D. M., and Tanner, W. P., Jr. On the width of critical bands. J. Acoust. Soc. Amer., 1962, 34, 108-113.

Tanner, W. P., Jr. and Birdsall, T. G. Definitions of d' and η as psychophysical measures. J. Acoust. Soc. Amer., 1958, 30, 922-928.

Tanner, W. P., Jr., Swets, J. A., and Green, D. M. Some general properties of the hearing mechanism. University of Michigan: Electronics Defense Group, 1956, Technical Report No. 30.

Watson, C. S. Masking of tones by noise for the cat. J. Acoust. Soc. Amer., 1963, 35, 167-172.

Watson, C. S., Rilling, M. E., and Bourbon, W. T. Receiver-operating characteristics determined by a mechanical analog to the rating scale. J. Acoust. Soc. Amer., 1964, 36, 283-288.

Wegel, R. L. and Lane, C. E. The auditory masking of one pure tone by another and its probable relation to the dynamics of the inner ear. Phys. Rev., 1924, 23, 266-285.

FOOTNOTES

[1]The expression of Eq. (2) can be obtained from the density functions for a bivariate normal distribution where x and y are independent and have the same standard deviation; i.e., σ_x and $\sigma_y = \sigma$. The joint density function is simply the product of the two independent density functions:

$$f(x, y) = 1/\sigma\sqrt{2\pi}\ \mathrm{Exp}[-x^2/2\sigma^2] \cdot 1/\sigma\sqrt{2\pi}\ \mathrm{Exp}[-y^2/2\sigma^2]$$

$$= 1/2\pi\sigma^2\ \mathrm{Exp}[-(x^2 + y^2)/2\sigma^2],$$

and letting $\rho^2 = x^2 + y^2$,

$$f(x, y) = 1/2\pi\sigma^2\ \mathrm{Exp}[-\rho^2/2\sigma^2].$$

This is the joint probability associated with a particular point (x, y) and not, of course, the density function for ρ, since there are an infinity of combinations of x and y that will yield the same value of ρ. To get the density function for ρ, we must first determine the probability, $P(\rho)$. The increment involved is a ring of radius, ρ, and width, $d\rho$. We must therefore multiply our expression by $2\pi d\rho$, and integrate:

$$P(\rho) = \int_0^\rho 2\pi\rho/2\pi\sigma^2\ \mathrm{Exp}[-\rho^2/2\sigma^2]d\rho = 1 - \mathrm{Exp}[-\rho^2/2\sigma^2].$$

Differentiating this expression yields

$$f(\rho) = \rho/\sigma^2\ \mathrm{Exp}[-\rho^2/2\sigma^2],$$

which is Rayleigh's distribution.

[2] The rationale for resorting to vector triangles is not always obvious to non-engineers, and for their benefit the following bit of trigonometry is provided:

Let us treat the near-sinusoid, the narrow band of noise, as if it were a sinusoid of the same frequency as the signal, but differing in phase at the moment, by the angle φ. The amplitude of the noise is _a_, and of the signal, _b_. Now, adding two sinusoids of the same frequency, but different phase, will yield another sinusoid of the same frequency, but usually different both in phase and in amplitude from the other two. Let us call the new amplitude, _c_, and the new phase angle, θ. We have then that

$$c \sin(2\pi ft + \theta) = a \sin(2\pi ft + \varphi) + b \sin 2\pi ft.$$

This expression holds for all values of _t_, and hence for $t = 0$; therefore

$$c \sin \theta = a \sin \varphi.$$

The expression also holds for _t_ such that $2\pi ft = \pi/2$; hence

$$c \sin(\pi/2 + \theta) = a \sin(\pi/2 + \varphi) + b \sin(\pi/2)$$

but $\sin(A + \pi/2) = \cos A$, and $\sin \pi/2 = 1$; hence

$$c \cos \theta = a \cos \varphi + b.$$

Squaring the two expressions:

$$c^2 \sin^2 \theta = a^2 \sin^2 \varphi \text{ and}$$

$$c^2 \cos^2 \theta = a^2 \cos^2 \varphi + 2ab \cos \varphi + b^2;$$

adding and combining terms:

$$c^2(\sin^2\theta + \cos^2\theta) = a^2(\sin^2\varphi + \cos^2\varphi) + 2\ ab \cos\varphi + b^2$$

and remembering that $\sin^2 A + \cos^2 A = 1$, we have

$$c^2 = a^2 + b^2 + 2ab \cos\varphi.$$

This is the familiar law of cosines, with the sign changed because of the fact that φ, as we have measured it, is the supplement of the included angle and that $\cos\varphi = -\cos(\pi - \varphi)$.

The demonstration tells us that we can find the new sinusoid by simply solving a triangle involving the two original amplitudes and the phase angle; such a triangle is shown in Fig. 8.

[3]It should be noted that this approach avoids several assumptions about which there have been arguments in the literature. We have not had to assume Fourier-series, band-limited noise, nor the applicability of sampling theory, and we have not employed likelihood ratios. These things are needed to obtain the analytical expressions, but we could have obtained our ROC curves by the brute-force method described. Our only assumption has been that the noise voltage is normally distributed.

[4]Tables of e^{-x} and $I_o(x)$ are to be found in the Handbook of Chemistry and Physics, Chemical Rubber Publishing Co., Cleveland, Ohio. Derivations and tables associated with the circular-normal distribution are to be found in the Handbook of Probability and Statistics with Tables, by Burlington and May, Handbook Publishers, Inc., Sandusky, Ohio, 1953.

[5]Figure 9 is taken from Jeffress (1964), who also presents a condensed version of Marcum's table, containing probabilities to three decimal places

for values of a_i from 0.2 to 4.0 by steps of 0.2 units, and values of $\underline{A}$, from 0.0 (noise alone) to 5.0, by steps of 0.5 units.

[6] Recently, Green (1965) has pointed out that increasing the size of the pedestal to where it, rather than the noise, is the dominant factor in masking, brings us under the jurisdiction of Weber's law. We are discriminating a change in the level of a tone, and $\Delta I/I$ should be nearly constant. There was nothing in the foregoing approach to predict this fact. If our noise is just noise ($A = 0.0$) and our signal has $A = 2.0$, we should obtain a d_s of about 1.3 (Jeffress, 1964). If we add a pedestal of $A = 2.0$ and employ a signal of $A = 4.0$, we should obtain the same d_s, and similarly for a pedestal of $A = 10.0$ and a signal of $A = 12.0$. As Green points out this is obviously in violation of Weber's law. It is like expecting a voltmeter to have the same accuracy in millivolts on the 100-volt scale as on the 1-volt scale. Green shows that we can get Weber's law back into operation either by assuming a "self-noise" which is proportional to the stimulus, or by assuming that, like the voltmeter, our detector has an error which is proportional to the stimulus. The latter idea appears more reasonable physiologically--the firing rate of the "amplitude" fibers of the VIIIth nerve is a quasi-logarithmic function of amplitude.

CHAPTER 15

BINAURAL SIGNAL DETECTION: VECTOR THEORY

Lloyd A. Jeffress

Department of Psychology and

Defense Research Laboratory

The University of Texas

TABLE OF CONTENTS

CHAPTER 15

BINAURAL SIGNAL DETECTION: VECTOR THEORY

Lloyd A. Jeffress

Department of Psychology and

Defense Research Laboratory

The University of Texas

I. INTRODUCTION

Many of the facts of binaural listening are summarized in the commonplace, "Two ears are better than one." Two ears provide a spare, they permit us to localize sound quickly and accurately, and they help us to detect a signal in noise--for example, speech in a background of other speech, the "cocktail-party effect," or a tonal signal in a background of thermal noise. Most of the present chapter will be concerned with this last, which can be rather strikingly exemplified by the following demonstration: We supply noise to one ear (earphone) at a comfortable listening level, then add a signal consisting of a 500 cps tone interrupted every quarter-second, and adjusted in level until it is just inaudible. If we now add the same noise to the other earphone, we find that the signal has become clearly audible. The signal again disappears when it too is added to the channel for the second earphone, making the sounds at the two ears alike. Now if we reverse the connections of either the noise input or the signal input (but not both) to one ear, we discover that the signal becomes loud and clear, and that it can be reduced in level by many decibels before it again becomes inaudible.

