UNCLASSIFIED
SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

AD-A199 832

	1b. RESTRICTIVE MARKINGS
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	3. DISTRIBUTION / AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S) AFOSR-TR- 88-0985
6a. NAME OF PERFORMING ORGANIZATION University of Florida	6b. OFFICE SYMBOL (If applicable)
7a. NAME OF MONITORING ORGANIZATION AFOSR/NL	
6c. ADDRESS (City, State, and ZIP Code) 219 Grinter Hall Gainesville, FL 32611	7b. ADDRESS (City, State, and ZIP Code) Bldg. 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION AFOSR	8b. OFFICE SYMBOL (If applicable) NL
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER AFOSR-85-0374	
8c. ADDRESS (City, State, and ZIP Code) Bldg 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
61102F	2313	A6	

11. TITLE (Include Security Classification)
Complex Auditory Signals

12. PERSONAL AUTHOR(S)
Dr. Green

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 15 Sep85 TO 14 Sep 88	Sep 88	

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	

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

Human detection of complex sounds were examined experimentally and theoretically. Three separate phenomena were studied: comodulation effects, perception of nonstationary spectra and detection of changes in static spectra. Results are briefly outlined and detailed in a number of attached preprints.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT ☐ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASS	
22a. NAME OF RESPONSIBLE INDIVIDUAL Dr. John Tangney	22b. TELEPHONE (Include Area Code) (202) 767-5021	22c. OFFICE SYMBOL NL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

88 10 5 205

AFOSR-TR- 88-0985

COMPLEX AUDITORY SIGNALS

Final Report: Complex Auditory Signals AFOSR-85-0374.

The following represents a summary of research of the research for the period, September 15, 1985 to September 1, 1988. We summarize our past research in terms of three major themes: 1)synchrony detection, 2) perception of nonstationary spectra, and 3) basic properties of profile analysis.

1. Synchrony Detection

One major research theme of our past research has been the topic of synchrony detection pursued largely by Dr. V. M. Richards. Briefly, she claims that the perception of many complex acoustic stimuli depends on simultaneous comparisons of dynamic changes occurring in different spectral regions. Specifically, Dr. Richards believes that envelope comparisons (Ref. 11, 12) can be made in different spectral bands and that the correlation between these envelopes is a major cue to the coherence and grouping found in many complex acoustic stimuli.

The idea of cross-spectral comparison has been around for some time in the acoustics community. Interest in this idea was considerably stimulated by the experiment of Hall, Haggard, and Fernandes 1984. Before that time, cross-spectral comparison was generally considered to be of no importance. The basis for that opinion was an unpublished technical report by Schubert and Nixon (1970). In their experiments, subjects were asked to distinguish between correlated and uncorrelated noise bands. They found that such discrimination was impossible. Richards (Ref. 10–see list presented below) found that such discrimination was possible and traced the earlier failure to a poor choice of frequency location and duration of the noise bands.

This effort is probably one of the most interesting current developments in psychoacoustics. Whereas, previously we had thought such comparisons were impossible, we now know that they are possible and can be made with some precision. In effect, they suggest that a new kind of auditory process must be carefully considered in explaining the perception of any complex auditory signal.

Dr. Richards is finishing the last year of her NIH post-doctoral fellowship and has made application for a FIRST award from NIH to further support this research. We presume, for the purposes of this proposal, that such support will be forthcoming. Thus, we will not request future funding in this proposal to support research on this very important topic. Although she is welcome to stay at Florida and pursue her research, she is naturally looking for a full faculty position and, undoubtedly, will eventually secure one.

2. Perception of Nonstationary Spectra

A second general theme of our past research has been the exploration of the perception of nonstationary spectra. The specific research involves complex auditory spectra containing components that are amplitude modulated. The resulting paper (Ref. 8) should appear shortly. Although amplitude modulation is known to greatly increase the saliency of individual components of a complex spectra, such modulation does little to increase the detectability of amplitude changes in these components. Except for the highest frequency components (f > 2000 Hz), amplitude modulation tends to make changes in level of components less detectable. One experimental condition allowed us to estimate the upper rate for which the relative phase of the modulation was important. That rate appears to be about 40 Hz and to be the same for frequency regions as diverse as 250, 1000, and 4000 Hz. For modulation rates above the value, the phase of modulation between the various components of the complex can be ignored; only the power spectra of the stimuli are important.

A second aspect of nonstationary spectra is the detection of amplitude variation occurring over the entire spectrum, such as the amplitude modulation of noise, or a silent gap inserted in the ongoing noise. In a recent paper (Ref. 7), we compared human performance in such tasks with a modification of a model first proposed by Viemeister (1979). Viemeister's model, typical of a wide class of model, accounts for such detection by using a decision rule that computes the variance in intensity fluctuation at the output of his detection process. Our modification was to compute the maximum-to-minimum ratio observed in the same output. It is, however, possible to argue that such detection depends on comparison of level fluctuation

07 SEP 1988

across different frequency channels, that is, synchrony detection. The reason is that amplitude modulation or the presence of a gap occurs at the same time for all frequency locations. Thus, one might also explore the extent to which gap detection can be explained on the basis of detecting simultaneous patterns of output observed over several spectral channels (synchrony detection). Gap detection and amplitude modulation detection are major components of one of our future research initiatives and will be discussed in greater detail later in this report.

3. Basic Properties of Profile Analysis

The third and final area of effort has concerned the basic properties of profile analysis. Since this will be a major theme of our proposed research, we will describe it only briefly here. We know, from a number of previous studies, that the smallest detectable increment in the intensity of a single component of a multi-component complex occurs when the frequency of the incremented component lies in the middle of the spectrum (Ref. 9). The detection of complex amplitude changes throughout the spectrum is currently not understood in terms of simple integration of the detectability of the change in single components (Ref. 5). However, it is possible to suggest a simple computational scheme to account for the detectability of most complex changes (Ref. 3). Unfortunately, this computational scheme can be shown to systematically fail to account for one class of stimulus change that involves changes in the spectral density of the components. The systematic exploration of this variable, the number of components used to represent the complex spectrum, indicates that the apparent analysis band of the listener (profile critical band) is about the same size as the conventional critical band (Ref. 2).

In addition to these substantive papers, we have also contributed summaries of this research area. Such efforts help to organize our own thinking about the area, and provide succinct summaries of this research for active researchers in this and neighboring fields. Ref. 1, 4, and 6 are illustrations of such efforts.

4. Past Publications

1) Bernstein, L. R., Richards, V., and Green, D. M. (1987) "Detection of spectral shape changes" in a book edited by Yost, W. A. and Watson, C. S. entitled Complex Auditory Detection (Plenum Publication)

2) Bernstein, L. R. and Green, D. M. (1987) "The Profile-analysis bandwidth" J. Acoust. Soc. Am. 81, 1888-1895.

3) Bernstein, L. R. and Green, D. M. (1987) "Detection of simple and complex changes in spectral shape." J. Acoust. Soc. Am. 82, 1587-1592.

4) Green, D. M. (1988) Profile Analysis: Auditory Intensity Discrimination. Oxford University Press, New York and Oxford

5) Green, D. M. (1986) "Frequency and the detection of spectral shape change" in a book edited by Moore, B.C.J. and Patterson, R. D. entitled Auditory Frequency Selectivity (Plenum Publishing Corp.)

6) Green, D. M. and Bernstein, L. R. (1987) "Profile Analysis and Speech Perception" in a book edited by M.E.H. Schouten entitled The Psychophysics of Speech Perception (Martinus, Nijhoff Publishers, Dordrecht, Boston and Lancaster.)

7) Green, D. M. and Forrest, T. G. "Detection of amplitude modulation and gaps in noise" (1988) A paper accepted for the VIII International Symposium on Hearing Research, Groningen, the Netherlands.

8) Green, D. M. and Nguyen, Q. T. (1988) "Profile analysis: Detecting dynamic spectral changes. to appear in Hearing Research.

9) Green, D. M., Onsan, Z. A., and Forrest, T. G. (1987) "Frequency effects in profile analysis and detecting complex spectral changes." J. Acoust. Soc. Am. 81, 692-699.

10) Richards, V. M. "Monaural envelope correlation perception" (1987) J. Acoust. Soc. Am. 82, 1621-1630.

11) Richards, V. M. "Component of monaural envelope correlation perception" (1987) submitted for publication in Hearing Research 1987.

12) Richards, V. M. "Aspects of monaural synchrony detection" (1988) A paper accepted for the VIII International Symposium on Hearing Research, Groningen, the Netherlands.

The following papers have been submitted for publication and are still under review.

13) Green, D. M. and Forrest, T. G. "Temporal gaps in noise and sinusoids" submitted to the J. Acoust. Soc. Am.

14) Raney, J. J., Richards, V. M., Onsan, Z. A. Onsan, and Green, D. M. "Signal uncertainty and psychometric functions in profile analysis" submitted to the J. Acoust. Soc. Am.

15) Richards, V. M., Onsan, Z. A., and Green, D. M. "Auditory profile analysis: Potential pitch cues" submitted to Hearing Research.

5. Personnel

The major technical people are listed below, along with comments on their present status.

Dr. Les Bernstein January, 1986 – January, 1988. Dr. Bernstein is now at the University of Connecticut, Medical Center. We have nearly completed negotiations with Dr. Bruce Berg, Ph.D. University of Indiana, 1987, to replace Dr. Bernstein. He is presently a research fellow in the radiology section of the Harvard Medical School and will join the laboratory in July, 1988.

Dr. Virginia M. Richards –NIH postdoctoral fellow, June 1985–present. She has been invited to continue her research at the laboratory, but is actively seeking a 'real' job.

Dr. Timothy Forrest–assistant in psychoacoustics, October, 1985–present. Dr. Forrest is an entomologist by training and is actively seeking an academic position in that area.

Mr. Quang Nguyen (B.S. Electrical Engineering, University of Florida, 1986).

Ms. Zekiye Onsan (B.S. Astronomy, University of Istanbul, Turkey, 1977).

Mr. Richard Newton (B.A. Computer Science, University of Florida, 1988) Now working for Intel Corporation, Santa Clara, California.

Mr. Timothy Tucker (B.S. expected May, 1988, Electrical Engineering, University of Florida).

Ms. Jill Johnson Raney–graduate student, 2nd year.

Ms. Cheryl Williams–secretary and laboratory coordinator.

THE DETECTION OF SPECTRAL SHAPE CHANGE

Leslie R. Bernstein, Virginia Richards,
and David M. Green
Psychology Department, University of Florida,
Gainesville, Florida
32611 U.S.A.

Introduction

We describe several experiments involving the detection of a change in the spectral shape of a complex auditory signal, what we call profile-analysis. All of the experiments are discrimination tasks involving a broadband "standard" spectrum and some alteration of that spectrum produced by adding a "signal" to the standard. For all of the experiments described here, we used a standard composed of a set of equal-amplitude sinusoidal components. The spectrum of the standard was, therefore, essentially flat. In different experiments, various waveforms were added to this standard to create changes in its spectral shape, and the ability to detect such changes was measured. In the first experiments, we describe how the relative phase among the components of the standard waveform influences the detection of a signal. The results are very simple. Phase seems to play no important role. The detection of a change in spectral shape appears to depend only on changes in the power spectrum of the signal and is independent of the temporal waveform. Next, we describe how the detection of an increment in a single component depends on the frequency of that component. These results provide the basic data to evaluate complex changes in the whole spectrum, such as a sinusoidal ripple in the amplitudes of the components over the entire spectrum. Our data indicate that there is a sizable discrepancy between the ability to detect changes occurring over the entire spectrum and the ability to detect changes in single components.

Procedure

We used a two-alternative, forced-choice procedure to evaluate the detectability of the change in spectral shape. In one interval, the listener heard the "stardard" sound; in the other interval, the listener heard the "standard plus signal". The signal component was always added at a fixed phase relation to the standard component, generally in-phase. An adaptive two-down, one-up rule was used to estimate 70.7 % correct detection. The thresholds reported are the signal amplitude re the component of the standard to which the signal is added. A threshold of 0 dB means that the signal and standard components are equal in amplitude. Typically, the average threshold was based on at least 12 runs of 50 trials. Each sound was generated digitially and presented for about 100 msec.

The standard spectrum was composed of a sum of sinusoidal components. Except for one experiment where the number of components is varied, there were 21 components extending in frequency from 200 to 5000 Hz. The ratio of the frequencies between sucessive components was constant; that is, the frequencies were spaced equally on a logarithmic scale. Because distance along the basilar membrane is proportional to the logarithm of frequency, our components provided a roughly uniform stimulus over the linear receptor surface of the cochlea.

One final experimental feature must be clearly understood. Because we are interested in the detection of a change in spectral shape, we must ensure that the observer is not simply discriminating a change in intensity at a single frequency region. To do this, we randomly varied the overall level of the sound on each and every presentation. The level of the sound was chosen from a rectangular distribution of intensity covering a range of 20 or 40 dB in 1 dB steps. The median level was about 50 to 60 dB SPL. Thus, while the "flat" standard might be presented at 71 dB, the altered spectrum, the "signal plus standard", might be presented at 34 dB on a given trial of the forced-choice procedure. The observer's task was to detect the sound with the altered spectral shape despite the difference in overall level.

Effects of phase

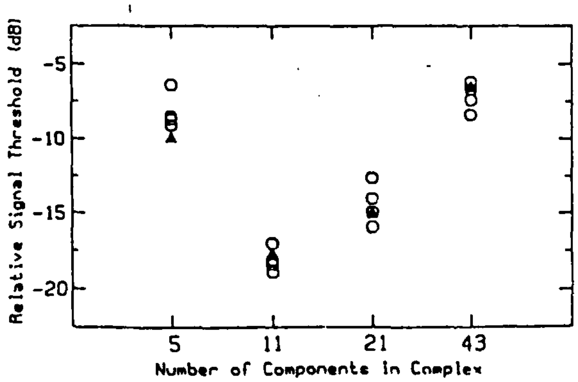
In most of the experiments concerning profile analysis, the phase of each component of the multitonal complex has been chosen at random and the same waveform (except for random variation of level) is presented during each "non-signal" interval. Therefore, the logical possibility exists that observers might recognize some aspect or aspects of the temporal waveform. If this were true, then discrimination could be based on some alteration of the temporal waveform during the "signal" interval rather than by a change in the spectral shape of the stimulus per se.

Green and Mason (1985) investigated this possibility directly with the following experimental manipulations. Multicomponent complexes were generated which consisted of 5, 11, 21, or 43 components spaced logarithmically. In all cases, the frequency of the lowest component was 200 Hz, the highest was 5 kHz. The overall level of the complex was varied randomly over a 40 dB range across presentations with a median level of 45 dB SPL per component. The signal consisted of an increment to the 1-kHz, central component of the complex.

In what Green and Mason termed the "fixed-phase" condition, four different complexes were generated for <u>each</u> number of components (5, 11, 21, and 43) by <u>randomly</u> selecting the phases of each component. Note that for these fixed-phase conditions, the same waveform (except for random variation of overall level) occurred during each non-signal interval.

In what Green and Mason called the "random-phase" conditions, 88 different phase-randomizations of the multicomponent complex were generated. On each interval of each trial, one of the 88 waveforms was selected at random (with replacement) for presentation. Thus, the <u>temporal</u> waveforms generally differed on each presentation. The amplitude spectra, however, were identical.

Figure 1. Signal threshold (dB) as a function of the number of components in the complex. Open circles: data obtained for each of the four phase-randomizations when the phase of each component was fixed throughout a block of trials ("fixed-phase" condition). Filled triangles: data from the "random-phase" condition in which the phases of the components were chosen at random on each presentation.



The results are presented in Figure 1. For each value of component number, the open circles represent the thresholds obtained for each of the four randomizations in the fixed-phase condition. The triangles represent the data obtained in the random-phase conditions. The results indicate that changing the phase of the individual components and thus the characteristics of the temporal waveform has little, if any, effect on discrimination even if the waveform is chosen at random on each and every presentation. These data are consistent with those obtained by Green, Mason, and Kidd (1984) who generated waveforms utilizing a procedure similar to the fixed-phase condition described above.

The inability of changes in the phase of the individual components, and thus changes in the characteristics of the temporal waveform, to affect discrimination supports the view that, in these tasks, observers are, indeed, basing their judgements on changes in spectral shape.

The form of the function relating threshold to the number of components in the complex is one that has been replicated many times in our laboratory. In general, as the number of components and thus the density of the profile is increased from 3 to 11 or 21 performance improves. An intuitive explanation for this result is that as the number of components which compose the profile is increased, additional independent bands or channels contribute to an estimate of the "level" of the profile.

Further increases in the density of the profile lead to decrements in performance and this trend is, for the most part, explained by simple masking. When the components are spaced so closely such that several components fall within the "critical band" of the signal, the addition of the signal produces a smaller relative increase in intensity and thus becomes more difficult to detect. In future publications we will present a more detailed analysis of these effects.

Frequency Effects

The results discussed above suggest that detection of an increment to a single component of a multi-component complex is based on changes in spectral shape. The phase relation among the components appears

Model Paper for 15% Reduction
to 20 x 40 picas type page

to have little, if any, effect on performance.

In exploring the nature of this process, one fundamental question is whether the frequency of the component which is incremented (the frequency region where the change in the power spectrum occurs) greatly influences the ability to detect a change in spectral shape.

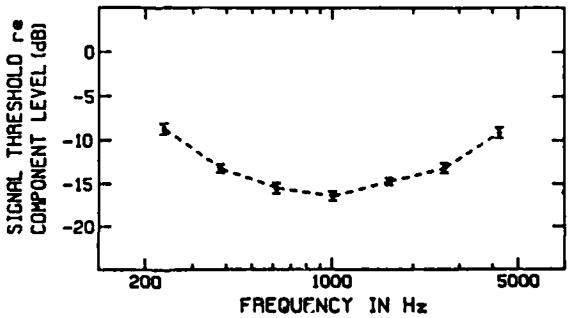
This question also bears on that of how the auditory system codes intensity. There are, at least, two different mechanisms that have been proposed as the basis for detecting changes in the intensity of sinusoidal components. One is what we will call the "rate" model. It assumes that changes in acoustic intensity are coded as changes in the rate at which fibers of the eighth nerve fire. One limitation of this model is the fact that the firing rates of practically all auditory fibers saturate as the intensity of the stimulus is increased (Kiang 1965; Sachs and Abbas, 1974; Evans and Palmer, 1980). The dynamic range of firing rate for many fibers is only about 20 to 30 dB. On the other hand, it is possible that there is some residual information in small changes of rate even at the highest stimulus levels where the amount of change produced by increasing the intensity of the stimulus is small. There is also the question of how one should regard saturation when one considers the entire population of fibers which may respond to a given stimulus in that different populations of fibers may saturate at different intensities.

A second view of intensity coding stresses the temporal characteristics of neural discharges. Sachs and Young (1979) and Young and Sachs(1979) have demonstrated that "neural spectograms" based on neural synchrony measures preserve the shape of speech spectra better than those based on firing rate. We were, therefore, particularly interested in how well observers could detect a change in spectral shape at very high frequencies. At the highest frequencies, above 2000 Hz, neural synchrony deteriorates and, if that code were used to signal changes in spectral shape, then the ability to detect such alterations in the acoustic spectrum should also deteriorate.

In one previous study, Green and Mason (1985), we made some measurements of how the locus in frequency affects the ability to detect a change in a complex spectrum. Our results suggested that the mid-frequency region, 500 to 2000 Hz, yielded the best performanace but variability among the different observers was sizable. Also, those data may have been contaminated by the listeners having received substantial prior practice with signals which were in the middle of the range.

The results of our most extensive experiment (Green, Onsan, and Forrest, 1986) on this issue are shown in Figure 2. The standard spectrum is a complex of 21-components, all equal in amplitude and equally spaced in logarithmic frequency. The overall level of the standard was varied over a 20-dB range with a median level of 40 dB SPL per component. The signal, whose frequency is plotted along the abscissa of the figure, was an increment in the intensity of a single component. The ordinate, like that of Fig. 1, is the signal level re the component level to which it was added. The results show that best detection occurs in a frequency range of 300 to 3000 Hz, with only a mild deterioration occurring at the higher and lower frequencies. If detection of an increment in this task were mediated by changes in neural synchrony, one would expect to observe considerably poorer performance at the highest frequencies as compared to the middle and low frequencies. This did not occur.

Figure 2. Signal threshold (dB) as a function of the frequency of the signal. Twenty-one-component complexes were employed. The signal was added in-phase to the corresponding component in the complex.



One other result from this recent study also deserves mention. The experiment described immediately above was repeated with one important exception. The median level of the standard was 60 rather than 40 dB SPL. This higher intensity level would be expected to produce firing rates at or close to saturation in nearly all fibers. Despite this fact, the thresholds obtained were, in almost all cases, lower than those obtained at the lower intensity level.

In conclusion, these two results do not afford a determination of the underlying neural code which mediates the detection of a change of spectral shape in our experiments.

Complex Spectral Changes

The experiments described above involve changes in the intensity of a single component of the multi-component profile (a "bump" in the spectrum). We now turn our attention to more complicated manipulations, experiments in which the intensities of several components of the spectrum were altered simultaneously. A primary goal of these experiments was to determine whether listeners' ability to detect these complex changes could be predicted on the basis of their sensitivity to changes in the intensity of a single component in the profile.

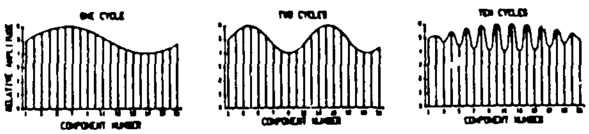


Figure 3. Three different frequencies, k, using sinusoidal variation. The signal amplitude at each component frequency is given by Eq. 1 and is added to the standard with a relative amplitude about 1/5 the standard amplitude.

Once again, a flat, "standard" composed of logarithmically spaced components ranging from 200 to 5000 Hz was used. The signal, however had an amplitude-spectrum that varied sinusoidally. The amplitude of the ith component, a[i], was given by

$$a[i] = \sin(2 * \pi * k * i/M) \qquad i=1,M \qquad \text{Eq. 1}$$

where k represents the "frequency" of the variation and M is the number of components presented. We refer to this variation in amplitude as a "sinusoidally rippled" spectrum, and to k as the "ripple frequency". Figure 3 illustrates the result of in-phase addition of the "standard" and the "signal" of for case M=21. The three values of k are as indicated. Cosinusoidally rippled amplitude spectra have also been examined. Such signals are generated as described above, except that the sine term of Eq. 1 is replaced by cosine.

Two points deserve note. The first is that k, the frequency of the ripple, is restricted by the number of components. This value must be smaller than one half the number of components ($k < M/2$). Second, changing the value of k does not alter the signal's root-mean-square (RMS) amplitude. All values of k produce the same a[i]'s, only their order is changed.

Thresholds were measured as the RMS amplitude of the signal re the RMS amplitude of the standard. Values of k ranged from 1 to 10. Thresholds were virtually constant for all values of k (ripple frequency) and type of varialtion (sine or cosine), with an average of -24.5 dB across all conditions (Green, Onsan and Forrest, 1986).

These data define a modulation transfer function (MTF). Interestingly, this function is flat rather than exhibiting the low-pass characteristic that is typically observed in sensory psychophysics. Because k may not exceed 10 for this 21-component complex, we were unable to investigate higher ripple frequencies and thus to assess more completely the form of the MTF. Undoubtedly, thresholds would increase if the ripple frequency were sufficiently large. We are currently examining the effect of greater ripple frequencies by using profiles composed of a greater number of components. These data will allow us to describe more fully the MTF i.e., the relation between the frequency of the ripple and detectability.

Finally, let us compare the rippled specrtum thresholds with predictions based on the ability to discriminate a bump in the spectrum; data obtained using increments to a single component of the profile. Because the ability to detect an increment in a single component of a 21 component spectrum is, to a first approximation, independent of the frequency of the signal (Fig. 2), one may predict the threshold for these 21 component rippled spectra. If we assume that the information concerning changes in the intensity of each of the signal's 21 channels is processed

Model Paper for 15% Reduction
to 29 × 46 picas type page

independently and that d' is proportional to pressure, then the optimal combination is the one in which the squared d' for the complex stimulus is equal to the sum of the squared d's associated with the each of the channels (Green and Swets, 1966). This leads to the expectation that the detectability will be improved by the square root of 21.

The process is as follows. The detection of a bump in a flat profile leads to thresholds of about -16 dB. This translates to a pressure of 0.16 relative to the standard. Thus, we would expect that the average pressure per component for a 21 component signal to be $0.16/\sqrt{21}$ or 0.035 (relative to the standard) which is equivalent to an RMS amplitude of -29 dB. This value is 4.5 dB smaller than the mean of -24.5 dB observed. Thus, performance on the complex spectral shape discrimination task is poorer than expected based on the data collected using changes in the intensity of a single component in the spectrum.

One could argue, of course, that there are less than 21 independent estimates of the spectrum. This is certainly possible, but two points argue against it. The first is that only six or seven independent channels across the 200 to 5000 Hz range are needed in order to acheive the level of performance found in using the rippled spectra. Second, if the different components are not processed independently, then increasing the ripple frequency would be expected to produce increases in discrimination thresholds. Rather, we find that ripple frequency does not affect threshold levels over the range of values tested, and that the thresholds obtained using complex, rippled spectra fall short of those expected based on the results of discrimination of changes in a single component of the profile.

Acknowledgement

This research was supported in part, by grants from the National Institutes of Health and the Air Force Office of Scientific Research. Dr. Richards was supported, in part, by an NIH post-doctoral fellowship.

References

Evans, E. F. and Palmer, A. R. (1980). Relationship between the dynamic range of cochlear nerve fibres and their spontaneous activity. Experimental Brain Research, 40, 115-118.

Green, D. M. and Mason, C. R. (1985). Auditory profile analysis: Frequency, phase, and Weber's Law. Journal of the Acoustical Society of America, 77, 1155-1161.

Green, D. M., Mason, C.R. and Kidd, G., (1984). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. 75, 1163 - 1167.

Green, D. M., Onsan, Z. A. and Forrest, T. G. (1986). Frequency effects in profile analysis. Paper in preparation for the Journal of the Acoustical Society of America.

Green, D. M. and Swets, J. A. Signal detection theory and psychophysics. New York: Wiley, 1966 (Reprinted by R. E. Krieger, Huntington, N. Y., 1974.)

Kiang, N. Y.-S. (1965). Discharge patterns of single fibers in the cat's auditory nerve. Research Monograph 35, Cambridge, MA: MIT Press.

Sachs, M. B. and Abbas, P. J. (1974). Rate versus level functions for auditory-nerve fibers in cats: Tone-burst stimuli. Journal of the Acoustical Society of America, 56, 1835-1847.

Sachs, M. B., and Young, E. D. (1979). Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. Journal of the Acoustical Society of America 66, 470-479.

Young, E. D. and Sachs, M. B., (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerver fibers. Journal of the Acoustical Society of America, 66, 1381-1403.

Additional Reference

Green, D. M. Profile Analysis:Auditory Intensity Discrimination. In Press, Oxford University Press, 1986.

The profile-analysis bandwidth

Leslie R. Bernstein and David M. Green
Department of Psychology, University of Florida, Gainesville, Florida 32611

(Received 22 December 1986; accepted for publication 23 February 1987)

Detection of a change in spectral shape, or profile analysis, appears to be mediated by comparisons across widely separated frequency "channels" rather than by local comparisons among adjacent frequency regions [e.g., Green *et al.*, J. Acoust. Soc. Am. **73**, 639–643 (1983)]. Two experiments were conducted in order to determine the "resolution bandwidth" of these channels. The first involved detection of an increment to a single component of a multicomponent background as a function of the number of components in the background. Performance improved as the number of components was increased from 3 to 21. Further increases yielded *poorer* performance and the estimate of the "resolution bandwidth" from these data suggests that this poorer performance was due simply to masking. The second experiment involved discrimination of a multicomponent complex having a flat amplitude spectrum from one having a sinusoidally "rippled" amplitude spectrum. The latter experiment yielded somewhat larger estimates of the "resolution bandwidth" than did the former. Finally, profile analysis was investigated under a dichotic condition that precluded peripheral masking of the signal. Our results, like those of Green and Kidd [J. Acoust Soc. Am. **73**, 1260–1265 (1983)], suggest that, although spectral analysis can be achieved using information across ears, performance is inferior to that obtained with diotic stimuli.

PACS numbers: 43.66.Fe, 43.66.Jh, 43.66.Rq, 43.66.Yw

INTRODUCTION

A wide variety of experimental data reported in previous publications suggests that detection of a change in spectral shape, or profile analysis, is a "global" process. The detection process appears to depend upon simultaneous comparisons across wide separations in frequency, i.e., across widely separated "channels" rather than on local comparisons among adjacent frequency regions (e.g., Green *et al.*, 1983; Green *et al.*, 1984). Consideration of the nature of this process has led to two related questions. The first question concerns the bandwidth of each of these channels. The second question concerns how information is combined across the individual channels. We choose to refer to these channels as "resolution bands" rather than as critical bands because, although they are probably closely related, it is unclear, *a priori*, whether they are indeed identical. The first two experiments we will report address these questions. Two quite different experiments were employed in order to determine the width of the "resolution bands." In the first, we measured listeners' thresholds for an increment to a single component of a multicomponent background as a function of the number of components in the background. In the second, listeners discriminated between a flat, multicomponent background and one which was characterized by a sinusoidally "rippled" spectrum. Thresholds were determined as a function of the number of "ripples" or the "frequency" of the ripple.

A related question concerns the extent to which the detection process is limited to peripheral processes. Certainly, some aspects of profile analysis appear to suggest some central comparison because increasing the density of components that define the profile leads to improvements in performance (e.g., Green *et al.*, 1983; Green *et al.*, 1984; Green and Mason, 1985). However, peripheral aspects are also apparent because, if the components which compose the multicomponent background or "standard" are spaced so closely that several components fall near the frequency of the signal, a decrement in performance results which appears to be due, at least in part, to simple masking (Green and Mason, 1985).

In almost all of the experiments concerning profile analysis reported to date, the stimuli have been presented diotically. That is, the stimuli were identical at each ear. In the third experiment, we wished to compare performance obtained with diotic stimuli to that obtained when the stimuli were presented *dichotically*. That is, the profile (except for the component at the signal frequency) was presented to one ear and the component to which the signal was added was presented to the contralateral ear. Green and Kidd (1983) also used this dichotic configuration and found performance to be substantially inferior to that obtained with diotic stimuli. However, it is unclear to what extent this result was influenced by their listeners having received substantial prior training with the diotic presentation. In the third experiment, we again attempted to determine whether profile analysis can be achieved when the stimuli are presented dichotically. More specifically, we wished to assess (1) how efficiently "profile" information could be integrated across the ears and (2) the form of the function relating detection threshold to the number of components which compose the background when the possibility of peripheral masking of the signal is removed.

I. EXPERIMENT 1—EFFECTS OF SPECTRAL DENSITY

A. Procedure

Multicomponent complexes consisting of 3, 5, 11, 21, 41, or 81 components with logarithmic frequency spacing between components were utilized. For each complex, the lowest frequency was 200 Hz; the highest was 5 kHz.

All stimuli were generated and presented via a PDP 11/73 which also controlled the experimental timing and collection of responses. The stimuli were played through a 16-bit D/A at a sampling rate of 25 kHz and were low-pass filtered at 10 kHz. The duration of each stimulus was 100 ms with 10-ms $\cos^2$ rise/decay ramps. The stimuli were presented diotically over TDH-50 earphones.

A two-alternative, forced-choice procedure was used. Each trial consisted of two 100-ms observation intervals separated by 500 ms. Intervals were marked by a visual display at the listener's response box. Feedback was provided for 200 ms after the listener responded.

During one observation interval, the multicomponent background was presented with all components at equal amplitude. The other interval contained the background plus the signal. The signal consisted of an in-phase addition to a single component of the complex. The signal occurred with equal *a priori* probability in the first or second interval.

Three different frequencies were selected for the signal: 380, 1000, and 2626 Hz (except in the case of the five-component complex where frequencies of 447, 1000, and 2236 Hz were employed. For the three-component complex, only one frequency of the signal, 1 kHz, was employed). Different frequencies of the signal were utilized in order to determine whether the resolution bandwidth depended on center frequency; e.g., the bandwidth might be a constant ratio of center frequency (signal frequency). If this were so, then decreases in performance (presumably due to masking) produced by increasing spectral density ought to be similar regardless of the region of the spectrum which contains the signal.

The level of the signal was varied adaptively in order to estimate that level which would produce 70.7% correct (Levitt, 1971). The level was decreased by 4 dB following two correct responses and increased by 4 dB following one incorrect response. After four "reversals," this step size was reduced to 2 dB. Threshold was defined as the mean of the signal level across all reversals, excluding the first four. Trials were run in blocks of 50 and each run produced approximately 12 to 16 reversals. The frequency of the signal was fixed over each block of trials. Twenty-four estimates of threshold were obtained for each listener and condition. The mean of these estimates, averaged across listeners, is reported as threshold.

The overall level of the stimuli was varied over a 20-dB range in 1-dB steps. A value was chosen randomly on each and every presentation in order to preclude the listeners' basing their judgments on absolute level rather than on the spectral shape. The median level was 50 dB SPL per component. The dependent variable (threshold) is the ratio in dB of the level of the signal (the size of the in-phase addition) to the level of the corresponding component in the background. For example, if the level of the signal were equal to the level of the component in the background, then we say the signal-to-background ratio is 0 dB.

Five paid observers with normal hearing participated in this experiment.

B. Results and discussion

Figure 1 contains the data obtained when the frequency of the signal was 1 kHz. The number of components in the profile is plotted logarithmically along the abscissa; signal threshold in dB is displayed along the ordinate. Each point represents the mean of the thresholds obtained from the five listeners. The error bars represent the mean of the standard errors computed for the individual listeners. The solid lines represent our theoretical predictions and will be discussed in detail later. The data indicate that as the number of components is increased from 3 to 21, threshold decreases monotonically (the signal becomes more detectable) from about − 11 to − 20 dB.

As the number of components is increased beyond 21, threshold increases monotonically to about − 8 dB for an 81-component complex. The sharp minimum in the function at 21 components is also characteristic of the individual data. These data are entirely consistent with those of earlier investigations (Green and Mason, 1985; Green *et al.*, 1983, 1984).

As mentioned in the introduction, these trends can be explained in a rather straightforward manner. Assume that the listener detects the presence of the signal by comparing the *relative* level in the resolution band containing the frequency of the signal to the level of the remaining bands across the spectrum. As the number of components which compose the profile is increased from 3 to 21, additional independent bands or channels contribute to an estimate of the mean "level" of the profile. As the number of components and thus the density of the profile is increased beyond 21, additional components fall into the "resolution band" of the signal. The addition of the signal then produces a rela-

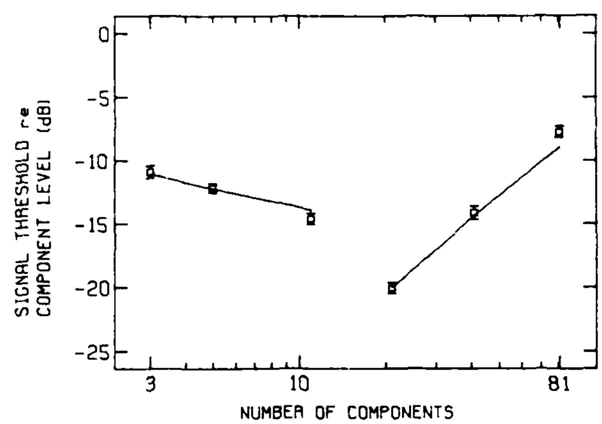


FIG. 1. Threshold for detection of an increment to the 1-kHz "signal" component of a multicomponent background as a function of the number of components in the background. Squares represent thresholds averaged across listeners. Error bars represent the mean of the standard errors computed for individual listeners. Solid lines represent theoretical predictions as discussed in the text.

tively smaller increase in power within the band and thus becomes more difficult to detect. According to these notions, the monotonic decrease in threshold in the left-hand portion of Fig. 1 is due to integration of information across bands. The monotonic increase in the right-hand portion is due to simple masking. We will now examine these explanations more formally.

The monotonic decrease in threshold in the left-hand portion of the figure was modeled by assuming that as components are added to the complex, they yield additional *independent* estimates of the level of the profile. Under this assumption, threshold would be expected to decrease at a rate of 1.5 dB per doubling of the number of components, that is, with the $\sqrt{n}$. A line having this slope was fit to the data by eye and appears to predict the decrease in threshold quite well.

