JUNE 01 1971

Stimulus Features in Signal Detection

Al Ahumada, Jr.; John Lovell

J. Acoust. Soc. Am. 49, 1751–1756 (1971)
https://doi.org/10.1121/1.1912577

Received 27 October 1970 4.7, 4.11

Stimulus Features in Signal Detection

AL AHUMADA, JR.

University of California, Irvine, California 92664

JOHN LOVELL

University of California, Los Angeles, California 90024

Short bursts of computer-generated Gaussian noise were rated by observers for the presence or absence of a 500-Hz signal tone burst. A multiple regression analysis found for each observer the linear combination of the energies in narrow bands around the tone frequency that best predicts his total ratings. The estimates of the regression coefficients provide graphs of the "frequency responses" of the observers. Most of the reliable variance in the total ratings was accounted for by the regression analysis in terms of energy in narrow bands. Differences among observers are explained in terms of differential weighting by observers of features labeled "tone presence," "pitch," and "loudness."

INTRODUCTION

Many researchers have been concerned with the measures or features of the auditory stimulus that a human observer uses to detect the presence or absence of a signal tone masked by Gaussian noise. Jeffress (1968), for example, has shown that a wide range of experimental detection results can be explained if it is assumed that the relevant stimulus feature is the integrated energy or amplitude passed by a bandpass filter centered at the signal tone frequency. The frequency response of this hypothetical filter has been measured by Greenberg and Larking (1968). The experiments reported here can be regarded as attempts to replicate their measurements.

Greenberg and Larkin used occasional probe signals differing in frequency from the usual signal tone. They plotted the percentage of detection responses to these probe signals as a function of their frequency to obtain "frequency response" functions for their observers. When the filter model assumptions are made, these "frequency response" functions can be transformed into filter response functions, if psychometric functions relating signal amplitude to percentage of correct responses are provided, as Greenberg (1969) did in a later paper. The probe signals can be regarded as a means of introducing known amplitude variations into the masking noise at different frequencies so that the result of amplitude variations at different frequencies can be assessed.

If computer-generated reproducible masking noise is used (Pfafflin and Mathews, 1966), amplitude variations in the noise at different frequencies can be measured and their effects on the observer's responses assessed.

To simplify the measurement of amplitude variations at different frequencies, we used bursts of Gaussian noise constructed by adding together n orthogonal gated sine waves having independent Rayleigh-distributed amplitudes, A_i, with amplitude parameter a defined so that a^2 is the average value of A_i^2. On trials when only the masking noise is presented to the observer (N trials), the voltage $V_N(t)$ applied to the observer's earphones may be represented as

$$V_N(t)=\sum_{i=0}^{n} A_i \sin(2\pi f_i t+\phi), \quad 0\leq t\leq T, \qquad (1)$$
$$=0, \qquad \text{elsewhere.}$$

The ϕ_i are independent uniform random variables on the interval $(0,2\pi)$. The orthogonality of the sine components is ensured by having the f_i be integral multiples of $1/T$.

On signal trials (SN trials), the stimulus waveform is augmented by a sinusoid of constant amplitude s, frequency f_j, and zero phase angle, so that

$$V_{SN}(t)=V_N(t)+s\sin 2\pi f_j t, \quad 0\leq t\leq T,$$
$$=0, \qquad \text{elsewhere.} \qquad (2)$$

The signal tone frequency f_j is the same as the most central component of the noise so that the SN stimuli could equally well be described as the same as the N stimuli, except that on SN trials the distribution of A_j is a times a noncentral chi variable with 2 degrees of freedom (df) and noncentrality parameter η equal to the signal-to-noise ratio,

$$\eta = s^2/a^2. \tag{3}$$

This latter interpretation of the A_i's as including the effect of the signal on SN trials is assumed in the rest of the paper, although Eq. 2 was actually used to construct the stimuli.

We assess the relative contributions of the amplitudes of different frequency components by finding the linear combination of squared amplitudes that best predicts the observer's average rating responses R_i to all m stimuli. A multiple regression analysis finds coefficients c_i and intercept constant b, which minimizes the total squared error of prediction,

$$P = \sum_{k=1}^{m} [R_k - (\sum c_i A_{ik}^2 + b)]^2. \tag{4}$$

If the observer based his ratings on the energy passed by a single linear filter, the c_i's would provide estimates proportional to the energy spectrum of the filter's response, and hence would have to be all of the same sign. We do not place this restriction on the c_i's, allowing for the possibility that increased energy at different frequencies might have opposite effects.