A. Terminology

To make the discussion of these phenomena as concise and as explicit as possible, let us follow common usage. Any improvement in detection that results from using two ears instead of one we will call masking level difference (MLD); this will be expressed in decibels. When the stimuli to both ears are the same in all respects--level, frequency, and phase--we will refer to the stimulus condition as diotic. The diotic condition is one homophasic condition; others result from altering both the signal and the noise to one ear in the same way--whether by reversing the phase of both, or by delaying both in time by the same amount. If the phase (or time) at one ear for the signal or for the noise (but not both) is altered relative to the other ear (for example, by reversing the connection for the noise to one ear) the condition is called antiphasic. If the noise for one ear is independent of the noise for the other (uncorrelated) the condition is called heterophasic. Any binaural condition which is not diotic is dichotic. The stimuli may be dichotic in phase, in time, in level, in frequency, and in many other ways.

To make our notation more specific and more complete, let us adopt the following symbols for the various combinations of noise and signal, using N for noise and S for signal, π to indicate a phase reversal at one ear relative to the other, 0 to indicate no phase or time difference between the ears, u to indicate that the noises at the two ears are uncorrelated (from separate sources), and m to indicate that the noise or the signal is monaural. We can list a number of the combinations as follows:

Monaural (Monotic)

Nm Sm	Noise and signal both monaural (same ear)

Homophasic

N0 S0	Noise and signal both in phase at the ears (Diotic)
$N\pi$ $S\pi$	Noise and signal both reversed in phase at one ear relative to the other (Dichotic)

The remaining conditions are all dichotic.

Mixed

NO Sm	Noise in phase, signal monaural
Nπ Sm	Noise reversed in phase, signal monaural

Antiphasic

NO Sπ	Noise in phase, signal reversed in phase at one ear
Nπ SO	Noise reversed in phase, signal in phase

Heterophasic

Nu SO	Noise uncorrelated, signal in phase
Nu Sπ	Noise uncorrelated, signal reversed in phase

B. A Typical MLD

Hirsh (1948) found that for a 500 cps signal the NO Sπ condition yielded thresholds about 11 dB lower than those for the diotic condition (NO SO). Since later studies have shown that the diotic condition does not differ reliably from the monotic (e.g., Blodgett, et al 1958), we may refer to Hirsh's 11 dB as a masking level difference (MLD).

The fact that Hirsh's subjects showed an MLD as large as 11 dB means that most of the signals to which they were responding were too weak to be detected by monaural means. In terms of the theory of signal detectability (TSD), if the 50% thresholds for Hirsh's subjects corresponded to d's of about 1.5 (see Jeffress, 1964, p. 772), the d's for monaural detection of the same signal would be about 0.25 or less. This tells us that the signal in the antiphasic case must be detected in almost every instance by means of some binaural detection mechanism. The present chapter is concerned with the nature of such a mechanism and with the attempt to formulate a model which will account for such results as Hirsh's.

II. THE STIMULUS FOR BINAURAL DETECTION

A. Webster's Hypothesis

To find what the stimulus may be that makes binaural detection possible, let us physically compare the voltages to the two earphones under the diotic condition and under the antiphasic. We may do this by connecting the wires for one earphone to the horizontal plates of an oscilloscope, and those for the other to the vertical plates. In this way we shall obtain a Lissajous pattern. For the diotic condition, the pattern will be a very narrow ellipse (almost a straight line) running from lower left to upper right and indicating an almost perfect positive correlation. If our equipment were perfect, the correlation would be perfect and the pattern would be a straight line.

If we reverse the connections for both the noise and the signal to one earphone ($N\pi$ $S\pi$), producing the other homophasic condition, we obtain a narrow ellipse running from lower right to upper left--an almost perfect negative correlation. Now if we reverse the connections for the signal but not for the noise ($N0$ $S\pi$), we obtain a drastically different pattern, spreading across the face of the scope. Even when we reduce the signal to a very small value, we still see some spreading of the pattern from what is the case where only the in-phase noise is present, or in-phase noise and in-phase signal. The signal-to-noise ratio needed for detection is vanishingly small for the antiphasic condition. For the diotic condition, we can detect the signal only as a change in the length of the narrow Lissajous pattern, and a substantial signal is required to produce a definite change; for the antiphasic condition, however, the addition of even a very small signal will produce enough spreading to be easily noticed.

The difference in the stimuli for the diotic and the antiphasic cases is obviously great; how do the ears make use of it? The spread of the Lissajous pattern in the antiphasic case suggests phase (the phase relation between noise plus signal at one ear and noise plus signal at the other) as the possible basis.

Webster (1951) proposed that the band of noise that masks a tone, the critical band, can be thought of as resembling a sinusoid which varies slowly in phase and in amplitude, and that adding a tonal signal will yield a resultant that generally will differ in phase from the original noise. If, in a particular instance, adding the signal at one ear advances the signal plus noise in phase, adding the signal reversed in phase at the other ear will retard it. There will be a phase difference between the signal plus noise at one ear and the signal plus noise at the other. It is this interaural phase difference which provides the basis for binaural detection. Webster assumed that the interaural phase difference was represented to the central nervous system as an interaural time difference. He computed interaural time differences for the case where the signal and the narrow band of noise were in quadrature at the moment of addition, using vector triangles to illustrate the process.

B. Vector Diagram for Hirsh's Data

Let us follow Webster, but employ the procedure of Chapter 13 to arrive at our vector triangles. Figure 8 of Chapter 13 gives an example. The signal and the narrow band of noise may be in any phase relation at the moment of addition, and the noise may have any amplitude. To simplify making an appropriate drawing, let us choose typical values for phase and for amplitude. For phase, 45° is a convenient mid-value, and for amplitude, the median is typical. From Eq. (3) of Chapter 13 we can determine the median by substituting 50% (0.50) for $P(a > a_i)$ and solve for a_i. We find it to be 1.177σ. Let us use this value for the length of our noise vector.

For the NO $S\pi$ condition at 500 cps, Hirsh employed a noise having a spectral level of 59.1 dB. If we assume a critical bandwidth of 50 cps, we obtain an effective level for the noise of 76.1 dB (59.1 + 10 log 50), Hirsh found the 50% threshold to require a signal level of 64.1 dB. If we take the rms noise voltage, σ, to be unity (see Chapter 13, p. 11), the signal voltage (76.1 - 54.1 = 12 dB less) will be 0.25 units (-12 = 20 log 0.25).

Figure 1 is a vector diagram illustrating the addition of signal to noise for this typical instance. The noise vector is drawn as having a length of 1.177 units, since the median value of the noise amplitude is 1.177σ. The signal is drawn as having a length of 0.25 units, and at a typical angle, 45°. The resultant noise-plus-signal amplitude is 1.36 units and the phase angle between the noise vector and the noise-plus-signal vector is 7.5°. The small change of amplitude from 1.177 to 1.36 would rarely be detected monaurally.

Figure 2 is the vector diagram for the NO $S\pi$ condition. The noise vector is the same for both ears, but the signal vector for the left ear is drawn in the opposite direction from that for the right because of the interaural phase reversal for the signal. Relative to the noise, the right ear leads by 7.5° and the left ear lags by 10°. The right ear leads the left, therefore, by 17.5°. The equivalent difference in the time of corresponding events at the two ears will be 97 μsec (17.5°/360° times the period of a 500 cps tone, 2.0 msec). It is this time difference which is responsible for the binaural detection of the signal.