For increases in the number of components beyond 21, an intuitively appealing explanation is that simple masking causes an increase in threshold. If the density of the complex is such that the addition of components causes one or more to fall within a common "resolution band," then they provide no additional information as to the level of the profile. Rather, their presence causes the signal to produce a smaller relative increase in power within the resolution band and thus a less effective signal. Because the minimum of the function lies at 21 components, this value appears to be a good estimate of the spectral density at which this occurs.

The increase in threshold in the right-hand portion of Fig. 1 was modeled in the following manner. For simplicity, we assumed that for the 21-component complex, only the 1-kHz signal component lies within the resolution band. Next, we calculated the increase in power within the band produced by the signal at threshold for this condition. The resolution band was modeled as a triangular filter symmetric in log space whose bandwidth we wished to determine. For the 41- and 81-component complexes, multiple components would, presumably, fall within this resolution band. As a function of the bandwidth of the filter, we calculated the level of the signal necessary to produce a constant increment in power within the band, i.e., the same increase in power produced by the signal at threshold for the 21-component complex. The predicted thresholds are plotted as the solid line in the right-hand portion of the figure for a triangular filter extending from 852–1174 Hz. Our predicted thresholds lie remarkably close to the data. Most important, the *equivalent rectangular bandwidth* of our filter, 162 Hz, compares favorably with accepted estimates of the critical band around 1 kHz.

The reader may be puzzled (as were we) that thresholds increase at a rate of about 6 dB per doubling of the number of components rather than at 3 dB per doubling. Note that we have calculated the size of the increment which must be added in-phase *to the single component at the frequency of the signal* in order to produce a constant increment in power within the band. Our calculations reveal that as the number of components is increased beyond 21 and multiple components begin to fall within the resolution band, the slope of the line relating threshold to the number of components is predicted to be about 5.5 dB per doubling, very close to that actually obtained. (The predicted slope eventually asymptotes to 3 dB per doubling but only for very, very large numbers of components, i.e., 10 000 or more!)

In summary, the data in Fig. 1 are described well by considering the improvement in performance as the number of components is increased from 3 to 21 to be due to integration of information across independent bands or channels, and the decrement in performance for increases beyond 21 to be due to masking. To the degree that the data depart from these predictions, they do so largely because the monotonic decrease in threshold does not exhibit a uniform slope. Rather, as noted above, there appears to be a sharp drop between 11 and 21 components. At present, we have no satisfactory explanation for this trend which is also exhibited in the individual listener's data.

Figure 2 is similar to Fig. 1 and contains the data obtained for frequencies of the signal of 380 Hz, 1 kHz, and 2.626 kHz. The parameter of the plot is the frequency of the signal. Three of the listeners from the original group of five participated in this portion of the experiment. The 1-kHz data have been replotted from Fig. 1. Each point represents the mean of their thresholds. Note that in the case of the five-component background, the low and high frequencies employed were actually 447 Hz and 2.236 kHz, respectively.

Curiously, the thresholds obtained with signals above and below 1 kHz do not exhibit a sharp minimum at 21 components. However, for all three frequencies of the signal, thresholds increased rapidly when the number of components was increased beyond 21. This finding suggests that the width of the resolution band is a constant ratio of center frequency over the range of values tested. Note that the thresholds for the 380-Hz and 2.626-kHz signals are elevated relative to those obtained when the signal was added to the central, 1-kHz component. The average increase in threshold relative to the 1-kHz signal is 4.7 and 6.1 dB for the 380-Hz and 2.626-kHz signals, respectively.

This elevation of threshold for high- or low-frequency signals is consistent with previous studies. Green and Mason (1985), who used stimuli similar to those employed here,

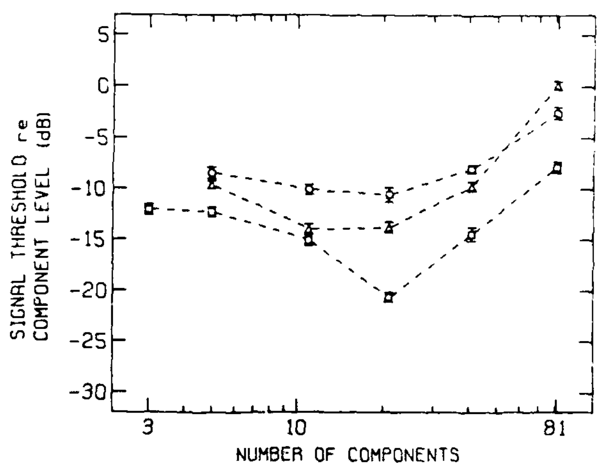


FIG. 2. Similar to Fig. 1. The parameter of the plot is the frequency of the signal; triangles: 380 Hz; squares: 1 kHz; circles: 2626 Hz. Note that for the five-component background, signal frequencies of 447, 1000, and 2236 Hz were employed.

also observed that detection thresholds were lowest when the signal was added to the central, 1-kHz component of a 21-component complex. In addition, a more recent investigation in our laboratory (Green *et al.*, 1987) also revealed this trend. However, the results of this latter study also indicated that (1) when listeners receive substantial practice, thresholds for low-frequency signals are extremely close to those obtained with a 1-kHz signal, differing by only about 3 dB and (2) above 1 kHz, threshold increases slowly with frequency reaching + 6 dB relative to that obtained at 1 kHz. The data in Fig. 2 are in agreement with these findings in that thresholds are, in general, lowest for the 1-kHz signal and highest for the 2.626-kHz signal.

II. EXPERIMENT 2—VARIATION OF SINUSOIDAL RIPPLE

The purpose of this experiment was to provide an additional, independent estimate of the resolution bandwidth. In this experiment, the signal produced a sinusoidal change in the amplitudes of the flat profile rather than an increment to a single component. That is, the addition of the signal produced what we refer to as a "rippled" spectrum. We wished to investigate how detection would be affected as a function of the number of ripples.

The results of a previous investigation (Green *et al.*, 1987) showed that thresholds were virtually constant as the "frequency" or number of ripples was varied from one to ten. We reasoned that if the frequency of the ripple was increased further, a point would be reached where the relatively high-frequency sinusoidal variation in amplitude would not be detectable because individual "cycles" would fall *within* single resolution bands. The "internal" spectrum would thus be flat and indistinguishable from the background. That is, we expected the data to exhibit a low-pass characteristic. The point at which sensitivity begins to decline would indicate the spacing at which a peak and valley of the ripple begin to fall within a single band and, hence, would provide an estimate of the resolution bandwidth.

A. Procedure

The standard waveform was a 161-component flat spectrum that ranged in frequency from 200–5000 Hz. The successive components were spaced equally on a logarithmic scale. The ratio between successive frequencies was 1.0203; there were 34.5 components per octave. The addition of the signal produced a power spectrum whose amplitude varied sinusoidally as a function of the logarithm of frequency. Figure 3 shows this manipulation graphically for a 21-component complex. The first panel shows a single cycle of sinusoidal variation in amplitude over the spectrum; the next one shows two cycles of amplitude variation; and, finally, the last panel shows ten cycles, the greatest variation that can be achieved with 21 components because alternate components increase and decrease in amplitude.

Specifically, the "signal" waveform was produced by setting the amplitude of successive components $a(i)$ according to the following equation:

$$a(i) = \sin[2\pi k(i/M)] \quad i = 1,2,...,M,$$

where i is the number of the component, ranging in this case from 1 to 161, $a(i)$ is the amplitude of the *ith* component of the signal spectrum, and k is frequency of the ripple. Recall that the first component, $i = 1$, corresponds to a frequency of 200 Hz, and the last component, $i = 161$, corresponds to a frequency of 5000 Hz.

The "depth" of the ripple resulting from the addition of the signal to the standard waveform depends upon the ratio of the amplitudes of the signal components to those of the standard's equal-amplitude components. The depth of the ripple is, of course, monotonically related to the signal-to-standard ratio. We scaled the amplitude of this "signal" and added each component in-phase (respecting sign) to the corresponding component of the flat standard spectrum to produce the change in the spectrum, such as that shown in Fig. 3. In that figure, the signal amplitude is about 20% of the standard amplitude.

It should be noted that by constructing the signal in the manner described above, the root-mean-square (rms) of the amplitudes across components is independent of the frequency of the ripple k because the 161 values for any set of $a(i)$ are the same; only their order within the set has been changed. If the maximum value for $a(i)$ is 1, the rms value is 0.707. We refer to the signal-to-standard ratio as the rms signal amplitude to the amplitude of any component of the standard.

Note that one disadvantage of this technique is that we cannot determine if the decline in detection performance as the number of ripples is increased is dominated by any specific frequency region(s). The data from experiment 1 suggest that the resolution bandwidth is, roughly, a constant proportion of center frequency. If this were true, then because our spectra are rippled sinusoidally as a function of the *logarithm* of frequency, the predicted decline in detection

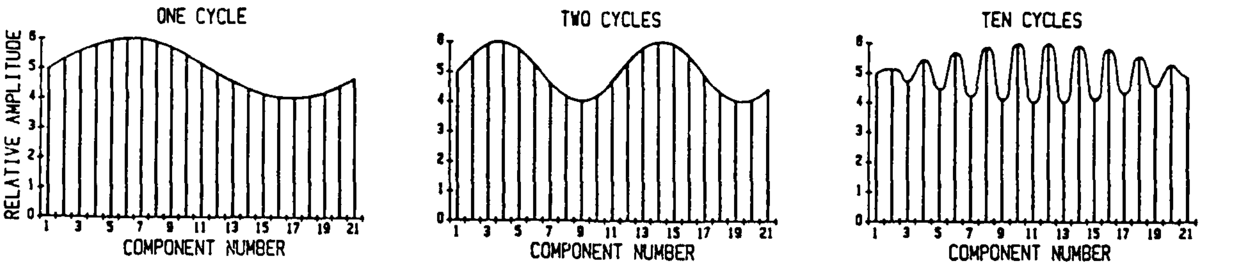


FIG. 3. Three different "frequencies" of ripple k, using sinusoidal variation. The signal amplitude at each component frequency is given by Eq. (1) and is added to the standard with a relative amplitude of about 1/5 that of the standard.

performance as the frequency of the ripple increases would be mediated by a uniform loss of resolution across the spectrum. In any case, our estimate of the resolution bandwidth yielded by this procedure must be a relative one, in effect, an estimate of Q.

The five observers who participated in experiment 1 also participated in this experiment.

B. Results and discussion

The data are displayed in Fig. 4 where the number of ripples imposed on the flat spectrum is plotted logarithmically along the abscissa. The threshold of the signal is displayed in the usual manner along the ordinate. Recall that we have reported the rms value as our measure of the amplitude of the signal.

The average data obtained from three of the five listeners (group 1) are plotted as squares; triangles represent the average data from the remaining two (group 2). Error bars represent the mean of the standard errors computed for the individual listeners whose data are displayed. Data averaged across all listeners are plotted along the solid line.

The data in Fig. 4 indicate that thresholds remain fairly constant as the number of ripples is increased from 1 to 10, a result which was also obtained by Green *et al.* (1987). As the "frequency" of the spectral ripple is increased beyond 10 to 80, thresholds increased monotonically for group 1 at a rate of about 6 dB/oct, reaching about − 10.5 dB at 80 ripples. Thresholds for the two listeners in group 2 are higher than those of group 1 over the entire range of values tested. In addition, they do not increase monotonically between 10 and 80 ripples. Rather, the thresholds obtained in the 40-ripple condition are higher than those obtained with 80 ripples.

At first, we thought this anomaly was either due to random fluctuations in the data or was artifactual. We reran several of the conditions after generating new signals for the 40-ripple condition and found that the elevation in threshold at 40 ripples persisted for these two listeners. Finally, we ran a third group of three listeners which included one naive listener and two highly trained listeners employed in our laboratory. Interestingly, both the naive listener as well as one of the highly trained listeners exhibited a nonmonotonicity similar to that displayed for group 2 in Fig. 4. Their thresholds for 40 ripples were higher than those for 80. We find this trend, which occurred in four of eight listeners tested, to be most perplexing. We are unable to offer a satisfactory explanation for its existence.

In general, the data of Fig. 4 exhibit the expected low-pass characteristic described earlier. When the number of ripples is increased from one to ten, thresholds remain essentially constant. As the number of ripples is increased beyond ten, thresholds increase. We attempted to use the data of Fig. 4 to determine the "low-pass cutoff" of the function and ultimately an estimate of the resolution bandwidth.

We used several procedures to fit two straight lines to each group's data and took their respective intersections as estimates of the "corner-" or 3-dB-down-point of the implied resolution band. Depending on the details of the procedure used to fit the data, the intersection occurred between 10 and 14 ripples. Considering that our components span 4.64 octaves, a 3-dB point of 14 ripples implies that the resolution band spans 0.33 octaves. At 1 kHz, the resolution band would be about 230 Hz wide. Similar calculations for a cutoff of ten ripples yield a bandwidth of about 320 Hz. There is a considerable discrepancy between these estimates of the resolution bandwidth and that of 160 Hz obtained from the data of experiment 1. In addition, they are quite a bit larger than usual estimates of the critical band in this region.

It should be noted that, for the data of Fig. 4, a cutoff of slightly greater than 20 ripples would have had to have been observed in order for the estimate of the resolution bandwidth to match that of 160 Hz obtained from experiment 1. No reasonable fit to the data of Fig. 4 would yield such an estimate. Thus we are confident that the discrepancy in our estimates of the resolution bandwidth is not a result of the particular details of the procedures we employed to fit the data.

On the other hand, if the resolution bandwidth was *not* a constant proportion of center frequency, then our single estimate of the resolution bandwidth of 230 Hz around 1000 Hz which corresponds to a Q of 4.34 could be dominated by any frequency region which was characterized by a proportionately small resolution bandwidth. However, because the data from experiment 1 suggest that the resolution bandwidth is roughly a constant proportion of center frequency and, because our spectra are rippled sinusoidally as a function of the *logarithm* of frequency and, hence, should occupy equal spatial intervals along the basilar membrane, we are reasonably sure that the decline in detection performance as the frequency of the ripple increases is accompanied by a uniform loss of resolution across the spectrum.

One possible, but unappealing, explanation for our disparate estimates of the resolution bandwidth is that the detection processes employed by the listeners to perform the tasks of experiments 1 and 2 are sufficiently different as to be mediated by *different* resolution bandwidths. It is interesting

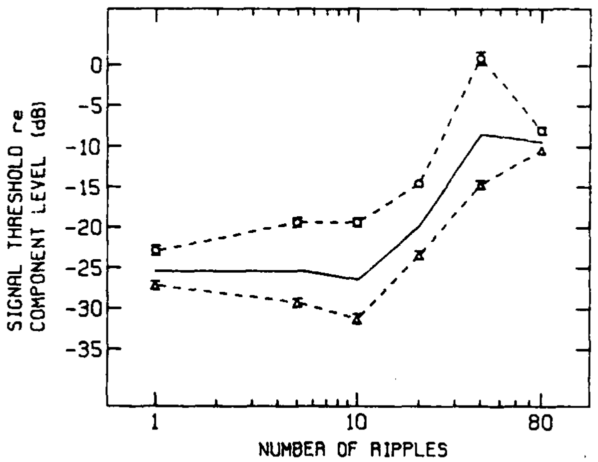


FIG. 4. Detection threshold as a function of the number of sinusoidal "ripples" imposed on the flat spectrum by the signal. Triangles: average data for group 1; open circles: average data for group 2; solid line: average data for all listeners. Error bars represent the mean of the standard errors computed for individual listeners.

to note that our stimuli with rippled spectra are in some sense similar to those employed by others (Bilsen and Ritsma, 1970; Yost and Hill, 1978) who investigated the discrimination of flat spectra from those with a linear spectral ripple. Furthermore, the thresholds obtained in these studies are similar to ours. It is quite possible then that our listeners employed pitch cues similar to those employed by listeners in these previous studies in order to detect the presence of the ripple. Such cues would not be expected to be available in the case of an increment to a single component (experiment 1). This suggests one way in which the tasks of experiments 1 and 2 may be different.

III. EXPERIMENT 3—DIOTIC/DICHOTIC COMPARISONS

The results of experiments 1 and 2 strongly suggest that detection of a change in spectral shape is *limited* by peripheral processes which produce peripheral masking. We noted in the introduction (as did Green and Kidd, 1983) that certain aspects of profile analysis appear to suggest some central comparison process(es). The present experiment was designed to investigate profile analysis in the absence of peripheral masking of the signal in an effort to assess more adequately the role of central processes.

A. Procedure

The procedure was the same as that employed in experiment 1 with a few important exceptions. All stimuli were generated and presented via an IBM-PC which also controlled the experimental timing and collection of responses. The stimuli were played through a 12-bit D/A at a sampling rate of 14.286 kHz and were low-pass filtered at 10 kHz.[1] The duration of each stimulus was 200 ms with a 5-ms, $\cos^2$ rise/decay ramp. The frequency of the signal was 1 kHz. Detection was measured as a function of the number of components in the multicomponent background.

Each trial consisted of two 500-ms observation intervals separated by 300 ms. The first 250 ms of each observation interval contained a visual warning display on the IBM's monitor. Feedback was provided for 400 ms. Eighteen estimates of threshold were obtained for each listener and condition.

Signals were presented either diotically as in experiment 1, or dichotically. When the dichotic configuration was employed, the "flat" profile (except for the component at the signal frequency) was presented to one ear and the component to which the signal was added was presented, in isolation, to the contralateral ear. Three paid observers with normal hearing (who had not participated in the previous experiments) participated in this experiment.

B. Results and discussion

Figure 5 displays the results for the diotic and dichotic conditions. The data for the diotic conditions are quite similar to those presented in Fig. 1 and, like those data, exhibit a minimum at 21 components. The dichotic thresholds are larger than the diotic for all numbers of components tested and do not exhibit any pronounced minimum. Indeed, they

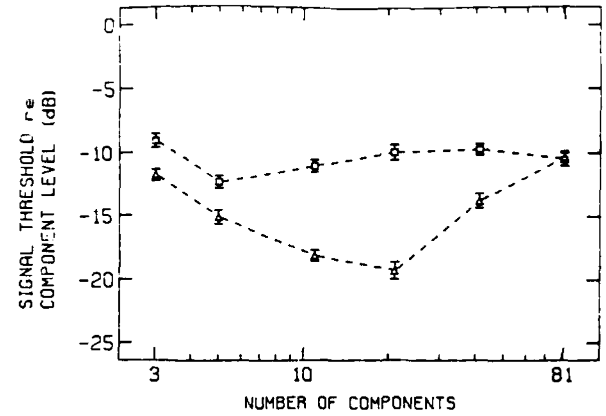


FIG. 5. Threshold for detection of an increment to the 1-kHz "signal" component of a multicomponent background as a function of the number of components in the background. The parameter of the plot is the interaural configuration of the signal. Triangles and squares represent diotic and dichotic conditions, respectively.

show little variation, having a mean of − 10.4 dB and ranging from − 9.0 dB at three components to − 12.3 dB at 5 components. It is important to note, that *all* of the dichotic thresholds displayed in Fig. 5 are superior to that which would be expected had the listeners ignored the profile information and used only the information in the contralateral ear to which the single sinusoidal component was presented. If such were the case and the listener based his/her decision on the interval containing the more intense tone, then one could calculate that the expected threshold, given the 20-dB random variation in overall level, would be − 3 dB (Green, 1986). The average dichotic threshold is about 7 dB lower than this value.

On the one hand, one would not expect the dichotic thresholds to *increase* as the number of components in the profile is increased beyond some critical value because the "flat" profile (except for the component at the signal frequency) was presented to one ear and the component to which the signal was added was presented, in isolation, to the contralateral ear. Thus there was no opportunity for peripheral masking to degrade performance as was true for the diotic conditions, and the data are consistent with such expectations.

On the other hand, if profile information could be combined across ears without loss, as the number of components increased from some small number, dichotic thresholds would be expected to decrease in a manner similar to that observed for the diotic thresholds. Furthermore, because this process would not be limited by masking, the dichotic thresholds could, theoretically, decline to some asymptotic value *at or below* that obtained in the most sensitive of diotic conditions. This, clearly, was not the case.

That the dichotic thresholds do not decline as the number of components is increased beyond five appears to suggest that there is some limit to the extent to which profile information can be combined across ears, which renders further increases in the number of components ineffective. Why this is so remains obscure.

In summary, the data of Fig. 5 do not appear to constitute a strong test of the extent to which profile analysis is mediated by peripheral and/or central processes. It may be the case that if central processes are involved, then they are limited by the efficiency with which information across the ears can be combined, a process that, presumably, precedes the determination of spectral shape.

We should observe that our dichotic thresholds are about 7–9 dB lower than those obtained by Green and Kidd (1983), who utilized a similar presentation with 3- and 21-component backgrounds. These investigators also found performance with dichotic stimuli to be substantially inferior to that obtained with diotic stimuli. However, it is unclear to what extent the discrepancy between the dichotic thresholds obtained by Green and Kidd and those obtained by us was influenced by their listeners having received substantial prior experience with the diotic presentations.

We also considered how binaural interaction may have influenced our dichotic thresholds. From the viewpoint of the majority of models of binaural hearing (see, for example, Colburn and Durlach, 1978), one prerequisite for binaural interaction is the presence of energy in corresponding frequency bands or channels at the two ears; i.e., neural events can only be compared across pairs of fibers with similar characteristic frequencies.

Recall that for our dichotic stimuli, the 1-kHz signal component was absent from the ear which contained the multicomponent background. To the extent that components in the profile which surrounded the 1-kHz region fell within a common "binaural critical band" with the 1-kHz component in the opposite ear, binaural interaction could occur. If such were the case, the listener could, theoretically, detect the presence of the signal by comparing the interaural intensitive disparities (IIDs) in the two intervals. During a signal interval, the IID would favor (relatively) the ear which contained the 1-kHz component to which the increment was added. This logical possibility exists regardless of the fact that the waveforms in each 1-kHz peripheral filter would not be highly correlated. Thresholds for IIDs have been shown to be as small as 0.4 dB (− 26.5 dB, using our dependent measure) with interaurally uncorrelated signals (Nuetzel, 1982).

The likelihood that a listener could utilize such a cue would increase as the number of components in the profile, and thus the proximity of components to the 1-kHz region increased. For example, in the case of the three-component complex, the "profile" channel contained the frequencies 200 Hz and 5 kHz. There is little possibility that either of these components could interact binaurally with the 1-kHz component in the opposite ear regardless of the value of the "binaural critical band" one chooses to accept (e.g., Bourbon, 1966; Sever and Small, 1979; Sondhi and Guttman, 1966). In contrast, for the 81-component complex, the components closest to 1 kHz in the "profile" ear are 960 and 1041 Hz, which lie within accepted values of a critical bandwidth.

The data of Fig. 5 do not support the notion that listeners were utilizing binaural cues because the number of components that compose the profile appears to have little, if any, systematic effect on the thresholds.

IV. GENERAL DISCUSSION

We have stated that profile analysis, or the detection of a change in spectral shape, appears to be a "global" process that relies on simultaneous comparisons across a wide range of independent frequency channels. Experiments 1 and 2 were designed to yield independent estimates of the bandwidth of these channels, what we have called the "resolution bandwidth." The estimate of 160 Hz around 1 kHz, obtained in experiment 1, is consistent with the accepted values of a critical bandwidth, and thus supports our contention that the decrements in performance observed with increasing spectral density (Fig. 1) are, in fact, due largely to masking.

The indirect estimate of the resolution bandwidth around 1 kHz, yielded by the data of experiment 2, was some 1.5 to 2 times larger than that obtained in experiment 1. This disparity leads us to speculate that the tasks of detecting an increment to a single component in a "flat," multicomponent background and that of discriminating a "rippled" from a "flat" spectrum differ in ways that are quite complex. One may not be able to simply extrapolate from one to the other.

This finding is not unique and, in one respect, these data are consistent with those obtained in an earlier investigation (Green, 1986). In that study, we reported that we were unable to predict listeners' ability to discriminate flat from rippled spectra on the basis of their sensitivity to changes in the intensity of a single component unless we assumed that performance in the former task was mediated by a *small number of widely spaced channels* across the spectrum, a conclusion which is entirely consistent with the data obtained in the present study. One way in which an effectively small number of channels may be produced is if the channels are not independent but rather are correlated in a manner suggested by Durlach *et al.* (1986). Regardless of the mechanism, it is difficult to understand why the number of effective channels for detecting the presence of spectral ripple is smaller than that which appears to be utilized when the task involves detecting increments to a single component. The data from experiment 1 show that thresholds continue to decline as the number of components is increased from 3 to 21, a fact which suggests that, for that task, there are many more than five or so effective bands.

It is difficult to understand why the effective bandwidth in these two tasks appears to differ so greatly. We are currently in the process of investigating, in greater detail, the nature of these two tasks.

The results of experiment 3, like the data of Green and Kidd (1983), suggest that although spectral analysis can be achieved using information across ears, performance is inferior to that obtained with diotic stimuli. However, our dichotic thresholds were somewhat smaller than those obtained by Green and Kidd. The present data do not support the notion that listeners were utilizing any binaural cues. In future publications we will report how binaural cues may affect the discrimination of spectral shape when a variety of dichotic configurations is employed.

ACKNOWLEDGMENTS

This research was supported by a grant from the Air Force Office of Scientific Research. The authors wish to thank Virginia M. Richards and T. G. Forrest for their helpful comments on earlier versions of this manuscript.

[1]A low-pass filter with a cutoff at or below the Nyquist frequency of 7.14 kHz should have been employed. Because the cutoff was 10 kHz, one or two "image" components in the region of 9 kHz were present in the passband of the filter. We reran several of the conditions employing a low-pass cutoff of 6 kHz. The thresholds we obtained were quite close to and did not differ in any systematic fashion from those presented in Fig. 5.

Bilsen, F. A., and Ritsma, R. J. (1970). "Some parameters influencing the perceptibility of pitch," J. Acoust. Soc. Am. 47, 469–476.

Bourbon, W. T., Jr. (1966). "Effects of bandwidth and level of masking noise on detection of homophasic and antiphasic tonal signals," unpublished dissertation, University of Texas, Austin, TX.

Colburn, H. S., and Durlach, N. I. (1978). "Models of binaural interaction," in *Hearing*, Vol. IV, *Handbook of Perception*, edited by E. C. Carterette and M. P. Friedman (Academic, New York).

Durlach, N. I., Braida, L. D., and Ito, Y. (1986). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. 80, 63–72.

Green, D. M. (1986). "Frequency and the detection of spectral shape change," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, New York).

Green, D. M., and Kidd, G., Jr. (1983). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. 73, 1260–1265.

Green, D. M., and Mason, C. R. (1985). "Auditory profile analysis: Frequency, phase, and Weber's Law," J. Acoust. Soc. Am. 77, 1155–1161.

Green, D. M., Kidd, G., Jr., and Picardi, M. C. (1983). "Successive versus simultaneous comparison in auditory intensity discrimination," J. Acoust. Soc. Am. 73, 639–643.

Green, D. M., Mason, C. R., and Kidd, G., Jr. (1984). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. 75, 1163–1167.

Green, D. M., Onsan, Z. A., and Forrest, T. G. (1987). "Frequency effects in profile analysis," J. Acoust. Soc. Am. 81, 692–699.

Levitt, H. (1971). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.

Nuetzel, J. M. (1982). "Sensitivity to interaural intensity differences in tones and noises measured with a roving-level procedure," J. Acoust. Soc. Am. Suppl. 1 71, S47.

Sever, J. C., and Small, A. M. (1979). "Binaural critical masking bands," J. Acoust. Soc. Am. 66, 1343–1350.

Sondhi, M. M., and Guttman, N. (1966). "Width of the spectrum effective in the binaural release of masking," J. Acoust. Soc. Am. 40, 600–606.

Yost, W. A., and Hill, R. (1978). "Strength of pitches associated with ripple noise," J. Acoust. Soc. Am. 64, 485–492.

Detection of simple and complex changes of spectral shape

Leslie R. Bernstein and David M. Green
Department of Psychology, University of Florida, Gainesville, Florida 32611

(Received 3 March 1987; accepted for publication 7 July 1987)

In most of the previous studies (see Green, 1987) concerning the detection of a change in spectral shape, or "profile analysis," the listener's task was to detect an increment to a single component of an otherwise equal-amplitude, multicomponent background. An important theoretical issue is whether listeners' sensitivity to more complex spectral changes can be predicted from these results. In the present investigation, the sensitivity of a single group of listeners to a wide variety of simple and complex spectral changes was determined. After collecting the data, it was noted that almost all the thresholds could be predicted by a simple calculation scheme that assumed detection of a change in spectral shape occurs when the addition of the signal to the flat, multicomponent background produces a sufficient difference in level between *only two* regions of the spectrum. Unfortunately, this scheme, while successful for our limited set of data, fails to account for other "profile" data, namely, those obtained when the number of components is altered.

PACS numbers: 43.66.Fe, 43.66.Ba, 43.66.Jh

INTRODUCTION

A number of previous publications by Green and his colleagues (see Green, 1987, for a review) have described listeners' ability to detect changes in spectral shape, a process termed "profile analysis." In most of those studies, the standard or background stimulus consisted of a number of equal-amplitude components spaced at equal logarithmic intervals in frequency. The signal, when added to the standard, produced a change in this spectrum, an increment to a single component of the multicomponent background. Thus a common spectral change was simply a "bump" in the otherwise flat spectrum.

An important theoretical issue is whether listeners' sensitivity to more complex spectral changes can be predicted from their sensitivity to increments to a single component of a 21-component background. While some limited attempts to address this question have been made in the past (Green, 1986), we wished to address it more thoroughly by determining the sensitivity of a single group of listeners to a wide variety of simple and complex spectral changes. By doing so, we hoped to describe a model that would account for all the data. Although a simple calculation scheme predicts the present data rather well, it clearly fails as a general model of profile analysis.

I. GENERAL PROCEDURE

In all the experiments described below, the stimuli were 21-component complexes with equal logarithmic frequency spacing between adjacent components. The lowest frequency was 200 Hz; the highest was 5000 Hz.

All stimuli were generated and presented via a PDP 11/73, which also controlled the experimental timing and the collection of responses. The stimuli were played through 16-bit D/A's at a sampling rate of 25 kHz and were low-pass filtered at 10 kHz. The duration of each stimulus was 100 ms with 10-ms $\cos^2$ rise/decay ramps. The stimuli were presented diotically over TDH-50 earphones to three listeners with normal hearing, who were seated in separate sound-treated rooms.

A two-alternative, forced-choice procedure was used. Each trial consisted of two 100-ms observation intervals separated by 500 ms. Intervals were marked by a visual display at the listener's response box. Feedback was provided for 200 ms after the listener responded.

During one observation interval, the multicomponent background was presented. All components of this standard were equal in amplitude. The other interval contained the standard plus the signal. The signal altered the amplitude of one or more components of the standard and occurred with equal *a priori* probability in the first or second interval.

The level of the signal was varied adaptively in order to estimate the level that would produce 79.4% correct (Levitt, 1971). The level was decreased by 4 dB following three correct responses and increased by 4 dB following one incorrect response. After four "reversals," this step size was reduced to 2 dB. Trials were run in blocks of 50 and each run produced approximately 10 reversals. Threshold was defined as the mean of the signal level across the last even number of reversals, excluding the first four. Twenty-four such estimates were obtained for each listener and condition. The mean of these estimates, averaged across listeners, is the dependent variable in all these experiments.

The overall level of the stimuli was varied over a 20-dB range in 1-dB steps. A value was chosen randomly on each and every presentation in order to preclude the listeners' basing their judgments on absolute level rather than on spectral shape. The median level was 50 dB SPL per component. Sensitivity to a change in the spectrum is reported as the ratio in dB of the level of the signal to the level of the corresponding component or components in the background. For example, if the amplitude of the signal component were equal to that of the corresponding component in the background, then we say the signal-to-background ratio is 0 dB. If the signal changes the amplitude of more than one component, we report the root-mean-square (rms) amplitude of the signal *re*: the standard amplitude.

II. EXPERIMENT I: SINGLE-INCREMENT THRESHOLDS

As noted in Sec. I, we wished to determine whether listeners' sensitivity to complex changes in spectral shape could be predicted from their sensitivity to increments of a single component in a "flat," multicomponent background. The first problem one encounters is that previous data (Green *et al.*, 1987) indicate that the detectability of an increment to a single component in the spectrum varies greatly as a function of the frequency of the component. Why this is so is an interesting issue in itself. The result has been known for some time (Green and Mason, 1985), but, as yet, we know of no satisfactory theoretical explanation for its existence.

The data obtained in this first experiment served as a basis for predicting listeners' sensitivity to complex changes in spectral shape. In addition, obtaining these data allowed us to determine whether the performance for this particular group of listeners was typical of that observed for the many listeners who have been tested previously.

The standard was composed of 21 equal-amplitude components ranging from 200–5000 Hz spaced equally distant on a logarithmic scale of frequency. The signal consisted of an in-phase addition to a single component of the standard. Five different frequencies were selected for the signal: 234, 525, 1000, 1903, and 4256 Hz. The frequency of the signal was fixed during a block of trials.

A. Results and discussion

The data are presented in Fig. 1. The frequency of the signal is plotted logarithmically along the abscissa; signal threshold in dB is displayed along the ordinate. Each point represents the mean of the thresholds obtained from the three listeners. The error bars represent the standard error of the mean computed across the data from all listeners. The dotted line represents the thresholds estimated by Green *et al.* (1987) in an extensive study of the effects of the frequency of the signal.

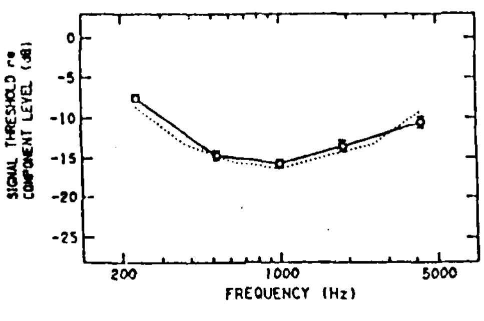


FIG. 1. Threshold for detection of an increment to a single component of a multicomponent background as a function of the frequency of the component. Circles represent thresholds averaged across listeners. Error bars represent the standard error of the mean computed across the data from all listeners. The dotted line represents data obtained by Green *et al.* (1987). Note that a three-down, one-up procedure (79.4% correct) was employed in the present study, whereas a two-down, one-up procedure (70.7% correct) was employed in the previous study.

Our average data indicate that the listeners were most sensitive to increments in the middle region of the spectrum. Thresholds differed by less than 2.5 dB for the 525-, 1000-, and 1903-Hz signals. The greatest sensitivity was observed with the 1000-Hz signal, which yielded a threshold of − 15.75 dB. The 234- and 4256-Hz signals yielded somewhat poorer thresholds of − 7.58 and − 10.51 dB, respectively.

These data are entirely consistent with those obtained previously (Green and Mason, 1985; Green *et al.*, 1987). Note that Green *et al.* used a two-down, one-up adaptive procedure that estimates the level of the signal that would yield 70.7% correct. In the present study, we employed a three-down, one-up procedure that estimates the level for 79.4% correct. Thus our listeners are, on average, somewhat more sensitive than those who participated in the previous study. Because d' is approximately proportional to the energy of the signal (Green *et al.*, 1987), the difference is equivalent to a change in signal level of about 3.6 dB.

The thresholds reported above are for the simplest spectral manipulation, the addition of an increment to a single component of an otherwise flat, multicomponent background and have established that our listeners' data are typical of those obtained previously. We now turn our attention to a series of experiments in which we measured listeners' sensitivity to more complex spectral manipulations.

III. EXPERIMENT 2A: STEP SPECTRA

In this experiment, the standard was composed of the equal-amplitude, 21-component background. The signal caused a change in the amplitude distribution of the components over the entire frequency range. Two types of change were studied. In the "step-up" condition, the amplitudes of all components above some critical frequency were increased by the same amount while, below that frequency, the amplitudes were decreased by the same amount. At the critical frequency, what we call the "step frequency," the amplitude was left unaltered. In the "step-down" condition, the same procedure was employed but the frequency scale was reversed. Five frequencies, the same as those employed in experiment 1, were chosen as the step frequencies.

A. Results and discussion

Figure 2 displays the average thresholds for the step-up (triangles) and step-down (inverted triangles) signals as a function of the step frequency. It is important to note that, for these signals, as well as for those described below, we have plotted the rms level of the signal *re*: the amplitude of a single component in the background. The solid line represents our calculated thresholds and will be discussed later, as will the data from the other experimental conditions that appear in the left-hand portion of the graph.