Results from two experiments are presented. The first experiment has been reported previously (Ahumada and Lovell, 1969), but the data were analyzed in a less informative way. The second experiment included more stimuli so that finer frequency resolution was possible.

I. METHOD

A. Experiment I

The observers were 10 students and young faculty at the University of California, Los Angeles, or the University of California, Irvine. The observer was asked to make a rating response of 1, 2, 3, or 4 to each stimulus burst. A response of 4 was to indicate a signal tone certainly present, 3 to indicate the signal tone probably present, 2 to indicate the tone probably not present, and 1 to indicate the tone certainly not present.

The stimuli were generated using a version of the Bell Telephone Laboratories Music IV program. The noise bursts consisted of 32 components spaced 10 Hz apart from 350 to 660 Hz. The signal tone frequency was 500 Hz, the same as the 16th noise component. Signal tone and noise-burst durations were 0.1 seconds. The values of s and a were chosen so that $s^2/a^2 = E/N_0 = 15.125$. Fifty stimuli were constructed, 25 of which had signal tones. The computer program generated eight random sequences of these 50 stimuli separated by 3 sec of silent response time. In addition, every block of 10 stimuli was preceded by a signal tone without noise as an aid to the observer.

The stimuli were converted to analog form by a 12-bit sample-and-hold digital-to-analog converter at a sampling rate of 10 kHz. The converted stimuli were bandpass filtered from 120 to 3000 Hz and then recorded on magnetic tape by an Ampex PR-10 at $7\frac{1}{2}$ in./sec.

Each of the eight sequences was presented four times to each observer and the total rating responses for each of the 50 stimuli were computed for each observer. The sound-pressure level (SPL) of the stimuli at the binaural earphones was 90 dB. Observers AA, KM, and ML listened to the stimuli through TDH 39 earphones from an Ampex AG-500 tape recorder. The rest listened through PDR 600 earphones from a Tandberg 74 tape recorder.

B. Experiment II

Seven observers were each presented 16 sequences of 200 stimuli, of which 100 contained signal tones. The stimuli had 31 components from 350 to 650 Hz and were generated by an assembly language program on a small computer (Varian 620 I). All stimuli were recorded on an Ampex AG-500 tape recorder and played back by the same recorder into TDH 39 earphones. All other details were as in Expt. I.

II. RESULTS

For each observer, the rating responses were totaled over the 32 or 16 presentations of the stimulus. If each rating response can be assumed to consist of a component dependent on the stimulus waveform plus components that vary from trial to trial (e.g., tape recording variations, sensory system fluctuations, response criterion variations, etc.), summing the ratings over presentations of the same waveform should form a total rating whose variations over stimuli are more related to the stimulus waveform than are any of the individual

TABLE I. Data for 10 observers in Expt. I. The proportion of correct stimulus classifications by rating response totals is represented by P_c; the corresponding detectability index, $d_s = 2z(P_c)$; the square of the multiple correlation coefficient predicting rating responses from component amplitudes, R^2; and squared multiple correlation coefficients for predicting response totals on SN and N trials separately, R_{SN}^2 and R_N^2.

Observer	P_c	d_s	R^2	R_{SN}^2	R_N^2
KM	0.96	3.50	0.79	0.26	0.23
HS	0.92	2.81	0.82	0.61	0.29
DN	0.92	2.81	0.85	0.69	0.25
JD	0.92	2.81	0.90	0.84	0.72
AA	0.88	2.35	0.82	0.69	0.22
ML	0.80	1.68	0.72	0.65	0.53
JL	0.80	1.68	0.72	0.59	0.66
CD	0.80	1.68	0.68	0.50	0.67
JB	0.80	1.68	0.58	0.33	0.51
IL	0.68	0.94	0.60	0.38	0.69

ratings. The ability of the observers' response totals to separate the stimuli having signals (SN) from those having noise alone (N) is shown in Tables I and II. The percentage correct P_c and the detectability index d_s are based upon locating a criterion response total for each observer such that the number of N stimuli having totals above the criterion equals the number of SN stimuli having response totals below the criterion. For example, in Expt. I, 24 of KM's response totals to the 25 SN stimuli were greater than 81, and 24 of the 25 N stimuli were given total ratings less than 81. Thus KM has $P_c=24/25=0.96$ and $d_s=2z(0.96)=3.5$, where $z(p)$ is the inverse of the cumulative (standard) normal distribution function.