III. NEURAL MECHANISMS

A. Peripheral Mechanism

In order to make use of the fact that the stimulus to one ear leads the stimulus to the other by a ten-thousandth of a second, the nervous system must provide two mechanisms--a peripheral one which preserves the temporal information

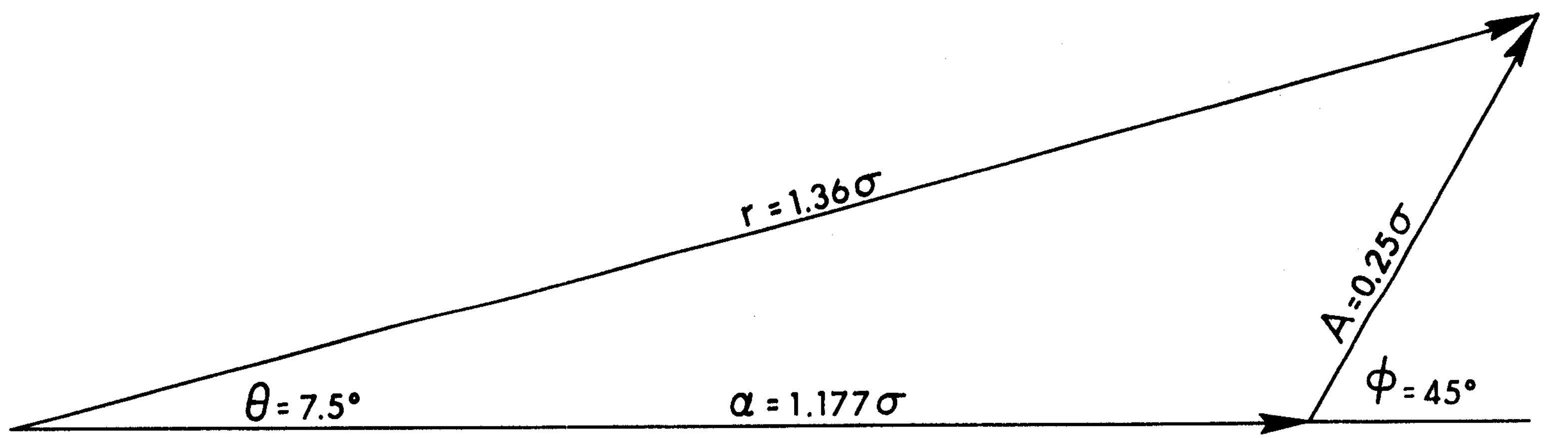

Fig. 1. Vector diagram for the NO SO homophasic condition.

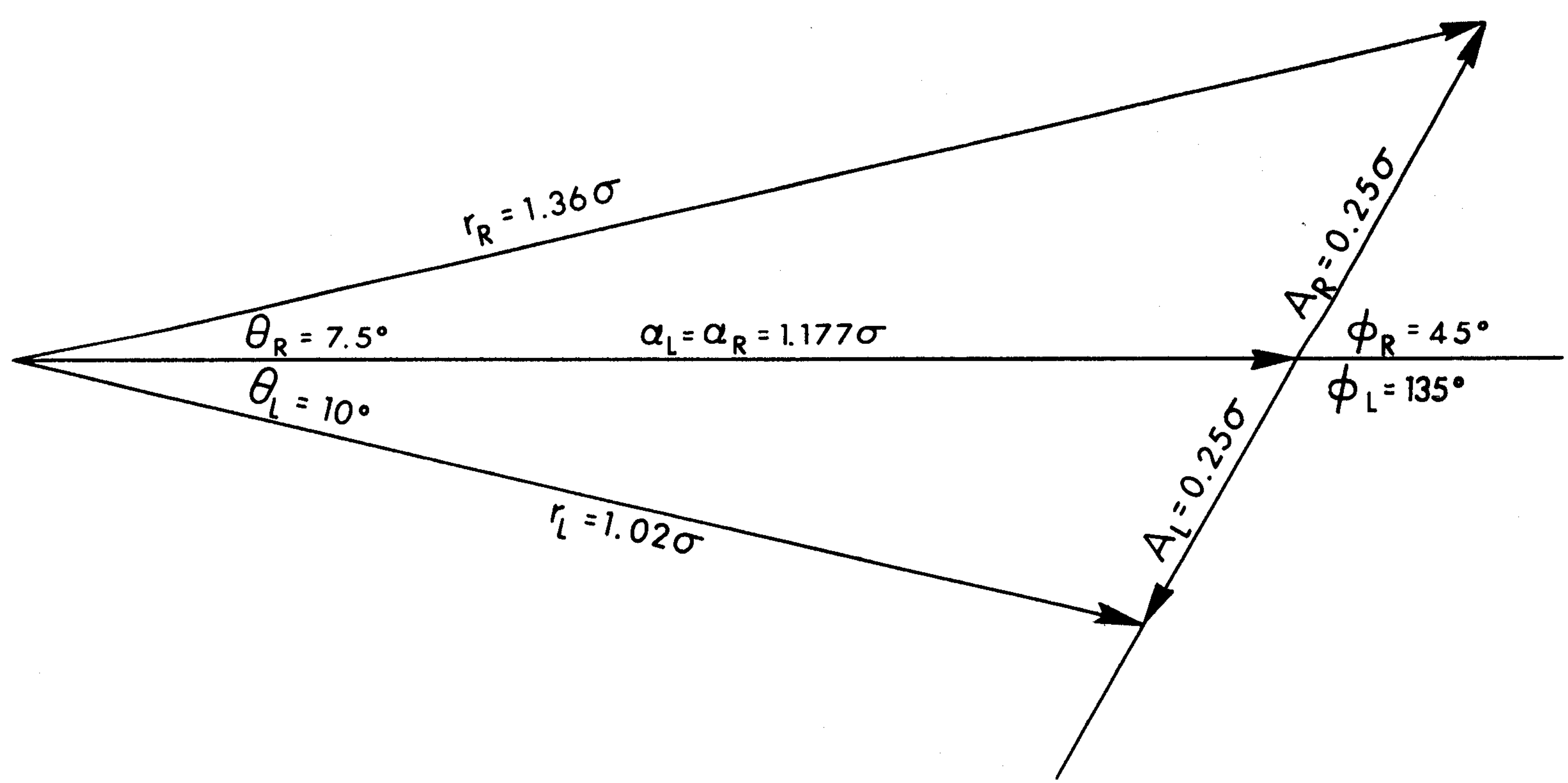

Fig. 2. Vector diagram for the NO Sπ antiphasic condition.

in the stimulus and transmits it to a higher center, and a central one where the temporal information from the two sides is compared. Rutherford (1886), who was apparently the first to realize that such machinery was essential to an explanation of our ability to localize tones, incorporated it into his "telephone" theory. The discovery by Adrian (1914) that neurons operate on the all-or-none principle, and that they are refractory for a moment after conducting an impulse, confused the issue for a time, but on the basis of their experimental work, Wever and Bray (1930) affirmed the ability of the auditory nerve to "follow" the sound up to frequencies as high as any involved in the localization of tones by time or phase differences.

Recent work (see Chapter 7) has shown that there are neurons in the auditory nerve which fire in step with the stimulus, with interspike intervals equal to the period of the tone or to multiples of the period. They appear to be the fibers employed in localizing tones, and if so, they also provide the basis for the binaural detection of tonal signals.

B. Central Mechanism

Jeffress (1948) described a hypothetical central mechanism for converting a difference of time into a difference of place. It depended for its operation upon neural summation (spatial summation at a synapse) of simultaneously arriving impulses from the two cochleas, and achieved the necessary simultaneity by neurally delaying the impulses from the ear that led in time. The neural delay introduced into the leading channel matched the acoustical delay in the lagging. Jeffress thought of the mechanism as having left-to-right in space distributed along one spatial dimension in the nervous system, and low-to-high in pitch, along another. In modern usage, the mechanism would be described as a cross-correlation matrix, with each synapse serving as a multiplier or "AND" gate. While the device was originally proposed for localization, it has been employed (see Jeffress, Blodgett, Sandel, and Wood, 1956) to account for many of the binaural phenomena of masking.

Galambos (1957), Galambos, Schwartzkopff, and Rupert (1959), and Moushegian, Rupert, and Whitcombe (1964) (see Chapter 20), have described cells located in the accessory olivary nucleus (medial superior-olivary nucleus) which are, as Galambos put it, "...exquisitely sensitive to whether or not the sounds have been presented simultaneously." They differ from the cells hypothesized by Jeffress in showing inhibition rather than summation, in being "Not AND" gates rather than "AND" gates, but they do show frequency selectivity, and do appear to be distributed in the times they require for left-right resolution.

The population of cells studied to date is not large enough to justify a very elaborate model, but we may safely assume that a region (or regions) exists in the central nervous system where time differences are represented in some form to higher centers. Whether this representation is spatially distributed as Jeffress suggested, or is dependent upon differences in activity of left and right on-going channels (Békésy, 1930, van Bergeijk, 1962) is not crucial to our present discussion. We may also safely assume that the region is distributed in frequency.