The data of Fig. 2 are similar to those of Fig. 1 in that a change in spectral shape appears to be most detectable when

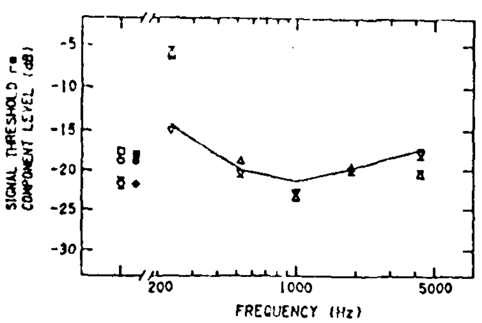


FIG. 2. Data and predictions for experiment 2. Average thresholds for step-up (triangles) and step-down (inverted triangles) are plotted as a function of the frequency of the step. Error bars represent the standard error of the mean computed across the data for all listeners. The solid line represents predictions from the calculation scheme. Obtained and predicted thresholds for tilt-up (squares), tilt-down (circles), and alternation (diamonds) are displayed in the left portion of the graph. Open symbols represent obtained thresholds; solid symbols represent predictions derived from the calculation scheme.

the step occurs in the midfrequencies and is least detectable at the extremes. Thresholds for the step-up and step-down conditions were virtually identical for three of the five frequencies tested. The greatest sensitivity was, once again, obtained at 1 kHz and yielded a threshold of about − 23 dB. At 4256 Hz, the threshold of − 20.5 dB for step-up was slightly lower than that of − 17.9 dB obtained with the step-down signal. A considerably larger discrepancy occurred at 234 Hz where the threshold for step-down was − 15.0 dB, while that for step-up was − 6.0 dB. We will discuss these discrepancies in greater detail after presenting a simple calculation scheme for predicting these complex spectral changes.

Except for the step-up condition at 234 Hz, the thresholds obtained with each step frequency are lower than those obtained for single increments at the same frequency (Fig. 1). An important clue to understanding this difference comes from consideration of the changes in level produced by adding the signal to the standard for each case.

In the case of an increment to a single component of an otherwise flat, multicomponent background, if the level of the signal *re*: the component to which it is added is − 15 dB, a relative increase of 1.42 dB is produced at the frequency of the signal. In the case of a step signal, recall that the components below the step frequency are decreased (increased), while those above the step frequency are increased (decreased). In that case, a − 15-dB signal will cause the components that are incremented to be raised 1.42 dB above the nominal background level and those that are decremented to be lowered by 1.7 dB. The total *difference* in level then would be 3.12 dB, a larger difference than that obtained for the single increment. The reader should note that the random variation of the level of presentation within and across trials encourages the listener to compare the *relative* levels in spectral regions *above and below* the step frequency.

B. The two-channel, level-difference calculator

In attempting to explain quantitatively the step-up/step-down data, we considered a simple calculation scheme that assumes detection of a change in spectral shape occurs *when the addition of the signal to the flat multicomponent background produces a reliable and sufficient difference in level between only two regions of the spectrum*. The single-increment data of Fig. 1 were used as the basis for our calculations.

Assume a level, x_i ($i = 1,2,...,21$), is measured in each of 21 frequency channels (corresponding to the frequencies of the discrete components of our stimuli). Each of these measures is assumed to be contaminated by independent Gaussian noise with mean zero and variance σ_i^2. Thus each x_i is assumed to have a normal distribution with mean m_i and variance σ_i^2. In a given experiment, two channels, i and j, are selected and the decision is based on the *difference* in level $(x_i - x_j)$ between only those two channels on each and every presentation.

When the equal-amplitude standard is presented, $m_i = m$ for all channels and, therefore, $m_i - m_j = 0$ for all i,j. In this case, $(x_i - x_j)$ is drawn from a normal distribution with mean zero and variance $(\sigma_i^2 + \sigma_j^2)$.

When the standard plus the signal is presented, $\Delta = (m_i - m_j)$ and $(x_i - x_j)$ is drawn from a normal distribution with mean Δ and variance $(\sigma_i^2 + \sigma_j^2)$. The detectability of the signal can be expressed as:

$$d'_{ij} = \frac{\Delta}{(\sigma_i^2 + \sigma_j^2)^{0.5}}. \qquad (1)$$

To calculate the d' for a given experimental condition, we assume it is maximized over the two combinatorial 21 or 210 possible pairs (i and j) of channels. For many conditions, this choice is simple. For example, when the signal consists of an increment to a single, say the k th, component, then $m_i = m$ for all $i \neq k$. In this case, $\Delta = 0$ except for $m_k - m_j$. Because the value of $m_k - m_j$ is constant regardless of the channel to which j corresponds, d' is maximized by choosing j so that $(\sigma_k^2 + \sigma_j^2)^{0.5}$ is minimized.

In order to apply the scheme described above, it was necessary to estimate a single parameter—the variance associated with the 1-kHz channel. Because the 1-kHz signal (signal 11) yielded the lowest threshold, the channel containing this frequency must be assumed to have the smallest variance. We denote its standard deviation as σ_j. Therefore, consistent with the strategy for maximizing d' described above, when a single increment occurs at any of the 20 frequencies other than 1 kHz, the level in its channel is compared to that at 1 kHz. Once the value for σ_j (corresponding to the 1000-Hz channel) is chosen, then all the σ_i's corresponding to each of the other frequencies are determined, because, for each other frequency, d'_{ij} and Δ are known.

The value of σ_j was calculated in an iterative fashion by minimizing the rms error of our calculated thresholds for all 21 signal conditions presented in these experiments. The value of d', at threshold, is 1.16, which corresponds to 79.4% correct level of performance estimated by our adaptive task. The value of σ_j, determined by the iteration, was 0.854 dB.

The exact value of this constant does not have a very large effect on the predicted thresholds because σ_j is the smallest standard deviation among all 21 channels. Thus, in calculating a d' value, the other standard deviation, σ_i, tends to dominate [see Eq. (1)].

Using the σ_i^2's so derived, linear interpolation was used to estimate σ_i's for frequencies other than the five for which data were obtained and thresholds for the step stimuli were calculated. The results of these calculations are represented by the solid line in Fig. 2.

As Fig. 2 shows, the results of the calculations fit the data quite closely except for the thresholds obtained with the step-up signal at 4256 and 234 Hz. Our analysis predicts that, for all frequencies, step-up and step-down thresholds should be identical. As mentioned earlier, this was clearly not the case for step frequencies of 4256 and 234 Hz. While the relatively small discrepancy at 4256 Hz may be due to random variation in the data, the large disparity between the step-up and step-down thresholds at 234 Hz cannot be dismissed so easily. This trend, which occurred for all three listeners, also occurred for another group of listeners who were run previously in a pilot experiment. We know of no logical explanation for its existence.

IV. EXPERIMENT 2B: TILTED SPECTRA

Once again, the standard or background consisted of the equal-amplitude, 21-component profile. Two different signals were used; each contained the same 21 components as the standard. The addition of the signal to the standard produced a spectrum that either tilted up or down about the central, 1-kHz, component. In the first case, the addition of the signal to the standard caused the amplitude of each successive component to increase linearly with component number. In the second case, the amplitude of each successive component decreased with component number. By varying the relative level of the signal with respect to the standard, we were able to vary the magnitude or slope of the spectral tilt. We should note that the "tilt" was linear on a pressure or amplitude scale. That is, it was not linear in dB. For the small amplitudes of the signal required for detection by our listeners, however, the tilt was, in fact, essentially linear in dB as well as amplitude. The listener's task was to discriminate the flat from the tilted spectra. Separate thresholds were estimated for both the positive and negative tilts.

A. Results and discussion

The results for the tilt-up and tilt-down stimuli are displayed in Fig. 2 as the open squares and circles, respectively. There is little difference in the thresholds for the two conditions, with tilt-down yielding a slightly lower threshold of -18.8 dB as compared to -17.5 dB obtained for tilt-up. Our calculated thresholds for the tilt stimuli are shown by the open and closed circles. As Fig. 2 indicates, our calculated values of -17.9 and -18.9 dB for the tilt-up and tilt-down conditions, respectively, mirror this trend and are within 0.4 dB of the thresholds obtained.

These tilted spectra provide perhaps the most illustrative demonstration of the operation of the calculation scheme. If one were to ignore the notion of different magnitudes of variance affecting the measurement of level on each channel then, clearly, the two channels that would yield the greatest difference in level would be those at the extremes of the spectrum—200 and 5000 Hz. As shown by Eq. (1), however, the maximal d' involves a trade-off between difference in level and the inherent variability of its measurement. According to the scheme, d' is maximized for the tilt-up spectrum when the difference in level between the 6th (447 Hz) and 21st (5000 Hz) components is employed. For tilt-down, d' is maximized when the 7th (525 Hz) and 21st channels are used. The thresholds reported above, which are within 0.4 dB of those actually obtained, were based on the use of these differences in level.

V. EXPERIMENT 2C: ALTERNATING SPECTRUM

In this experiment, we employed a signal essentially identical to that used by Green and Kidd (1983). The signal was such that, when it was added to the background, it caused successive components to be alternately incremented and decremented.

A. Results and discussion

The average threshold of -21.7 dB obtained for this condition is plotted as the open diamond in Fig. 2. This value is about 6 dB lower than that obtained for the single increment (experiment 1) at 1 kHz. Green and Kidd (1983) also measured a 6-dB difference for these two conditions.

Our calculated threshold of -21.8 dB for the alternating spectrum is plotted as the solid diamond in Fig. 2. Once again, our calculations were based on the assumption that detection of a change in spectral shape occurs when the addition of the signal to the flat, multicomponent background produces a reliable and sufficiently large difference in level between only two regions of the spectrum. For the alternating spectrum, the difference in level between any two adjacent components is identical. Recall, however, that these differences are most detectable in the 1-kHz region. Therefore, the difference in level between only two components in this frequency region is used in calculating the threshold.

Once more the calculation scheme provides an easy way to understand why this result occurs. The level of the signal required for detection of a change in spectral shape is 6 dB less in the case of the alternating spectrum than for a single increment. Again, the essence of the explanation is that the addition of the signal to the background produces simultaneously *both* increments and decrements to the flat background. Hence, a given level of the signal produces a greater dB difference between two regions of the spectrum and a larger d' than does a single increment. What is counterintuitive about this account is that the use of only a single difference in level describes the data despite the opportunity for many such comparisons—ten independent pairs of comparisons for the alternating spectrum.

VI. EXPERIMENT 3: SINUSOIDALLY RIPPLED SPECTRA

In this experiment, the signal produced a sinusoidal change in the amplitudes of the flat profile. That is, the addition of the signal produced what we refer to as a "sinusoidally rippled" spectrum.

A. Procedure

The standard waveform was the 21-component flat spectrum that ranged in frequency from 200–5000 Hz with successive components spaced equally on a logarithmic scale. The addition of the signal produced a power spectrum whose magnitude varied sinusoidally as a function of the logarithm of frequency. We measured thresholds for ripples of 1, 5, and 10 cycles.

Specifically, the "signal" waveform was produced by setting the amplitude of successive components, $a(i)$, according to the following equation:

$$a(i) = \sin[2\pi k(i/M)] \quad i = 1,2,...,M,$$

where i is the number of the component, ranging in this case from 1 to 21, $a(i)$ is the amplitude of the ith component of the signal spectrum, and k is the "frequency" of the ripple. Recall that the first component, $i = 1$, corresponds to a frequency of 200 Hz, and the last component, $i = 21$, corresponds to a frequency of 5000 Hz.

The "depth" of the ripple resulting from the addition of the signal to the standard waveform depends upon the ratio of the amplitudes of the signal components to those of the standard's equal-amplitude components. The depth of the ripple is, of course, monotonically related to the signal-to-standard ratio. We scaled the amplitude of this "signal" and added each component in-phase (respecting sign) to the corresponding component of the flat standard spectrum to produce the change in the spectrum.

It should be noted that, by constructing the signal in the manner described above, the rms of the amplitudes across components is independent of the frequency of the ripple, k, because the 21 values for any set of $a(i)$ are the same; only their order within the set has been changed. If the maximum value for $a(i)$ is one, the rms value is 0.707. We refer to the signal-to-standard ratio as the rms signal amplitude to the amplitude of any component of the standard.

B. Results and discussion

The results are shown in Table I. The data indicate that thresholds are fairly constant at about −23 dB for the three ripple frequencies tested. The 5-cycle ripple appears to have yielded slightly better performance than the two other values. These data are consistent with those obtained earlier by Green *et al.* (1987) and by Bernstein and Green (1987).

The calculated thresholds are quite close to the obtained values. The largest discrepancy occurred at ten ripples where the obtained threshold was −21.8 dB, while the predicted threshold was −24.7 dB. Our calculations exhibit a slight monotonic decrease in threshold as the number of ripples is increased from one to ten. This is due to the fact that, as the number of ripples increases, the spectral "peaks" and "valleys" increase in number and become more closely spaced. Thus the two frequency channels whose levels are compared to produce the maximal d' fall in an increasingly narrow region around 1 kHz where the variances (σ_i^2) are

TABLE I. Obtained and predicted thresholds (dB) for the sinusoidally rippled spectra.

Threshold	Number of ripples 1	5	10
Obtained	−22.4	−25.0	−21.8
Predicted	−22.2	−22.8	−24.7

smallest. This leads to increased sensitivity. Interestingly, in the two previous studies mentioned above, this predicted trend was obtained. However, in the present study, the lowest obtained threshold occurs at five ripples.

VII. GENERAL DISCUSSION

It is important to note at the outset that two restrictions were present in all the experimental conditions we have reported in this article. First, the standard spectrum from which changes in spectral shape were detected was always flat. Second, and probably more important, the standard was defined by a fixed number of components (21). For this set of restricted conditions, all the thresholds can be predicted with good accuracy by calculating the difference in level between *only two* frequency channels. The rms difference between the calculated and obtained thresholds is only about 2.2 dB; that is, the calculations account for slightly greater than 80% of the variance across all thresholds obtained. If the prediction for the step-up signal at 234 Hz is excluded from the analysis, then the rms difference drops to about 1.2 dB.

The problem with this approach is its lack of generality. For example, it does not predict the *decrease* in threshold that is observed for single increments as the number of components in the profile is increased from 3 to about 21 components (Green *et al.*, 1984; Green and Mason, 1985; Bernstein and Green, 1987). The systematic addition of components from 3 to 21 improves detection performance by approximately 13 dB! The standard explanation for this phenomenon is that, because of integration of information across channels, the greater number of components leads to a better estimate of the level of the flat spectrum. Such an argument may be correct, but is completely inconsistent with the calculation scheme we have used in this article. Consider a signal consisting of an increment to the central, 1-kHz, component of the 21-component background. The scheme assumes that such a signal is detected on the basis of the difference in level between the channel containing the 1-kHz component and a single adjacent channel. Removing all but these two components should, in principle, leave detection performance unaffected. The data from previous studies, particularly those of Green *et al.* (1984), demonstrate that this is clearly not the case.

The comparison of these two sets of results leads to a paradox. On the one hand, if the signal is a single increment at 1 kHz, then the results indicate that components far removed from the frequency of the signal enhance the detectability of this increment. On the other hand, when the signal produces changes in the standard that are widely distributed across the spectrum, such as in the case of the alternating or rippled spectra, only two components of the signal appear to contribute to its detection. In short, it appears that the entire spectrum contributes to an estimate of the flat, standard spectra, but only two channels contribute to the detection of the signal.

Finally, we must compare this calculation scheme with other models of profile analysis. There are, unfortunately, few alternatives. Durlach *et al.* (1986), in a recent article, suggest an optimum model to combine information across different frequency channels. This model, by introducing special assumptions, could be reduced to one that considers only the difference in level between two channels. However, their general model is considerably more complicated than the simple calculation scheme presented here.

For the restricted conditions of the present experiment, our simple calculation scheme provides better predictions than more complicated, and generally more efficient, detection procedures. As an example, compare the case of the single increment (experiment 1) to the alternating or rippled spectrum (experiment 2C). According to our calculations, no advantage is gained from this increased number of changes; only the level of a single pair of components is compared. Green (1986) has explored a more efficient way of combining information (vector summation of d') over different frequency channels with little success. Using an optimum combination of 21 statistically independent channels, his predictions were about 7 dB less than the obtained thresholds.

Interestingly, our scheme is similar, in several respects, to Zwicker's "excitation pattern" model for intensity discrimination (for an overview, see Zwicker, 1970). Even a stimulus with a very narrow spectrum, such as a sinusoid, produces excitation in a number of critical bands. Changes in the intensity produce changes in their corresponding "excitation patterns." According to Zwicker's model, the listener detects a change in intensity when the greatest change in excitation level in a single critical band exceeds some threshold value. Zwicker's excitation-pattern model was constructed to account for the data obtained when the listener was required to make *successive* comparisons of intensity. In our case, the change in level between two channels or bands is observed during a single presentation of the stimulus, and must, perforce, be a *simultaneous* comparison of intensity.

Zwicker's model assumes that only the change in level within a single channel is relevant despite the fact that many channels may exhibit a change, a process not unlike that which we have proposed. Florentine and Buus (1981) have suggested an alternative version of Zwicker's model in which successive changes in level within several critical bands are integrated statistically (vector summation of d') and used as the basis for detecting a change in the intensity of the stimulus. As noted previously, Green's application of a similar model to *simultaneous* discriminations greatly overpredicts the data.

In summary, we have presented a simple calculation scheme that predicts the detectability of complex changes in spectral shape. Despite its success for the limited experimental conditions presented in this article, it clearly fails as a general model. Understanding the nature of the experimental restrictions in more detail may suggest ways to modify the calculation scheme.

ACKNOWLEDGMENTS

This research was supported by a grant from the Air Force Office of Scientific Research. The authors wish to thank Dr. Virginia M. Richards and Dr. Timothy G. Forrest for their helpful comments on earlier versions of this manuscript.

Bernstein, L. R., and Green, D. M. (1987). "The profile-analysis bandwidth," J. Acoust. Soc. Am. 81, 1888–1895.

Durlach, N. I., Braida, L. D., and Ito, Y. (1986). "Towards a model for discrimination of broadband signals," J. Acoust. Soc. Am. 80, 63–72.

Florentine, M., and Buus, S. (1981). "An excitation-pattern model for intensity discrimination," J. Acoust. Soc. Am. 70, 1646–1654.

Green, D. M. (1987). *Profile Analysis: Auditory Intensity Discrimination* (Oxford U.P., New York).

Green, D. M. (1986). "Frequency and the detection of spectral shape change," in *Auditory Frequency Selectivity*, edited by B. C. J. Moore and R. D. Patterson (Plenum, London).

Green, D. M., and Kidd, G., Jr. (1983). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. 73, 1260–1265.

Green, D. M., and Mason, C. R. (1985). "Auditory profile analysis: Frequency, phase, and Weber's Law," J. Acoust. Soc. Am. 77, 1155–1161.

Green, D. M., Mason, C. R., and Kidd, G., Jr. (1984). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. 75, 1163–1167.

Green, D. M., Onsan, Z. A., and Forrest, T. G. (1987). "Frequency effects in profile analysis," J. Acoust. Soc. Am. 81, 692–699.

Levitt, H. (1971). "Transformed up–down methods in psychoacoustics," J. Acoust. Soc. Am. 49, 467–477.

Zwicker, E. (1970). "Masking and psychological excitation as consequences of the ear's frequency analysis," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp and G. F. Smoorenburg (Sijthoff, Leiden), pp. 376–396.

From: AUDITORY FREQUENCY SELECTIVITY
Edited by Brian C. J. Moore and Roy D. Patterson
(Plenum Publishing Corporation, 1986)

'FREQUENCY' AND THE DETECTION OF SPECTRAL SHAPE CHANGE

David M. Green

Psychology Department
University of Florida
Gainesville, Florida 32611, U.S.A.

INTRODUCTION

In several recent papers, we have investigated the detection of a change in spectral shape of a complex auditory signal. The discrimination task involves a broadband 'standard' spectrum and some alteration of that spectrum produced by adding a 'signal' to the standard. For most of the experiments, we have used a standard that is composed of a set of equal-amplitude sinusoidal components. The standard spectrum is, therefore, essentially flat. In different experiments, different waveforms have been added to this standard spectrum to create a change in spectral shape, and the detectability of such changes has been measured. A signal commonly used in these experiments was a single sinusoid added in-phase to some component of the standard. Since this signal increases the intensity at only one frequency region, we describe this situation as detecting a 'bump' in an otherwise flat spectrum. One experimental question is whether a bump at one frequency region is easier to hear than a bump at some different frequency region. Also, we might consider more complicated changes in the spectra such as a signal that produces changes in the amplitudes of several components of the standard. How well are such alterations of the acoustic spectra detected, and how is the detectability of these general changes related to the detectability of an increment at one frequency?

FREQUENCY EFFECTS

Before reporting on the detectability of more complicated signals, we must begin by determining how changes in the frequency locus of a single, sinusoidal signal affect the ability to detect a spectral change. The experimental task is as follows. The standard is a 21-component complex composed of equal-amplitude sinusoids spaced equally on a logarithmic frequency scale, ranging from 200 to 5000 Hz. The ratio of successive frequencies in the complex is, therefore, 1.175. Such a "uniform" standard was selected because we may regard the cochlea as a linear receptor array, where distance along the array is roughly proportional to the logarithm of sound frequency. Our uniform standard then produces excitation at roughly equal spatial intervals. We have also tested non-uniform standards with unequal amplitude or non-uniform frequency spacing between components (Kidd, Mason, and Green, 1986), but it appears that the detection of an increment in a single component of the complex is always more difficult for those standards than when the same increment is made in the "uniform" standard.

Before presenting our results, we should make clear one other important detail of the experimental procedure. To insure that the observers are actually listening to a change in the shape of the spectrum, rather than a change in absolute intensity level at some limited frequency region, we randomly vary the overall level of the sound presented. Each trial of the two-alternative forced-choice task contains two sound presentations, the standard and signal-plus-standard. The overall level for each presentation is determined by selecting from a uniform distribution of amplitude levels, typically having a range of 20 dB, in 1 dB steps. The median of this distribution in the present experiment is 60 dB SPL, but the exact value matters little (Mason, Kidd, Hanna, and Green, 1984). Thus, the standard might have components presented at a level of 64 dB, whereas the signal-plus-standard might be presented at an average component level of 52 dB. The correct answer is the less intense sound. The duration of the presentations was about 100 ms, and the onset and offset have short, 5 ms cosine ramps to diminish audible transients.

Figure 1 shows data on the detectability of an increment in a single component of the uniform standard at several different signal frequencies. The abscissa is the frequency of the signal, that is, the frequency of the

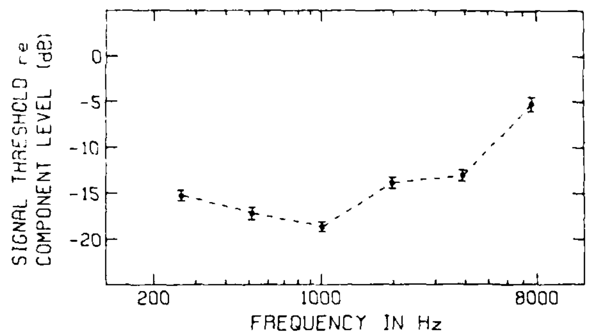


Fig. 1. Threshold for a spectral bump as a function of frequency.

component that is added in-phase to the corresponding component of the standard. The ordinate is the threshold for that signal as determined in a two-alternative forced-choice task. The signal is adapted in level using a two-down, one-up rule and thus estimates a 0.707 probability of being correct. The value plotted as the threshold is the ratio of the signal amplitude to the amplitude of that component of the standard in decibels. If the signal amplitude is one-eighth the amplitude of the component of the standard at threshold, then the threshold value is about -18 dB. This value corresponds to a Weber fraction of 1.03 dB. This value appears to be near the minimum of the data shown in Fig. 1. At frequencies lower or higher than about 1000 Hz, the signal becomes a bit more difficult to hear, but the effects of frequency are not very great. Only at the highest frequency tested, 4256 Hz, is the threshold elevated by more than five dB.

We should also make clear that this minimum at the middle frequency region depends both on the absolute frequency value and the relative position of the increment within the complex spectrum. Other experiments have shown a minimum in the function at lower frequencies for a complex occupying a frequency range 200 to 2000 Hz. The minimum, however, is not solely dependent on context: absolute frequency value is also important. If the complex consists of frequencies ranging from 1000 to 10,000 Hz, the smallest threshold occurs when the signal is presented at the lowest frequency component (Green, Onsan, and Forrest, 1986).

For the remainder of this paper, we will consider more complicated alterations in this standard spectrum. The standard spectrum will always occupy the frequency range from 200 to 5000 Hz. For this frequency range, Fig. 1 shows that changes in such spectra are approximately equal in detectability as long as the frequency of such a change is less than 3000 Hz.

SINUSOIDAL VARIATION IN THE SPECTRA

One way to learn something about the mechanisms responsible for detecting these alterations in the acoustic spectra is to use sinusoidal changes in the power spectra and to vary the frequency at which the sinusoidal variation occurs. Because our spectra are defined at only a finite number of points, the 21 frequencies of the components of our standard spectra, we can alter the frequency variation only over a limited range. Figure 2 shows how we carried out this experimental manipulation. The spectrum displayed at the left of the diagram shows a signal that produces a single cycle of sinusoidal variation over the amplitude of our successive components. Recall that the frequencies of the components are equally spaced along a logarithmic frequency axis. The next spectrum shows two cycles of variation. As the frequency of variation is increased, we finally reach the spectrum shown on the right side of the figure. In this spectrum, successive components alternately increase or decrease in amplitude, and no higher rates of variation can be achieved.

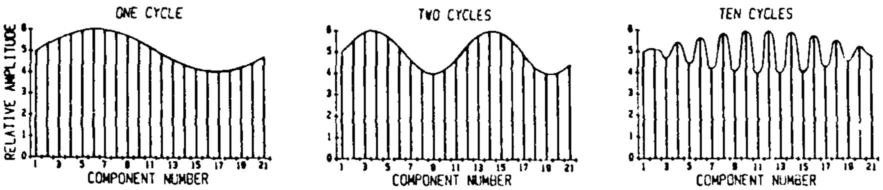


Fig. 2. Three different frequencies, k, of sinusoidal variation.

The following equation expresses how the variation was achieved. Let a[i] be the amplitude of the ith component of the signal spectra, where i ranges from the first component, a frequency of 200 Hz, to the last component, M. In this experiment M=21, and the frequency of the last component is 5000 Hz. We set the amplitude of the ith component as follows:

$$a[i] = \sin(2 * pi * k * i/M) \quad i = 1,2..M \qquad \text{Eq. 1}$$

where k represents the 'frequency' of the variation in the amplitude spectra (k=1,2..10). If we scale the amplitude of this 'signal' and add each component in-phase to the corresponding component of the 'standard' complex, in which each component is equal in amplitude, we produce the change in spectral shape shown in Fig. 2. We speak of this variation in amplitude as producing a ripple in the spectrum and refer to the parameter, k, as the frequency of the ripple.

Before presenting the results, we should make some general comments on this method of constructing the rippled spectrum and how changes in the frequency parameter, k, affect various parameters of the resulting spectrum. First, the root-mean-square (rms) of the amplitudes across the 21 components of the signal is independent of k, the frequency of the ripple. If the maximum value of a[i] is unity, this rms value is 0.707. This value is independent of k because for any frequency of ripple, the 21 values for the set a[i] are the same; only their order is changed. This is true because of the modular nature of the sine function. Thus, the value of any function whose domain is the set of 21 amplitude values, a[i], is independent of k, the frequency of the ripple. Next, we should note that a cosine ripple can be achieved by using the cosine rather than the sine function in Eq. 1. Naturally, the cosine ripple has the same rms amplitude as the sine ripple, and that value is also independent of the frequency of the ripple, k.

To construct the ripple spectrum actually presented to the listeners, we first scale the signal and add it, in-phase, to the components of the standard spectrum. The depth of the resulting ripple depends on the ratio of the amplitude of the signal components to the amplitude of the standard components. The components of the standard spectrum are all equal in amplitude. It is convenient to use the rms amplitude of the 21-signal components as our measure of signal amplitude. We will then refer to the signal-to-standard ratio, meaning the ratio of the rms signal amplitude to the amplitude of any component of the standard. Obviously, the depth of the ripple produced in the resulting spectra is monotonic with this signal-to-standard ratio. The spectra shown in Fig. 2, for example, were constructed with a signal-to-standard ratio of 0.1414.

Finally, we should note that, for any signal-to-standard ratio, the sum of the difference in amplitude between successive components increases monotonically with k. This occurs simply because the difference in amplitudes between successive components approximates the derivative of a[i] (see Eq. 1) which is proportional to k. Thus, the larger the values of k, the more ragged the spectra, and, if we measure raggedness as the rms difference between the amplitude of successive components, it changes from 0.21 to 1.41 as k changes from 1 to 10.

RESULTS

Table 1 presents the data on the threshold for changes in the spectra, using either sine or cosine ripples and different values of k. The threshold for the signal is measured in terms of the signal-to-standard ratio and is nearly constant and independent of k, the frequency of the ripple. Different changes in spectral shape produced by either the sine or cosine version of Eq. 1, for all frequencies of ripple, are equally detectable. With the exception of a single point, the k=9 cosine ripple, the thresholds are within 2 dB of the same value, namely, -24 dB, for all the conditions tested.

This is a most unusual result in sensory psychophysics. In almost all studies of sensory systems, some change in the modulation transfer function with frequency is evident. In this study, the modulation transfer function is essentially flat. Obviously, if the total number of components in the standard signal were increased, then one would reach a point where the high frequency variation was not evident. This point would occur when successive components fell within the same critical band and, hence, could not be resolved. This result would indicate a simple low-pass filter behavior for the system and is consistent with the fact that the frequency resolution for the ear is limited.

Table 1. Average threshold value for different frequencies of ripple. Entry is signal rms to standard ratio in dB.

Frequency of Ripple, k	1	2	3	4	5	6	7	8	9	10	mean
Sine	-24.6		-24.0		-24.6		-23.9		-25.0	-24.6	-24.5
Cosine	-24.1	-23.0	-23.8	-23.2	-24.7	-25.7	-25.7	-23.0	-29.3	-22.8	-24.5

Our results were obtained with 21 components. The ratio of the frequencies of successive components is 1.175. Thus, presumably, each component falls in a separate critical band. If neighboring critical bands were linked with an excitatory center and inhibitory surround, as a simple lateral inhibition model might suggest, then we would expect to see better thresholds at some frequencies of ripple than at others. Our results imply that the local interactions between different critical bands produce no apparent resonance as a function of the frequency of ripple. It is unlikely that finer frequency spacing between the components of the standard will reveal such resonance, because we are nearing the frequency resolution limits of the cochlear array, the critical band. Our spacing is already narrower than one might expect for a lateral inhibition network. If we look at physiological studies of inhibition, then we find that the inhibitory side bands are at least one or two critical bands away from the central components (Sachs and Kiang, 1968).

Before leaving these results, we should also compare the detection of these ripple signals to the detection of a change in amplitude of a single component of the standard, a spectral bump. As the results indicate, the thresholds for the two signals are about 6 dB apart; the threshold for the single bump is -18 dB, whereas the threshold for ripple is -24 dB. This difference is far short of what one might expect on the basis of several models. Perhaps the simplest idea is to assume that a rippled change in a spectrum is easier to detect than a change in amplitude of a single component, because the rippled change allows one the opportunity to combine the output of several independent channels. The standard way to combine such channels, assuming statistical independence, predicts that the combined detectability (d'_c) will equal the square root of the sum of the squares of the detectabilities for each separate channel (d'_i). Since there are 21 separate components, this leads to the expectation that the combined detectability will be

$$(21)^{1/2} * 0.707 = 3.23$$

better than the detectability of the single channel. The factor 0.707 arises because it is the average amplitude of the signal in a rippled spectrum compared to the amplitude of the signal for the single component bump. Since detectability (d') for a single component signal is roughly proportional to signal voltage, not power, that factor translates to a difference of about 10 dB, some 4 dB greater than the empirically determined result. One might, of course, argue that there are not 21 independent channels but some lesser number, even though the spacing between successive components is about two critical bands. To achieve the difference of 6 dB, however, one must assume that only 8 independent channels are combined, a value that seems unbelievably low.

To more fully appreciate the problem inherent in the differences in detectability of these two spectra, let us consider the Weber fraction at each component frequency for the two spectral changes, a spectral bump and a one-cycle ripple. Figure 3 shows a plot of the Weber fraction plotted as a function of the component numbers, from 1 to 21, for these two spectral changes. The two threshold values used in constructing these plots are the average threshold values measured for these two cases.

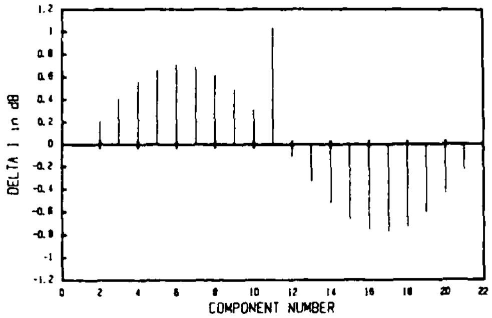


Fig. 3. DELTA I in dB at each component for two spectral changes.

The signal-to-standard level is -18 dB for the single-component increment at the eleventh component. This is about the value seen in Fig. 1 for single, low-frequency increments in a single sinusoid. The rms signal-to-standard level is -24.5 dB for the rippled spectrum. This is the average value found for such rippled spectra; see Table 1. The Weber fraction for the single-component signal is just over one dB. The rippled spectrum produces different Weber fractions ranging from about -0.7 to +0.7 dB. The rms value for the Weber fraction is 0.525 dB for the rippled spectrum. One does not need to have an elaborate theoretical structure to be puzzled by the fact that these two patterns of spectral change are nearly equal in detectability. The combination of two of the larger Weber fractions for the rippled spectrum will easily exceed the Weber fraction for the single increment, and the rippled spectrum will have 19 Weber fractions remaining. One can only conclude that the Weber fractions at the different component frequencies appear to contribute little to the detectability of spectral change in the case of the rippled spectra. Nor does this kind of discrepancy exist only for sinusoidal changes in the spectrum.

Green, Kidd, and Picardi (1983) measured the detectability of sizable changes in a 21-component spectrum. They compared the detection of a single bump at 950 Hz, with a 'downward' or 'upward' step in which all components above or below 950 Hz were increased or decreased in amplitude. They also used a rippled spectrum in which alternate components were increased or decreased in amplitude by the same amount. The average difference in thresholds between the single component change and the change produced at all 21 components was about 5 or 6 dB. In this case, the 0.707 value does not come into play and the expected difference for a 21-component combination is 13 dB. Once again, the obtained difference in threshold is much less than one would expect if the channels could be combined in an optimum statistical manner.

The preceding analysis is premised on the amplitude of the ripple, or some similar quantity, being the important stimulus feature. It is easy to believe, a priori, that it is not the signal amplitude, but rather the difference in amplitude between successive components that is really important in detecting the spectral change. This idea suggests that it is the 'step' created by the elevation of a single component that gives it some advantage in detectability that is not enjoyed by the smoother sinusoidal ripples. But this line of argument is mistaken. First, as we observed above, the difference between successive components in the sinusoidal ripple changed by nearly an order of magnitude as we increased the ripple frequency. The ragged, high-ripple-frequency spectra are no easier to detect than the single cycle of sinusoidal variation. Also, as Green and Kidd found, the total number of steps in the spectra plays essentially no role in determining the detectability of a spectral change. A single step, either upward or downward, was not significantly different in detectability from a spectrum that stepped up and down on successive components - a 20 step spectrum. We simply do not understand enough about the process of detecting a spectral change to account for these discrepancies.

LOW FREQUENCY RIPPLE AND MORE DENSE SPECTRAL PATTERNS

Lastly, we report on an experiment in which we varied the number of components used to define the spectra. A low-frequency ripple was used, two cycles of variation over the range from 200 to 5000 Hz, either sine or cosine. The independent variable of the experiment was the number of components used to define this low-frequency ripple. The components were always of equal amplitude for the standard spectrum. For the sine rippled spectra, the amplitude of the components is given by the Eq. 1. For any value of the parameter, M, the spectra were constructed so that the ratio of frequencies between successive components of the pattern was a constant. The specific value of this ratio can be determined from the formula; the ratio, R, is equal to ten raised to the power 1.3979/(M-1). Thus, for M=81 components, the ratio is 1.041. This means that the nearest components at 1000 Hz are 1041 and 961 Hz. For M=3, the three components of the spectra are 200, 1000, and 5000 Hz. For the three-component pattern, the ripple was simply an elevation in the 1000 Hz component.