Least-squares estimates of the c_i were obtained for each observer from a multiple regression program (Cooley and Lohnes, 1962) with the response totals entered as the criterion variable and the values of A_i^2 for each stimulus entered as the predictor variables. These estimates are graphed in arbitrary units in Figs. 1 and 2. The accuracy of the estimates increases with the the predictor variables and it increases with the number of stimuli, but it decreases with the number of components which are used as predictors. When all 32 or 31 components were used as predictors, the estimates of the c_i were too variable to see a frequency response curve. To stabilize the estimates of the c_i, the predictors are five sums of five adjacent A_i^2 in Expt. I and nine sums of three adjacent A_i^2 in Expt. II. Adding A_i^2 values in groups of five or three is equivalent to finding the least-squares estimates of the c_i given the constraint that all c_i within a group must be the same. These graphs directly show the relative size and direction of the effect of increased energy in each frequency region upon the response totals.

Although some of the coefficient curves are flat or resemble simple high-pass filter response curves, most of the observers show a definite peaked response near the signal tone frequency. Also, most of the curves are definitely asymmetrical. The majority have high-frequency coefficients higher than the lower-frequency coefficients, but two observers show the opposite asymmetry (AA and JH in Fig. 2). Four observers (HS and

TABLE II. Data for 7 observers in Expt. II. The proportion of correct stimulus classifications by rating response totals is given as P_c; the corresponding detectability index, $d_s=2z(P_c)$; the square of the multiple correlation coefficient for predicting rating response totals from component squared amplitudes, R^2; the square of the multiple correlation coefficient for predicting rating response totals from the three features of Fig. 5, R_f^2; the predictable variance in the response totals, V_p; r_{12}; and squared multiple correlation coefficients for predicting response totals on SN and N trials separately, R_{SN}^2 and R_N^2.

Observer	P_c	d_s	R^2	R_f^2	V_p	r_{12}	R_{SN}^2	R_N^2
CC	0.950	3.29	0.85	0.83	0.96	0.93	0.58	0.41
CV	0.900	2.56	0.81	0.80	0.94	0.88	0.51	0.47
AA	0.895	2.51	0.86	0.85	0.97	0.94	0.63	0.53
JH	0.860	2.16	0.79	0.72	0.95	0.91	0.65	0.47
CH	0.774	1.50	0.52	0.51	0.76	0.61	0.32	0.37
MS	0.652	0.78	0.45	0.43	0.60	0.43	0.30	0.42
TC	0.585	0.43	0.08	0.04	0.78	0.62	0.16	0.09

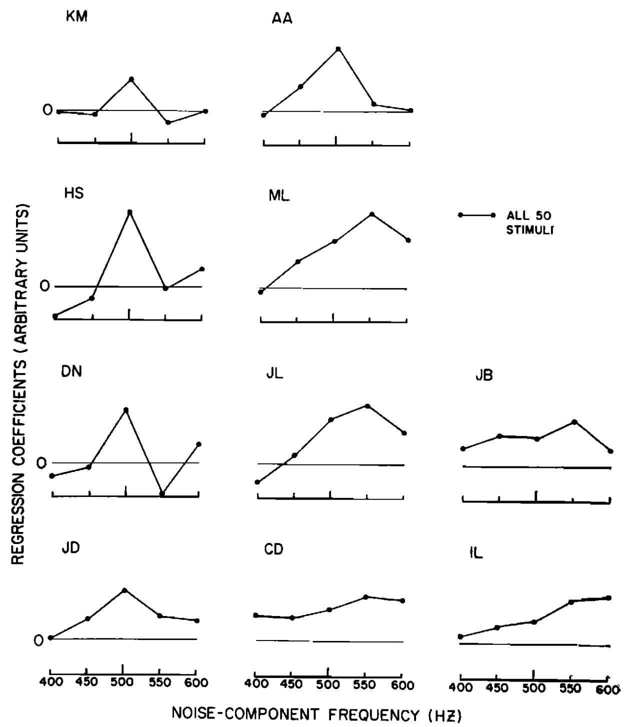

FIG. 1. Linear regression coefficients (c_i) for predicting observer rating response totals from five sums of five adjacent component squared amplitudes (A_i^2) in Expt. I.