On the basis of psychophysical findings, we appear to be justified in making the further assumption that the cell populations associated with the median plane are larger than those associated with regions lying to the right or left. The greater accuracy of localization near the median plane as compared with more lateral positions (see Chapter 17) supports this assumption.

The two assumptions, frequency selectivity and greater cell density for the median plane, are given additional support by the following rather striking demonstration: The subject listens to a narrow band of noise centered at 500 cps and lagging at one ear by 2.0 msec. He hears the noise as located in the middle of his head, much as if it were a 500 cps tone delayed by one period and hence in phase again. Now we increase the bandwidth of the noise. The subject hears the sound spread out from the center. As the bandwidth is widened to a few hundred cycles per second, the subject hears the sound move toward the leading side, and by the time the bandwidth is two- or three-thousand cycles per second, he hears the sound as tightly bunched at the leading ear. He is no longer aware of the 500 cps narrow band of noise, which when it was the only noise present, he heard clearly as in the center of his head.

Our model explains the foregoing demonstration rather simply. The broad-band noise, lagging in its arrival at one ear by 2.0 msec, produces neural activity in the channels for that ear which can be matched with corresponding activity in the channels for the other ear by introducing a 2.0-msec neural delay in those channels. The coincidence thus established yields the perception of sidedness, and because all frequencies require the same delay for temporal coincidence, they will all have the same lateral position: 2.0 msec toward one side. At the same time, the 500 cps part of the noise is achieving coincidence in the middle of the head and with no neural delays, through matching the first cycle from the lagging ear with the second cycle from the leading ear, and so on. This sound, however, is not heard because the preponderance of lateral activity where all frequencies are localized obscures it. When the bandwidth is reduced, leaving only a narrow band centered at 500 cps, the denser population associated with the median plane prevails and dominates perception. The sound is heard as centered.

To make this description more explicit, let us assume that there are a thousand cells associated with a given frequency-band at the median plane, and only ten with the same band for a neural delay of 2.0 msec. A 500-cps, narrow-band noise lagging by 2.0-msec at one ear would, therefore, stimulate a thousand cells (through coincidences between the second cycle from the leading side and the first cycle from the lagging) for every ten cells that were stimulated through coincidences between the acoustically delayed impulses from one side and neurally delayed impulses from the other. The thousand will prevail over the ten, and the sound will be heard as centered. When we increase the range of frequencies, we increase the number of tens proportionately, but not the thousands. Only the 500-cps band will be centered. Other nearby frequencies will lie to the right and to the left. More distant frequencies will lie farther to the right and to the left; for them neural delays will be required for coincidence between the lagging first impulses from one side and the second impulse from the other. Only for 500 cps will these coincidences fall in the median plane. But a neural delay of 2.0 msec will produce coincidence between the acoustically lagging first cycles from one side and the neurally delayed first cycles from the other for <u>all</u> frequencies. If we make the band of noise wide enough, the many tens will prevail over the few thousands, and will dominate perception. We will hear the sound as sided.

C. <u>Huggins's Pitch</u>

The fact that our hypothetical central mechanism is distributed in the frequency/pitch domain prepares us for another phenomenon: that pitch can be created by binaural interaction where none exists with either monaural stimulus. Huggins and later, Cramer and Huggins (1958) shifted the phases of the frequencies in a wide band of noise, moving them in one direction when they were above 600 cps and in the opposite direction when they were below 600 cps. The resulting noise was still "white" and sounded exactly like the unaltered white noise when heard alternately with it. But when the altered noise was applied to one ear and the unaltered to the other, the subject heard a 600 cps "pitchiness" in the middle of his head. The sound had the quality of a 600 cps narrow band of noise. This is exactly what we should expect. The unshifted 600 cps component of the noise will achieve cycle-by-cycle coincidences in the median plane, but the phase-shifted 700 cps component will be located to one side where the neural population is smaller, and the 500 cps component to the other side, where again the population is smaller. The larger 600-cps population will dominate perception and we will hear a centered, 600-cps band of noise, having the pitch associated with 600 cps.

D. Central Mechanism and Binaural Detection

Let us assume that we are listening to the noise of Fig. 2. The part of the noise involved in masking is a narrow band centered at 500 cps. It is sending a series of impulses to the 500 cps region of the central mechanism. Pairs of these impulses arising from corresponding parts of the waveforms at the two ears will arrive more or less simultaneously at the center. If the ears were perfect transducers and if the coincidence devices of the central mechanism were also perfect, the resulting localization would be perfect and we would hear a sound precisely centered in the median plane. We know, however, (see Chapter 20) that the neural following of the sound is not perfect; there will be some fluctuation, and impulses will be generated at one ear a little earlier in the cycle, or a little later, than the corresponding impulses at the other ear. This will produce slight fluctuations in the times of arrival of the impulses at the central mechanism, and consequently, a fluctuation in the left-right position of the sound image. We can think of this fluctuation as being a new kind of "noise"; as being, in fact, the "noise" that is going to do the masking. If it did not exist, there would be no masking. The acoustic noise, the stimulus, would provide a precise center, a reference from which the left-right departures that result from adding the signal would stand out sharply.

Now let us add the signal. Figure 2 shows us one of the possible results. For the particular combination of phase and amplitude shown there, the impulses arising at the right ear will lead the corresponding impulses from the left ear by 97 μsec. Before the signal was added, the impulses were arising more or less simultaneously. As a consequence of adding the signal, the image will move from the median plane, and the noise-plus-signal image will appear 97 μsec to the right. The next signal, involving a new combination of amplitude and phase, may yield a noise-plus-signal image which moves farther to the right or to the left, or which moves so little from the median plane as to fail to escape from the "noise" and so go undetected.

Recalling that the localization mechanism is distributed in the frequency-pitch domain as well as in the time, we will expect the image to acquire pitch when it moves from the median plane. Anything that makes the 500-cps region conspicuous, different in activity from other frequency regions, will arouse the pitch associated with 500 cps. We will <u>detect</u> the signal because of the movement from the median plane that it produces, but we will <u>hear</u> it as a 500-cps "pitchiness."

Recently, Hafter, Bourbon, Blocker, and Tucker (1964) have reported a study in which the subjects positioned a slider, similar to that used by Watson, <u>et al</u> (1964), to indicate where in the head they heard the signal. The ends of the left-right travel of the slider represented the ears, and the center, the middle of the head. When a very weak signal was employed, the responses were bunched toward the center, but with a stronger signal they were more spread out. For very strong signals, the responses were bunched near the ears. This is exactly as predicted from our theory. Any given signal is equally likely to be to the left or to the right, and the amount of movement will be a function of the momentary interaural phase difference, and hence of the length of the signal vector.

IV. TSD AND BINAURAL DETECTION

A. ROC Curves

In the same way that the ROC curves for human observers were used in Chapter 13 to infer something about the underlying distribution functions for noise and for noise plus signal, so they may be used here for binaural detection. Watson, Rilling, and Bourbon (1964) repeated the rating-scale experiment described in Chapter 13, this time for the NO $S\pi$, antiphasic stimulus condition. The resulting ROC curves were very different from those they had obtained earlier with the same subjects, but under the diotic condition. Just as the earlier curves were different from curves derived from overlapping normal distributions, and suggested instead the Rayleigh function for amplitude, so our present curves are different from either, and suggest still different distribution functions for noise and plus signal. This should not surprise us, since we believe that our present stimuli are interaural time differences; our "noise" is neural "noise" and our "signal" is the interaural time difference which results when we add a tonal signal to the acoustic-noise stimulus. To distinguish which noise and which signal is being discussed, let us continue to employ quotation marks when we refer to the neural "noise" or "signal" and omit them when we are referring to the physical stimuli.

B. Distribution Function for "Noise"

We have pictured the "noise" with which we are concerned as a random fluctuation around the median plane, caused by irregularities in the arrival times of corresponding impulses from the two ears. The distribution of these time differences should be symmetrical around zero. We should also expect it to be normal, since the differences are essentially "errors" and are due to a multiplicity of causes. The subject's responses should therefore be some sort of transform (power function?) of half of a normal distribution, since he is responding to the magnitude of the interaural time difference, not to its sign. Let us take half of a normal distribution then as representing the "noise."