Figure 4 shows the data as a function of M, the number of components in the spectra. As can be seen, the threshold for the pattern is elevated

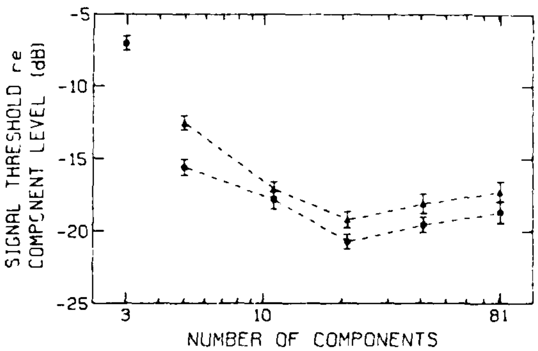


Fig. 4. Threshold for a two cycle ripple as a function of the number of components in the ripple.

if there are fewer than about 11 components in the spectra. As the number of components increases, the threshold decreases and becomes nearly independent of the exact number used. This result can be summarized by saying that the spectral profile becomes better defined as the number of components in the spectrum increases. This result is consistent with some previous studies where we increased the number of components in the spectral pattern (Green, Kidd, and Picardi, 1983, and Green, Mason, and Kidd, 1984). In the previous studies, however, the signal was an increment in a single component of the spectrum. Once the density of components exceeds a certain value, the additional components cause masking at the signal frequency as Green and Mason (1985) have shown. In the present case, the signal is imposed on all the components of the pattern and no masking can occur. In this case, the threshold remains constant and independent of the number of components, once some optimal frequency spacing is exceeded.

SUMMARY

The thresholds for changes in a spectral pattern were measured for several patterns of change. The frequency of sinusoidal variation in the spectral pattern has little, if any, effect on the detectability of such spectral change, at least over the frequency range studied. There is no evidence of any sort of lateral inhibition network. The density of components used to define the sinusoidal pattern plays no role, at least for a very low frequency of variation. A major puzzle of the experimental findings was the relatively small difference in threshold for the detection of a sinusoidal change in the spectrum and the detection of an increase in amplitude of a single component. No explanation of this discrepancy was completely convincing.

ACKNOWLEDGEMENT

This research was supported, in part, by a grant from the National Institute of Health, the Air Force Office of Scientific Research, and by special funds contributed by the Dean of the College of Liberal Arts and Science. The author is particularly indebted to Zekiye A. Onsan and Timothy G. Forrest for their help in collecting the data and the preparation of this report and to Dr. Leslie R. Bernstein who read and commented on an earlier version of this paper.

REFERENCES

Green, D. M., Onsan, Z. A. and Forrest, T. G. (1986). Frequency effects in profile analysis, J. Acoust. Soc. Am., submitted.

Green, D. M. and Mason, C. R. (1985). Auditory profile analysis: Frequency, phase, and Weber's Law, J. Acoust. Soc. Am., 77, 1155-1161.

Green, D. M., Kidd, G., Jr. and Picardi, M. C. (1983). Successive versus simultaneous comparison in auditory intensity discrimination, J. Acoust. Soc. Am., 73, 639-643.

Green, D. M., Mason, C. R. and Kidd, G., Jr. (1984). Profile analysis: Critical bands and duration, J. Acoust. Soc. Am., 73, 1163-1167.

Kidd, G., Jr., Mason, C. R. and Green, D. M. (1986). Auditory profile analysis of irregular sound spectra, J. Acoust. Soc. Am., submitted.

Mason, C. R., Kidd, G., Jr., Hanna, T. E. and Green, D. M. (1984). Profile analysis and level variation, Hearing Res., 13, 269-275.

Sachs, M. B. and Kiang, N. Y. S. (1968). Two-Tone Inhibition in Auditory-Nerve Fibers, J. Acoust. Soc. Am., 43, 1120-1128.

DISCUSSION

YOST

When the sinusoidal ripple added to a complex sound is on linear frequency as it is in rippled noise, listeners are as sensitive to detecting the ripples as they are in the experiments of Green when the ripple is on log frequency (Yost and Hill, 1978 and Bilsen and Ritsma, 1970). In addition, results with rippled noise on linear frequency indicate best sensitivity when the spacing between the spectral peaks is 200 to 500 Hz. Thus, with linear sinusoidal ripple there is evidence for a resonance that may be the result of lateral inhibition.

REFERENCES

Bilsen, F.A. and Ritsma, R.J. (1970). Some parameters influencing the perceptibility of pitch, J. Acoust. Soc. Am., 47, 469-476.

Yost, W.A. and Hill, R. (1978). Strength of pitches associated with ripple noise, J. Acoust. Soc. Am., 64, 485-492.

CARLYON

You rightly conclude from your rippled spectrum experiment that lateral inhibitory networks do not appear to be operating in the conditions of your experiment. However, I would like to point out that a lateral inhibitory model such as Shamma's would not work effectively with your stimuli. This is because your components are not harmonically related (and therefore have varying phases), whereas the speech-like stimuli on which Shamma bases his model are harmonic and have non-random phase. Therefore we can only conclude that lateral inhibitory networks were not influencing your results, and not that such networks do not operate in other circumstances, such as when processing speech.

PROFILE ANALYSIS AND SPEECH PERCEPTION*

David M. Green and Leslie R. Bernstein

Psychology Department, University of Florida, Gainesville, Florida 32611, USA

INTRODUCTION

There is, unfortunately, a wide gulf between research in psychoacoustics and research on speech perception. These differences arise, in part, because of the different objectives of the investigators. Understanding how the auditory system functions and understanding the speech code are different and distinct goals. But there are some areas and topics where one might expect a communality of interest. The topics of auditory perception and the limits of certain basic auditory discrimination processes are both areas that should enjoy mutual interest and concern. But, even here, wide differences are apparent in the way these topics are approached by the speech scientist and by the psychoacoustican. These differences are especially evident in the choice of stimulus materials. The psychoacoustic stimuli are simple; the speech stimuli are complex. The just-noticeable-difference in the frequency or the amplitude of an isolated pure tone appears to have little to do with how we recognize differences between vowels or broadband consonants.

The simplicity of psychoacoustic stimuli is understandable, given the considerable emphasis placed by that discipline on the control of stimulus intensity. Psychoacoustic stimuli are presented at specific sound pressure levels, and considerable time and effort are devoted to ensuring that these levels fall within some small tolerance. A typical limit is some fraction of a decibel, since the Weber fraction for intensity of a single sinusoidal stimulus is about 1 dB. The absolute sound level of speech, on the other hand, is seldom a variable of much concern. Obviously, the sound must be intense enough to ensure that the listener can hear the utterance. But that condition can be met over a large intensity range, and 10 or 20 dB differences between presentation levels may well be regarded as secondary. The reason for such broad limits is simple to explain: the speech code involves a change in spectral composition over time and seldom depends on an absolute intensity level. Relative intensity levels at different regions of the spectrum, the definition of peaks and valleys in the spectrum, and the frequency region where the energy is present are thought to be the most important aspects of the speech code. Indeed, intensity

*The research was supported, in part, by a grant from the National Institute of Health and the Air Force Office of Scientific Research. Our thanks to Dr. Virginia M. Richards whose extensive comments on an earlier draft considerably improved this one.

level per se is generally not part of the speech code; rather, it is used to accent or embellish the utterance.

The preceding observations provide sufficient background for why we find it interesting to study the ability of the human observer to discriminate changes in the shape of the spectrum of a complex auditory stimulus. Such studies, we hope, will provide us with basic information as to how the auditory sense operates and will begin to contribute to our understanding of speech perception, which, after all, is the primary function of the auditory process. In order to understand our research on the discrimination of changes in spectral shape and, in particular, how it differs from the previous studies of intensity discrimination, we must first consider in some detail the intensity discrimination task.

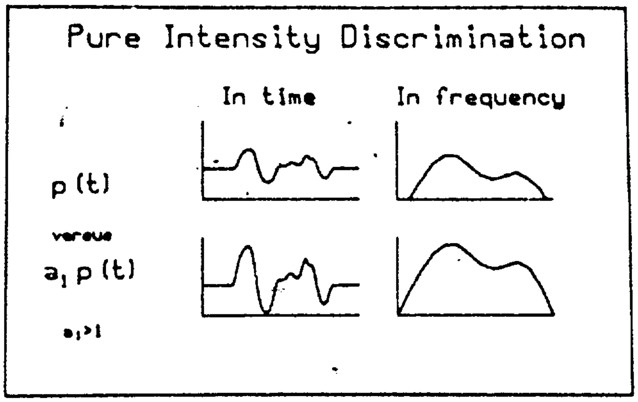


FIGURE 1, Pure intensity discrimination in which the observer discriminates a standard stimulus [p(t)] from a scaled version [a1*p(t)]. In the frequency domain (right side of the figure), the effect of scaling is simply to displace the spectrum along the ordinate. The temporal waveforms (left side) are identical except for the scaling factor.

Let us first consider the simplest intensity discrimination task, what we might call "pure" intensity discrimination. The two sounds used in the discrimination task are either one pressure wave, p(t), or a scaled version of that same wave, a*p(t), where the constant, a, is not unity. In the frequency domain, the two spectra are simply displaced from one another along the ordinate, assuming we have plotted the spectra on a logarithmic intensity scale, such as decibels. The discrimination problem is to select between the two spectra. Pure intensity discrimination, such as that illustrated in Figure 1, may be contrasted with a different task, that of discriminating a change in spectral shape, what we call "profile analysis". The stimuli to be discriminated in this task are illustrated in Figure 2. The two pressure waves, $p_1(t)$ and $p_2(t)$, may be completely unrelated. Since the waveshapes are different, the spectra of the two sounds will also be different, as illustrated in the right hand portion of Figure 2. Although the shapes of the spectra differ, the listener might use differences in intensity at a particular frequency region in order to achieve the discrimination between the two stimuli. Unless some special precautions are taken, there is nothing to prevent the listener from discriminating a change in spectral shape on the basis of some difference in intensity at some particular frequency region. Thus, the experimenter could not, in general, guarantee that the observer's performance in discriminating a change in spectral shape is in any way different from discriminating a change in intensity.

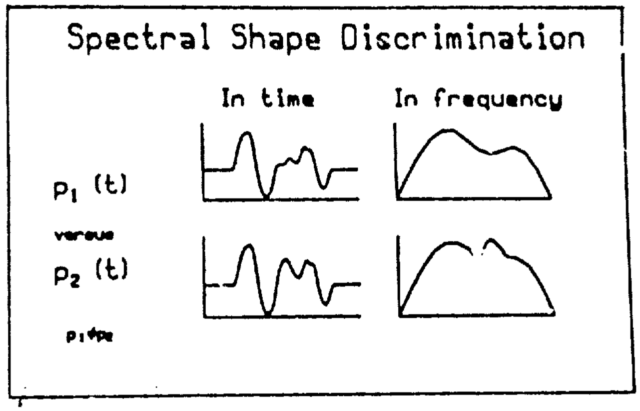


FIGURE 2, Spectral shape discrimination. The stimuli to be discriminated [p1(t),p2(t)] may be completelv unrelated. Thus, in the frequency domain, their spectral shapes differ. The temporal waveforms also differ.

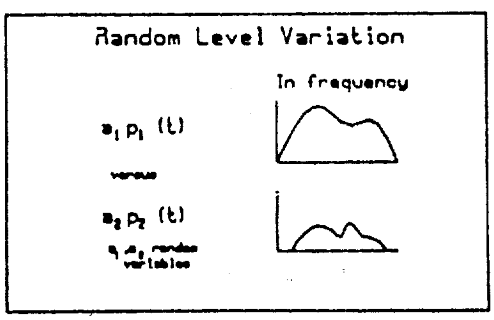


FIGURE 3, Stimuli to be discriminated are scaled by random variables (a1, a2) on each and every presentation to ensure that discrimination is based on spectral shape rather than intensity.

The special experimental manipulation that ensures that shape, not absolute intensity, is the critical cue in the case of spectral shape discrimination is illustrated in Figure 3. It is randomizing the overall intensity level. On each and every presentation of the stimulus, the level at which they are presented is chosen at random. Thus, the scale constants, a_1 and a_2, are random variables as the figure indicates. If the range of these random variables is sufficiently large, the stimuli heard in the discrimination task will clearly differ in intensity, and the observer will be forced to compare some other aspect of the stimuli to distinguish between them. In our case, that difference is the shape of the auditory spectra. The minimal comparison that must be made to achieve such discrimination is that the listener measures the sound levels on two or more different parts of the spectra and simultaneously compares them. The absolute sound level of these measurements is largely irrelevant, because the stimuli change in absolute level on each and every presentation.

The differences in the structure of these two discrimination tasks force the observer to use somewhat different discrimination processes. In pure-intensity discrimination, the listener must construct some estimate of absolute intensity level and either compare two such estimates made at different times or compare a single estimate with some long term standard. In spectral shape discrimination, a simultaneous comparison of two or more spectral regions must be made, and from this comparison an estimate of relative level on any single presentation is largely irrelevant, because it is confounded by the randomization of overall level.

What we would like to do in this paper is review some of our research on this topic and especially emphasize what we have learned about how such spectral comparisons operate. As psychoacousticans, our primary interest is on how the auditory sense works, but we feel these experiments may provide some insight about how complex spectral discriminations are made in speech waveforms.

Procedure and Stimulus Conditions

Before proceeding with a description of the individual experiments, let us outline something about the procedure and stimulus conditions used in the research and why these experimental conditions were chosen. For almost all of the studies, we use a multitonal complex. The stimuli generally cover the speech range, from 200 to 5000 Hz. The frequencies of the individual components are not, however, harmonic, as they are in speech. The tones are chosen so that successive components are equally spaced on a logarithmic frequency axis. Thus, the frequency ratio of successive components is a constant. The reason for choosing logarithmic spacing is as follows. We know the cochlea achieves a rough Fourier analysis of the stimulus in which different places along the basilar membrane are maximally sensitive to different frequencies. Roughly, this linear array is arranged so that equal spatial extent is coded as equal differences in logarithmic frequency. Our tones, therefore, provide a uniform stimulus over the linear receptor surface of the cochlea.

A typical discrimination task involves two stimuli, a "standard" complex and some alteration of the standard complex which we achieve by adding a "signal" to the standard. The signal itself consists of the in-phase addition of energy at one or more components to the corresponding component or components in the standard complex. We use equal-amplitude tones for our standard because the observers learn this standard easily. Thus, little training is needed in order to study various alterations from this standard. We use a two-alternative forced-choice procedure, and adaptively change the level of the signal to estimate the level which would yield 70.7% correct. Overall intensity is typically chosen at random over a 40-dB range in 1 dB steps. The median level is usually about 50 dB SPL per component.

In the studies reported here, the dependent variable is the level of the signal (the size of the increment) re the level of the corresponding component or components in the background. For example, if the level of the signal is equal to the level of the corresponding component(s) of the standard, then we say the signal-to-standard ratio is 0 dB. In that case, the component to which the signal is added would be increased in level by 6 dB. In many studies, the signal is simply an increase in the intensity of a single component. But other changes have been studied as well, such as a variation in the amplitudes of all components of the standard. In the following, we recount some of the things we have learned about the perception of a change in the shape of such a complex auditory spectrum.

Effects of phase

In most of the experiments concerning profile analysis, the phase of each component of the multitonal complex has been chosen at random and the same waveform (except for random variation of level) is presented during each "non-signal" interval. Therefore, the possibility exists that observers may recognize some aspect or aspects of the temporal waveform. If this were true, then discrimination could be based on some alteration of the temporal waveform during the "signal" interval rather than by a change in the spectral shape of the stimulus per se.

Green and Mason (1985) investigated this possibility directly. Multicomponent complexes were generated which consisted of 5, 11, 21, or 43 components spaced logarithmically. In all cases, the frequency of the lowest component was 200 Hz, the highest was 5 kHz. The overall level of the complex was varied randomly over a 40-dB range across presentations with a median level of 45 dB SPL. The signal consisted of an increment to the 1-kHz, central component.

In what Green and Mason termed the "fixed-phase" condition, for each number of components (5, 11, 21, and 43), four different standard waveforms were generated by randomly selecting the phases of each component of the complex. Each of these standards was fixed for a block of trials and signal thresholds were obtained for each of the different randomizations. Note that for these fixed-phase conditions, the same waveform, except for changes in overall level, occurred during each non-signal interval.

In what Green and Mason called the "random-phase" conditions, for each value of the number of components (5, 11, 21, and 43) 88 different standard waveforms were generated by randomly selecting the phase of each component of the complex. On each presentation of every trial, pairs of these 88 waveforms were selected at random (with replacement). Thus, the temporal waveforms generally differed on each presentation. The amplitude or power spectra of the stimuli were, however, identical.

The results are presented in Figure 4. For each value of component number, the open circles represent the thresholds obtained for each of the four randomizations in the fixed-phase condition. The solid triangles represent the data obtained in the random-phase conditions. The results indicate that changing the phase of the individual components and thus the characteristics of the temporal waveform has little, if any, effect on discrimination. This is true whether the same phase is used for a block of trials or if the waveform is chosen at random on each and every presentation. These data are consistent with those obtained by Green, Mason and Kidd

(1984) who had generated their waveforms using a procedure similar to the fixed-phase condition. The form of the function relating threshold to the number of components which compose the multicomponent background will be discussed in detail in a subsequent section.

The inability of changes in the phase of the individual components, and thus changes in the characteristics of the temporal waveform, to affect discrimination supports the view that, in these tasks, observers are, indeed, basing their judgements on changes in the power spectra of the stimuli.

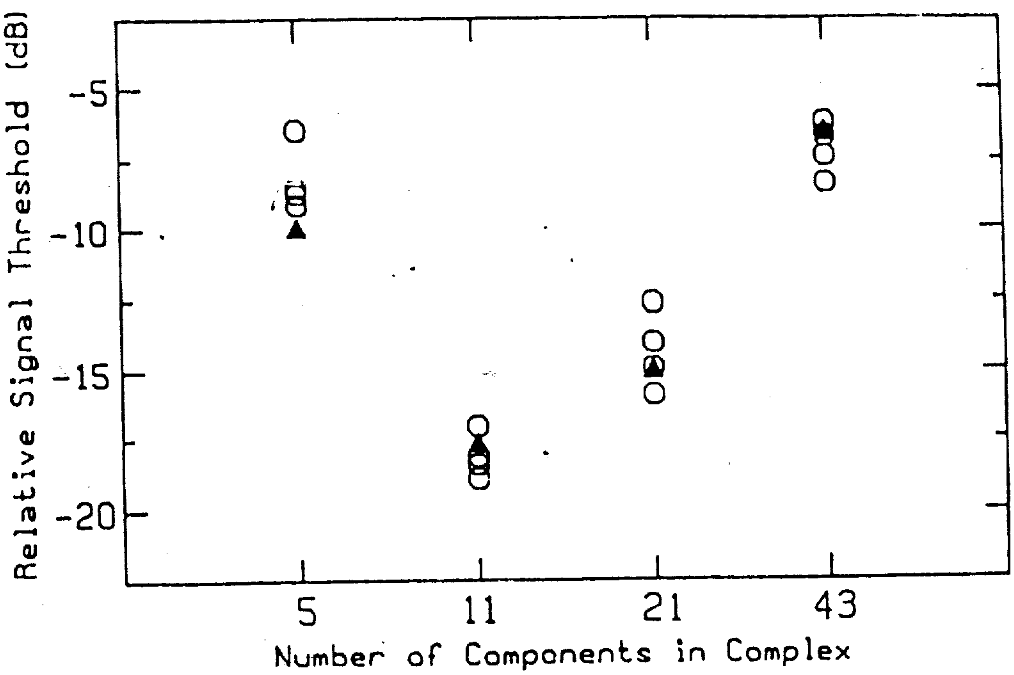


FIGURE 4, Signal threshold (dB) as a function of the frequency of the number of components in the complex. Open circles represent the data obtained for each of the four phase-randomizations when the phase of each component was fixed throughout a block of trials ("fixed-phase" condition). Filled triangles represent the data from the "random-phase" condition in which the phases of the components were chosen at random on each presentation.

Frequency Effects

So far we have demonstrated that the detection of changes in the shape of a complex auditory spectrum is based on changes in the power spectrum of the stimulus; the phase relation among the components is unimportant. The next question we consider is whether the ability to detect a change in the power spectrum is greatly influenced by the frequency region where the change occurs. Consider our complex standard composed of a number of sinusoidal components. Suppose we alter that standard spectrum by increasing the intensity of a single sinusoid. A natural question is--does the frequency locus of the change greatly affect the ability to detect the change? The answer to this question settles an important practical issue-- to what degree are different frequency regions homogeneous? In speech, at least for vowels, the significant spectral changes typically occur within the range of 500 to 2000 Hz. As far as we are aware, there is no claim that small alterations of the spectrum are better detected at one frequency region rather than some other. Thus, we would be surprised to find that the ear's ability to detect a small change in the spectrum differs greatly as a function of frequency.

This question is also of basic interest in psychoacoustics, because it bares on the question of intensity coding and whether or not temporal factors, such as the synchrony of discharge patterns, are utilized as part of the intensity code. Sachs and Young (1979) and Young and Sachs (1979) have demonstrated that 'neural spectograms' based on neural synchrony measures preserve the shape of speech spectra better than those based on firing rate codes. We were, therefore, particularly interested in how well the observers could detect a change in spectral shape at higher frequencies. At the highest frequencies, above 2000 Hz, neural synchrony deteriorates and, if that code were used to signal changes in spectral shape, then the ability to detect such alterations in the acoustic spectrum should also deteriorate. Certainly, differences among vowels are not signaled by changes in the location of higher frequency formats. But, in the case of speech, this frequency limitation may be the result of the production system, that is, the coding system, not the decoding system.

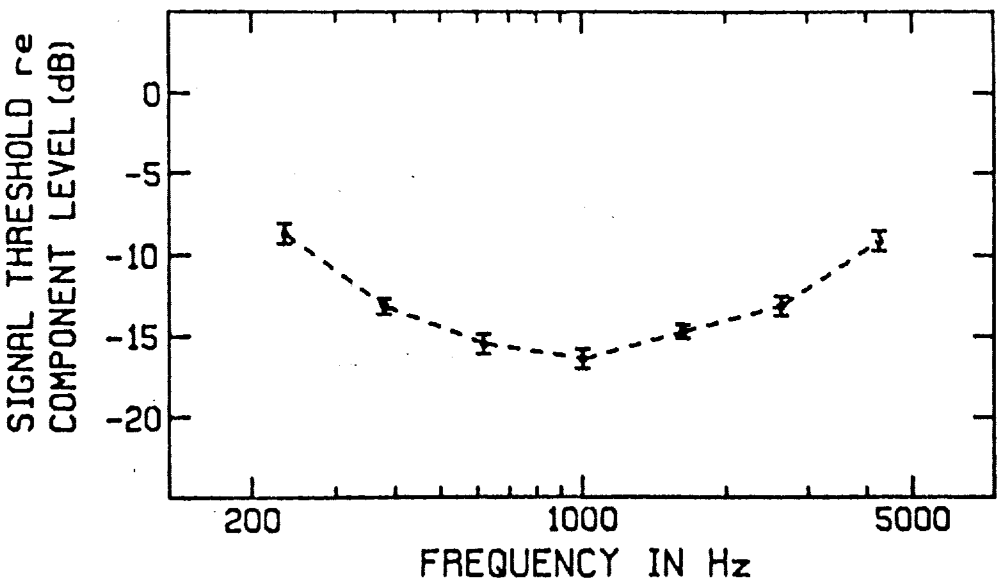


FIGURE 5, Signal threshold (dB) as a function of the frequency of the signal. A twenty-one-component complex was used as the standard. The frequency of the lowest component was 200 Hz; the frequency of the highest component was 5000 Hz. The signal, whose frequency is indicated on the abscissa, was added in-phase to the corresponding component in the complex.

In a previous study, Green and Mason (1985), we had measured how the locus of the frequency changes affects the ability to detect a change in complex spectra. These results suggested that the mid-frequency region, 500 to 2000 Hz, was the best, but variability among the different observers was sizable. Also, those data were taken after a previous experiment in which the signals were in the middle of that range. Although extensive training was given in the later experiment to all the different frequencies tested, it is conceivable that some of the data were influenced by the preceding experiment. In any case, the recent move of our laboratory provided an opportunity to recruit a new set of listeners that were truly naive with respect to the parameter of interest.

The results of our most extensive experiment (Green, Onsan, and Forrest, 1987) on this issue are shown in Figure 5. The standard spectrum was a complex of 21 components, all equal in amplitude and equally spaced in logarithmic frequency. The overall level of the standard was varied over a 20 dB range with the median value of 60 dB. The signal, whose frequency is plotted along the abscissa of the figure, was an increment in the intensity of a single component. The ordinate, like that of Figure 4, is again the signal level re the level of the component to which it was added. The results show that best detection occurs in a frequency range of 300 to 3000 Hz, with only a mild deterioration occurring at the higher and lower frequencies. These results give little support to the idea that neural synchrony is used to estimate intensity level, because, were such the case, there should be a more marked deterioration in the ability to hear a change in the spectrum as a function of frequency.

PROFILE-ANALYSIS AND THE CRITICAL BAND

The evidence presented thus far suggests the detection of a change in spectral shape, or profile-analysis, is a "global" process relying on simultaneous comparisons in two or more regions of the spectrum. An issue of central concern is the width of the spectrum over which these comparisons can be made. If one were to invoke classical "critical-band" notions, which pervade much of psychoacoustic research, it would be expected that only frequencies close to the frequency of the signal could be used in detecting an increment.

Green, Mason, and Kidd (1984) obtained data which address this issue. In their experiment, the signal consisted of an increment to the 1-kHz, central component of a multitonal complex. The multitonal complexes consisted of equal-amplitude, logarithmically-spaced components. In the first condition, what we will refer to as the "range" condition, the standard consisted of a three-component complex. The parameter was the range of frequencies spanned by the standard, that is, the separation in frequency between the two components which flanked the central, 1-kHz component.

In the second condition, what we will call the "range/number" condition, the number of components as well as the range was varied. Additional flanking components were added to the complex resulting in multitonal complexes of 3, 5, 7, 9, and 11 components. These additional components increased the range of frequencies covered by the standard.

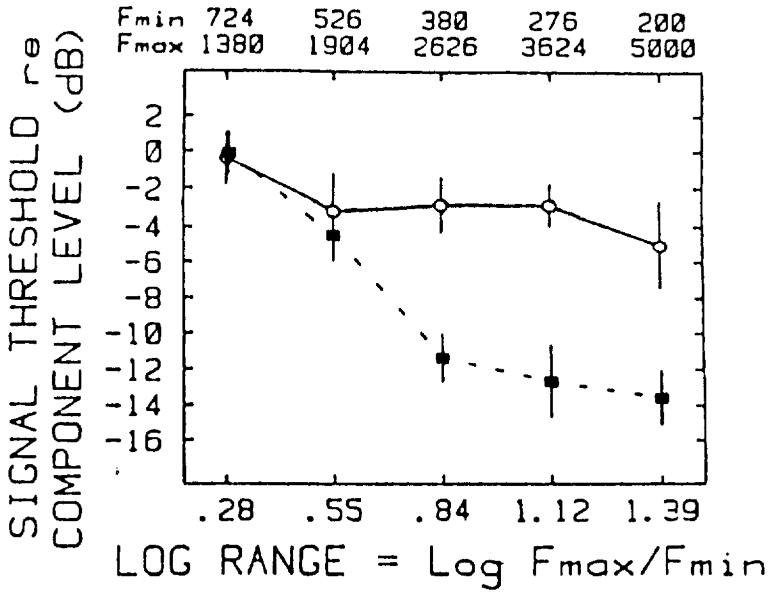


FIGURE 6, Signal threshold (dB) as a function of the logarithm of the ratio of frequencies spanned by the complex. Open circles represent the data obtained from the "range" condition, in which each complex comprised three components. The signal was always added to the central component of the complex, a 1000 Hz component. The numbers at the top of the graph give the frequency of the other two components of the complex. Solid squares represent data obtained from the "range/number" condition in which the number of components in the complex and the range was varied. Again, the signal is an increment in the central component. From the left-most portion of the graph, the squares represent complexes comprising 3, 5, 7, 9, and 11 components respectively.

The results of these two conditions are presented in Figure 6. The abscissa is the logarithm of the ratio of the highest to the lowest component in each complex. The data obtained in the range condition, with the three-component complexes, are plotted as open circles. The solid squares represent the data obtained in the range/number condition when the range of frequencies spanned by the complex and the number of components covaried. Each point is the mean of six estimates of threshold obtained from the three subjects who participated. The error bars represent the mean of the standard error computed for each observer.

Focusing on the data obtained in the range/number condition (solid squares), it is clear that as the number of components is increased, performance improves by 10 dB or more. Although only a small improvement is realized when the number of components increases beyond seven, the data obtained with seven components indicate that tones almost 1.5 octaves away from the central, 1-kHz component (2626 Hz and 380 Hz) have a dramatic effect on performance. This result is in conflict with "critical-band" notions which would predict that energy at frequencies remote from the signal would have little effect on its detection.

The data obtained with the three-component complexes (open circles) also indicate that increasing only the range of the complex improves performance but not to the extent found when the number of components is also increased.

In short, for a given frequency range, performance is improved when the number of components which compose the profile, that is, its density is increased. This result was also obtained by Green, Kidd, and Picardi (1983). Their data showed, in addition, that if the density of components in the complex is great enough, then several components fall very close to the frequency of the signal and detection performance will deteriorate. Such an outcome is explained by simple masking and its existence supports the critical band concept. In such a case, the additional components fall within a critical-band surrounding the frequency of the signal component, and thus an increment to the signal component produces a relatively smaller increase in power in its region of the spectrum.

In summary, the conflict with classical "critical-band" concepts arises because energy at frequencies remote from that of the signal influences performance in these tasks. The data confirm the notion that profile analysis is a global process which relies upon the integration of information across many critical-bands.

Profile Analysis versus Simple Intensity Discrimination

In the concluding section of this paper, we compare the acuity of discriminating a change in the shape of a complex spectrum to the acuity of detecting a change in absolute intensity level. As reviewed in the first section of this paper, one may distinguish two separate processes for comparing intensity in a complex spectrum. The first we called pure-intensity discrimination; this process detects a change in absolute intensity level. The acuity of this process can be measured in tasks where the spectrum of the signal does not change its shape, but is simply altered in level. We have contrasted this process with the detection of a change in the shape of the complex auditory spectrum, what we have called profile analysis. In detecting a change in spectral shape, the process must be one of simultaneous comparisons of intensity levels at different regions of the spectrum, because random variation in the overall level of the spectrum on successive presentations renders the use of absolute level on any presentation an ineffective strategy. For a fixed change in intensity, can one hear that change best in a pure-intensity discrimination task or in a profile task? A clear answer to this question is of some practical importance, because for many naturally-occurring stimuli such as complex speech spectra, both processes are potentially available. Presumably, the observer uses either a combination of the two systems or the more sensitive system alone. To predict performance in a variety of realistic situations, one would have to know the relative sensitivity of the two systems.

Comparison of detection performance in the two situations is, however, complicated by the issue of prior training and experience. The situation is not unlike that of testing the ability of observers to hear some phonemic distinction in a particular language. If one uses a group of subjects whose natural language uses this distinction, then one may expect finer discrimination capacity from that group than from another group of subjects whose native language does not use this distinction. Similarly, we have observed that listeners with a long history of training in pure intensity-discrimination experiments often do poorly when first confronted by a task involving the detection of a change in spectral shape. It is also true that observers who are well-practiced in detecting changes in spectral shape often find detection of simple intensity changes initially difficult. Recently, a well-trained profile observer complained, when asked to discriminate a change in the intensity of a single sinusoid, that the only thing he could listen for was a change in loudness!

A second factor that makes the comparison of the detection performance in the two tasks difficult is that there is more range in the ability of different people to hear simple changes in intensity level than is usually admitted in the literature. The impression that the Weber fraction is nearly constant over individuals is created largely by the use of a very compressive measure of the Weber fraction in dB [$10\log(1+\Delta I/I)$]. One often reads that the Weber fraction is about 1 dB. What is not appreciated is that a change from 0.5 to 1.5 dB corresponds to a 10 dB change on the scale of signal-to-background level which we have commonly used in profile experiments. Individual differences among listeners are sizable. Using our scale of signal-to-component level, then we often find differences of 10 dB among individuals in both pure-intensity discrimination tasks as well as profile tasks.

A final complication is that the observers we use in most of our profile tasks are not a random selection from the population; rather, they are selected on the basis of previous listening performance. Some observers find it extremely difficult to hear the change in shape of a complex spectrum. While they improve with practice, it does not appear likely that they will ever be useful participants in a series of experiments involving the comparison of thresholds obtained in a variety of different experimental conditions. Our usual procedure is to train and test subjects over a period of one or two days (two hours of listening per day) on the detection of an increment in a 1000 Hz tone in an 11- or 21-component complex. For the listener to continue in these experiments, we require that detection performance reach the -10 to -20 dB range at the end of two or three days. In general, we believe that practically all subjects could be trained to reach this level of performance, but if more than three days are required we feel that such observers would require an excessive amount of training throughout the various conditions of the experiment.

A direct comparison of the relative sensitivity of two groups of listeners was recently made by Green and Mason (1985). They compared two groups of observers--five experienced in profile listening, five who were not. The five inexperienced profile listeners had considerable training in tasks that could be classified as pure-intensity discrimination tasks. The thresholds for the ten observers were measured in two detection tasks, a pure intensity-discrimination task and a profile task.

The pure-intensity discrimination task was the detection of a change in the level of a 1000 Hz sinusoid. The sinusoid was fixed in level at 40 dB SPL. The profile task was the detection of the change in the intensity of that same component, but the 1000 Hz component was surrounded by 10 other, equal-amplitude components. We used the familiar, 11-component complex (200-5000 Hz). To make the tasks comparable, the level of all the components was also fixed on each and every presentation, at 40 dB SPL. The ratio of the frequencies between successive components of the complex was approximately 1.38. Thus, the two neighbours to the 1000 Hz components had frequencies of 1379 and 724 Hz. The signal duration in both tasks was 100 msec. The thresholds were estimated from the mean of 6 runs of 50 adaptive trials (two down-one up). Table 5-1 presents the thresholds estimated in the two tasks for the ten observers.

Table 5-1

Entry is the relative signal threshold in dB
(standard error of estimate)

Observers	Single Sinusoid	Profile	Diff (SS-P)
Profile Experienced			
1	-10.5 (1.4)	-18.6 (1.7)	8.1.
2	-6.4 (2.0)	-13.6 (0.6)	7.2.
3	-12.0 (0.8)	-18.5 (1.3)	6.5
4	-11.2 (1.3)	-15.8 (1.2)	4.6
5	-18.0 (1.5)	-22.7 (2.3)	4.7
mean	-11.6 (1.4)	-17.8 (1.4)	6.2
Profile Inexperienced			
6	-20.0 (1.6)	-10.9 (2.2)	-9.1
7	-13.2 (2.0)	-12.3 (1.6)	-0.9
8	-19.7 (1.0)	-9.2 (1.3)	-10.5
9	-14.0 (1.0)	-10.0 (1.6)	-4.
10	-17.4 (0.8)	-20.2 (1.4)	+2.8
mean	-16.9 (1.6)	-12.5 (1.6)	-4.3

As can be seen in the table, there is almost a perfect interaction between thresholds in the two tasks and previous training. The best average detection performance is about -17 dB for both groups, but it occurs for different conditions. For the experienced profile listeners, it occurs in the profile conditions. For the inexperienced profile listeners, it occurs in the single sinusoid condition. The average difference between performance on the favored and unfavored task is also very similar in the two groups, about 5 dB. The pattern of interaction between past listening experience and the two detection tasks is reflected by nearly every individual observer with one singular exception (Observer 10). That observer, whose performance level is good on both tasks, is somewhat better on the profile task, despite the lack of previous experience. Note the range of thresholds obtained for either group within each task. Such differences among individuals are typical.