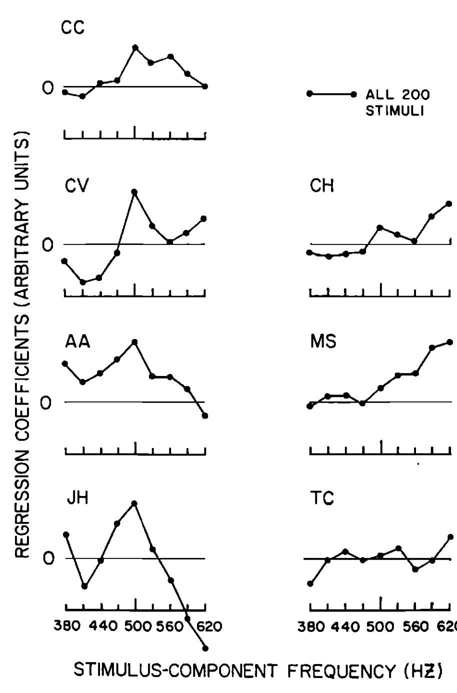

FIG. 2. Linear regression coefficients (c_i) for predicting observer rating response totals from nine sums of three adjacent component squared amplitudes (A_i^2) in Expt. II.

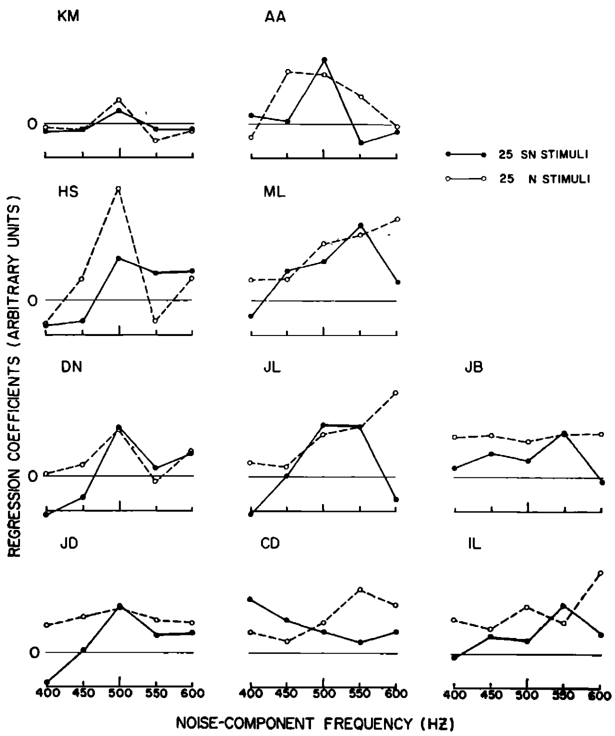

FIG. 3. Linear regression coefficients, as in Fig. 1, computed separately for SN and N trials.

DN in Fig. 1 and CV and JH in Fig. 2) have definitely negative coefficients.

The adequacy of the linear prediction of the response totals by the A_i^2 is shown in Tables I and II. The values labeled R^2 give the multiple correlation coefficient squared, the proportion of the response total variance that is perfectly predicted by the estimated linear combination of the A_i^2. Values of R^2 must be greater than 0.22 (Expt. I) and 0.08 (Expt. II) to be significantly different from zero at the 0.05 level. Table II also contains split-half correlations, r_{12}, between the response totals from alternate tape presentations in Expt. II. If the split-half correlation can be assumed to reflect the correlation of two quantities each having the same predictable part but each having independent random parts, the predictable proportion of the variance of the response total, V_p, is given by

$$V_p = r_{12}/[r_{12}+\tfrac{1}{2}(1-r_{12})]. \quad (5)$$

Table II gives values of V_p for each observer in Expt. II. If these values are compared with the values of R^2 for each observer, it can be seen that most of the predictable variance in the observer's response can be accounted for by the linear combination of the A_i's plotted in Fig. 2.