C. Distribution Function for "Signal"

The "signal" for binaural detection results (as in Fig. 2) from the phase difference between the vector sum of noise plus signal at one ear and that at the other. To determine how the corresponding time differences are distributed, let us use the construction of Fig. 2 with the data from the experiment of Watson, Rilling, and Bourbon. Their signal had a level of 54 dB; their noise had a spectral level of 50 dB. In Fig. 2 we assumed a critical bandwidth of 50 cps, but recent work by Langford and Jeffress (1964) indicates that 100 cps would be a better estimate of the bandwidth associated with binaural detection at 500 cps. Let us use that value. The effective level of the noise will then be $50 + 10 \log 100 = 70$ dB. If we take the rms voltage of the effective band of noise corresponding to this as unity, the rms signal voltage will be 0.158 units ($20 \log S = 54 - 70 = -16$ dB). We will use 0.158 as the length of our signal vector. If the rms noise voltage, σ, is taken as unity, we can determine from the Rayleigh distribution a series of equally probable values of the noise amplitude to use as our noise vectors (see Chapter 13).

By taking ten mid-decile values of the noise amplitude and twelve equiprobable phase angles (every 15°), we obtain 120 equally probable values of the interaural phase angle and 120 corresponding time differences. Figure 3 is a histogram of these time differences. The shape of the distribution is characteristic, and appeared when other values were chosen for the signal amplitude. The "hump" in the tail of the distribution is due to the frequently-occurring large interaural time differences which result when the noise amplitude is small, and the signal, therefore, relatively large.

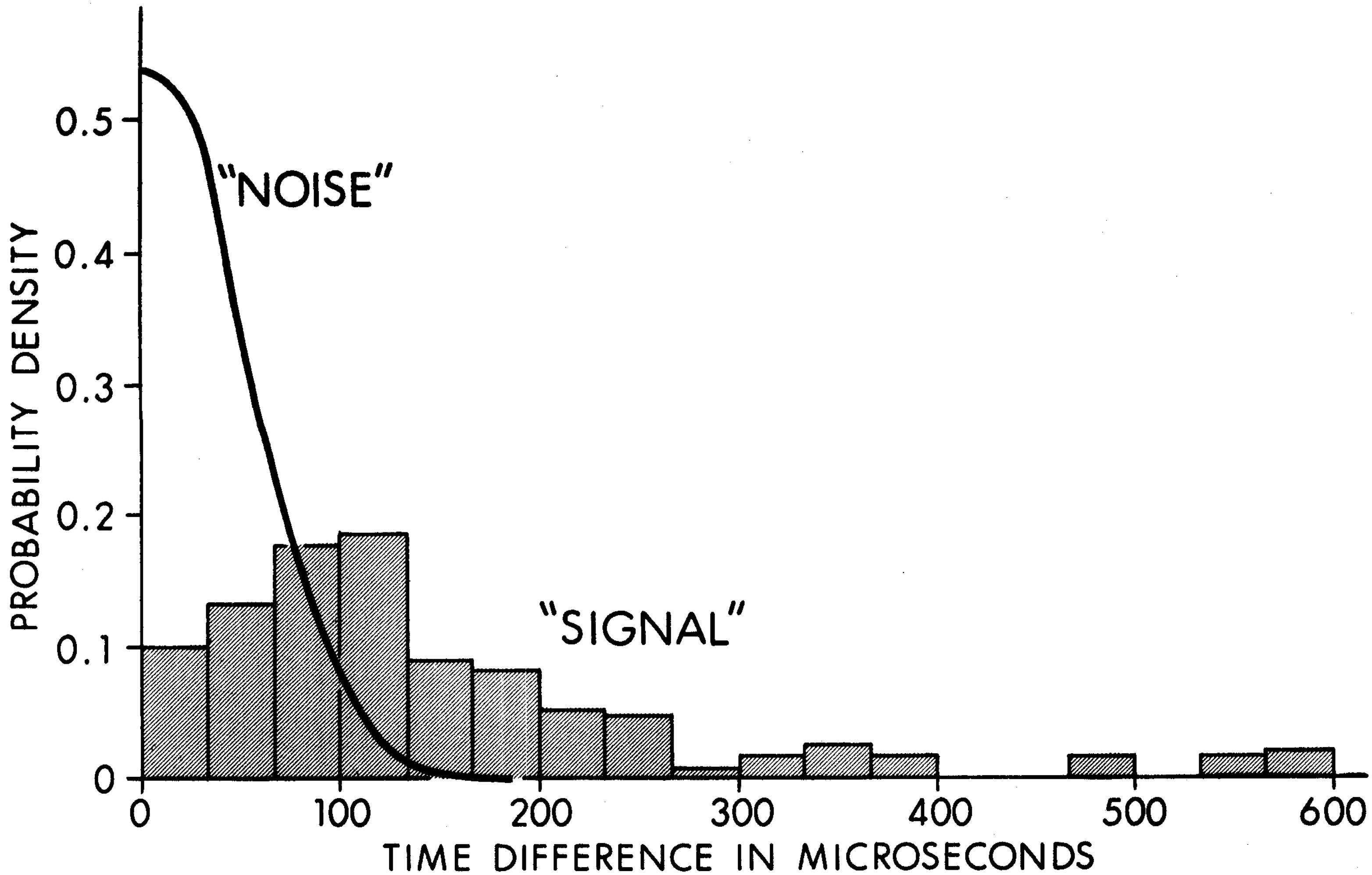

Fig. 3. Probability density functions for "noise" and "signal" for the NO $S\pi$ antiphasic condition.

D. Distribution for "Signal" plus "Noise"

The subject's decision is based upon the difference in magnitude between the "noise" and the "signal-plus-noise." We need therefore to know the nature of the SN distribution. Adding "noise" to the "signal" means adding (or subtracting) a small, normally distributed, time difference to the generally larger, oddly distributed, time difference that is the "signal." The effect of this addition will be to make the variance of the SN distribution slightly greater than that of the S distribution, and to make it a little more nearly normal. If we disregard the, probably slight, change of shape, we can obtain the SN distribution from the S distribution by simply a change of scale.

E. ROC Curve for NO Sπ

Figure 4 presents the data for one of the subjects of the study by Watson, Rilling, and Bourbon, and a curve derived from the distribution for S (SN) of Fig. 3, and half of a normal distribution for N. The ratio of standard deviations for the two distributions was chosen to yield a point on the negative diagonal that was in agreement with the experimental data at that point. The remaining points on the theoretical curve were obtained from the areas under the two distributions. The computed value of the standard deviation for the S (SN) distribution was 167 μsec, and the standard deviation of the N distribution required to yield the desired point proved to be 51 μsec.

An attempt was made to fit the data of Fig. 4 by employing half of a normal distribution for SN and chosing the ratio of standard deviations in the same way as before. The result was the dashed curve of Fig. 4. It will be seen that while the fit is about as good as that for the solid curve in the region to the right of the negative diagonal, the fit to the left (high criterion region) is poor. The normal distribution yields the prediction of too few false alarms for a given detection level.