Presumably, with enough training, both groups would improve on the unfamiliar task, but, unfortunately, we have no firm data to support that conjecture. Informally, we tried to improve the performance of the inexperienced profile listeners in the profile tasks, but their thresholds, after an additional 2000 trials, did not improve very much. We are still uncertain about how best to interpret this result. The interaction present in the data reflects either a difference in training or real individual difference among observers. It may be that differences in past experience can simply not be overcome by a few thousand trials. One could argue that it is like trying to hear an acoustic distinction that is not used in one's native language. Alternatively, it is possible that there are simply two different types of observers. One type good at discriminating changes in absolute intensity, another good at discriminating changes at spectral shape.

While it is unlikely that a random sample of ten individuals would divide so perfectly, one cannot claim that the profile group was a completely random sample. As described previously, some preliminary testing was completed before selecting this group and such tests could indeed have biased the group to be good 'profile' listeners. The difference in their performance in the two tasks is reasonably uniform over all the observers experienced in profile listening. The group inexperienced in profile listening was probably a more random selection from the general population, and their results are more mixed. Observers 7 and 10 show little difference in their performance in the two tasks (their difference scores are -0.9 and +2.8 dB, respectively). Whether there are really different types of observers or simply differences in past experience remains a fascinating, but unsettled, question.

REFERENCES

1. Green, D.M., & Mason, C.R. (1985). Auditory profile analysis: Frequency, phase and Weber's Law. Journal of the Acoustical Society of America, 77, 1155-1161.
2. Green, D.M., Kidd, G., Jr., and Picardi, M.C. (1983). Successive versus simultaneous comparison in auditory intensity discrimination. Journal of the Acoutical Society of America, 73, 639-643.
3. Green, D.M., Mason, C.R. and Kidd, G., Jr. (1984). Profile analysis: Critical bands and duration. Journal of the Acoustical Society of America, 75, 1163-1167.
4. Green, D.M., Onsan, Z.A., & Forrest, T.G. (1987). Frequency effects in profile analysis. Journal of the Acoustical Society of America, in press.
5. Sachs, M.B., & Young, E.D. (1979). Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate. Journal of the Acoustical Society of America, 66, 470-479.

6. Young, E.D., & Sachs, M.B. (1979). Representation of steady-state vowels in the temporal aspects of the discharge patterns of populations of auditory-nerve fibers. Journal of the Acoustical Society of America, 66, 1381-1403.

Detection of amplitude modulation and gaps in noise

David M. Green and Timothy G. Forrest

Psychoacoustics Laboratory, Department of Psychology, University of Florida, Gainesville, Florida, 32611, USA

Introduction

Nearly ten years ago, Viemeister (1979) proposed a model to explain how human observers detect amplitude modulation of a noise signal. In this paper, we modify that model slightly and extend its application to the detection of brief temporal gaps in noise. Measuring gap detection has become an increasingly popular way of assessing temporal properties of the auditory system (Penner, 1977; Fitzgibbons, 1983; Green, 1985). Simulation using a modified model provides excellent predictions of the thresholds obtained with partially filled gaps as well as their psychometric functions. Despite this success, the computer simulations indicate that the gap threshold is not strongly influenced by the two major variables of the model, namely, filter bandwidth and integration time. Thus, the applicability of this model to the understanding of hearing impairment remains unclear.

The paper begins with a brief description of the detection tasks and Viemeister's original model. Next, modifications of the original model are described and our reasons for their adoption are explained. The applicability of this modified model to partially filled noise gaps is then described. Finally, we explore the model's predictions about how gap threshold should change as a function of the two major parameters of the model. We begin with a description of the task.

Detection task

All the detection data we will discuss were based on a choice between two stimulus alternatives. One stimulus alternative was an uninterrupted or continuous noise which we refer to as the standard. The other stimulus alternative was noise which was interrupted or altered in amplitude in some fashion. One such alteration was a temporal gap in the noise process, illustrated at the top left (see Fig. 1). The second alteration was amplitude modulation of the noise waveform, illustrated in the bottom left of Fig. 1. These two alterations define two detection tasks called gap detection and

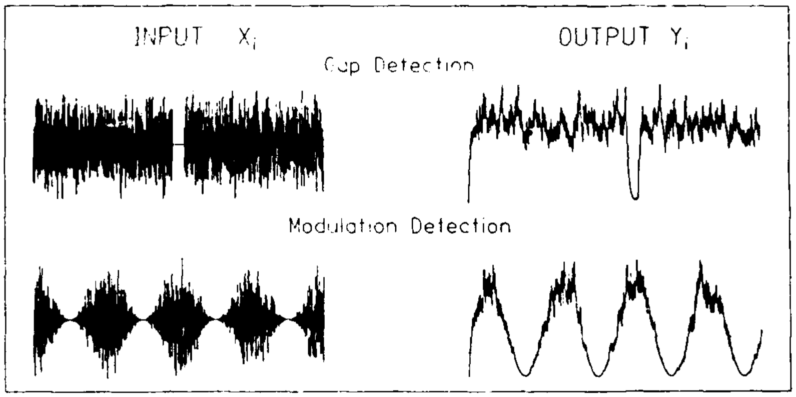


Figure 1. Input and output waveforms for broadband noises with a gap (top) or sinusoidally-amplitude modulated noise (bottom).

modulation detection. Disregard the right column of Fig. 1, it represents the output of a model that we will discuss shortly. All the data reported in this paper were obtained from one of these two detection tasks. The following is a brief summary of the details of the stimulus.

Two-alternative forced-choice procedures were used to estimate all thresholds. The standard was either continuously present or was presented for 500 ms and occurred in one of the two stimulus intervals. The signal was presented in the other interval of the forced-choice task. A two-down one-up adaptive procedure was used to estimate threshold. We generally report the mean of three listeners' thresholds.

Broadband noise was computer generated and presented over 12-bit D to A converters at a rate of 25,000 points per second, and lowpass filtered at 10,000 Hz. More details of the stimulus procedure can be found in Forrest and Green (1987).

The simulations reported in this paper were obtained by programming the model to act as a human observer. The input to the model was the digital version of the signals heard by the observers. The model analyzed two sound buffers corresponding to the two intervals of the forced-choice procedure (standard and signal) and made a choice between the two. The signal level was adjusted adaptively to estimate a threshold for the model, just as it had been for the human observers. All computations were carried out on a microcomputer (IBM AT or equivalent). The human observers took about 5 minutes to run 50 trials and to obtain a threshold estimate with about 10 to 15 reversals. The model took about 3 to 10 times longer.

Gap detection procedure

The two stimulus alternatives of the gap detection task were: 1) the standard waveform or 2) the signal waveform. The standard waveform was a 500-ms burst of noise of constant average level. The signal waveform was also a 500-ms burst of noise, except each sample from the temporal center of the noise burst was scaled by an amount equal to (1-k). The duration of this attenuated segment was called the noise gap. The task problem was to discriminate between these two alternatives. If the value of $k = 0.5$, then the noise was reduced in level by 6 dB for the duration of the gap. If $k = 1.0$, then the noise was fully cancelled during the gap, a condition typical of that used in most studies of gap detection.

An atypical part of the procedure used in these experiments was that we randomized the level of each sound as it was presented. The level was selected from a rectangular distribution with a range of 10 dB. The median level of the noise was about 65 dB overall, 25 dB spectrum level. We randomized the presentation level because the introduction of the gap reduces the total energy in the noise waveform by an amount that depends on the size of the gap and the amount of the attenuation. Randomization discourages observers from trying to use overall level as a detection cue, and makes the primary detection cue one of temporal variation of noise level present within the half-second sample.

Modulation detection procedure

The two stimulus alternatives of the modulation detection procedure were: 1) the standard waveform or 2) a signal waveform. The standard waveform was a continuous noise presented throughout the 50 trials of the two-alternative forced-choice task. The signal waveform, 500 ms in duration, was a set of noise sample multiplied by a sinusoid. Thus, the signal waveform may be described as

$$s(t) = [1 + m\cos(2\pi f_m t)]\cdot n(t) \qquad (1)$$

where $n(t)$ is the unmodulated or standard noise waveform, f_m is the rate of modulation in Hertz, and m is the degree of modulation.

A somewhat atypical part of the procedure used in these experiments was that the signal waveform was adjusted in power, so that the average power of the signal and standard waveforms were equated. When noise is amplitude modulated, the modulated waveform is increased in average power by an amount that depends on the degree of modulation. The expected, or average, power of $s(t)$, $\langle S\rangle$, is given by

$$\langle S\rangle = (1+m^2/2)\cdot\langle N\rangle \qquad (2)$$

where $\langle N\rangle$ is the expected power of the standard noise. Thus, unless $m = 0$, a potential cue for detecting the presence of modulation is the increase in overall power caused by amplitude modulating the noise. This potential

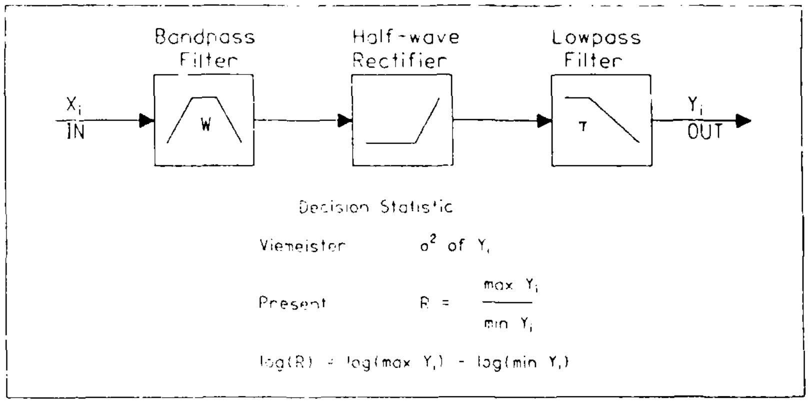


Figure 2. Viemeister's three stage model. Samples of the input waveforms (X_i) pass through an initial bandpass filter with bandwidth (W), then through a half-wave rectifier and a simple lowpass filter. A decision statistic (bottom) is computed from samples of the output waveform (Y_i).

artifact was appreciated by Viemeister (1979) and is responsible for the asymptotic value of the threshold for high modulation frequencies, where m is large (see Fig. 7 of Viemeister, 1979). In all our experiments, we scaled the signal waveform, so that the expected power of the signal was exactly $\langle N \rangle$, independent of the value of m.

Viemeister's MTF model

The three stages of Viemeister's MTF model are shown in Fig. 2. The incoming signal is first bandpass filtered, with a filter of bandwidth, W. Next the signal is half-wave rectified. Finally, the output of the rectifier is smoothed with a simple one-stage (6 dB per octave) lowpass filter. The output of the model, Y_i, provides the input to a decision stage that selects which of the two stimulus alternatives is correct. In Viemeister's original work, he used the variance of the Y values. Such a decision statistic will be larger, on average, when the noise has been amplitude modulated, as shown in Fig. 1. The figure shows the output of the model to either a gap (top) or amplitude modulated (bottom) input.

Our modification of the original model consisted of changing the decision statistic. Instead of using the variance of the output number, Y_i, we used the ratio, R, of the maximum of Y_i to the minimum of Y_i observed during the bulk of the observation interval. Specifically, we considered all values of Y_i that occurred after the initial three time-constants of the 500-ms observation interval. After determining the maximum and minimum values of Y_i that occurred during that interval, we computed the ratio, R. The decision rule

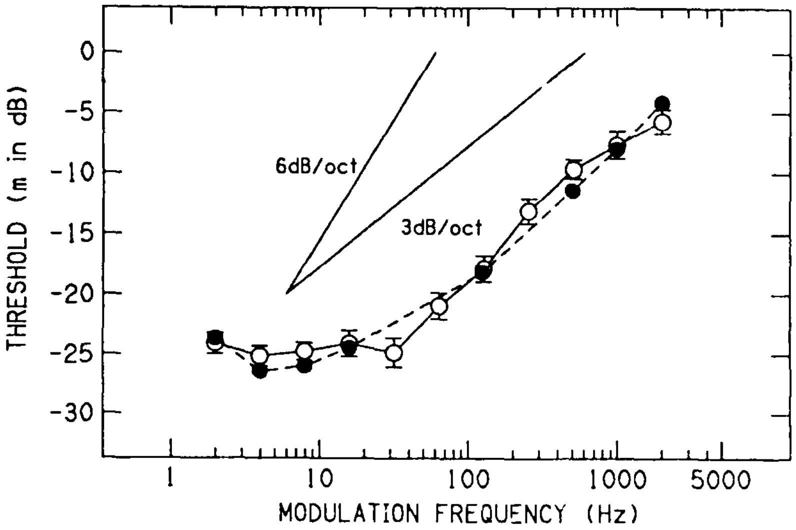


Figure 3. Temporal modulation transfer function for human subjects (solid symbols) and a three-stage model simulation (open symbols). Model parameters are W = 4000 Hz and τ = 3 ms.

assumed that the stimulus with the larger value of R was the signal. Such a decision rule is somewhat inefficient compared to the calculation of the variance of Y_i because it is based on only two of the many samples of Y_i present during an observation interval. We adopted this rule for several reasons.

First, we wanted a decision statistic that would function sensibly even with changes in overall level of the sound, as was true in our gap detection experiment. The variance statistic would change systematically with overall level, whereas the expected value of R is independent of the overall sound level. Second, the R statistic produces the 3-dB-per-octave slope observed in psychophysical data for large modulation rates, as we shall now demonstrate.

Modulation detection data

Figure 3 shows the average data of our observers, solid symbols, as well as the data from our simulation, open points. For these simulations, the first stage bandwidth was 4000 Hz and the time constant of the lowpass filter was 3 ms. As can be seen, the data appear to fall along a 3 dB per octave line at high frequencies. This is somewhat unexpected, since the final lowpass filter has an attenuation skirt of 6 dB per octave. We believe that this shallow slope arises because, as the frequency of modulation increases, a greater number of

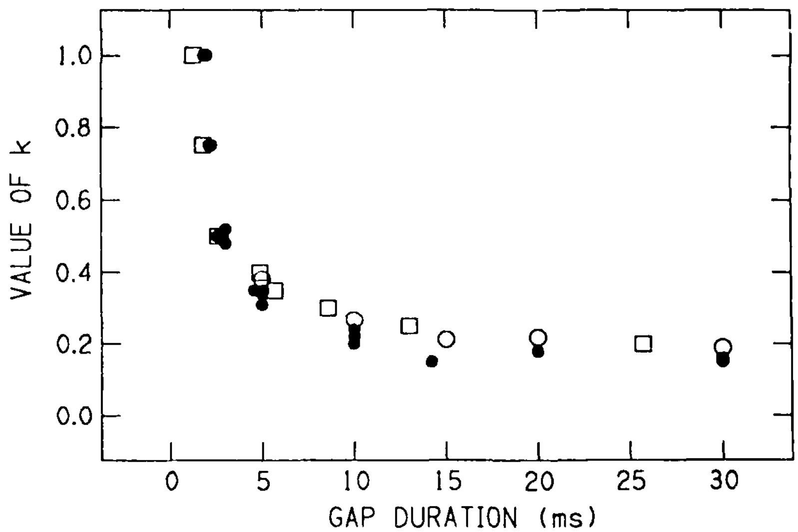


Figure 4. Gap detection data for partially filled gaps in noise for human subjects (solid symbols) and the three-stage model (open points). Model parameters are W = 4000 Hz and τ = 3 ms.

potential maxima and minima are produced. This increase in the number of extrema increases the number of potential signals observed in any observation interval and ameliorates the rapid fall in sensitivity that one would expect from the attenuation produced by the lowpass filter. A problem with this explanation is that sensitivity below the cut off frequency is constant and independent of modulation rate (Fig. 3). Viemeister's model, which uses the variance as the decision statistic, will produce a 6 dB per octave decline at high frequencies, if the noise samples are equalized in overall power, as we have shown elsewhere (Forrest and Green, 1987). Viemeister's original data do not show the 6 dB per octave slope because the noise was not equalized in power (ibid, see Fig. 9). We are now in a position to compare the computer simulations and data obtained from human observers in the gap experiments.

Gap detection data

Figure 4 shows the data for the detection of partially filled gaps in noise. The figure presents data obtained from human observers, solid points, and computer simulations for corresponding conditions, open symbols; again, the parameters of the simulation were W = 4000 Hz and τ = 3 ms. As can be seen, the fit of the model to the data is very satisfactory. Thus, a single model produces reasonably good predictions of both the gap (Fig. 4) and the modulation detection data (Fig. 3).

One of the striking characteristics of gap thresholds is their stability. This stability arises in part because of the steepness of the psychometric function. For k = 1.0 (complete cancellation) the psychometric function for both the human observers and the model has a range of only 1 ms! For smaller values of k (0.50 and 0.35) the psychometric functions are less steep and human detection performance is actually superior to that obtained with the model. We are presently exploring a variety of different ideas on how to alter the computer simulation so that its predictions will better mimic the human data. The urgency of this project will be apparent when we consider how gap detection thresholds change with the other parameters of the model, namely, bandwidth and integration time.

Gap detection as a function of W and τ

Because estimates of gap thresholds are stable, they are often touted as an excellent way to assess temporal parameters of hearing-impaired listeners. Clinical investigators have often reported that gap thresholds for hearing-impaired listeners are appreciably greater than those obtained from normal listeners. Gap thresholds for the hearing-impaired may be in the 10- to 20-ms range when measured with k = 1.0 in broadband noise (Formby, personal communication). Other experiments show that gap thresholds increase systematically, for both normal and hearing-impaired listeners, as the bandwidth of the noise is decreased (Fitzgibbons and Wightman, 1982; Fitzgibbons, 1983; Shailer and Moore, 1983; Buus and Florentine, 1985). The gap thresholds found in these experiments are factors of 3 to 5 larger than the typical gap threshold value of 2-3 ms found with most normal observers in broadband noise. We naturally wondered if we could alter the parameters of the computer model to produce data that would simulate such large gap thresholds. The following table shows how the simulated gap threshold changes as the two major parameters of the model are altered. The columns

Table 1. Effect of bandwidth (W) and tau (τ) on simulated gap detection thresholds. Entry is the mean value of a silent gap needed to achieve about 70% correct in an adaptive task. The standard deviation of these estimates is about 17% of the mean.

	Bandwidth (Hz)			
Tau(ms)	400	800	1600	3200
1.5	6.20	3.92	2.01	1.31
3.0	6.16	3.76	2.28	1.40
6.0	7.57	4.67	2.73	1.95
12.0	7.62	5.21	3.17	2.55

show the variation in bandwidth and the rows are different time-constant values.

The first thing to note about the table are the relatively small changes in gap threshold caused by alteration of the time-constant value. A change of nearly an order of magnitude in the value of the time-constant increases the gap threshold by only a factor of two, and then only at the largest bandwidth. For the smaller values of bandwidths, which are needed to produce any significant increase in gap threshold, the changes with τ are minuscule. To achieve gap thresholds approaching the measured values of 10 to 20 ms, we would need to assume totally unreasonable parameter values for the model.

We are now exploring how the introduction of internal noise at different stages of the model will alter this situation. At present, we can only say that the model gives reasonably good predictions for normal hearing listeners, but is not particularly useful in interpreting the results obtained from listeners with abnormally large gap thresholds. Indeed, changes in the major temporal parameter of the model, τ, produce surprisingly little variation in the size of the gap threshold.

Acknowledgments

This research was partially supported by grants from the National Science Foundation and the Air Force Office of Scientific Research. We thank Dr. Craig Formby and Dr. Virginia Richards for their comments on earlier versions of this article.

References

Buus, S. and Florentine M. (1985). "Gap detection in normal and impaired listeners: the effect of level and frequency," in *Time Resolution in Auditory Systems*, edited by Axel Michelsen (Springer, Berlin), pp. 159-179.

Fitzgibbons, P. J. (1983). "Temporal gap detection in noise as a function of frequency, bandwidth, and level," J.Acoust.Soc.Am. 74, 67-72.

Fitzgibbons, P. J., and Wightman, F. L. (1983). "Gap detection in normal and hearing-impaired listeners," J.Acoust.Soc.Am. 72, 761-765.

Forrest, T. G., and Green, D. M. (1987). "Detection of partially filled gaps in noise and the temporal modulation transfer function," J.Acoust.Soc.Am. 82, 1933-1943.

Green, D. M. (1985). "Temporal factors in psychoacoustics," in *Time Resolution in Auditory Systems*, edited by Axel Michelsen (Springer, Berlin), pp. 122-140.

Penner, M. J. (1977). "Detection of temporal gaps in noise as a measure of the decay of auditory sensation," J.Acoust.Soc.Am. 61, 552-557.

Shailer, M. J., and Moore, B. C. J. (1983). "Gap detection as a function of frequency, bandwidth, and level," J.Acoust.Soc.Am. 74, 467-473.

Viemeister, N. F. (1979). "Temporal modulation transfer functions based upon modulation thresholds," J.Acoust.Soc.Am. 66, 1364-1380.

Comments

Patterson:

It is difficult to understand the motivation behind a model of hearing that ignores cochlear filtering and assumes that a wideband signal is passed directly through to the broad lowpass process that determines the MTF. Would it not be better to assume that bandwidth effects are the result of combining the outputs of sets of adjacent auditory filters, and thereby make the model a lot more realistic?

Reply by Green:

I have always believed that the primary test of a theory was its ability to predict the data, not whether the assumed processes mimicked our current understanding of how the system functions on a more molecular level. Indeed, it seems to me that the theory or model must be simpler than the process it hopes to explain at least in some respects, otherwise, it achieves no economy of understanding. The present model has only two free parameters (bandwidth and integration time) and predicts with fair accuracy the results of normal-hearing listeners in two experimental situations, see Fig.3 and Fig.4. But, as Table 1 indicates, it does not provide much understanding of hearing-impaired listeners.

For that reason we have been exploring a model much like that described in your comment. That model, a series of parallel, narrow-band channel, raises the issue of how the output of these several independent channels are combined. This is not an issue where more molecular investigations provide much insight. We are presently exploring a number of different decision rules but, as yet, have nothing to report.

Hearing Research, 32 (1988) 147–164
Elsevier

HRR 01032

Profile analysis: Detecting dynamic spectral changes

David M. Green and Quang T. Nguyen

Psychoacoustics Laboratory, Psychology Department, University of Florida, Gainesville, Florida, U.S.A.

(Received 27 August 1987; accepted 17 November 1987)

This paper explores how amplitude modulation influences the detection of changes in spectral shape. We generally used a complex of 21 equal-amplitude components, the lowest frequency was 200 Hz, the highest 5000 Hz, with equal logarithmic spacing between components. The signal was an increase in level of one or more components of the complex. The overall level of the sound varied randomly over a 20-dB range. Three experiments are reported. In the first, we determined how the modulation of a single-frequency component influenced the detection of amplitude change at that region. In the second experiment, the signal was an alteration of the entire spectrum and that alteration was subjected to various forms of amplitude modulation. In neither experiment did modulation generally increase the detectability of the signal. Finally, in the third experiment, we determined the effects of modulating the 'signal' and 'nonsignal' parts of the spectrum in different relative phases. The results of this experiment showed that the relative phase is important only for modulation rates slower than about 40 Hz. For faster rates, the temporal structure of the spectrum is unimportant. Thus, for modulation rates above 40 Hz, only the power spectrum of the stimulus is critical.

Psychoacoustics; Intensity discrimination; Amplitude modulation; Profile analysis

Introduction

The salience of any component of a stationary, multicomponent complex is greatly enhanced by a brief change in practically any parameter of that component – amplitude, phase, or frequency. A component or set of components, previously unnoticed in a stationary spectrum, suddenly becomes prominent when those components are briefly varied in amplitude. Such amplitude variation, as long as it differs from the remainder of the spectrum, is an effective way to highlight or segregate a particular set of components. Although the effects are obvious from casual observation, there is remarkably little experimental evidence documenting these claims. Viemeister (1980) has published one of the few systematic investigations of these phenomena. Summerfield et al. (1987) have exploited this idea to produce vowel-like spectra. McAdams (1984) has used both amplitude and frequency modulation to produce a number of dramatic demonstrations that clearly establish the saliency of temporal variation in a multicomponent complex.

In our recent work, we have been exploring the detection of amplitude changes in a multicomponent complex—what we call profile analysis (Green, 1987; Bernstein et al., 1987). We wondered how amplitude modulation of the altered components might affect the detection of such changes. In the first two experiments, we explore how such amplitude variation influences the detection of changes in spectral shape for such multicomponent complexes. In the first experiment, a single component of a 21-component complex was changed in level. We wished to determine whether amplitude variation of this component would affect the detection of a small change in its average amplitude. In the second experiment, the signal was a more complex change in the shape of the spectrum, for example, alternate components were increased or decreased in amplitude. Again, we wished to determine how such amplitude variation

Correspondence to: D.M. Green, Psychoacoustics Laboratory, Psychology Department, University of Florida, Gainesville, FL 32611, U.S.A.

would influence the detection of such spectral changes.

The third experiment explored another facet of how amplitude variation of the spectrum may affect the ability to detect a change in spectral shape. In most profile experiments, the overall level of the sounds is varied on each and every presentation in an effort to insure that the primary detection cue is a change of relative level at different spectral loci, rather than a change in absolute level at some single-frequency location. Suppose the spectral change occurs at a single-frequency locus. The detection of such change requires the observer to compare the level at the signal part of the spectrum (where the amplitude change may occur) with some other nonsignal part (where no amplitude change can occur). We have described this comparison as a simultaneous comparison of level, to distinguish it from the successive comparison of level common to many other psychophysical tasks. Suppose the signal and nonsignal parts of the spectrum are now sinusoidally modulated in amplitude at the same frequency but with different relative phases. We can, for example, make the signal and nonsignal parts of the spectrum wax and wane, either in-phase or out-of-phase. How will this relative phase influence detection of the signal, and how will the threshold for in-phase and out-of-phase conditions vary with the frequency of modulation? These are the main questions of the last set of experiments.

General methods and procedures

All stimuli were generated using an IBM-XT microcomputer and a Data Translation DT-2801A interface board for D to A conversion. The stimuli were all digitally computed and played over the 12-bit D to A converter at a sample rate of 25 000 points per second. All stimuli were lowpass filtered; the filtered output was 3 dB down at 6000 Hz, and 20 dB at 6750 Hz.

The observers listened binaurally to Sennheiser model HD414SL earphones; both phones driven in-phase. The listeners were seated in sound-treated rooms and responded using the computer's keyboard. Events within the trial cycle, as well as feedback after each response, were signaled via the computer's monitor. Three listeners served in each of the three experiments. One of the authors, QN, participated in all three experiments. The other observers were students at the University. One of the students observed in all three experiments. A second observed in only the first and second experiments and was replaced by a third student who observed only in the third experiment. The observers listened for about two hours daily. The student observers were paid an hourly rate plus a bonus upon the completion of the experiment.

In all the detection tasks, the signal was an alteration in the spectrum of some multicomponent signal, which we call the 'standard'. Typically, the standard was a 21-component complex; in one experiment a 7-component complex was used. The components of the standard were always equal in amplitude and their frequencies were spaced equally on a logarithmic scale that extended from 200 Hz–5000 Hz. A signal consisted of an increase in the intensity of one or more components of the standard. The sound was presented for 500 ms. The onsets and offsets were shaped by a 20-ms cosine-squared envelope. In different experiments, the signal component and/or some components of the standard were amplitude modulated. Later, we will describe these dynamic conditions in more detail and will also describe how the signal was measured.

An adaptive two-alternative forced-choice procedure was used to estimate the signal threshold. The adaptive procedure (3-down/1-up) estimates a signal level corresponding to a probability of being correct equal to 0.794. The initial step size for the signal was 4 dB and was reduced to 2 dB after three reversals. The threshold value of the signal was estimated from the average of the last even number of reversals in a 50-trial block, excluding the first three reversals. An average of about 11 reversals was obtained. We report the signal threshold as the level of the signal re the amplitude of the component of the standard to which it is added. Thus, a typical signal threshold of about −18 dB corresponds to an increment at the signal component of 1 dB re the other components of the standard. At the start of each adaptive run, the signal level was equal to the level of the component of the standard to which it was added, that is, the component after addition of the signal was 6 dB re the level of the other components of the standard. The overall intensity level of the sound presented in each interval was chosen randomly from a rectangular distribution, ranging from −10 dB to 10 dB re the median level. This procedure discourages the listener from using absolute intensity as a cue for detecting the signal. The median level of each component of the standard was 62 dB SPL; the overall level of the 21-component standard was, therefore, 75 dB SPL.

Single-component signal

Stimuli

In our first experiment, the signal to be detected was an alteration in the amplitude of a single component of a 21-component complex. The frequency of the 'signal' component was either: a low frequency (235 Hz), the second component of the complex; a medium frequency (1000 Hz), the middle component of the complex; or a high frequency (4257 Hz), the penultimate component of the 21-component complex. We compare the threshold for three 'dynamic' conditions with a 'stationary' condition in which the signal component was constant in amplitude during the entire observation interval and somewhat larger in amplitude than the amplitude of the 20 other components. This 'stationary' condition was the one used in most previous experiments on profile analysis. We know from previous research (Green and Mason, 1985; Green et al., 1987) that such a change in spectral shape will be most easily detected when the 'middle' component of the spectrum is the signal component. In the dynamic conditions, three experimental manipulations were used to assess how temporal variations at the signal component influenced the detectability of spectral alterations of the standard. All three conditions involved some form of amplitude modulation and the frequency of modulation, f_m, was the major independent variable, ranging from 2–640 Hz. The three conditions are precisely described in the appendix, where equations used to generate the waveforms are presented. Fig. 1 provides a graphic presentation of the three basic manipulations. Let us describe the three dynamic conditions depicted in Fig. 1.

The three panels of Fig. 1 illustrate the three experimental (dynamic) conditions. We used a logarithmic scale of amplitude and frequency in order to roughly represent the effective spectra of these stimuli, as processed in the auditory periphery. The standard 21-component complex is represented in the panel to the left and the effects of adding the signal to the complex are represented in the panel on the right. To improve the clarity of the figure, we show only five components of the 21-component complex: the first, sixth, eleventh, sixteenth, and twenty-first. The signal is always represented as affecting the middle, or 1000-Hz, component. In constructing these illustrations, we have selected a single value, 50 Hz, for the frequency of modulation, the main independent variable of the dynamic conditions. We also selected a signal level of −18 dB as a representative threshold value for the chief dependent variable; the signal is about 12% of the amplitude of any other component in the complex. When the signal component is added (in-phase) to the component of the standard, it makes that component about 1 dB larger. (In terms of the equations of the Appendix, $20\log(a_j/a) = -18$ dB.)

In Condition 1, the standard (top left, Fig. 1) is a set of 21 sinusoidal components, all equal in amplitude. The signal is sinusoidally modulated in amplitude $[a_j(1-\cos(2\pi f_m t))]$ before being added to the corresponding component of the complex. The resulting spectrum (top right, Fig. 1) is a set of 20, stationary, equal-amplitude components and one amplitude-modulated component, what we call the 'signal component', f_s. In the long-term amplitude spectrum, there is a small increase in level of the signal component as well as the addition of two sidebands located in frequency at $f_s + f_m$ and $f_s - f_m$. The sidebands are low in amplitude (24 dB below the signal component). At low rates of modulation, this condition is virtually identical to the stationary profile condition, except that the amplitude of the signal component varies slightly, from $(a + 2a_j)$ to (a), at the frequency of the modulation.

In Condition 2, the standard (middle left, Fig. 1) is a set of 21 sinusoidal components, all equal in amplitude, but the signal component, f_s, is also 100% amplitude modulated $[1-\cos(2\pi f_m t)]$. This produces two sidebands at frequencies $f_s + f_m$ and

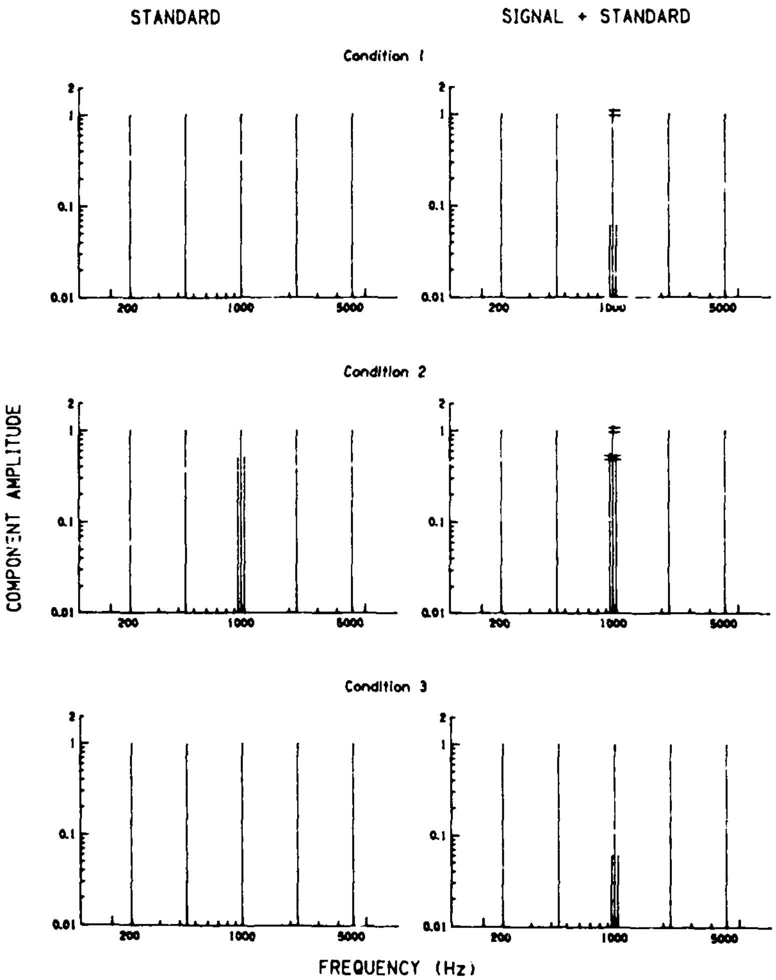


Fig. 1. Schematic of the spectra used in the three dynamic conditions of the single-component-signal experiment. Only five components of the 21-component spectra are represented. The left panels are the spectra of the standard alone. The right panels are the spectra of the signal-plus-standard. The amplitude modulation condition is represented by sidebands near the central component. The signal amplitudes depicted are typical of those measured in the experiment.

$f_s - f_m$. The level of the sidebands is 6 dB less than the level of the component which is modulated. For this condition, the signal is also amplitude modulated $[a_j(1 - \cos(2\pi f_m t))]$ before being added to the corresponding component of the complex (see middle right, Fig. 1). At low rates of modulation, this is similar to the stationary profile condition, except that the amplitude of the signal component, f_s, varies considerably, from $(2a + 2a_j)$ to (0), at the frequency of modulation. In the long-term amplitude spectrum, the signal produces a small increase in the amplitude of the signal component, as well as two relatively large sidebands located symmetrically around that component.

In Condition 3, the standard (bottom left, Fig. 1) is again a set of 21 sinusoidal components, all equal in amplitude. For this condition, the signal is multiplied by a sinusoidal component $[a_j \cos(2\pi f_m t)]$ (so-called 'suppressed carrier modulation') before being added to the corresponding component of the complex. The resulting spectrum (bottom right, Fig. 1) is similar to the standard spectrum, but with two small sidebands located at frequencies $f_s + f_m$ and $f_s - f_m$. At low rates of modulation, the signal component waxes $(a + a_j)$ and wanes $(a - a_j)$ and has an average amplitude of a, equal to that of the signal component of the standard spectrum. The sidebands are low in amplitude (24 dB down from the average amplitude of the signal component). The long-term amplitude spectrum is the same as the flat spectra of the standard, except for the presence of some slight energy in the two sidebands located symmetrically around the signal frequency.

Results and discussion

The results are presented in Fig. 2. For each experimental condition, the threshold for the signal was determined at nine different modulation frequencies, ranging from 2–640 Hz. The signal threshold is plotted as a function of modulation frequency for each of three signal frequencies in the separate panels of the figure. The threshold reported is the average over three observers. Although the observers often differ from each other, the major trends are well represented in the average data. The error bars represent the standard error of the mean threshold (twelve threshold determinations from each of the three observers).