As a further check on the validity of Eq. 4 for predicting response totals, regression analyses were done separately for the SN stimuli and the N stimuli. If our regression model were adequate, estimates of c_i for the two cases should be the same, except for possible scale factor differences arising in the category scaling process. These estimates appear in Figs. 3 and 4, and they suggest that most observers are not adequately described by the model. For these observers, it appears that A_i^2 of components further from the signal tone frequency are given more positive weight on N trials than they are given on SN trials. In many cases (AA, ML, JL, and JD in Fig. 3 and CC and CH in Fig. 4), the lower coefficients away from the tone frequency on SN trials are actually negative. The proportion of response variation that these linear combinations account for is shown in the columns labeled R_{SN}^2 and R_N^2 in Tables I and II. These proportions must be greater than 0.42 (Expt. I) or 0.17 (Expt. II) to be significantly greater than zero at the 0.05 level. In Expt. I, four of the R_N^2 values do not reach this level of significance, which means that corresponding individual estimates of c_i values are not reliably different from zero. The fact that R_N^2 is almost always smaller than R_{SN}^2 should not necessarily be interpreted as showing that the observers are using a measure more consistent with the model on SN trials. The addition of the signal increases the variability of the model's stimulus measure, and the observers used more response categories for the SN stimuli, so that R_{SN}^2 should be greater than R_N^2 if the only stimulus measure used by the observer were a linear combination of the component energies.

The shape of the N curves in Figs. 3 and 4 are especially interesting in that they represent the contribution of the different stimulus components ot the observers' response totals at the lowest possible signal level. The SN curves and the pooled curves show the contribution of grouped components in the presence of a grouped signal component, which has about four to six times as much energy as the other components.

III. AN EXPLANATORY MODEL

The variety of curves for different observers could arise if there are several features of the stimulus that observers could weight differently. The following three features seem adequate to explain the variability among observers: (a) tone presence, (b) pitch, and (c) loudness. Plots of coefficients c_i, which correspond to detectors for these features in the context of the regression model, appear in Fig. 5. It is evident by inspection that most of the apparently reliable features of the observers' c_i curves can be represented by linear combinations of these three orthogonal feature curves. The pitch and loudness curves allow a straight line of any slope and intercept to be fit to the curves, and the tone-presence curve allows the fitting of the peak that usually appears near the signal tone frequency.

The ability of the three features to account for the amplitude coefficient curves was assessed by a multiple regression analysis using the three features as predictor variables and the response totals as the criterion variable. The value for each feature was computed as a linear combination of A_i^2, using the coefficients plotted in Fig. 5. Multiple correlation coefficients, R_f^2, for these feature predictions are shown in Table II for the observers in Expt. II. They are mathematically constrained to be smaller than the over-all R^2, but the three features account for almost as much of the response total variance. A broader and a narrower tone-presence detector were tried, but did not predict as well as the one presented.

The feature curves are of course idealized and somewhat arbitrary. The problem of extracting features from data of this sort is similar to the factor analysis problem of identifying basic personality or intelligence features from the scores on a test (Harmon, 1960). The intuitive method we used is analogous to a "simple structure" criterion in that features were chosen so that some observers' curves could be explained in terms of only one or two features. Observer reports also played a role in the feature selection. Observers AA and JH, for instance, reported that SN trials appeared to be lower in pitch, while other observers generally reported the opposite in informal postexperiment questioning.

Differences between SN and N trials could appear if the combining rule for the features were nonlinear. A rule that observers might have followed is to ignore the loudness feature unless the other features have low outputs. A rule of this sort would lead to loudness having less effect on the SN trials when the tone-presence feature is more likely to have a high output and would result in lower SN coefficients.