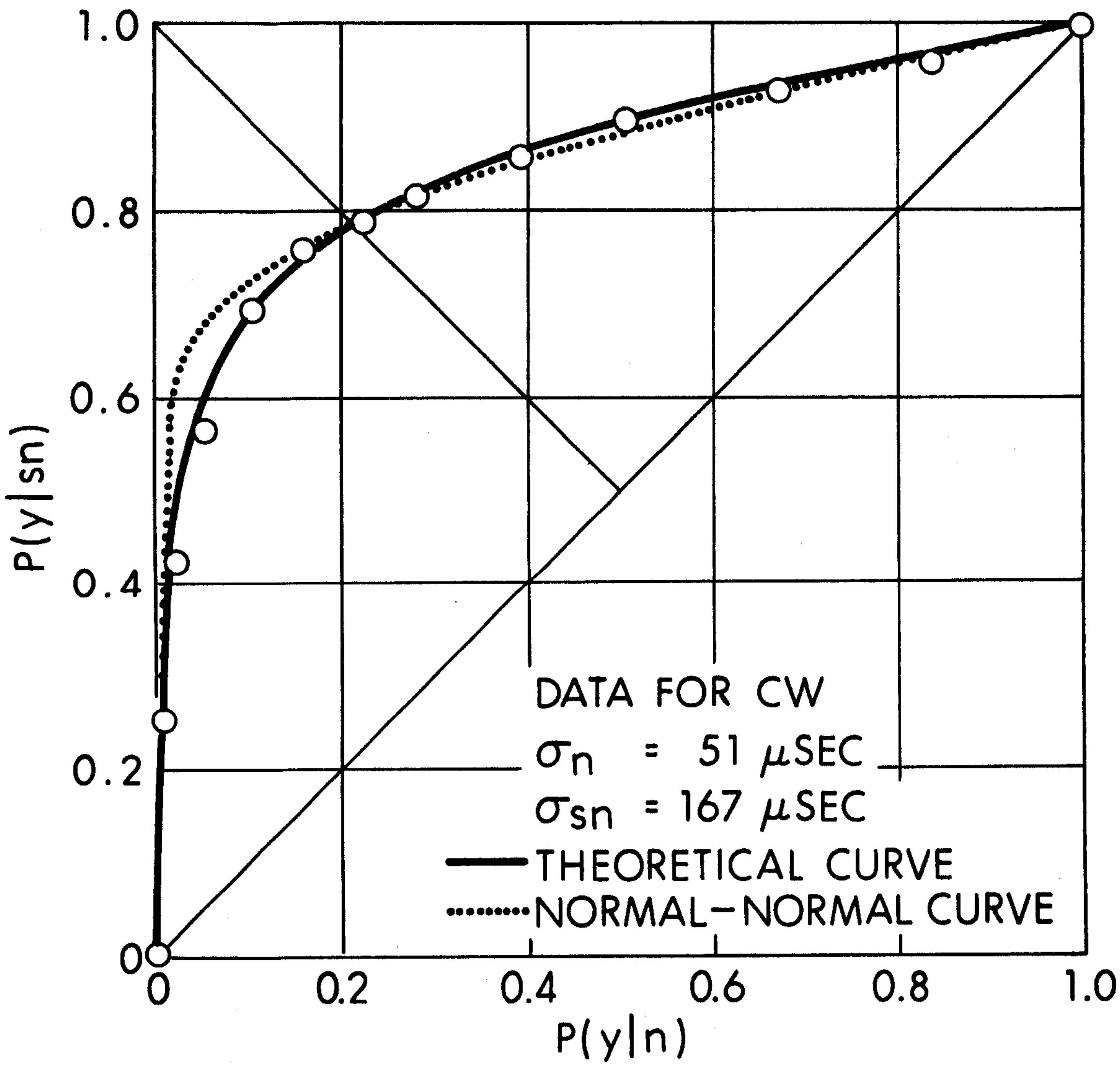

Fig. 4. ROC curve obtained from Fig. 3. Points are from Watson, Rilling, and Bourbon, (1964).

Data for the other two subjects of the experiment were not fitted so well by the theoretical curve as the data shown. The other two subjects exhibited, in the high criterion region, higher false alarms rates for similar detection levels. Attempts to fit their data by a different adjustment of the ratio of standard deviations did not improve matters greatly, nor did assuming a different value for the effective level of the signal. A good fit could be obtained only by assuming that the subjects' "noise" distribution was somewhat platykurtic (Pearson's β_2 ca 3.6). No theoretical basis for such a distribution is immediately apparent.

V. OTHER BINAURAL STIMULUS CONDITIONS

A. Binaural Noise, Monaural Signal

The case where the noise is binaural and in phase at the ears and the signal is monaural (NO Sm) is an interesting one. Hirsh (1948), in his pioneering study of binaural phenomena, investigated this case and found that it provided some release from masking compared with the homophasic conditions. The poorer detection in the homophasic cases he referred to as instances of "binaural inhibition." Later work, growing out of his, has shown that the two homophasic conditions are no worse than the monotic (see e.g., Jeffress, et al 1956) and that if we use the monotic condition as the base from which to determine masking level differences, the differences are positive (or zero). A possible exception occurs when the noise for the two ears comes from independent sources. Whitmore (1964) and Mulligan (1964) have both found small negative MLDs for the Nu Sm condition as compared to the Nm Sm. Miller (1964), on the other hand, has found either no difference or a slight positive one. In either case the difference is small, usually less than 1 dB, much smaller than the difference between NO Sm and Nm Sm.

Figure 1 can be used to illustrate the NO Sm condition if we take the noise vector as representing the noise stimulus to both ears. The signal vector is monaural, and the resultant phase angle, θ, is both the phase angle between the resultant and the original noise, and between the resultant at one ear and the noise at the other. It is therefore the interaural phase angle, and is the basis for the binaural improvement in detection that occurs under this stimulus condition.

The size of the interaural angle is independent of the length of the noise vector at the "off" (non-signal) ear so long as this noise is co-linear with the other noise vector. On this basis we should therefore predict that the MLDs obtained in the NO Sm condition would be independent of the noise level at the "off" ear, so long as there was <u>some</u> noise. That this prediction is not true is due to the presence of the neural "noise." Blodgett, Jeffress, and Whitworth (1962) varied the level of the noise in the "off" ear from a value equal to that in the ear receiving the signal down to zero voltage. They found a progressive reduction in the corresponding MLDs. When the noise in the "off" ear was more than 10 dB below that in the ear receiving the signal, the MLDs were zero. Recently, Egan (1964), using a more sensitive psychophysical method, has found a measurable MLD even with noise 40 dB below that in the ear receiving the signal. Under these conditions, Egan's subjects described the noise as being "monaural" but nevertheless gave MLDs indicative of some binaural detection.

The angle between the noise vector at one ear and the noise-plus-signal vector at the other is equally often positive and negative. The subject therefore has no way of knowing which ear is receiving the signal unless he detects it monaurally, which he will not often do with the weak signals employable in this stimulus condition (NO Sm). Egan and Benson (1964) tested this prediction by presenting the signal randomly to one ear, or the other ear, and asking the subject to indicate which ear received the signal. The subjects' performances on the left-right part of this task was near the chance level while they were still able to yield substantial detection scores. It was only when they made the signal large enough to yield good monaural detection scores, that they obtained reliable left-right information from their subjects.

B. Effect of Noise Correlation

Licklider (1948) showed that reducing the interaural correlation for the masking noise reduces the MLDs associated with reversing the phase of the signal. We may denote this stimulus condition as N+ Sπ, where the plus sign indicates that the noise at the two ears is positively correlated but with a correlation less than unity. Licklider achieved this reduction of correlation by adding noise from two additional noise generators, one for each ear. Part of the noise received by his subjects was therefore NO (the common part, perfectly correlated) and part was Nu. Licklider showed that the correlation coefficient is given by the ratio of the power in the correlated part to the total power, provided that the spectra of the three noises are the same, and that matching levels are used at the two ears.

Where Licklider had used speech as his signal, Jeffress, Blodgett, and Deatherage (1953) used a 500 cps tone. They found a reduction in the MLDs as the correlation was reduced from +1.0 to 0.0. Later, Robinson and Jeffress (1963) repeated and extended the work of Jeffress et al using a two-interval forced-choice procedure with a tonal signal. Both groups found a rapid reduction in MLDs as the correlation was reduced slightly, with a slower rate of reduction for further reductions of the correlation.