The threshold of the signal is expressed as 20 log a_j/a (see Appendix). In the modulation conditions, the amplitude of the signal component will wax and wane, but no attempt has been made to calculate an 'equivalent' signal value. The three different modulation conditions are coded in the figure: a square, a triangle, and a circle represent the thresholds obtained for Conditions 1, 2, and 3, respectively. The solid horizontal line represents the threshold obtained in the stationary profile condition, and its value is indicated below the line. In this stationary condition, the signal component is a when the standard is presented and $a + a_j$ when signal-plus-standard is present (see appendix). Because the overall level of the spectrum randomly varies over a 20-dB range, the observer must listen for some change in the shape of the spectrum, either dynamic or steady state, rather than any change in absolute amplitude.

Before beginning the discussion of the results, one should recall that the sounds were presented for a duration of one-half second. Thus, for the low modulation frequencies, only a few cycles of the modulation were presented. The relatively short observation interval probably inflates some of the threshold values for these lower modulation rates.

The first general observation is that the dynamic cues do not greatly improve the detectability of the signal; in fact, they make the detection of a spectral change more difficult. Practically all the data points from the various dynamic conditions lie above the horizontal line which represents the average threshold for the stationary condition. Consider in particular Condition 2. At least for the lower modulation rates, the signal component is very salient because it is always 100% amplitude modulated (both in the standard and signal-plus-standard). This amplitude fluctuation makes the signal frequency clearly evident in listening to the multicomponent complexes. Despite this increased saliency, the spectral alteration is generally harder to hear when it varies in time than when presented at a fixed level.

We should qualify this observation by noting that the relation between the threshold obtained in the dynamic conditions relative to the stationary threshold appears to depend on signal frequency. At the low and middle signal frequencies, the dynamic cues are largely detrimental. At the highest signal frequency, they do not notably impair the detection of the signal and may slightly aid detectability. In fact, for the 4257-Hz signal condition, temporal variation in signal amplitude produces somewhat lower thresholds (Conditions 1 and 3), at least for the slower and moderate rates of modulation. This result was observed for all three listeners and represents the only conditions where the dynamic presentation made the change in spectral shape easier to hear than the simple steady-state condition. These conclusions obviously depend on how the signal is measured.

One might argue that the dynamic thresholds should be 'corrected' by the change in power of the signal caused by the modulations, rather than simply $20\log(a_j/a)$. If this procedure is followed, then Conditions 1 and 2 (amplitude modulation) should be increased by 1.7 dB $[20\log(1 + \frac{1}{2})]$. Condition 3 (suppressed carrier modulation) de-

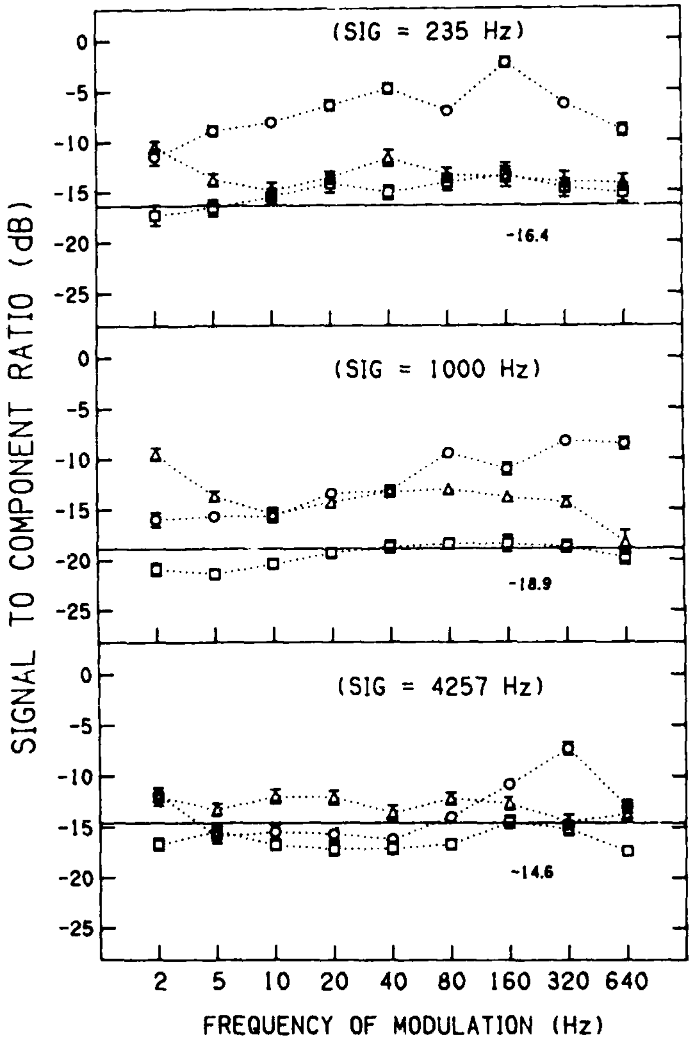


Fig. 2. Results of the single-component-signal experiment. The ordinate is the threshold for the signal, $20\log(a_j/a)$ (see Appendix). The abscissa is the frequency of modulation, f_m. The three signal frequencies are plotted in separate panels. The data for the various dynamic conditions are coded: square-Condition 1, triangle-Condition 2, circle-Condition 3. The horizontal line (and number) in each panel is the threshold for the stationary profile condition. Each data point is the average threshold for the three observers.

creases the signal level by 3 dB ($20 \log\frac{1}{2}$). Thus, one can mentally increase the thresholds of Conditions 1 and 2 by 1.7 dB and decrease the threshold determined in Condition 3 by 3 dB. If this 'correction' is made, then Condition 3 is slightly better than the steady-state condition at the highest signal frequency, about the same at 1000 Hz, and remains much poorer at the lowest frequency. Condition 2, when the data are translated upward, would never produce a threshold that is lower than that obtained in the stationary condition. Condition 1, when corrected upward, would always be worse at the lowest signal frequency, mostly worse at the middle signal frequency, and equal to or slightly better than the stationary condition at the highest signal frequency. Unfortunately, until we learn more, there is no way to know the best way to measure the stimulus.

We should comment on the threshold levels obtained in Condition 3. In this condition, the task is to detect amplitude modulation at the signal component, with a number of other non-modulated components present in the spectrum. If the nonmodulated components were not present, then this task would be equivalent to detecting amplitude modulation of a single sinusoid. For this condition, our measure of threshold, $20\log(a_j/a)$, is equal to $20\log(m)$, where m is the degree of amplitude modulation – $\{[1 + m\cos(2\pi f_m t)]\cos(2\pi f_c t)\}$. Zwicker (1959) has measured the threshold for the detection of amplitude modulation with a single sinusoid. He finds, at low modulation rates, a threshold of about −25 dB for a low frequency carrier, about −30 dB for a 1000-Hz carrier, and about −35 dB for a high-frequency carrier. For all three carrier frequencies, he finds that the thresholds increase about 10 dB as the rate of modulation increases from 2–80 Hz. If our observers are as sensitive to modulation as Zwicker's, then one is forced to conclude that the nonmodulated components of Condition 3 exercise a considerable amount of masking. Therefore, in a separate experiment, we measured the threshold of our listeners for detecting amplitude modulation of a single-frequency component in isolation. If the sounds are gated for one-half second durations, the thresholds are about 5–10 dB lower than those found in Condition 3. If the tone is continuously present, as it was in Zwicker's study, then the thresholds for detecting a half-second of modulation are another 5–10 dB lower, depending on carrier frequency and modulation rate. At the lowest modulation rates, our listeners are also less sensitive than Zwicker's, by between 2 and 7 dB, the largest discrepancies being for the 1000 Hz carrier. In short, some of the differences between Condition 3 and existing data on the detection of amplitude modulation of single sinusoids are due to the presence of the other, nonmodulated, components. But a larger part of the difference is due to individual differences in the observers and the mode of stimulus presentation—gated versus continuous listening. We are presently studying these differences in greater detail.

We can also compare the results obtained in the different conditions in an attempt to infer the most effective cues to the presence of the signal. In the stationary profile condition, only the amplitude of the signal components is different from the amplitude of the other components of the multicomponent standard. Let us call this a 'relative-level' cue. The detectability of this cue is of central concern for most profile-analysis experiments. Condition 1 adds a dynamic component to this simple situation, because both the modulation of the signal and the relative-level cue are present. Condition 2 makes the dynamic cue less important, since amplitude modulation is present in the standard stimulus, as well as the standard-plus-signal. Condition 3, at least at the higher rates of modulation, makes the relative-level cue unimportant, since, at the highest rates, the signal component is equal in amplitude to all other components of the complex.

Comparison of these conditions, however, does not reveal any general rules about the relative effectiveness of different cues, at least for all three signal frequencies. Compared to the stationary profile condition, simple amplitude modulation of the signal (Condition 1) does not greatly change the detectability of the signal. Condition 2, which reduces the importance of the modulation cue, raises the signal threshold considerably for the middle frequency signal (1000 Hz), but does not greatly affect the highest and lowest signal frequency. Condition 3, which emphasizes the purely temporal cue, at low rates of modulation, is generally ineffective at the lowest and middle signal frequency. At the highest modulation rates, Condition 3 generally becomes ineffective. In those conditions where this is not true, for example the 640-Hz modulation at the 4257-Hz signal frequency, detection of the sidebands probably has occurred.

Multiple-component signals

Stimuli

In our second experiment, we altered the amplitude of many components of the 21-component complex. Once again a stationary profile condition was compared with two dynamic conditions. In this stationary condition, alternate components of

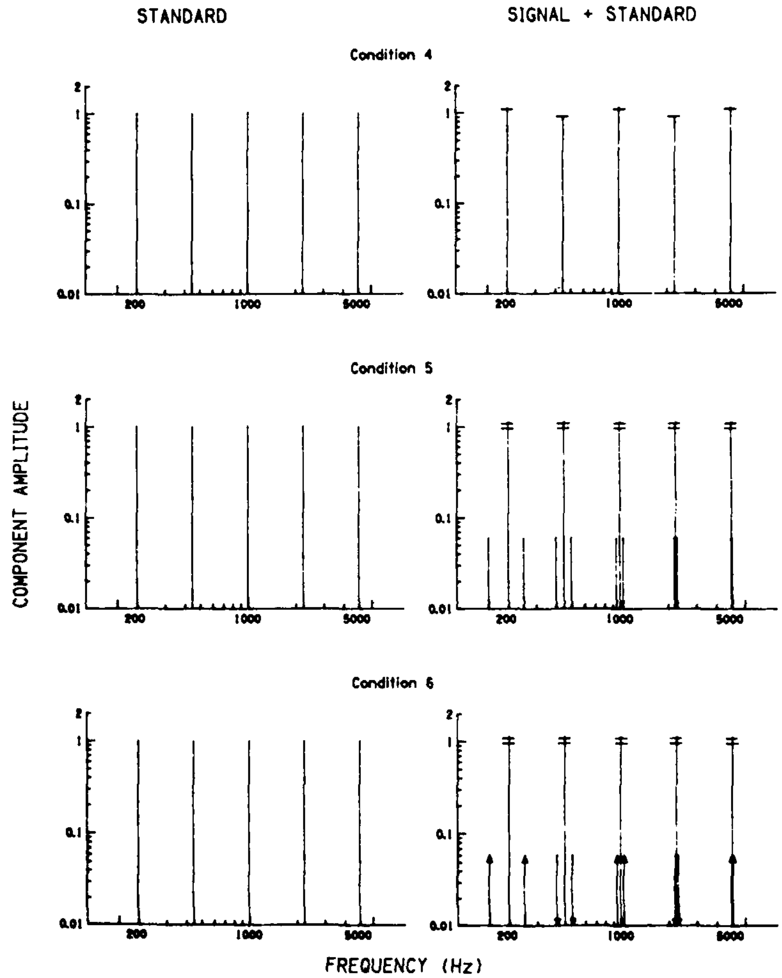


Fig. 3. Schematic of the spectra used in the multiple-component-signal experiment. Only five components of the 21-component spectra are represented. The left panels are the spectra of the standard alone. The right panels are the spectra of the signal-plus-standard. The amplitude modulation condition is represented by sidebands near the central component. The arrows on the sidebands represent the phase of the signal modulation re the standard modulation (see Appendix). The signal amplitudes indicated are typical of those found in the experiment.

the spectrum were increased or decreased in amplitude. Thus, the listener's task was to discriminate a flat multicomponent spectrum from a 'serrated' spectrum, one in which the amplitudes of successive components were alternately higher and lower than the average. We know from previous research that listeners can detect such alternations over the entire spectrum more easily than when the same amplitude change occurs at only a single component. In the dynamic conditions, two experimental manipulations were employed. The two conditions are precisely described in the Appendix, where equations used to generate the waveforms are presented. Fig. 3 provides a graphic presentation of the stationary condition and the two dynamic stimuli.

Once more, we use a logarithmic scale of amplitude and frequency. The standard 21-component complex is represented in the panel to the left and the effects of adding the signal to the complex are represented in the panel on the right. Once more, only five components of the full 21-component complex are represented: the first, sixth, eleventh, sixteenth, and twenty-first components. The value of frequency of modulation is chosen to be 50 Hz, and the threshold value of the signal −18 dB.

In Condition 4, the stationary condition, the standard (top left, Fig. 3) is a set of 21 sinusoidal components, all equal in amplitude. The addition of the signal to the standard causes the alternate components of the standard to be increased and decreased in level (top right, Fig. 3). Thus, the listener must discriminate a flat from an alternating spectrum. No amplitude modulation of either the standard or signal is present in this condition.

In Condition 5, the standard (middle left, Fig. 3) is a set of 21 sinusoidal components, all equal in amplitude. For this condition, the signal is simply an amplitude modulation version of the standard $[k(1 + \cos(2\pi f_m t)]$. Adding that signal to the standard (in-phase) produces the resulting spectrum (middle right, Fig. 3). The result is the standard spectrum with a slight amplitude modulation of all components. In the long-term spectrum, there is a small increase in the amplitude of all components and two sidebands located about each component frequency, plus and minus the modulation frequency, f_m. The sidebands are low in amplitude (−24 dB re the carrier). At low rates of modulation, this condition is similar to the stationary profile condition, except that the amplitude of all components varies from $(a + 2k)$ to (a), at the frequency of modulation. Note that the components of the complex are spaced at equal distances on a logarithmic frequency scale and the sidebands are a constant linear distance from these components. Thus, the relative logarithmic separation of the sidebands changes systematically with frequency as illustrated in the figure.

In Condition 6, the standard (bottom left, Fig. 3) is a set of 21 sinusoidal components, all equal in amplitude. For this condition, the signal is again an amplitude-modulated version of the standard. However, the alternate components of this complex are modulated in different phases. Even components are modulated by $[k(1 + \cos(2\pi f_m t)]$, whereas odd components are modulated by $[k(1 - \cos(2\pi f_m t)]$. The resulting long-term spectrum (bottom right, Fig. 3) is a small increase in the amplitude of the signal component and two sidebands for each component that alternate in-phase (shown by the arrows pointing up or down for alternate components). At low rates of modulation, the 21 components of the complex wax $(a + 2k)$ and wane (a) in amplitude, with alternate components out-of-phase.

Results and discussion

The results are presented in Fig. 4. For each experimental condition, the threshold for the signal was determined at nine different modulation frequencies, ranging from 2–640 Hz. The signal threshold, average over three observers, is plotted as a function of modulation frequency in the separate panels of the figure. Again, the major trends are well represented in the average data. The error bars represent the standard error of the mean threshold (twelve threshold determinations from each of the three observers).

The threshold of the signal is expressed as the level of the amplitude of the signal component, k (see Appendix), re the level of the standard com-

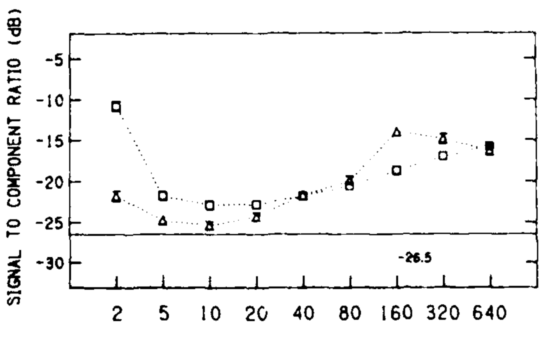


Fig. 4. Results of the multiple-component-signal experiment. The ordinate is the threshold for the signal, $20\log(k/a)$ (see Appendix). The abscissa is the frequency of modulation, f_m. The data for the two experimental conditions are coded: square-Condition 5, triangle-Condition 6. The horizontal line (and number) is the threshold for the stationary profile condition. Each data point is the average threshold for the three observers.

ponent to which it is added in-phase (20log k/a). We have not corrected this threshold value to take account of the change in signal power created by the modulation. One should be wary in comparing the thresholds for the multiple signal conditions with those of the single component conditions reported in the preceding experiment. In this multiple signal experiment, the same 'signal' amplitude, k, has been used to alter 21 components, rather than a single component, a_j. Thus, there are, in effect, a total of 21 signals rather than the single component of the former experiment. If we measured the total signal energy or power, it would be 13 dB greater for the 21-component signal, but no attempt has been made to 'equalize' the signal threshold. Indeed, there is still some doubt as to how to accomplish that objective (Green, 1986; Green et al., 1987; Bernstein and Green, 1987).

The two modulation conditions are coded in the figure: a square and a triangle represent the thresholds obtained for Conditions 5 and 6, respectively. The solid horizontal line represents the threshold obtained in the stationary profile condition (Condition 4). Once more, we remind the reader that the sounds were presented for a duration of one-half second. This brief duration may inflate the threshold values for the slower modulation rates.

Again, the results show that a change in spectral shape, presented in a dynamic mode, does not improve the ability to detect a change in spectral shape over the same amount of change presented in a stationary spectrum. Generally, the thresholds are best for the stationary, saw-tooth condition, represented in the figure by the solid horizontal line (-26.5 dB). If one increases the signal power created by the amplitude modulation of the signal (1.7 dB), thus elevating all the threshold points by 1.7 dB, the discrepancy widens. The dynamic conditions produce higher signal thresholds, even if a 'corrected' threshold quantity is calculated.

As the modulation rate increases, the spectral changes become increasingly difficult to hear. In the first experiment, at the higher rates of modulation, audible sidebands probably did occur. In these experiments, at these high rates of modulation, the sidebands were largely inaudible since they moved into adjacent masking components. Thus, as the frequency of modulation increases, the shape of the long-term power spectrum of standard and signal-plus-standard is essentially the same, and the threshold for the signal increases. That the sidebands were of some importance is demonstrated at the highest modulation frequencies. Detection of the sidebands is probably the primary reason that the signal can be heard in Condition 5, since the spectrum is essentially flat once the temporal variation within a channel is lost. The phase of the relative modulation appears to lose importance at a relatively low modulation frequency, 20–40 Hz, since at that frequency Conditions 5 and 6 produce very similar thresholds. The different threshold estimates seen for these conditions at modulation rates of 160 and 320 Hz probably reflect the different interactions of sidebands of similar frequency but opposite phase. That temporal variation in the spectrum is effective only at relatively low rates of modulation (below 40 Hz) is a conclusion also suggested by the next experiment.

Time-varying signal and nonsignal

Stimuli

In this third experiment, we explore a slightly different aspect of the question of how temporal variation in the spectrum influences profile analysis. Here we concentrate on the relative coherence of modulation in what we call the 'signal' and 'nonsignal' part of the spectrum. Detecting a change in spectral shape, if the change occurs at a single spectral locus, requires a simultaneous comparison of intensity information across different frequency channels. An obvious question is, how does temporal variation within the signal and nonsignal channels influence such comparisons? Note that the nonsignal part of the spectrum can be considered as a kind of amplitude standard, against which comparisons of the amplitude in the signal channels can be made. If both signal and nonsignal channels vary coherently, then detection of a change in the relative level of the signal channel should proceed smoothly. But suppose the nonsignal part of the spectrum varies out-of-phase with the signal part. How will this lack of amplitude coherence influence the detectability of a change? The specific question we asked was, suppose the signal and nonsignal levels are both amplitude modulated, will the relative phase between the envelopes of the two modulation waveforms influence detection performance?

In these experiments, the standard is our usual multicomponent complex. In one experiment, it contained 21 components, in the other 7 components. The signal was a single component of the standard that was added in-phase to the standard, thus producing a small increment in that component of the standard. We call the component at the signal frequency the 'signal component.' The amplitude of this component will have the value a if standard alone is present and the value $a + a_j$ if the signal is added to the standard. The other components (20 or 6 in number) we call the 'nonsignal components.' They all have amplitude a, independent of whether or not the signal is present. We now multiply each of these two waveforms, the signal component and the nonsignal components, by a modulation waveform, $m(t)$, where

$$m(t) = 1 + \sin[2\pi f_m t + \Theta(s)].$$

The rate of modulation, f_m, is in cycles per second. The phase of the modulation, Θ, depends on the waveform being modulated, s. Suppose both signal and nonsignal components are modulated, using the same value of theta. We call this the 'in-phase' modulation condition. In that case, the level of the signal and nonsignal components wax and wane together. Likewise we can choose different values of thetas for the signal and the nonsignal components. A phase difference of 180° between the two thetas is referred to as the 'out-of-phase' condition. In that case, the signal amplitude increases while the amplitude of the nonsignal components is decreasing and vice versa.

Because of the randomization of presentation level, listening only for overall intensity at any component is a poor detection strategy. To obtain good detection performance, one must compare the level of the signal component with the nonsignal component on each particular presentation, although the levels will vary at the modulation rate during each stimulus presentation. The task is always the same: to determine whether there is a relative increment in a single component of an otherwise flat spectrum.

The relative detectability of an increment at three different frequency locations was measured using both the 21- and 7-component standards. The signal frequencies were the second, middle, and penultimate components of the complex—either 235, 1000, or 4257 Hz for the 21-component complex or 342, 1000, or 2924 Hz for the 7-component complex. The rate of modulation ranged between 2 and 160 Hz.

Results and discussion

Fig. 5 presents the results for the 21-component standard; Fig. 6 presents the results for the 7-component standard. The threshold for the signal is plotted along the ordinate, and the rate of modulation, f_m, is plotted along the abscissa. The threshold values are averages across three observers; two of the three had participated in the previous experiments, the third subject listened only to these conditions. The square symbols code the in-phase condition, and the triangles code the out-of-phase condition. The solid horizontal line in the figure represents the threshold for the signal in a stationary profile condition, the signal is unmodulated and simply increases the amplitude of the signal component to $a + a_j$ during the entire observation interval. The slight differences in average threshold value from those reported in the previous experiment arise for two reasons. First, one listener is different. Second, these two measurements were taken several months apart; thus, the observers common to both measurements have had more experience in the task, and their thresholds had improved slightly. One subject improved an average of about 4 dB, the other 2.5 dB. Such long-term improvement is characteristic of profile experiments (Kidd et al., 1987).

The general form of the results is sensible. At very slow rates of modulation, for example 2 Hz, it is difficult to detect the signal in the out-of-phase condition. We presume this occurs because the nonsignal components provide little basis for a simultaneous comparison of signal and nonsignal levels; the nonsignal components are nearly absent when the signal component is in the vicinity of a maximum, and the reverse. Without a simultaneous level cue, the observer is forced to listen to overall level which is a very poor cue, given the

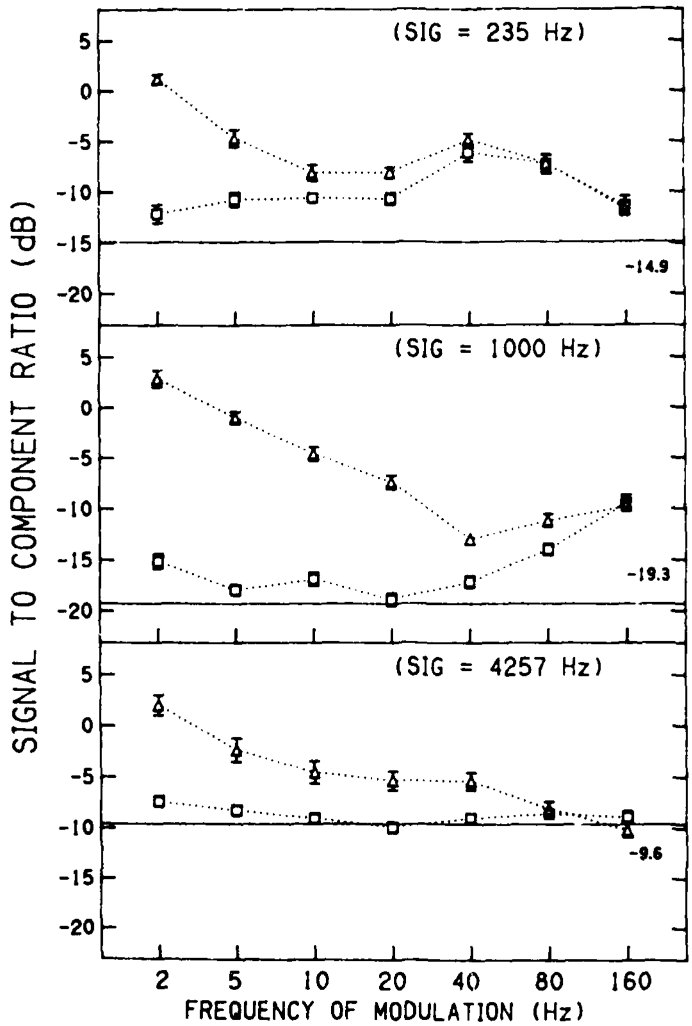


Fig. 5. Results for the experiment, using a 21-component complex, on the relative phase of signal and nonsignal components. The ordinate and abscissa are the same as those used in Figs. 2 and 4. The threshold values, when signal and nonsignal components are in-phase (square) and out-of-phase (triangle), are plotted. The signal frequencies are indicated. The horizontal line (and number) in each panel is the threshold for the stationary profile condition. Each data point is the average threshold for the three observers.

presentation level is selected at random from 20-dB range. If overall level is the only cue, then, given a 20-dB range, one can calculate that the threshold will increase to about +3 dB using the (3-down, 1-up) adaptive rule (Green, 1987, p. 20).

For the in-phase condition, the signal and nonsignal components wax and wane together, and, hence, a simultaneous comparison of levels is possible. As a general rule, the threshold for the in-phase condition is nearly the same as what one obtains with the stationary (unmodulated) condition, independent of the rate of modulation. The largest exception seems to be the lowest frequency signal (235 Hz) for the 21-component profile where, for modulation frequencies less than 20 Hz, all the in-phase thresholds seem to be about 5 dB

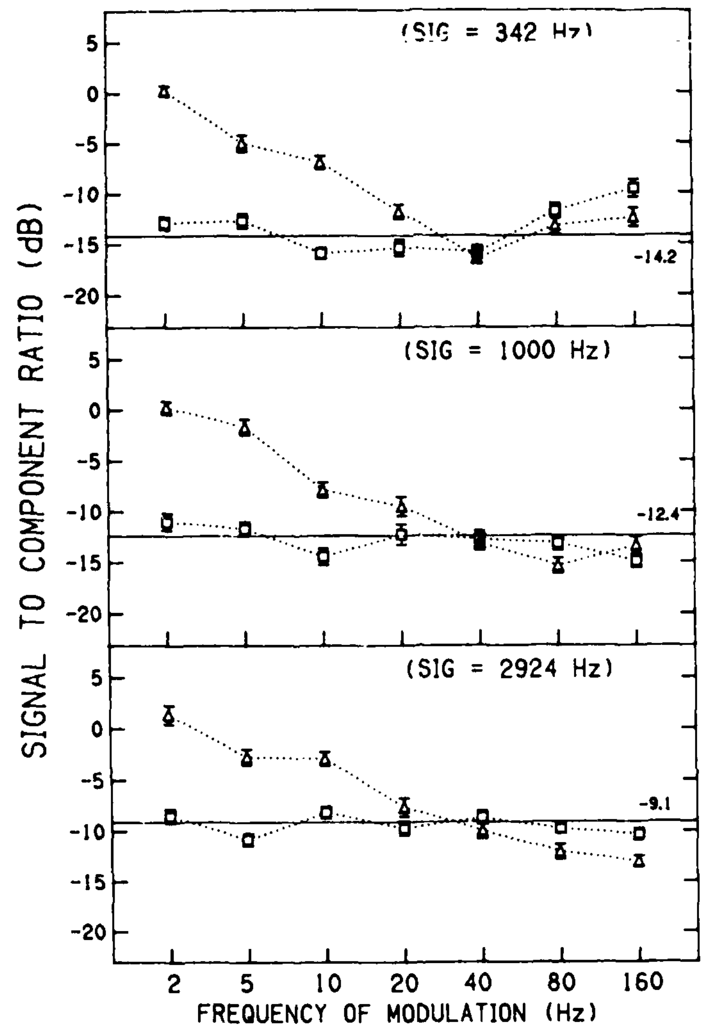


Fig. 6. Same as Fig. 5 except a 7-component complex was used.

poorer than the unmodulated condition.

In terms of detecting a spectral change, only when the sidebands produced by the modulation interact with the major components of the spectrum does there appear to be a difference in the threshold for the unmodulated and modulated conditions. When this interaction occurs, the spectrum sounds 'rough'. Depending on the rate of modulation, the signal frequency, and the density of the components in the spectrum, this roughness appears to cause a small elevation in threshold for the in-phase conditions. For example, in Fig. 5, the small bump in the data for the 235-Hz signal frequency (21-component) condition at a frequency of modulation of 40 Hz is where the sidebands of the first (200 Hz) and third (276 Hz) spectral components are very close in frequency to the signal (second) component (235 Hz). Similarly, the upturn in the data for the 1000 Hz signal frequency (21-component) condition at a modulation frequency of 160 Hz is where the sidebands of the nearest components fall close to the signal frequency. Indeed, the general lack of such effects in the 7-component data is consistent with this explanation. For the data of the 7-component complex (Fig. 6), there is little difference between the threshold measured in the in-phase condition and the stationary condition at any modulation frequency. The only departure from this rule occurs at the lowest signal frequency (342 Hz) where the data appear to drift upward for the highest rates of modulation. In that case, the explanation appears to be some interaction between the most closely spaced components, 200 and 342 Hz.

While there is a clear difference between the thresholds obtained in the in-phase and out-of-phase conditions for the lower modulation rates, once the frequency of modulation exceeds about 20–40 Hz, the thresholds for the two conditions are nearly the same. This result is easy to understand in terms of a simple filter model. Suppose the rate of modulation is so fast that the sidebands fall outside the filter located at the center frequency of the modulation. When such a condition occurs, the output of the filter is constant and shows no variation in amplitude produced by the modulation. When the amplitude variation is no longer present, the relative phase between the signal component and any other components of the spectrum ceases to be important. Thus, at the highest rates of modulation, the phase angle, theta, should be irrelevant and the thresholds for the in-phase and out-of-phase conditions should be nearly the same, as they are.

Note, however, that the region where frequency modulation makes phase irrelevant is about 20–40 Hz and is largely independent of the density of the spectrum or the signal frequency. Indeed, for the 7-component complex, the modulation rate at which the in-phase and out-of-phase conditions yield nearly equal threshold is about the same for all three signal frequencies.

This last observation raises a potential problem with the preceding filter explanation. A simple application of this filter idea would suggest that the modulation frequency at which the modulation phase becomes irrelevant should vary systematically with signal frequency. The reason for this expectation is straightforward. We know that the widths of the frequency channels in the auditory system vary systematically with center frequency. One estimate of the critical band (Zwicker, 1961) is about 16% of the center frequency, that is, about a bandwidth of 32 Hz at 200 Hz, 160 Hz at 1000 Hz, and 640 Hz at 4000 Hz. Thus, as frequency of modulation is varied, a critical band centered at lower frequencies, because it has a smaller bandwidth, will produce a near steady (carrier) tone at much smaller rates of modulation than a band centered at a higher frequency. If we assume the filter output is nearly constant when the sidebands are located at the filter bandwidth, then a frequency of modulation of 16 Hz at a center frequency of 200 Hz is equivalent to an 80-Hz modulation rate at a center frequency of 1000 Hz, or a 320-Hz modulation rate at a center frequency of 4000 Hz. The modulation rate producing an equivalent change in filter output is proportional to center frequency.

One should, however, be cautious in interpreting how this fact will affect the detectability of the signal in this experiment. A profile experiment must involve a comparison of the signal level with the nonsignal level. Previous work on profile analysis has shown that the comparison process is not restricted to the locally adjacent critical bands (Green et al., 1984). Thus, interpreting how modulation frequency and signal frequency will interact is complicated. Consider one condition, the threshold for the 2924-Hz signal frequency with a 7-component complex. The effects of modulation phase become negligible for a modulation rate of about 20 Hz. At that frequency, the critical band at the signal frequency is nearly 500 Hz wide. Thus, the temporal fluctuation in the signal channel should be considerable. Yet the relative phase of that modulation, between the signal and nonsignal channels, is irrelevant. The only explanation we can offer is as follows. Assume that the level of some nonsignal channels (presumably much lower in frequency) is being used as a basis of comparison with the level in the signal channel. The level in these channels is essentially constant, because the center frequency and hence bandwidths are much smaller. Therefore, because the nonsignal channel is not changing in level, the relative phase of fluctuations in the 2924-Hz channel is irrelevant.

Relative phase of the modulation

In this experiment, we systematically measure (in steps of 45°) how different phase angles between the two modulators would affect the detectability of the signal. While we have described the comparison of different levels in a profile task as simultaneous, it is possible that some small time is taken to actually compare the two levels. We have used the word 'simultaneous' to distinguish the process from the successive comparison of levels across the intervals of the forced-choice procedure, a process that involves time measured in seconds. The 'simultaneous' comparison of profile analysis may occupy a few milliseconds, even if it occurs within a single observation interval. We wished to determine if this time were measurable and, therefore, varied the relative phase of modulation between the signal and nonsignal components in much finer steps than the two used in our previous conditions. To measure the time of comparisons precisely, we would like to use a high frequency of modulation so that changes in phase would reflect small time differences. But our dependent variable is the difference in threshold for two phase conditions. Thus, we must use a frequency of modulation that produces some measurable difference. As our compromise, we selected the 1000-Hz signal (21-component) condition with a modulation rate of 5 Hz. That condition produces a difference in threshold of about 15 dB, between the in-phase or out-of-phase condition (Fig. 5, middle panel). Unfortunately, the 5-Hz modulation rate is equivalent to a period of 200 ms, so our time scale for phase effects is relatively gross.

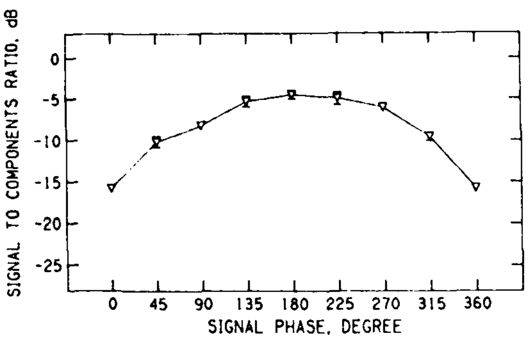


Fig. 7. Results for the relative signal-nonsignal phase experiment. A 21-component complex was used. The signal frequency was 1000 Hz and the frequency of modulation was 5 Hz. The ordinate is the threshold for the signal. The abscissa is the relative phase between the modulator for the signal and nonsignal components. Zero degree means the signal and nonsignal components are in-phase, 180° the out-of-phase condition. The average of the three observers is plotted.

Fig. 7 shows the results. We express the in-phase condition as zero or 360° and the out-of-phase condition as 180°. The signal threshold for each of these phase conditions is plotted along the ordinate. (The threshold at 360° is simply the zero point, replotted.) As can be seen, the in-phase condition appears to provide the lowest threshold, and the 180° phase angle produces the highest threshold. Intermediate phase angles fall along a relatively smooth curve. Any small delay in the comparison process must be either zero or smaller than about 125 μsec, the equivalent of a 45° phase change.

Conclusion

The detection task common to all these experiments is to discriminate a change in spectral shape of a multicomponent complex.

While the saliency of individual components of a multicomponent complex can be enhanced by amplitude modulation, the detectability of amplitude changes in such components is not greatly influenced. For amplitude changes in single components of a multicomponent complex, only at the highest signal frequency (4257 Hz) does amplitude modulation appear to provide a consistent benefit to the detection of such changes (Fig. 2). For amplitude changes over multiple components, dynamic conditions generally produce much poorer thresholds than a stationary condition.

The relative coherence of the signal and nonsignal components of the spectrum is important, but only at the lower modulation rates ($f_m < 40$ Hz). Above this rate of modulation, the relative phase of different parts of the spectrum is unimportant (Figs. 5 and 6).