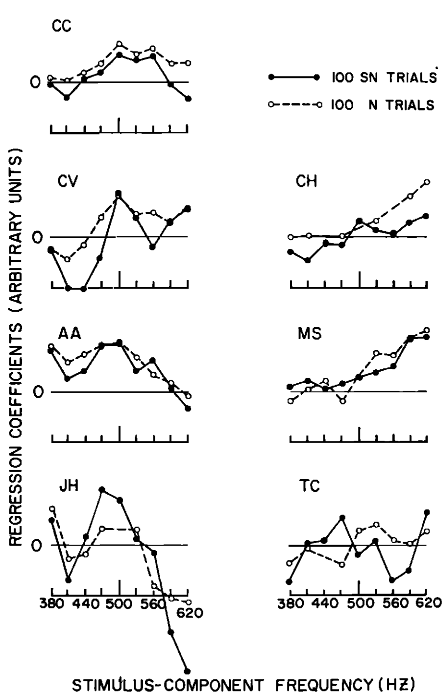

FIG. 4. Linear regression coefficients, as in Fig. 2, computed separately for SN and N trials.

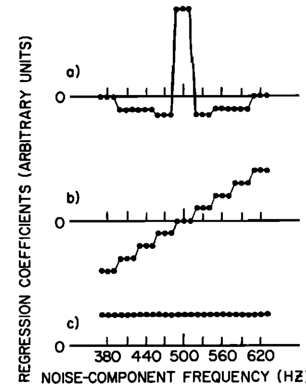

FIG. 5. Hypothetical linear regression coefficients that form feature detectors for (a) tone presence, (b) pitch, and (c) loudness.

IV. DISCUSSION

Our experimental results generally confirm those obtained by the Greenberg and Larkin (1968) probe stimulus method. Relatively narrow peaks occur in "frequency-response" curves obtained by both methods. These more direct measurements support those who have inferred relatively narrow critical bands (Fletcher, 1953; Schafer, Gales, Shewmaker, and Thompson, 1950; Swets, Green, and Tanner, 1962; and Jeffress, 1968) as opposed to those who have inferred wide bands (deBoer, 1962; van den Brink, 1964; Green and Swets, 1966) in tone detection experiments.

The observer's frequency selectivity may be somewhat sharper than our curves suggest, since, as Henning (1967) suggested, the observer may center his "filter" at different frequencies on different trials. If the filter is "captured" by the tone when it is distinct enough, Henning's variable filter model is the same as the filter bank model (Ahumada, 1967), which assumes that the observer simultaneously monitors several filters and uses the largest output from the bank to decide upon a response. This model predicts that the N curves will be wider and flatter than the SN curves, as most of them are in Figs. 3 and 4. For the filter bank model, the SN curves should approximate the filter response, while the N curves, as well as Greenberg and Larkin's probe stimulus curves, should indicate the range of the filter bank. We tried estimating the best fitting filter and bank width for the data in Expt. I, but the results were not encouraging. Narrow filters and narrow banks best fit the SN data, while wide filters and wide banks best fit the N data.

Some of the Greenberg and Larkin curves appear to dip below 50%, indicating that increased energy of these frequencies causes the observer to guess that a signal was not presented. Our corresponding result that regression coefficients for some frequencies are negative rejects a simple filter model and requires at least positive and negative combinations of the detector outputs of two or more filters. Carterette, Friedman, and Lovell (1969) discuss models that provide the excitatory and inhibitory combinations of filter outputs necessary to construct our feature detectors.

At another level of explanation, the negative coefficients can be regarded as showing that observers are using the pattern of spectral energy rather than the absolute level of energy in a particular region. Leshowitz (1971) has shown this to be the case when observers are discriminating single from double clicks. In this regard the gated noise detection situation is probably more similar to click and pulse discrimination experiments than to continuous-noise detection experiments, where an amplitude reference level is continually available. An analysis of the sequential effects in gated noise detection experiments suggests that observers base their response more on a comparison of the current stimulus with the previous stimulus than they do in the continuous noise situation (Sandusky and Ahumada, 1971).

Asymmetric "frequency-response" curves were reported by Greenberg (1969). Most of his observers were similar to most of ours in that high-frequency probe tones were responded to more like the signal frequency tones. Were it not for our two observers showing the opposite asymmetry, our effects could have been explained as low-frequency insensitivity. The ability of observers to react as high-pass and low-pass detectors calls into question the interpretation of bandlimiting experiments where cutoffs are varied on one side to measure the observers' critical bandwidth (Bourbon, Evans, and Deatherage, 1968).