Jeffress, Blodgett, and Deatherage (1952) reduced the correlation for the noise by a different method; by adding a time delay to the noise channel for one ear, and thus reducing the autocorrelation. Their findings suggested that the MLD drops more rapidly when the correlation is reduced by a time delay than when it is reduced by adding uncorrelated noise. The autocorrelation for a rectangular band of noise is given by

$$\rho = \frac{\sin \pi W\tau}{\pi W\tau} \cos 2\pi f\tau$$

where W is the bandwidth, f is the center frequency and τ is the time delay in one channel. Jeffress et al assumed a bandwidth of 50 cps in their computation of the correlations. Langford and Jeffress (1964) repeated and extended the work, carrying the time delays to 9 msec, and again basing the calculations on a bandwidth of 50 cps, found the same rapid drop in MLDs as the correlation was reduced from 1.0. In a later unpublished experiment, Langford and Jeffress found that a narrow band of noise (50 cps) yielded much larger MLDs for the NO $S\pi$ stimulus condition than those associated with a wide band of noise, where for the diotic condition (NO SO) detection was the same with the narrow-band masker and the wide-band. They inferred from this that the band of noise involved in binaural detection is wider than that for monaural or diotic detection, arguing that since narrowing the band to 50 cps did not appreciably improve detection in the diotic case but made about a 10 dB improvement in binaural detection (NO $S\pi$), the bandwidth involved in binaural detection must be considerably wider than 50 cps. When they employed 100 cps in their calculations of the autocorrelation function, the MLDs plotted against correlation fitted a similar plot of the data from the experiment by Robinson and Jeffress almost perfectly.[1]

The finding, that the bandwidth involved in binaural detection is different from that for monaural should not surprise us; the types of auditory-nerve fiber involved are different. Monaural detection appears to utilize cells that fire at a rate determined by the amplitude of the stimulus, while in binaural detection, the firing rate (or interspike interval) is determined by the frequency, and largely independent of amplitude. The bandwidth for monaural detection may depend on a combination of mechanical filtering and neural "funneling" (see Chapter 13), while for binaural detection, the bandwidth may conceivably be determined solely by the mechanical action of the cochlea.

C. Vector Description of Correlation

Let us examine the stimuli in two cases, one where the correlation is reduced by adding a one-period time delay to the noise for one ear, and the other by adding random noise from two additional noise generators. Figure 5 shows the noise vectors which could result from a high positive (ca 0.8) correlation for the two cases. The upper figure shows the case where a one-period time delay has been introduced. The noise vectors make a slight angle with one another due to the possible phase shift in the noise from one cycle to the next. The slight difference in the length of the vectors reflects the possible change in amplitude that can occur during one period.

The lower figure shows the case where the reduction of correlation has been brought about by adding random noise. Here there is a common vector representing the common, correlated part of the noise. To this for one ear has been added another noise vector, representing noise from a second generator. The noise for that ear is the resultant of the two noises. For the other ear, the noise is the vector sum of the common part and a new, uncorrelated part. This yields a second resultant which is the noise for the second ear. The two resultants can thus differ in phase and in amplitude. In both figures, the amount of difference in phase and in amplitude between the noise vectors for the two ears, depends upon how much the correlation has been reduced from unity, and for the high correlation shown, the two figures are necessarily very similar. The similarity of the noise vectors appears to justify the prediction of a similarity in masking---a prediction borne out by the finding of Langford and Jeffress.

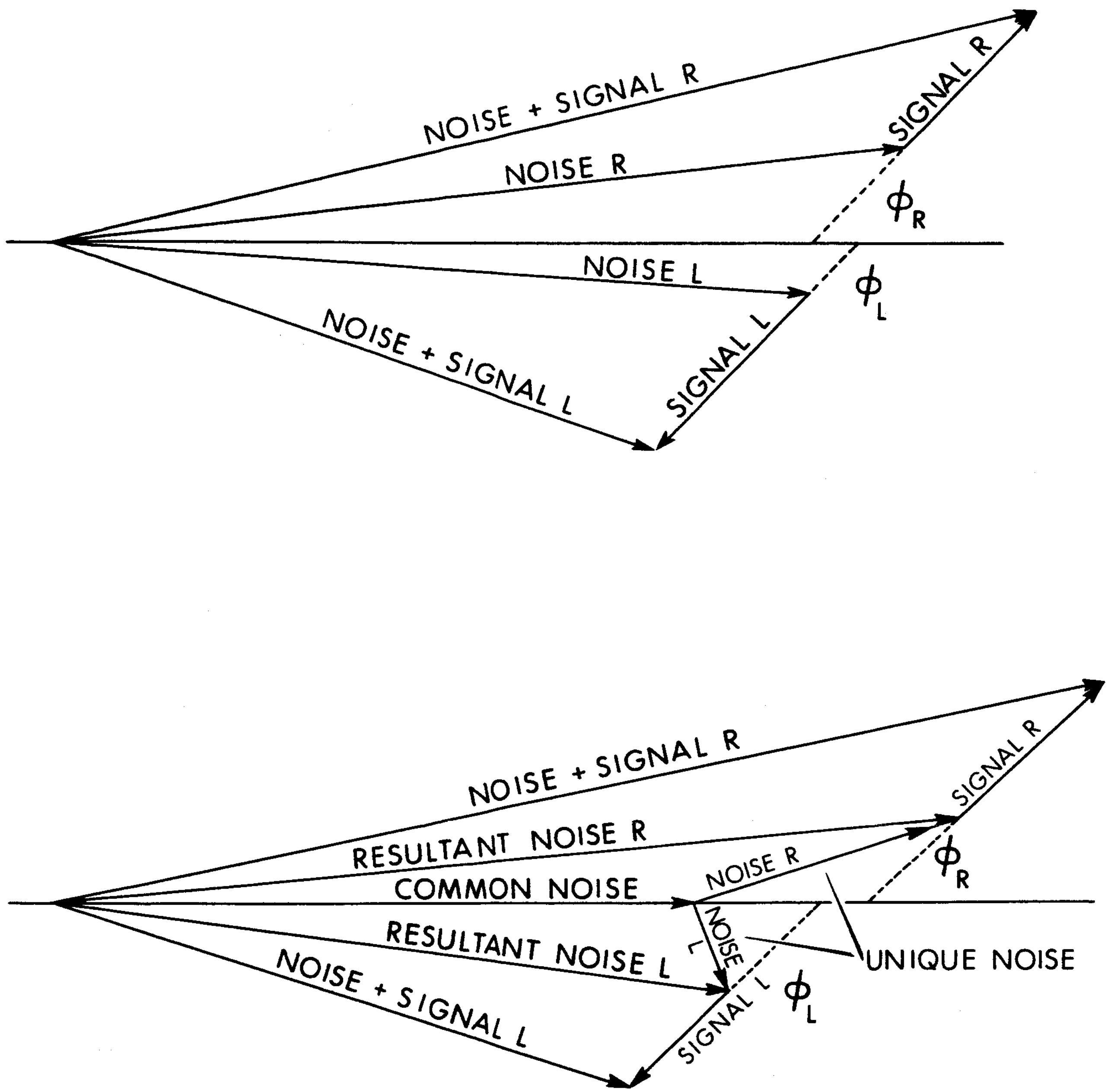

Fig. 5. Vector diagrams for NO Sπ where the noise correlation is reduced: (a) by adding a time delay in one channel, (b) by adding independent noise.

The subject's experience in the two stimulus conditions represented by the drawings of Fig. 5 are very different. When the correlation is reduced by adding a one-period time delay to the channel for one ear, the subject hears the noise as bunched at the undelayed ear, as described earlier in the present chapter. In the second case, he hears the sound as being in the median plane with a small spreading-out due to the uncorrelated part of the noise. In both cases, however, the detection of a signal is due to the additional interaural time difference resulting from adding the signal to the band of noise centered in the head. The presence of a mass of additional noise at the side of the head, in the time-delayed case, apparently has little or no effect upon the detection of the signal.

In the discussion earlier (Section III D), we postulated the existence of neural "noise" which would cause some spreading out of the diotic acoustical noise from the exact center of the head. Reducing the correlation for the noise will have the effect of increasing this spreading, so increasing the effectiveness of the "noise" as a masker. It is this spreading, then, which is responsible for the reduction of MLDs under the N+ $S\pi$ stimulus condition as the correlation is reduced from unity.

D. MLDs for High-Frequency Signals

Hirsh (1948) found masking level differences of the order of 3 dB at 2000 cps and 5000 cps. Similar MLDs were found in later work by Hirsh and Burgeat (1958), Webster (1951), and Durlach (1963), all at frequencies higher than 1500 cps. Since the hypothesis discussed in the present chapter is based on neural "following," and since such following apparently does not take place at frequencies much above 1500 cps (see e.g., Licklider and Webster 1950), the hypothesis is not adequate to explain the high-frequency MLDs.

Durlach (1964) proposes an extension of the hypothesis of the present chapter to account for the MLDs that cannot be explained in terms of interaural phase (time) differences. If we look again at Fig. 2, we see that adding an antiphasic signal to a diotic noise not only creates a phase difference between the ears (which we must assume is useless at high frequencies) but also creates a difference of amplitude. The SN vector for the right ear is longer than that for the left. Such an interaural difference of amplitude will occur each time the signal is added antiphasically, except for the times when the signal is in quadrature with the noise at the moment of addition. Durlach proposes that the central nervous system can detect a difference of level at the ears and that it is this detection added to the monaural detection of amplitude changes that is responsible for the MLDs obtained in the antiphasic case.