Appendix

The following are the equations used to generate the stimuli used in the first two experiments.

We refer to the standard as St(t), and the signal as S(t). These waveforms were sampled, digitized, and stored in buffers, which were then played to the listeners through the D to A devices. The signal-plus-standard waveform was created by adding (point by point) the two buffers, representing the signal and standard waveform.

Experiment 1. Single-component signal

For all conditions in this experiment, the parameter a_j is adjusted to estimate the signal threshold.

Stationary Profile Condition

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = a_j \cos(2\pi f_j t + \Theta_j)$$

Dynamic Conditions
Condition 1

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = a_j[1 - \cos(2\pi f_m t)] \cos(2\pi f_j t + \Theta_j)$$

Dynamic Spectral Changes
Condition 2

$$St(t) = \sum_{i<>j}^{20} a \cos(2\pi f_i t + \Theta_i)$$

$$+ a[1 - \cos(2\pi f_m t)]\cos(2\pi f_j t + \Theta_j)$$

$$S(t) = a_j[1 - \cos(2\pi f_m t)]\cos(2\pi f_j t + \Theta_j)$$

Condition 3

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = a_j \cos(2\pi f_m t)\cos(2\pi f_j t + \Theta_j)$$

Experiment 2. Multiple component signals

For all conditions in this experiment, the parameter k is adjusted to estimate threshold.

Stationary profile condition
Condition 4

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = \sum_{i=1}^{21} k(-1)^i \cos(2\pi f_i t + \Theta_i)$$

Dynamic conditions
Condition 5

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = \sum_{i=1}^{21} k[1 - \cos(2\pi f_m t)]\cos(2\pi f_i t + \Theta_i)$$

Condition 6

$$St(t) = \sum_{i=1}^{21} a \cos(2\pi f_i t + \Theta_i)$$

$$S(t) = \sum_{i=1}^{21} k\left[1 + (-1)^i \cos(2\pi f_m t)\right]$$

$$\times \cos(2\pi f_i t + \Theta_i)$$

Acknowledgements

This research was supported by the National Institutes of Health and by the Air Force Office of Scientific Research. We are grateful to Drs. Leslie Bernstein, Timothy G. Forrest and Virginia Richards for their comments on an earlier draft of this paper. The comments of two anonymous reviewers were also helpful in revising the paper.

References

Bernstein, L.R. and Green, D.M. (1987) Detection of simple and complex changes of spectral shape. J. Acoust. Soc. Am. 82, 1587–1592.

Bernstein, L.R., Richards, V. and Green, D.M. (1987) The detection of spectral shape change. In: W.A. Yost and C.S. Watson (Eds.), Auditory Processing of Complex Sounds, Erlbaum, New Jersey, pp. 6–15.

Green, D.M. (1986) Frequency and the detection of spectral shape change. In: B.C.J. Moore and R.D. Patterson (Eds.), Auditory Frequency Selectivity, Plenum, England.

Green, D.M. (1987) Profile Analysis: Auditory Intensity Discrimination. Oxford University Press, New York, Oxford.

Green, D.M. and Mason, C.R. (1985) Auditory profile analysis: Frequency, phase, and Weber's Law. J. Acoust. Soc. Am. 77, 1155–1161.

Green, D.M., Mason, C.R. and Kidd, G. (1984) Profile analysis: Critical bands and duration. J. Acoust. Soc. Am. 75, 1163–1167.

Green, D.M., Onsan, Z.A. and Forrest, T.G. (1987) Frequency effects in profile analysis and detecting complex spectral changes. J. Acoust. Soc. Am. 81, 692–699.

Kidd, G., Mason, C.R. and Green, D.M. (1986) Auditory profile analysis of irregular sound spectra. J. Acoust. Soc. Am. 79, 1045–1053.

McAdams, S. (1984) Spectral fusion, spectral parsing and the formation of auditory images. Doctoral dissertation, Stanford Univ.

Summerfield, Q., Sidwell, A. and Nelson, T. (1987) Auditory enhancement of changes in spectral amplitude. J. Acoust. Soc. Am. 81, 700–708.

Viemeister, N.F. (1980) Adaptation of masking. In: G. van den Brink and F.A. Bilsen (Eds.), Psychophysical, Physiological and Behavioral Studies in Hearing, Delft Univ. Press, Netherlands.

Zwicker, E. (1959) Die Grenzen der Hörbarkeit der Amplitudenmodulation und der Frequenzmodulation eines Tones. Acustica, Beihefte Ab125-AB133.

Zwicker, E. (1961) Subdivision of the audible frequency range into critical bands. J. Acoust. Soc. Am. 33, 248 (Letter to Editor).

Frequency effects in profile analysis and detecting complex spectral changes

David M. Green, Zekiye A. Onsan, and Timothy G. Forrest
Department of Psychology, University of Florida, Gainesville, Florida 32611

(Received 19 June 1986; accepted for publication 2 November 1986)

Seven experiments on the detectability of intensity changes in complex multitonal acoustic spectra are reported. Two general questions organize the experimental efforts. The first question is how the detectability of a change in a flat (equal energy) spectrum depends on the frequency region where a single intensive change is made. The answer is that frequency region plays a relatively minor role. Frequency changes in the midregion of the spectrum are the easiest to hear, but thresholds increase by only about 5 dB over the range from 200 to 5000 Hz. For all frequencies, the psychometric function is of the form $d' = k(\Delta p)$, where k is a constant and Δp is the change in pressure. The second question is how can we predict the detectability of complex changes over the entire frequency range from the detectability of change at each separate region. Thresholds for detecting a change from a flat spectrum to a spectrum whose amplitude varies in sinusoidal ("rippled") fashion over logarithmic frequency are measured at different frequencies of ripple. The thresholds are found to be independent of ripple frequency and are 7 dB higher than predicted on the basis of an optimum combination rule.

PACS numbers: 43.66.Ba, 43.66.Dc, 43.66.Fe, 43.66.Jh [RDS]

INTRODUCTION

In several previous papers, we have reported on the ability of listeners to detect alterations in the shape of complex acoustic spectra. Often the standard stimulus was a multicomponent spectrum composed of equal-amplitude sinusoids. The change was created by increasing the intensity of one component of the standard. In this paper, we systematically explore the question of how the frequency of the altered component affects the ability to detect the change. Is it easier to detect a low- or high-frequency change in the intensity profile of a complex stimulus? There are several reasons for asking such a question, but the one we will stress here is the empirical one. We will need this information to address the second question of this paper; namely, how do we use the data on the detectability of changes in local regions of the spectra to predict the detectability of more complex changes?

In a previous study, we measured how detectability of an increase in a single component changed as a function of the frequency of the component (Green and Mason, 1985). In general, those results suggested that a change in the intensity of the midfrequency region, 500 to 2000 Hz, produced superior performance, but the variability among observers was sizable. Also, those data may have been contaminated by prior practice because the subjects had participated in an earlier experiment in which the change occurred in this midfrequency region. Although extensive training was given for all frequency regions, it is conceivable that the effects of the earlier practice influenced the data. In the present study, we used the recent move of our laboratory as an opportunity to recruit a set of listeners who had no previous training at any one frequency region.

Once we have studied how the detectability of a change in a single region of the spectrum varies with component frequency, we are ready to consider spectral changes of a more complicated variety. Suppose the listener were trying to detect spectral change at many component frequencies. Is it possible to develop a simple rule to account for the detection of these more complicated changes? The more complicated spectral change that we investigate is a sinusoidal alteration over the entire spectral profile—a sinusoidal ripple over logarithmic frequency. The frequency of this ripple is then varied and detection performance assessed for a number of ripple frequencies. The results obtained with the various ripple frequencies are easy to summarize—they all produce about the same threshold. This threshold, however, is about 7 dB higher than would be expected on the basis of an optimum combination of the detectabilities at the local regions.

I. GENERAL PROCEDURE

In all the experiments, the listener's task was to detect a change in the spectral shape of a complex multicomponent waveform. The components of the standard were always of equal amplitude and always equally spaced on a logarithmic frequency scale. The phase of each component of the standard was chosen at random, and this phase was used for all presentations of this condition. The standard spectrum was altered in shape by changing the intensity of one or more of the sinusoidal components. This alteration can be thought of as adding a "signal" waveform to the standard. Thus the discrimination task was to distinguish between the standard alone and the standard with the signal added to it. The overall level of the sounds, standard or standard-plus-signal, was varied on each and every presentation according to a random schedule, so that the observers were forced to detect a change in the shape of the standard spectra, rather than simply a change in intensity at some region of frequency.

All waveforms were generated digitally, played over di-

 0001-4966/87/030692-08$00.80

gital-to-analog converters at a sample rate of 25 000 Hz, and low-pass filtered at 10 000 Hz. The duration of the sounds differed in the different experiments, but all were turned on and off with a 5-ms raised cosine window. The observers were seated in sound-treated (IAC double-walled) rooms and the stimuli were presented binaurally over TDH-39 earphones, both phones driven in-phase.

A two-alternative forced-choice procedure was used with an adaptive, two-down, one-up technique to estimate a signal level corresponding to a 0.707 probability of correct choice ($d' = 0.76$). The initial step size of 4 dB was halved after the first four reversals. Fifty trials were run in blocks, and the estimated threshold was computed as the average of the remaining pairs of reversals after excluding the first three reversals. Typically, 10 to 16 reversals occurred within each block. For a given stimulus condition, 6 runs of 50 trials were run in succession. Each trial lasted about 2 s, and it took about 15 min to complete 6 runs of 50 trials. All of the data reported here were based on two or three separate replications; that is, average thresholds were based on 12 or 18 fifty-trial blocks.

Normal-hearing observers participated in the experiments. They were college students recruited through advertisements placed in the student employment office and the music and speech departments. They were paid at an hourly rate for their services and were given a special bonus upon completing the entire sequence of measurements.

II. SINGLE SIGNAL COMPONENT—EFFECTS OF FREQUENCY LOCATION

A. Single component signal in 21-component profile

The "standard" for this experiment was a complex of 21 equal-amplitude components spaced equally on a logarithmic scale of frequency. The lowest frequency component was 200 Hz, the highest was 5000 Hz, and the ratio of the frequencies of successive components in the spectrum was 1.1746. The level of the standard varied between trials over a range of 20 dB and the median sound-pressure level of the standard was 40 dB per component. Because there were 21 components in the complex, the overall level was 13 dB higher (53 dB SPL).

The signal was a single sinusoid added in-phase to one component of the standard. A threshold was measured for detecting this increment at each of seven different frequencies: 234, 380, 617, 1000, 1620, 2626, and 4256 Hz. The stimulus duration was 100 ms. We report the average threshold over six listeners, based on twelve 50-trial determinations of threshold at each frequency

Figure 1 presents the results of this experiment. The value along the abscissa is the frequency of the component to which the signal was added. The value along the ordinate is the size of the signal at threshold measured as the signal amplitude *re*: the amplitude of the component of the standard to which the signal is added (in-phase). For example, if the signal were the same size as the component of the standard, the threshold would be reported as 0 dB. If the signal were 1/10 the amplitude of the component of the standard to

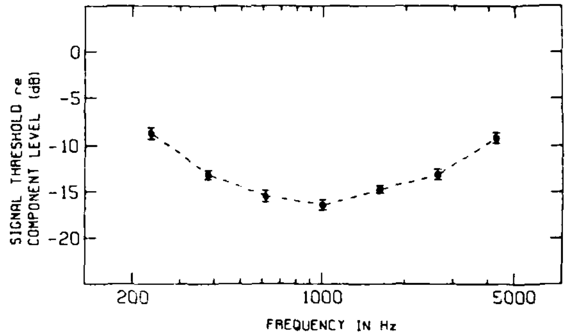


FIG. 1. Discrimination of an increment added to a single component of an equal-amplitude, 21-component standard waveform with a frequency range of 200 to 5000 Hz. The abscissa shows the frequency of the incremented component (signal) and the ordinate is the threshold for 70% correct discrimination of the signal. Thresholds are the ratio of the level of the signal increment to the level of a single component of the standard in dB. Error bars are the standard error computed over 6 subjects with 12 runs each.

which it is added, the threshold would be − 20 dB. As can be seen, the detection of the increment does vary some with signal frequency. The midfrequency region, 500 to 2000 Hz, produces the best detection. Increments in the flat spectrum outside this frequency region are somewhat more difficult to detect, but the difference never exceeds 10 dB. The error bars are the standard error of the mean computed over the 72 threshold estimates made at each frequency (6 observers and 12 threshold estimates per observer). The average data, shown in the figure, are typical of all the observers. The results are similar to those obtained by Green and Mason (1985). The function of Fig. 1 is smoother and shows slightly less variation with frequency than was found in the earlier study.

B. Effects of overall intensity and duration

The stimulus conditions were similar to those employed in experiment 1, except the overall intensity level of the stimuli was increased 20 dB and the median standard level was 60 dB SPL per component rather than the 40-dB level used in the previous experiment. Also, two presentation durations were studied: 100 ms as in the previous experiment, and 30 ms. Three observers participated in this experiment; only one had participated in the first experiment. These three observers participated in all the remaining experiments.

Figure 2 presents the results of this experiment. The quantities plotted on the ordinate and abscissa are the same as in Fig. 1. The threshold values for the 30-ms presentation duration are shown by open circles (the upper dashed curve), while the 100-ms data are plotted as open triangles (the lower curve). The solid line segments are the results obtained in the first experiment with a 100-ms duration and lower intensity level.

The 100-ms presentation duration produces lower signal thresholds than the 30-ms presentation duration at almost all frequencies. We are puzzled by the two thresholds being the same at the highest signal frequency (4256 Hz) and suspect this coincidence is chance fluctuation. At all other frequencies, except 380 Hz, the difference in threshold

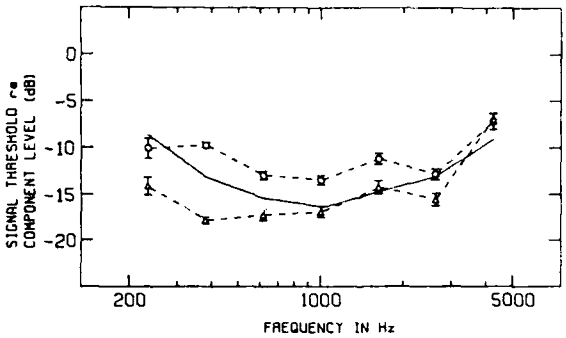


FIG. 2. Effect of level and signal duration on threshold. Conditions were the same as in Fig. 1 (shown as solid line), except intensity levels of the stimuli were increased 20 dB to a median level of 60 dB per component. In one condition, signal duration was the same as in experiment 1 (100 ms: triangles); in another condition, 30 ms was used for signal duration (circles). Error bars are standard errors computed over 18 runs for three subjects.

for the two durations is nearly the same. The average difference in threshold, over all frequencies, is 3.6 dB. This value is only slightly smaller than the value of 5 dB, which one would expect from an equal-energy rule. Our definition of the signal threshold is proportional to the level of the signal; thus a change in duration of a factor of three would necessitate a change in signal power of a factor of 3, or 5 dB, to hold signal energy constant. The equal-energy rule has received empirical support in a previous paper (Green *et al.*, 1984).

Detection thresholds for this experiment are generally similar as a function of frequency to those obtained in the first experiment (solid line segments). The only difference worth comment is that, whereas the first experiment showed a shallow bowl-like curve, the results of the second experiment show less of an increase in threshold for the lower frequencies. Averaging the data at the two durations would show a nearly flat function for the lower and midfrequency region and a slight increase at the highest frequency. Whether this difference in the two experiments arises because of differences in observers or because these observers have now had more practice in this detection task is unknown. We believe that training may play some role since, in our experience, there is a very slow improvement in the ability to hear the spectral change in the lower frequency region that is not evident for the higher frequencies.

One purpose of this experiment was to determine if the effects of frequency were altered appreciably by a presentation duration of 30 ms, which is shorter than the duration of an acoustic reflex. Such does not seem to be the case, and we can rule out the acoustic reflex as playing any significant role in the studies that employed longer presentation duration.

C. Frequency context

In the two preceding experiments, the best thresholds occur for the middle frequencies of the standard, between 500 and 2000 Hz. Does this reflect greater sensitivity for these frequencies, or are these lower thresholds because this region is in the center of the standard? To answer this question, we generated two 21-component standard stimuli with only slightly overlapping frequency ranges. The "low-frequency" standard ranged in frequency between 200 and 2000 Hz. The ratio of frequencies of successive components of the standard was 1.122. The "high-frequency" standard ranged in frequency between 1000 and 10 000 Hz and had the same ratio between successive frequency components as the low-frequency set. For each standard, we measured the threshold for an increment in a single component of the standard, in either a relatively low-, middle-, or high-frequency region of that standard. The three signal frequencies were 224, 632, and 1782 Hz for the low-frequency standard; the signal frequencies were 1122, 3162, and 8912 Hz for the high-frequency standard. The other conditions were similar to those used in the first experiment. The component level of the standard was 40 dB and the presentation duration was 100 ms. The thresholds were based on eighteen 50-trial runs.

Figure 3 presents the result of this experiment. The ordinate and abscissa are the same as those used in Fig. 1. The thresholds for the three signal frequencies in the lower frequency standard are shown as the open circles. The thresholds for the three signal frequencies in the higher frequency standard are shown as the open triangles. The curve depicted by the solid line segments is the result obtained in the first experiment (with frequency range from 200 to 5000 Hz).

For the lower frequency standard, the middle-frequency signal is the easiest to hear and either end of the frequency range produces higher thresholds, a result consistent with the finding in the first experiment. In fact, the two lower frequency signals have thresholds remarkably similar to those obtained with the wider frequency complex employed in the first experiment. The threshold for the upper frequency signal, 1782 Hz, however, is nearly 10 dB higher than that determined for the wider frequency complex. This result presumably reflects the effects of context, the relative loca-

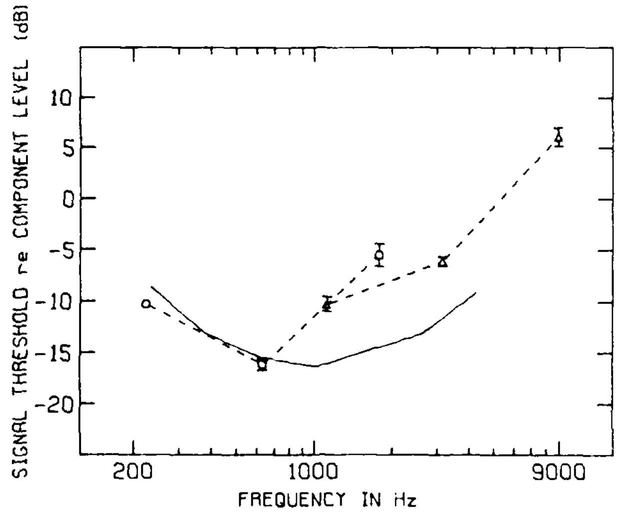


FIG. 3. Effect of frequency range on discrimination of an increment on a single sinusoid in a 21-component spectrum. The conditions were similar to those used in the first experiment (solid line) with the frequency range of the spectrum changed. The low-frequency standard, shown as circles, ranged in frequency from 200 to 2000 Hz. The high-frequency standard (triangles) ranged from 1000–10 000 Hz. Error bars are the standard errors computed over 12 runs for three subjects.

tion of the signal frequency within the standard complex. Similar effects of context have been reported in previous papers (Green and Mason, 1985).

For the high-frequency complex, the middle-frequency signal is not the easiest to detect, and it appears that at these higher frequencies, the frequency of the signal component per se exerts a stronger influence on the signal's threshold than does context. The effect of context is again evident if we compare the thresholds obtained in this experiment with those obtained in the first experiment (solid curve). The presence of components below 1000 Hz, as occurs in the 200- to 5000-Hz standard, produces lower thresholds for every component where comparisons can be made than those obtained for the complex extending from 1000 to 10 000 Hz.

As a simple summary, we may say that signals in the middle of a standard are generally easier to detect than signals located at the extremes, providing the entire range is located below at least 5000 Hz. Above this frequency, the absolute frequency of the signal may play a larger role than the effects of context.

D. Extended frequency range

In this experiment, we used a standard with as wide a frequency range as is practically possible. The standard in this experiment was a 30-component complex ranging in frequency from 200 to 10 000 Hz. The median level of the components was 50 dB SPL, and the ratio of the frequencies between successive components was 1.144. The signal presentation duration was 100 ms.

The addition of the "signal" produced a change in five adjacent components of the standard. If we number these five components successively starting with the lowest frequency, then 3 is the middle component of the set. The odd components of this set, 1, 3, and 5, were increased in amplitude and the even components, 2 and 4, were decreased in amplitude. Thus the observers were discriminating between two stimuli, the standard with a flat (equal amplitude) spectrum, or the signal-plus-standard with a five-component ripple located at some frequency region within the flat complex. The frequency region of this ripple was the independent variable of the experiment. The threshold for this ripple was measured at six different regions, which was specified by the frequency of the middle component of the five-component complex, namely, 261, 514, 1009, 1981, 3889, and 7635 Hz. The thresholds are based on twelve 50-trial runs.

Figure 4 presents the result of this experiment. The curve depicted by the solid line segments are the data from the first experiment. We should comment on how threshold values are computed for these five-component signals. We have plotted the threshold on a single-component basis, thus − 20 dB means that the amplitude of all five signal components is 1/10 the amplitude of the standard component to which it is added (or subtracted). We have never conducted a formal experiment comparing increments and decrements, but informal testing has convinced us that the detectability of a fixed signal amplitude is not very different whether we add or subtract it from a component of the standard.

If total signal energy were used as a measure of threshold, instead of our single-component measure, then the five-

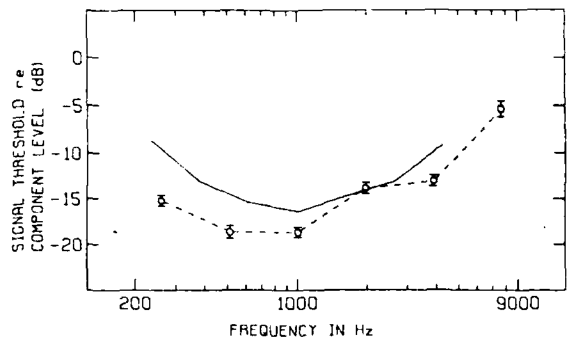


FIG. 4. Discrimination of a 5-component ripple as a function of the frequency of the center component of the ripple in a 21-component standard with a 200- to 10 000-Hz range. Ripples were on successive components of the standard with the phases such that, when added to the standard, the first, third, and fifth components of the ripple were incremented, while the other two components were decremented by a like amount. Thresholds are the level of the increment relative to the level of a single component of the standard in dB.

component ripple is about 7 dB greater in energy than any single component. If all the data points were increased by 7 dB, nearly all would fall above the solid line, which represents the threshold for the single-component signal used in that experiment. Thus we find that the single-component signal is the easiest signal to detect on an energy basis. A similar conclusion was reached by Green and Kidd (1983).

The general shape of the function is similar to what we found in the other experiments. The middle-frequency region is, once again, the easiest in which to detect a change in the spectrum, with the higher frequencies being much worse, especially at the extreme frequencies.

E. Summary of frequency location experiments

In general, all the results of these experiments exploring the frequency locus of the change in shape of a complex spectrum reveal no strong effects of frequency. When threshold is plotted as a function of frequency, the data resemble a shallow bowl with the minimum located in the moderate-frequency range, 500 to 2000 Hz. At the very highest frequencies, the signal can be more difficult to detect, by as much as 10 to 13 dB, but no abrupt changes in this function are evident. Only at frequencies as high as 7000 Hz does the ability to detect changes in a complex spectrum appear to deteriorate substantially.

III. DETECTION OF COMPLEX SPECTRAL CHANGES

In the next series of experiments, we turn our attention to the detection of complex alterations in spectral shape. To predict the detectability of such complicated changes, we hoped to use a rule based on the detectability of changes at individual components. To implement such a scheme, we first need to know the trading relation between signal amplitude and signal detectability; that is, we need to know the psychometric functions for changes in individual components.

A. Psychometric function

In this experiment, we estimate the psychometric functions for the detection of an increment in a single component at three frequencies, 380, 1000, and 2626 Hz. For a given frequency, the observer heard as the alternative of a two-interval forced-choice trial either the standard alone or one of three fixed-signal levels added to the standard. All three signal levels occurred with equal probability within a single listening session of 100 trials, so that signal level would not be confounded with trial block. The signal levels were chosen on the basis of a prior estimate of threshold obtained using the adaptive procedure. The middle signal level was set to produce about 75% correct and the other two signals set at level 6 dB above and below this value. Ten 100-trial runs were used to estimate the psychometric function at each frequency so that about 333 trials were used to estimate the percentage of correct judgments at the three signal levels.

Actually, two psychometric functions were estimated in two different experimental conditions at each of the three signal frequencies. One condition was a profile condition. In this case, the signal component was presented with 20 other components present and the overall level was randomly varied (50 dB $\pm$ 10 dB SPL). In the second condition, the single component was presented in isolation at a fixed level (60 dB SPL), so we are estimating the psychometric function for a simple intensity-discrimination task. Stimulus duration was 100 ms.

Psychometric functions for the simple pure-tone intensity discrimination task and profile tasks are shown in Figs. 5 and 6, respectively. The data for the three listeners and three signal frequencies are presented in each figure. The data shown in the figures were obtained from the following procedure. First, we converted the percentage of correct responses at each signal level to d'. Next, for each individual condition and listener, we plotted three data points, the value of 20 log d' versus signal level (20 log signal pressure). These data were then fit with a line having a slope of unity and one free parameter, the signal pressure that produced a $d' = 1$. For each listener and condition, we let this pressure be 0 dB. In this way, all of the data for all conditions and listeners could be plotted on a single graph, as in Figs. 5 and 6.

As these figures show, the detectability of the signal increases monotonically with the level of the signal. The average slope measured for the ten 100-trial runs for all conditions (subjects and frequencies) is 0.97 for the profile condition (Fig. 6) and 0.75 for the intensity-discrimination condition (Fig. 5). For the latter condition, previous experiments have found a slope value close to unity for well-practiced listeners (Green, 1960). The low value for the slope in this condition probably reflects a lack of sufficient training in that condition. Our listeners had spent most of the time listening to profile conditions. They all complained about the difficulty of the pure intensity-discrimination experiment. One observer summarized his frustration by saying "The only thing you can listen for is a difference in loudness."

Although the linear relation between d' and signal pressure provides a very good approximation to the data, other, alternative expressions are also consistent with the data. Another suggestion for the form of the psychometric function is that d' is proportional to the difference in level (ΔL) between the standard component and the standard-plus-increment (Rabinowitz *et al.*, 1976). In this formulation,

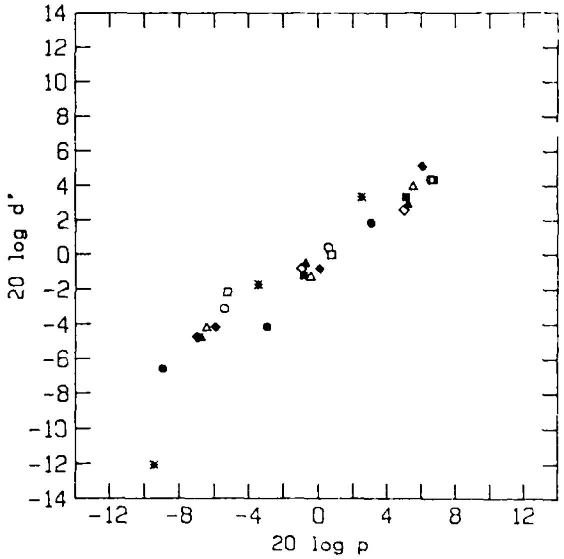


FIG. 5. Psychometric function for intensity discrimination. Plotted are d' for a single subject at each of three signal frequencies and levels. The functions were adjusted so that a $d' = 1$ occurs at 0 dB for each condition.

$$d' = k\Delta L = k\,10\log(1 + \Delta I/I) = k\,20\log(1 + \Delta p/p)$$

where k is a constant that depends on the experimental condition and the listener, I is the intensity (p the pressure) of the standard, and ΔI (or Δp) is the increment in intensity (or pressure). Recall that we always add the signal component in-phase to the component of the standard. For small values of $\Delta p/p$, ΔL is approximately equal to $8.686 * (\Delta p/p)$. The data of Fig. 6 show a linear relation between d' and $\Delta p/p$, but this could be interpreted equally well as implying a linear relation between d' and ΔL. The present data provide no way to choose between these different expressions for the form of the psychometric function.

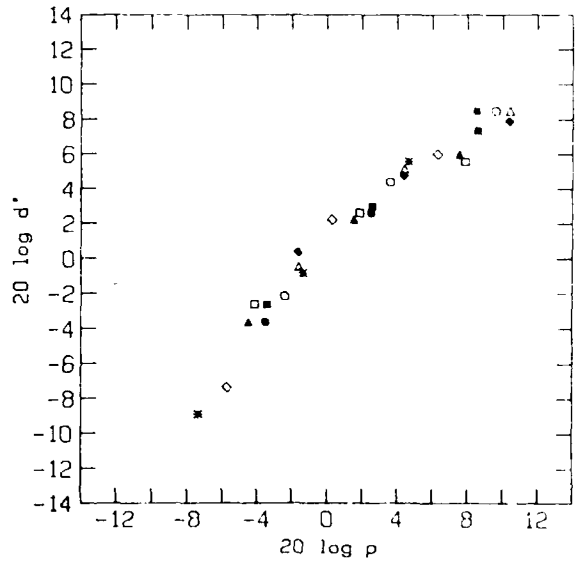


FIG. 6. Psychometric function for discrimination of a change in spectral shape, the addition of an increment on a single component in a 21-component standard flat spectrum. The coordinates are the same as Fig. 5

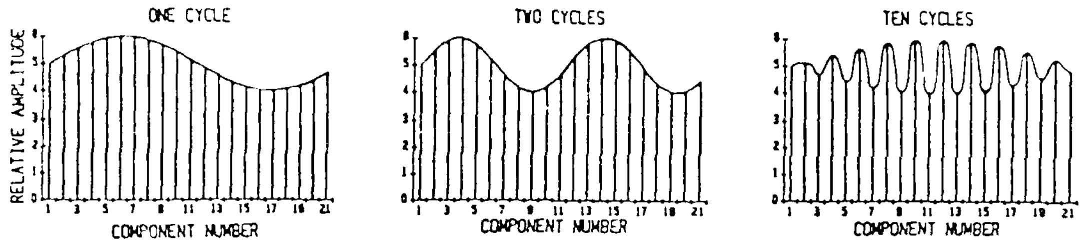


FIG. 7. Waveforms showing three different frequencies k of sinusoidal variation in component amplitudes.

B. Detection of rippled spectra

In this experiment, the standard waveform was the 21-component flat spectrum that ranged in frequency from 200 to 5000 Hz. The successive components were spaced equally on a logarithmic scale of frequency. When the signal waveform was added to the standard spectrum, it produced a resulting spectrum whose amplitude varied sinusoidally as a function of the logarithm of frequency, what we call a "rippled" spectrum. Figure 7 shows this manipulation graphically. The first spectrum shows a single cycle of sinusoidal variation in amplitude over our 21 components; the next spectrum shows two cycles of amplitude variation; and, finally, the last figure shows the highest rate of variation that can be achieved, since alternate components increase and decrease in amplitude.

Specifically, the "signal" waveform was produced by setting the amplitude of successive components, $a[i]$, according to the following equation

$$a[i] = \sin(2\pi k i/M), \quad i = 1,2,\ldots,M,$$

where i is the number of the component, ranging in this case from 1 to 21, $a[i]$ is the amplitude of the ith component of the signal spectrum, and k is frequency of the ripple. Recall that the first component, $i = 1$, corresponds to a frequency of 200 Hz, and the last component, $i = 21$, corresponds to a frequency of 5000 Hz. We scaled the amplitude of this "signal" and added each component in-phase (respecting sign) to the corresponding component of the flat standard spectrum to produce the change in the spectrum, as shown in Fig. 7. In that figure, the signal amplitude is about 20% of the standard amplitude. A cosine ripple can be constructed in the same manner by substituting the cosine function for the sine function in the equation.

It should be noted that by constructing the signal in this way the same 21 values occur for the set of amplitudes, $a[i]$, independent of the frequency of the ripple k. The parameter k simply reorders the set of 21 values. One result of this fact is that, for higher ripple frequencies, the spectrum appears to have a smoothing function imposed on the ripple (compare the ten-cycles ripple with the one- or two-cycle ripple in Fig. 7). Another consequence of this fact is that a quantity such as the root mean square (rms) of the amplitude values is independent of the frequency of the ripple k. If the maximum value for $a[i]$ is 1, the rms value is 0.707. A cosine ripple with the same scaling has the same rms value as the sine ripple, and this value is also independent of k.

The "depth" of the ripple resulting from the addition of the signal to the standard waveform depends upon the ratio of the amplitudes of the signal components to those of the standard's equal-amplitude components. Thus it is convenient to use the rms value of the 21-signal components as our measure of signal amplitude. We refer to the signal-to-standard ratio as the rms signal amplitude to the amplitude of any component of the standard. The depth of the ripple is, of course, monotonic related to the signal-to-standard ratio.

Table I lists the average threshold measured for these rippled (for some sine and all cosine) spectra at different frequencies of ripple k. The threshold for the signal is measured in terms of the signal(rms)-to-standard ratio and is nearly constant and independent of the frequency of the ripple. That is, the different changes in spectral shape caused by varying k were equally detectable and the thresholds were all about -24 dB. In only one instance, $k = 9$ cosine ripple, did the threshold differ from the mean by more than 2 dB.

Let us now explore the question of how we might try to predict these data. Can we account for the detection of a rippled spectrum (an intensity change in several components of the complex) on the basis of the listener's ability to detect an intensity change in each individual component? In other words, can we predict the detection of a broad spectral change on the basis of the detection at each point along the spectrum? In experiment 2, we obtained data for the three listeners in a task requiring them to detect a change in a single component of a 21-component complex (Fig. 2, open triangles). We attempted to use those data (extrapolating the threshold from the seven measured frequencies to the remaining 14) to see if the detectability of rippled spectrum could be predicted on the basis of the detectability of the individual components. One of the simplest rules is the opti-

TABLE I. Average threshold value for different frequencies of ripple. Entry is signal rms to standard ratio in dB.

Frequency of ripple k	Sine condition	Cosine condition
1	24.6	24.1
2		23.[illegible]
3	24.0	−23.8
4		23.2
5	24.6	24.7
6		25.7
7	23.[illegible]	25.7
8		−23.[illegible]
9	25.[illegible]	26.3
10	24.6	−22.8
Mean	24.5	24.5

mum combination rule, in which the d' for the combined signal is the square root of the sum of individual d''s squared (see Green and Swets, 1966, p. 239; also Green, 1958). In experiment 5, we found that d' is proportional to signal pressure (see Fig. 6); thus we can determine the pressure at each component needed to achieve the observed level of detectability for the rippled spectrum.

Figure 8 shows the results of that calculation. The abscissa is the component number, 1 represents the 200-Hz component, and 21 represents the 5000-Hz component. The ordinate is the relative signal pressure (ratio of signal pressure to pressure of that component of the standard). The average threshold data for the single increment task are plotted on the scale of relative pressure and shown by the open circles. This is the same data shown in Fig. 2 (open triangles) except the data in that figure are plotted on a decibel scale. In addition, in Fig. 8, we have interpolated or extrapolated the threshold values for the missing frequencies. The relative threshold value at each component of the rippled spectrum is shown by the sinusoidal function marked with crosses. We have chosen the threshold data for a 1-cycle sine ripple—the thresholds for the different frequencies of ripple are so similar it matters little which frequency we select (see Table I). The threshold predicted by the optimum combination rule is shown by the smaller amplitude sinusoid indicated by the solid line. The difference between the predicted and obtained value is about 7 dB.