The large percentage of the response total variation, which is predicted by the amplitude spectrum of the signals, suggests that there is little left for the phase structure to predict. Patterson, Ronken, and Green (1969) have demonstrated that pulses having identical power spectra are discriminable, but the nature of the features that observers use to discriminate them and the possible role of these features in masking experiments has yet to be studied in detail.

ACKNOWLEDGMENT

This work was supported in part by the National Institutes of Health.

REFERENCES

AHUMADA, A. (1967). "Detection of Tones Masked by Noise: A Comparison of Human Observers with Digital-Computer-Simulated Energy Detectors of Varying Bandwidths," Dep. Psychol., Univ. of California, Los Angeles, Calif., Tech. Rep. 29.

AHUMADA, A., and LOVELL, J. (1969). "Bandwidth Estimates in Human Signal Detection," J. Acoust. Soc. Amer. *45*, 327(A).

DE BOER, E. (1962). "Note on the Critical Bandwidth," J. Acoust. Soc. Amer. *34*, 985–986.

BOURBON, W. T., EVANS, T. R., and DEATHERAGE, B. H. (1968). "Effects of Intensity on 'Critical Bands' for Tonal Stimuli as Determined by Band Limiting," J. Acoust. Soc. Amer. *43*, 56–59.

VAN DEN BRINK, G. (1964). "Detection of Tone Pulses of Various Durations in Noise of Various Bandwidths," J. Acoust. Soc. Amer. *36*, 1206–1211.

CARTERETTE, E. C., FRIEDMAN, M. P., and LOVELL, J. (1969). "Mach Bands in Hearing," J. Acoust. Soc. Amer. *45*, 986–998.

COOLEY, W. W., and LOHNES, P. R. (1962). *Multivariate Procedures for the Behavioral Sciences* (Wiley, New York).

FLETCHER, H. (1953). *Speech and Hearing in Communication* (Van Nostrand, New York), 2nd ed.

GREEN, D. M., and SWETS, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York).

GREENBERG, G. Z. (1969). "Frequency Selectivity during Amplitude Discrimination of Signals in Noise," J. Acoust. Soc. Amer. *45*, 1438–1442.

GREENBERG, G. Z., and LARKIN, W. D. (1968). "Frequency-Response Characteristics of Auditory Observers Detecting Signals of a Single Frequency in Noise: The Probe-Signal Method," J. Acoust. Soc. Amer. *44*, 1513–1523.

HARMON, H. H. (1960). *Modern Factor Analysis* (Univ. of Chicago Press, Chicago, Ill.).

HENNING, G. B. (1967). "A Model for Auditory Discrimination and Detection," J. Acoust. Soc. Amer. *42*, 1325–1334.

JEFFRESS, L. A. (1968). "Mathematical and Electrical Models of Auditory Detection," J. Acoust. Soc. Amer. *44*, 187–203.

LESHOWITZ, B. (1971). "Measurement of the Two-Click Threshold," J. Acoust. Soc. Amer. *49*, 462–466.

PATTERSON, J., RONKEN, D. A., and GREEN, D. M. (1969). "Phase Perception of Transient Signals Having Identical Power Spectra," J. Acoust. Soc. Amer. *46*, 121(A).

PFAFFLIN, S. M., and MATHEWS, M. V. (1966). "Detection of Auditory Signals in Reproducible Noise," J. Acoust. Soc. Amer. *39*, 340–345.

SANDUSKY, A., and AHUMADA, A. (1971). "Contrast in Detection with Gated Noise," J. Acoust. Soc. Amer. *49*, 1790–1794.

SCHAFER, T. H., GALES, R. S., SHEWMAKER, C. A., and THOMPSON, P. O. (1950). "The Frequency Selectivity of the Ear as Determined by Masking Experiments," J. Acoust. Soc. Amer. *22*, 490–496.

SWETS, J. A., GREEN, D. M., and TANNER, W. P., JR. (1962). "On the Width of Critical Bands," J. Acoust. Soc. Amer. *34*, 108–113.