REFERENCES

Adrian, E. D. The all-or-none principle in nerve. J. Physiol., 1914, 47, 450-474.

Békésy, G. von. (1930) In Experiments in hearing. McGraw-Hill: New York, 1960, pp. 272-301.

Blodgett, H. C., Jeffress, L. A. and Taylor, R. W. Relation of masked threshold to signal duration for various interaural phase-combinations. Am. J. Psychol., 1958, 71, 283-290.

Blodgett, H. C., Jeffress, L. A., and Whitworth, R. H. Effect of noise at one ear on the masked threshold for tone at the other. J. Acoust. Soc. Amer., 1962, 34, 979-981.

Bourbon, W. T. and Jeffress, L. A. Effect of bandwidth of masking noise on detection of homophasic and antiphasic tonal signals. J. Acoust. Soc. Amer., 1965, 37, (A).

Cramer, E. M. and Huggins, W. H. Creation of pitch through binaural interaction. J. Acoust. Soc. Amer., 1958, 30, 413-417.

Durlach, N. I. Equalization and cancellation theory of binaural masking-level differences. J. Acoust. Soc. Amer., 1963, 35, 1206-1218.

Durlach, N. I. Note on binaural masking-level differences as a function of the interaural correlation of the masking noise. J. Acoust. Soc. Amer., 1964, 36, 1613-1617.

Egan, J. P. Masking-level differences as a function of interaural disparities in intensity of signal and of noise. J. Acoust. Soc. Amer., 1964, 36, 1992 (A).

Egan, J. P. and Benson, W. Masking-level differences for lateralization of a weak signal. J. Acoust. Soc. Amer., 1964, 36, 1992 (A).

Galambos, R., Microelectrode studies on the auditory nervous system. Ann. Otol. Rhinol. and Laryngol., 1956, 65, 1053-1059.

Galambos, R., Schwartzkopff, J., and Rupert, A. Microelectrode study of superior olivary nuclei. Amer. J. Physiol., 1959, 197, 527-536.

Green, D. M. Interaural phase effects in the masking of signals of different durations. J. Acoust. Soc. Amer., 1965, 37,

Hafter, E. R., Bourbon, W. T., Blocker, A. S., and Tucker, A. Detection of lateral movement under antiphasic masking conditions. J. Acoust. Soc. Amer., 1964, 36, 1991 (A).

Hirsh, I. J. Influence of interaural phase on interaural summation and inhibition. J. Acoust. Soc. Amer., 1948, 20, 536-544.

Hirsh, I. J. and Burgeat, M. Binaural effects in remote masking. J. Acoust. Soc. Amer., 1958, 30, 827-832.

Jeffress, L. A. A place theory of sound localization. J. Compar. Physiol. Psychol., 1948, 41, 35-39.

Jeffress, L. A. Stimulus-oriented approach to detection. J. Acoust. Soc. Amer., 1964, 36, 766-774.

Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H. Masking of tones by white noise as a function of the interaural phases of both components. I. 500 cycles. J. Acoust. Soc. Amer., 1952, 24, 523-527.

Jeffress, L. A., Blodgett, H. C. and Deatherage, B. H. Masking of tones by white noise as a function of interaural time or phase displacements. II. 167 cps. J. Acoust. Soc. Amer., 1953, 25, 190 (A).

Jeffress, L. A., Blodgett, H. C., Sandel, T. T., and Wood, C. L. III. Masking of tonal signals. J. Acoust. Soc. Amer., 1956, 28, 416-426.

Langford, T. L. and Jeffress, L. A. Effect of noise crosscorrelation for a narrow-band masking noise on the detection of a tonal signal. J. Acoust. Soc. Amer., 1964, 36, 1042 (A).

Licklider, J.C.R. Influence of interaural phase relations upon the masking of speech by white noise. J. Acoust. Soc. Amer., 1948, 20, 150-159.

Miller, J. D. Personal communication, 1964.

Moushegian, G., Rupert, A., and Whitcombe, M. A. Brain stem neuronal response patterns to monaural and binaural tones. J. Neurophysiol., 1964, 27, 1174-1191.

Mulligan, B. M. The effect of noise at one ear on the detection of a signal at the other ear. Unpublished Ph. D. dissertation, University of Mississippi, 1964.

Robinson, D. E. and Jeffress, L. A. Effect of varying the interaural noise correlation on the detectability of tonal signals. J. Acoust. Soc. Amer., 1963, 35, 1947-1952.

Rutherford, W. The sense of Hearing. J. Anat. Physiol., 1886, 21, 166-168.

Van Bergeijk, W. A. Variation on a theme of Békésy: a model of binaural interaction. J. Acoust. Soc. Amer., 1962, 34, 1431-1437.

Watson, C. S., Rilling, M. E. and Bourbon, W. T. Receiver operating characteristics by rating scale for antiphasic stimulation. J. Acoust. Soc. Amer., 1964, 36, 1991 (A).

Webster, F. A. Influence of interaural phase on masked thresholds. I. the role of interaural time-deviation. J. Acoust. Soc. Amer., 1951, 23, 452-462.

Wever, E. G. and Bray, C. W. The nature of acoustical response: the relation between sound frequency and frequency of impulses in the auditory nerve. J. Exp. Psychol., 1930, 13, 373-387.

Whitmore, J. K. Effect of interaural noise correlation on the detectability of monaural tonal signals. Unpublished Master's thesis, University of Mississippi, 1964.

FOOTNOTE

[1] The idea that the band employed in antiphasic detection is wider than the band used in monaural (or diotic) detection receives further support from signal-duration studies. Shortening a tonal signal to 10 msec degrades detection in the monaural or monotic case by about 2 dB more than it would simply on the basis of E/N_o (Green 1965, Blodgett, Jeffress and Taylor 1958). No such degradation is found at 10 msec for the NO $S\pi$ condition. The loss in the first case is usually explained as being due to the spreading of signal energy outside of the filter band. The absence of loss in the NO $S\pi$ case suggests a wider filter. Still further support for the hypothesis is found in a study by Bourbon and Jeffress (1965) in a band-narrowing experiment continuing the experiment by Langford and Jeffress mentioned above. They found that any narrowing of the bandwidth of the masking noise that produces even a slight improvement in monaural or diotic detection, produces a spectacular improvement in detection under the NO $S\pi$ condition.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)	2a. REPORT SECURITY CLASSIFICATION
Defense Research Laboratory The University of Texas Austin, Texas 78712	UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

MASKING AND BINAURAL PHENOMENA

4. DESCRIPTIVE NOTES *(Type of report and inclusive dates)*

Summary

5. AUTHOR(S) *(Last name, first name, initial)*

Jeffress, Lloyd A.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
14 June 1965	102	93

8a. CONTRACT OR GRANT NO. NObsr-72627 b. PROJECT NO. NObsr-93124 Nonr-3579(04) c. NASA R-129-09-030-17 d.	9a. ORIGINATOR'S REPORT NUMBER(S) Defense Research Laboratory Acoustical Report No. 245 9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None

10. AVAILABILITY/LIMITATION NOTICES

Unlimited

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	Bureau of Ships Office of Naval Research National Aeronautics and Space Admin.

13. ABSTRACT

The two chapters comprised in this report were written for publication in a book, Modern Foundations of Auditory Theory, to be edited by J. V. Tobias and E. D. Schubert, and published by The Academic Press. Work on the chapters was supported by the Bureau of Ships through Contracts NObsr 72627 and 93124 and by the National Aeronautics and Space Administration (Grant R-129) through Office of Naval Research Contract Nonr 3579(04).

The first chapter summarizes recent work on the subject of masking, particularly the masking of tonal signals by Gaussian noise. It discusses various theories of masking and of auditory function with considerable emphasis on the theory of signal detectability (TSD).

The second chapter presents the vector theory of binaural unmasking, with some discussion of the possible underlying neurophysiological machinery.

DD FORM 1 JAN 64 1473 0101-807-6800

UNCLASSIFIED
Security Classification