This is a sizable discrepancy, and it is not sensitive to details of the initial threshold values. To demonstrate the robust nature of this discrepancy, suppose the individual thresholds were all about equal and at a relative pressure value of 0.15 (a signal level − 16 dB below the level of the standard). There are 21 components, so that the square root of the sum of equal d''s is $(21)^{1/2} = 4.6$ or an expected improvement of 13.2 dB. The observed improvement is about 8

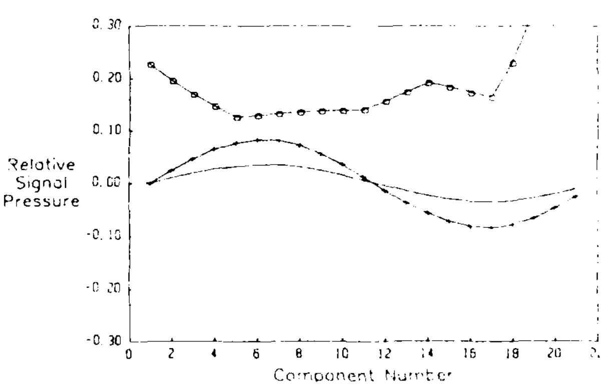


FIG. 8 Relative signal pressure to standard pressure as a function of the component number of the 21-component complex. The points marked by the open circles are the **average** threshold values for the detection of an increment on that component in a flat standard spectrum. The sinusoidal function marked by crosses is the threshold pressure for the detection of a ripple stimulus versus a flat spectrum. The sinusoidal function marked by the line is the predicted threshold pressure for this condition according to the optimum combination rule.

dB, from − 16 to − 24 dB, a discrepancy of 5.2 dB. Further, the direction of the discrepancy is similar to that found by Green and Kidd (1983). They compared an increment on a single component with increments added to all 21 components of the standard. The improvement in threshold was about 5 dB less than might have been expected by a optimum combination rule.

If one maintains the optimum combination rule, then the only avenue of escape is to argue that there are not 21 independent d's that contribute to the detectability of the complex spectral change, but some lesser number. The reduction needed to fit the data is sizable. If we want an 8-dB improvement, rather than 13 dB, then we conclude there are only 6.5 independent detectors contributing to the detection of the rippled spectrum. If we want only a 6-dB improvement, the number is 4 independent detectors. The assertion that there are only 4 or 6.5 independent detectors covering a frequency range of 4.6 oct (200 to 5000 Hz) would mean these bands span 1 oct each (assuming 4 detectors) or 3/4 oct each (assuming 6.5 detectors). The width of the widest critical band estimates is about 1/5 oct. This means the profile analysis band is between 4 and 2.5 times larger than a critical band. The assumption of such a wide profile analysis band, however is inconsistent with the mean threshold data. A 10-cycle ripple implies that a single cycle covers about 1/2 oct (4.6/10), or that both a peak and valley of the ripple fall in the same analysis band. The threshold for the 10-cycle ripple should, therefore, be elevated with respect to the thresholds for the all conditions with a lower frequency of ripple. The data (Table I) show the threshold is virtually independent of ripple frequency. We clearly need to make independent estimates of the width of these analysis bands before we can accept these numbers.

C. Effects of spectral density

The final question we wish to address is how the number of components in the spectrum might affect the ability to discriminate a flat from a rippled spectrum. In this last experiment, using the same three observers, we varied the number of components used to generate the spectrum for a single, low-frequency ripple, $k = 2$. As we varied the number of components in the spectrum, the logarithmic spacing was preserved; *that is, the ratio of successive components in* the spectrum was constant. This ratio can be computed from the formula: $\text{Ratio} = 10^{1.3979/(M-1)}$. We used the values $M = 3, 5, 11, 21, 41,$ and 81. For example, successive components of the 81-component waveform had ratios of 1.041, and the nearest components to the 1000-Hz component were 961 and 1041 Hz. The standard spectrum was always flat; that is, the components were all equal amplitude. Both a sine and cosine variation were used. For the three-component case, the ripple was simply an elevation at the 1000-Hz central component.

Figure 9 shows the data as a function of M, the number of components in the spectrum. The thresholds clearly decrease as the number of components increases to about 21, where the threshold value reaches about − 24 dB. As the number of components is increased, the function appears to

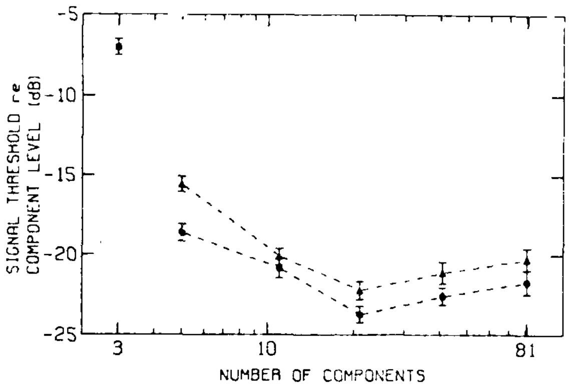


FIG. 9. The rms signal threshold for a 2-cycle ripple as a function of the number of components in the ripple M. Circles are thresholds for sine ripples and triangles for cosine ripples. The square represents the threshold for a three-component flat spectrum with an increment on the center, 1000-Hz component. Error bars are the standard errors calculated over 18 runs for three subjects.

rise only slightly. The cosine ripple appears to produce slightly poorer detection performance for all numbers of components, but the difference is small. A similar, but much smaller difference between the sin and cosine ripple is also apparent in Table I, at the lowest frequency of ripple.

The results obtained with our 21-component waveforms were taken with a sufficient number of components to obtain sensitive detection performance. For our frequency region, 200 to 5000 Hz, we suspect the ability to distinguish a flat spectrum from a rippled one is largely independent of the number of components used to define the spectrum, as long as at least 21 components are used. The small increase in threshold evident in Fig. 9 for 41 and 82 component densities is probably the result of the lack of extensive practice at those densities. Most of the observers' training in this sequence of experiments used a 21-component complex. Our 21 components span about 4.5 oct, a density corresponding to about 5 components per octave. The frequency spacing of the components near 1000 Hz is about 150 cycles, approximately one component per critical band. As Fig. 9 shows, once this density is achieved, the ability to detect a low-frequency ripple is essentially independent of component density.

IV. CONCLUSIONS

These studies have reported measurements on the ability to detect intensity changes in an equal-energy (flat) spectra. The intensity changes investigated were of two types. In one set of experiments, the change in intensity is limited to a narrow frequency region. In the other, the intensity changes occur over the entire spectrum.

For the changes confined to a narrow frequency region, the frequency at which the intensity change is produced influences the detectability of such a change only slightly. The midfrequency region (500 to 2000 Hz) appears to be where the smallest changes can be detected, but the extreme frequency region is worse by only a few decibels. Changes in spectral shape near 7000 Hz are more difficult to detect than the same type of change near 1000 Hz by about 12 dB (Fig. 4), and the change in threshold as a function of frequency is gradual.

The psychometric functions for a change in the intensity of a single component appear to be the same whether a single or multiple components are present. The function that approximates the form of the psychometric function is $d' = k$ (increment pressure).

For intensity changes occurring over a broader frequency region, the ability to detect a ripple over the range from 200 to 5000 Hz was essentially independent of the frequency of the ripple. Sine and cosine ripples were also nearly equal in detectability.

The comparison of the detection of broad changes versus narrow changes in the spectrum revealed an anomaly. The broader changes are more difficult to hear by some 5 to 7 dB than one would expect on the basis of a simple model that integrates the detectability over the separate frequency regions.

ACKNOWLEDGMENTS

This research was supported by grants from the National Institute of Health and the Air Force Office of Scientific Research. Dr. Virginia Richards and Dr. Les Bernstein helped improve the paper with their comments on an earlier draft. Dr. Robert Lutfi and Dr. Neal Viemeister materially contributed to the clarity of the paper.

Green, D. M. (1958). "Detection of multiple component signals in noise," J. Acoust. Soc. Am. 30, 904–911.

Green, D. M. (1960). "Psychoacoustics and detection theory," J. Acoust. Soc. Am. 32, 1189–1203.

Green, D. M., and Kidd, G., Jr. (1983). "Further studies of auditory profile analysis," J. Acoust. Soc. Am. 73, 1260–1265.

Green, D. M., and Mason, C. R. (1985). "Auditory profile analysis: Frequency, phase, and Weber's Law," J. Acoust. Soc. Am. 77, 1155–1161.

Green, D. M., Mason, C. R., and Kidd, G., Jr. (1984). "Profile analysis: Critical bands and duration," J. Acoust. Soc. Am. 75, 1163–1167.

Green, D. M., and Swets, J. A. (1966). *Signal Detection Theory and Psychophysics* (Wiley, New York) (reprinted by Krieger, Huntington, NY, 1974).

Rabinowitz, W. M., Lim, J. S., Braida, L. D., and Durlach, N. I. (1976). "Intensity perception. VI. Summary of recent data on deviations from Weber's law for 1000-Hz tone pulses," J. Acoust. Soc. Am. 59, 1506–1509.

Monaural envelope correlation perception

Virginia M. Richards
Department of Psychology, University of Florida, Gainesville, Florida 32611

(Received 24 February 1987; accepted for publication 16 July 1987)

The ability to discriminate between simultaneously presented 100-Hz-wide bands of noise with envelopes that were either similar or dissimilar was measured. The center frequencies of the noise bands, f_L and $f_L + \Delta f$ Hz, were systematically varied. When the bands of noise were separated by an octave, $\Delta f = f_L$, discriminations were at chance levels. For frequency separations less than an octave, $\Delta f < f_L$, discrimination was best for $f_L = 2500$ and 4000 Hz, somewhat poorer for $f_L = 1000$ Hz, and impossible for $f_L = 350$ Hz. Listeners were also asked to discriminate between bands of noise with envelopes that were either perfectly or partially correlated, and bands with envelopes that were either uncorrelated or partially correlated. The data suggest that, when transformed to an equal-variance scale (Fisher's z), equal changes in Fisher's z lead to equal changes in detectability, regardless of the correlation of the envelopes of the reference signal.

PACS numbers: 43.66.Nm, 43.66.Mk, 43.66.Fe [WAY]

INTRODUCTION

Although the importance of envelope synchrony has been studied extensively in the binaural domain, the importance of monaural envelope synchrony has received relatively little attention (see Goldstein, 1965, for a historical review). Two recent exceptions are the study of Schubert and Nixon (1970) and the comodulated masking release (CMR) studies introduced by Hall *et al.* (1984).

Schubert and Nixon (1970) examined the detectability of temporal synchrony for bands of noise of different center frequencies. In their experiment, low-pass noise was multiplied by two carriers, one at 350 Hz and a second one between 375 and 700 Hz. On some trials, the two bands were derived from a single source, and, on the remaining trials, the bands were derived from independent sources. Thus listeners were asked to indicate whether the simultaneously presented noise bands were synchronous (single generator) or independent (independent generators). Listeners were unable to make the discrimination.

More recent experiments have suggested that listeners are indeed able to compare envelopes extracted in different frequency regions. For example, Hall *et al.* (1984) showed that the masked threshold for a tone in a band of noise is reduced when a second, temporally synchronous band of noise is present at a frequency removed from the signal. They referred to this as comodulation masking release, or CMR. The extent of release varies from a few dB to as much as 10 dB depending on the separation in center frequencies of the two bands of noise. Further, CMR is observed for monaurally, diotically, and dichotically presented stimuli, and whether or not the second noise band is at a frequency higher than or lower than the noise band containing the signal (Hall *et al.*, 1984; McFadden, 1986; Cohen and Schubert, 1987).

A recurrent issue in CMR experiments is the importance of envelope correlation between the two bands of noise. In the "noise alone" interval, the two noise bands have identical envelopes. In the "signal plus noise" interval, the addition of the signal to one of the noise bands degrades the correlation between the envelopes. Thus the reduction in the masked threshold that defines CMR may reflect the discrimination of changes in correlation associated with the addition of a signal to one of the two noise bands.

Given the experiment of Schubert and Nixon, this may seem an unlikely explanation. However, Schubert and Nixon examined the discrimination of envelope correlation at relatively low frequencies (350 Hz), while the CMR experiments have yielded the largest effects at higher frequencies (1000 Hz and above). It seemed reasonable, then, to repeat the experiment of Schubert and Nixon at higher frequencies in order to determine whether envelope correlations are detectable at frequencies where CMR is measured.

Experiments 1 and 2 examine whether or not noise bands with identical envelopes can be discriminated from noise bands with envelopes that are statistically independent. A large portion of the auditory spectrum was examined in order to determine the effects of frequency region and frequency separation. Experiment 3 examines the auditory systems' sensitivity to changes in monaural envelope correlation. Finally, the data of experiment 3 are examined in an effort to determine the plausibility of the proposal that changes in envelope correlation are responsible for the reduction in masking observed in the CMR paradigm.

I. EXPERIMENT 1: MONAURAL ENVELOPE CORRELATION PERCEPTION AS A FUNCTION OF CENTER FREQUENCY AND FREQUENCY SEPARATION

A. Procedure

Listeners were asked to discriminate between two signals, one composed of two noise bands with identical envelopes (*paired* signal), the other composed of statistically independent noise bands (*independent* signal). For both signals, the two noise bands were centered at frequencies f_L (lower center frequency) and $f_L + \Delta f$.

1. Signal generation

In practice, the *paired* signals were the sum of two waveforms $w(t)$ and $w^*(t)$, each of which is the sum of several cosines:

$$w(t) = \sum_{i=-m}^{m} a_i \cos[2\pi(f_L + \delta i)t + \theta_i], \quad (1a)$$

$$w^*(t) = \sum_{i=-m}^{m} a_i \cos[2\pi(f_L + \Delta f + \delta i)t + \theta_i], \quad (1b)$$

where δ is the frequency separation of the tones added to generate the noise, f_L is the center frequency of the lower band, and Δf is the difference between the low and high center frequencies. Here, δ was 10 Hz and m was 5, yielding nominally 100-Hz-wide bands of noise. The *independent* bands were generated similarly, except that the a_i's and θ_i's of Eqs. (1a) and (1b) were independently chosen rather than being identical. Note that, for the *paired* signals, the noise bands are uncorrelated since they are centered at different (orthogonal) frequencies. The envelopes, however, are identical. In contrast, the bands that comprise the *independent* signals are statistically independent, both in their envelope and waveform characteristics.

All of the signals used were generated and played using an IBM PC microcomputer. For each $(f_L, f_L + \Delta f)$ frequency pair tested, 32 *paired* noise bands were generated. The low-frequency bands were generated first. The amplitude (a_i) of each tone was chosen at random from Rayleigh-distributed values and the phases (θ_i) from values uniformly distributed between zero and 2π. Next, the associated high-frequency bands were generated. These 11 tones had the same amplitudes and phases as their low-frequency counterparts, but the frequencies were increased by a value of Δf Hz.

After the 32 *paired* noise bands were computed and stored, the stimuli used in the experiment were selected in the following manner. First, one set of *paired* noise bands was chosen at random and the two bands were added in order to produce a *paired* signal. In order to generate an *independent* signal, a low- and a high-frequency band were chosen, and added if their envelopes were not the same (i.e., they did not form a pair).

Figure 1 illustrates two hypothetical signals, a *paired* signal indicated on the top, and an *independent* signal shown on the bottom. Also shown are the amplitude spectra and the waveforms associated with the two noise bands that are combined to make the signal. In this example, f_L was 350 Hz and $f_L + \Delta f$ was 630 Hz. Note that the bands of noise that comprise the *paired* signal have identical envelopes and the same relative amplitude spectra.

The stimuli were played through a 12-bit D/A converter at a sampling rate of 14.3 kHz and low-pass filtered (Kemo VBF/23) at 6 kHz. The signal duration was 100 ms, including 5-ms cosine-squared onset/offset ramps.

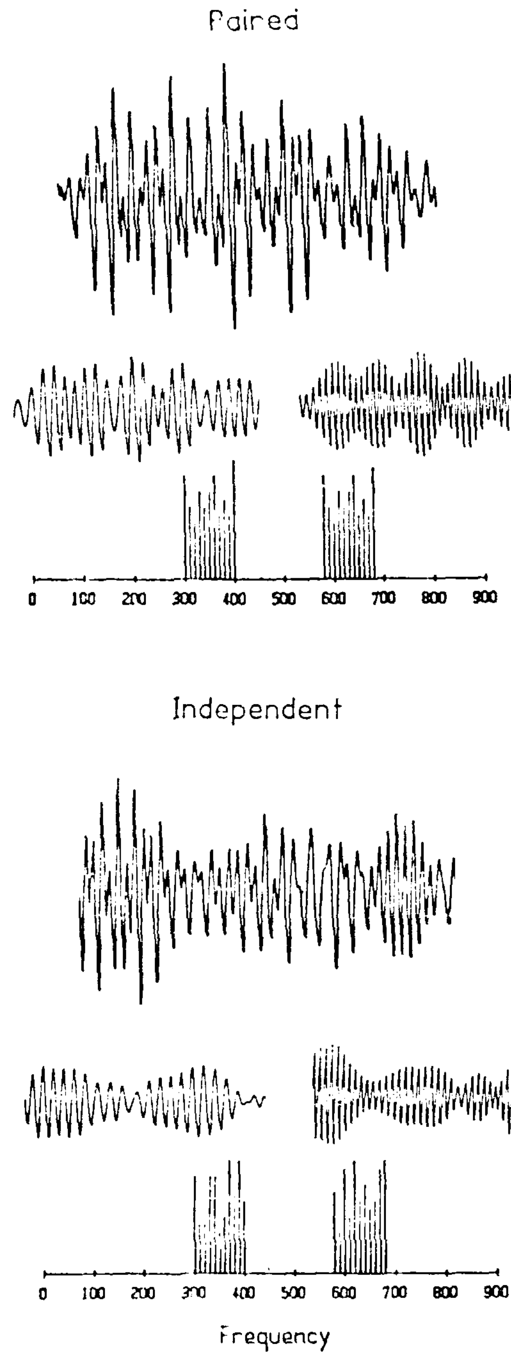


FIG. 1. Two stimuli are shown. The top panel presents the *paired* noise bands, whose envelopes are identical. The bottom panel shows two *independent* noise bands, whose envelopes are also independent. The summed waveform, amplitude spectra, and the waveforms of the individual bands of noise are drawn.

2. Method

In one interval of a 2IFC paradigm, the signal was *paired*, in the other interval the signal was *independent*. Listeners indicated which of the two intervals contained the *paired* bands. The dependent variable was the percent correct discriminations.

Four values of f_L were tested: 350, 1000, 2500, and 4000 Hz. Table I indicates the 17 $(f_L, f_L + \Delta f)$ pairs tested. The difference between the center frequencies of the two noise bands is in terms of $\Delta f/f_L$, the normalized frequency separa-

TABLE I. The center frequencies of the bands of noise used in experiment 1. The left-hand column indicates the lower center frequency f_L, and the body of the table indicates the associated higher center frequencies tested, $f_L + \Delta f$. Along the top, the relative frequency separation, $\Delta f / f_L$, is indicated.

		Relative frequency separation ($\Delta f / f_L$)							
		<0.1	0.1	0.2	0.3	0.4	0.6	0.8	1.0
f_L,	350				460	490		630	
Low	1000		1100	1200		1400		1800	
center	2500	2650	2750	3000		3500	4000	4500	5000
freq.	4000	4150	4400			5600			

tion. The extent of separation (Δf) ranged between 100 Hz (the noise bandwidth) and an octave.

Subjects listened in a sound-treated room. The stimuli were presented diotically via Sennheiser HD 414 SL earphones at an average level of 65 dB SPL per component, or a spectrum level of 55 dB SPL. The signal durations were 100 ms, the two listening intervals were separated by 300 ms, and both intervals were indicated on a display screen. Following the subject's response, feedback was displayed for 240 ms. Conditions were tested in 50 trial blocks, 6 blocks at a time. Thus each data point is based on 300 trials. Subjects typically completed 20 blocks a day and conditions were completed in random order.

Listeners were undergraduate students paid to participate, except GR who is the author. All had normal hearing. Listeners first heard the (2500, 2750)-Hz condition. After completing the first few sets of 50 trials, further practice was not needed. Throughout the experiment, several conditions were repeated. In no case was an effect of practice evident. For the repeated conditions, only the last 300 trials contributed to the data reported.

B. Results and discussion

Figure 2 shows the percent correct identifications as a function of normalized frequency separation, $\Delta f / f_L$, between the two bands of noise. Since the two noise bands were centered at f_L and $f_L + \Delta f$, a $\Delta f / f_L$ of 1 indicates that the two bands were separated by an octave. Data for two conditions, $f_L = 2500$ (circles) and $f_L = 4000$ (triangles) Hz, are presented. The data for each of the three subjects are plotted separately.

When the low-frequency noise band was centered at either 2500 or 4000 Hz, performance deteriorated as the separation between the two noise bands was increased. When the relative separation was at or below 0.1, performance was nearly perfect. As the separation approached an octave, performance approached chance levels. Extrapolating from the averaged data, discriminability reached 75% at $\Delta f / f_L \simeq 0.3$.

Figure 3 shows the data for $f_L = 350$ and 1000 Hz. Again, the parameter is f_L. These results differ from those observed at higher frequencies (Fig. 2). When f_L was 1000 Hz, performance did not change monotonically with Δf. Rather, the best performance was observed for frequency separations between 200 and 400 Hz. When f_L was 350 Hz,

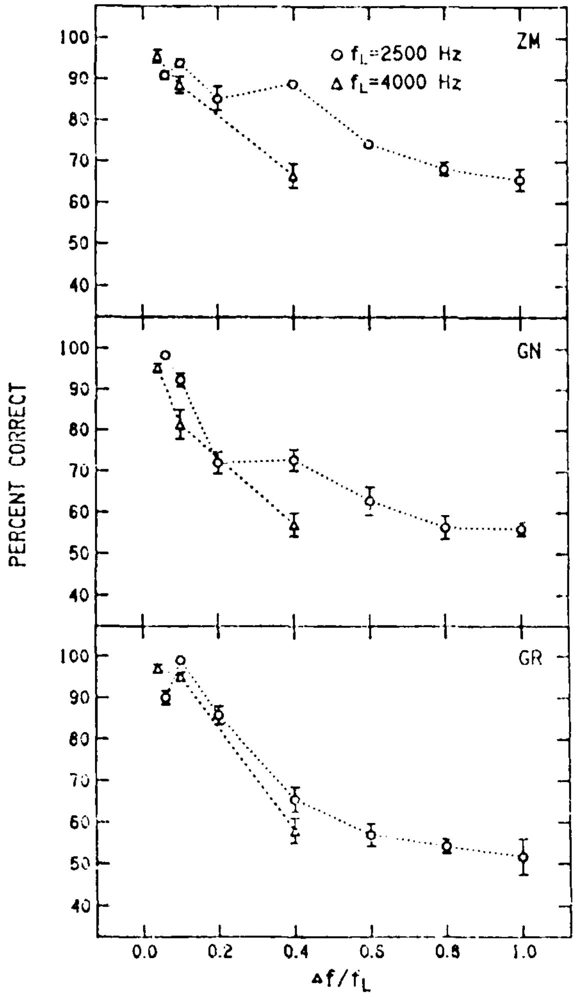


FIG. 2. Percent correct is plotted as a function of relative frequency separation, $\Delta f / f_L$. The lower center frequency was either 2500 (○) or 4000 (△) Hz. Data for the three subjects are plotted separately. Error bars indicate one standard error of the mean.

listeners performed at chance levels. The 350-Hz data provide a replication of Schubert and Nixon's (1976) findings.

It is clear that listeners are able to detect whether or not simultaneously presented noise bands have correlated or uncorrelated envelopes. But, as has been pointed out by McFadden (1987), we must be careful not to envision the proximal stimulus as two separately resolved bands of noise. For example, in the high-frequency conditions ($f_L = 2500$, 4000 Hz), performance was best when the bands of noise were closest, i.e., those conditions in which interactions within a single critical band were most likely to occur.

The fact that performance is poor when the noise bands are centered below 1000 Hz is consistent with the hypothesis that temporal envelopes are important for discrimination. Our results are in line with those of Henning (1980) and Henning and Ashton (1981), who found that the detectability of the interaural delay of the envelopes of sinusoidally amplitude-modulated tones was poor until the center frequency of the signal exceeded 1000 Hz. Further, Bernstein

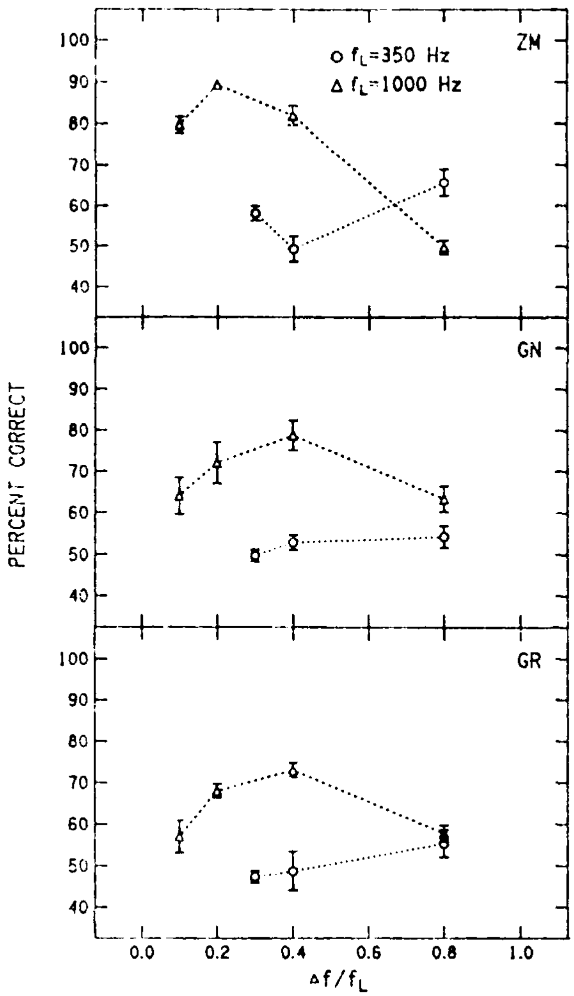


FIG. 3. Percent correct is plotted as a function of relative frequency separation, $\Delta f/f_L$. The lower center frequency (f_L) was either 350 (○) or 1000 (Δ) Hz. Data for the three subjects are plotted separately.

and Trahiotis (1985) showed that the envelope delay of sinusoidally amplitude-modulated signals contributed little to intracranial movement when the center frequency of the signal was smaller than 1000 Hz, but that envelope delays lead to relatively large intracranial movement when the center frequency exceeded about 1000 Hz.

Unfortunately, the comparison between the low- (f_L = 350, 1000 Hz) and high-(f_L = 2500, 4000 Hz) frequency regions may have been contaminated by the use of a fixed noise bandwidth. When the center frequencies are large, the 100-Hz band of noise is presumably narrow enough to pass through a single critical band. At lower frequencies, however, the noise bandwidths may have exceeded critical bandwidths. Thus, at low frequencies, the opportunity for envelope comparisons may have been limited by the fact that the output of no two critical bands had identical envelopes. For this reason, the experiment was repeated at the lower frequencies, with the change that narrower noise bands were used.

Additionally, the experiment was repeated in the presence of low-pass noise. This control was included in an effort to reduce the subjects' reliance on low-frequency distortion products that may have been present due to the nonlinear nature of either the signal generation apparatus or the auditory system.

II. EXPERIMENT 2A: THE EFFECT OF BANDWIDTH

A. Procedure

The procedure was nearly identical to that of the first experiment, except that relatively narrow bandwidths were used. Two frequency regions were tested, f_L = 350 and 1000 Hz. When the low-frequency band was centered at 350 Hz, three tones made up the noise band, yielding a nominally 20-Hz-wide band of noise. When the low-frequency band was centered at 1000 Hz, a nominally 40-Hz-wide band of noise was used. These bandwidths were chosen so that the bandwidth of the lower noise band relative to its center frequency was approximately that used in experiment 1 when f_L was 2500 Hz (i.e., a relative width of 0.04). As in experiment 1, the spectrum level was set at 55 dB SPL. The same listeners participated, and again little or no practice was needed.

B. Results and discussion

Figure 4 shows percent correct as a function of normalized frequency separation, $\Delta f/f_L$, of the two bands of noise. The comparison of interest is with the data presented in Fig. 3. For GN, there is little effect of reduced bandwidth for any of the conditions tested. For the other two listeners, ZM and GR, performance at the smallest frequency separation for f_L = 350 Hz appears to have exceeded chance. Otherwise, for f_L = 350 Hz, performance was unchanged. In the 1000-Hz condition, ZM and GR performed more poorly when the narrower bandwidths were used.

Across all listeners and center frequencies, the reduction in bandwidth did not lead to performance levels as good as those found at higher frequencies (Fig. 2). Thus the inability to detect changes in envelope correlations at relatively low frequencies does not appear to be due to limitations imposed by the bandwidth of the noise.

These data present two curious interactions for which we have no explanation. First, the between-listener differences (ZM and GR vs GN) are quite striking. Second, when there is an effect, reducing the noise bandwidth is marginally advantageous in one instance (f_L = 350 Hz), and detrimental in the other (f_L = 1000 Hz).

III. EXPERIMENT 2B: THE EFFECT OF LOW-PASS NOISE

A. Procedure

The procedure was similar to that of the first experiment, except that low-pass noise was continuously present (noise source: General Radio Company, #1382, low-pass filter: Kemo VBF/23). This experiment was included in order to determine whether low-frequency distortion products were affecting the discrimination of envelope correlations. The spectrum level of the low-pass noise was fixed at 35 dB SPL, and the cutoff frequency was located approximately

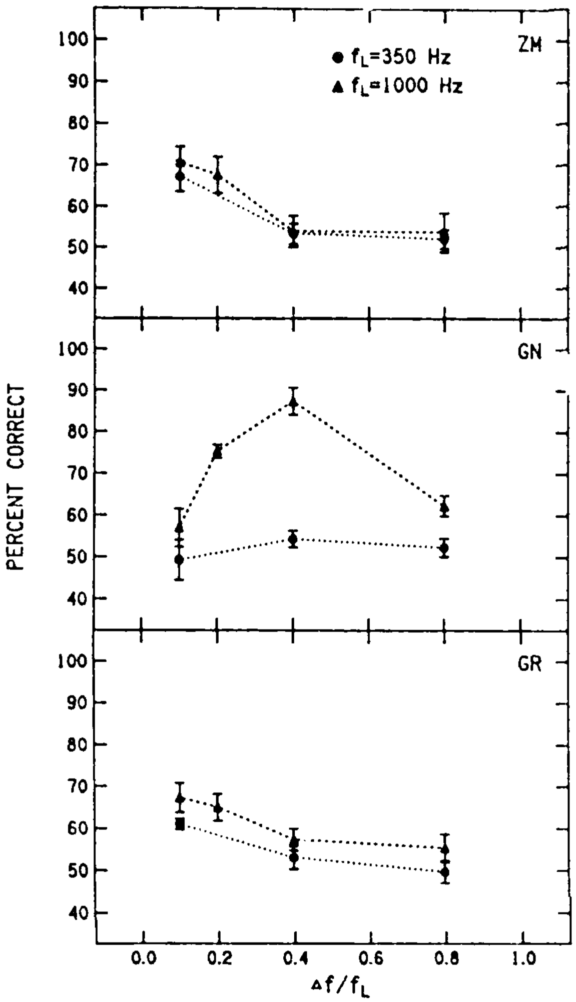


FIG. 4. The same as Fig. 3 except that narrower noise bands were used. The noise bands had bandwidths of 20 and 40 Hz for f_L's of 350 (●) and 1000 (▲) Hz, respectively. Data for the three subjects are plotted separately.

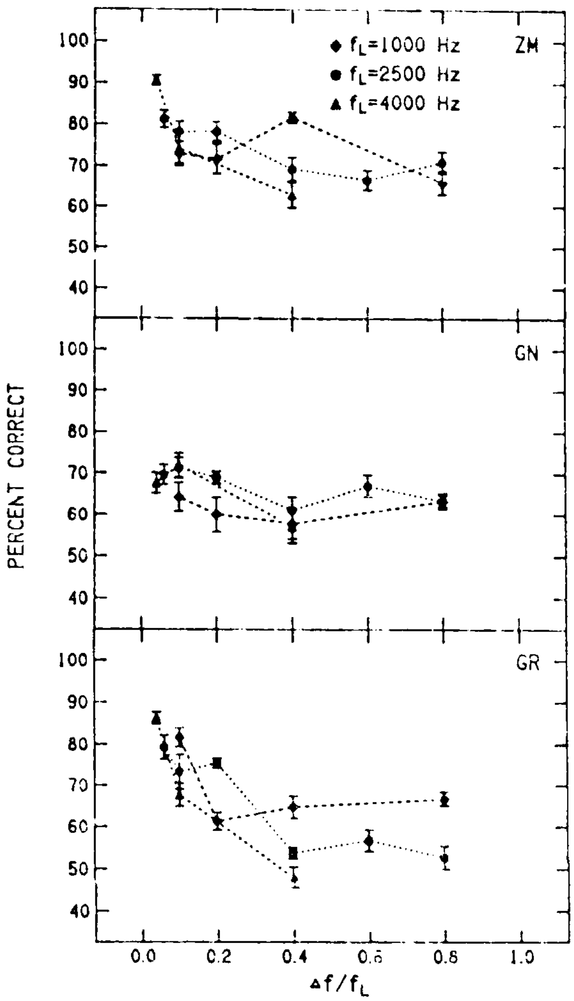


FIG. 5. The same as Figs. 2 and 3 except the data were collected in the presence of low-pass noise. Three f_L's were tested: 1000 (◆), 2500 (●), and 4000 (▲) Hz.

one-third of an octave below the lowest component of the lower noise band. At the frequency of the lowest component of the narrow band of noise, the low-pass noise was approximately 60 dB below the level of the narrow band of noise. Three frequency regions, f_L = 1000, 2500, and 4000 Hz, were tested. In this experiment, subjects improved with practice, and so several practice runs were needed in order for the subjects to achieve asymptotic levels of performance.

B. Results and discussion

Figure 5 shows performance obtained when low-pass noise was present. Three frequency regions were tested, f_L = 1000, 2500, and 4000 Hz. Comparing these data to the data of experiment 1 (Figs. 2 and 3), it can be seen that low-pass noise depresses performance, especially when the performance levels had been high. For the smaller frequency separations ($\Delta f/f_L$ <0.4), performance was on average 11% lower when low-pass noise was present. Whether this reduction reflects the fact that the added low-pass noise interfered with the lower noise band or the masking of nonlinear distortions is not obvious. In any event, information introduced via nonlinear distortions cannot be the sole cue used for discrimination since scores of 90% were obtained even when the experiment was completed in the presence of low-pass noise.

IV. EXPERIMENT 3: DETECTING CHANGES IN ENVELOPE CORRELATION

This experiment differs from those previously described in that the envelopes of the low- and high-frequency noise bands were no longer either identical (*paired* condition above) or independent. Instead, noise bands with envelopes that were partially correlated were also used. In this experiment, the function relating changes in envelope correlation and discriminability was determined. To this end, listeners discriminated between signals composed of two bands of noise whose envelopes were either partially correlated or unedependent and noise bands whose envelopes were either partially or fully correlated.

A. Procedure

1. Signal generation

The partially correlated stimuli were created in a manner similar to the "three generator case" described by Licklider and Dzendolet (1948; see also Jeffress and Robinson, 1962). The combination algorithm used to generate the stimuli was as follows. First, 32 *paired* noise bands were computed [Eqs. (1a) and (1b)]. As before, the *paired* bands had identical envelopes. Adding high- and low-frequency bands that did not form a pair yielded stimuli whose envelopes were independent.

The stimuli whose envelopes were only partially correlated were generated by combining four noise bands, two *paired* bands, $w(t)$ and $w^*(t)$, and two *independent* bands, $u(t)$ and $v^*(t)$:

$$S(t) = \alpha[w(t) + w^*(t)] + \beta[u(t) + v^*(t)]. \qquad (2)$$

In order to achieve a particular correlation, the *paired* high/low bands were multiplied by one constant (α), and the *independent* high/low bands were multiplied by a second constant (β). The constraint that $\alpha^2 + \beta^2 = 1$ ensured that stimuli whose envelopes were partially correlated retained the same average level as the primary signals.

The correlation coefficients, which depend on α and β, were computed empirically. They represent the correlation of the resultant high- and low-frequency envelopes, not the correlation of the high- and low-frequency waveforms (which is zero due to the different center frequencies). In order to evaluate the correlations, 100 simulations were completed. Waveforms were computed according to Eqs. (1)-(3), the high- and low-frequency envelopes were extracted, and the Pearson moment correlation coefficient (r) was computed. The correlation coefficients were then transformed to normally distributed z scores according Fisher's r to z transformation, $z = 1/2[\ln(1+r) - \ln(1-r)]$ (McNemar, 1969). The resulting 100 Fisher's z scores were then averaged, and the average value transformed back to r. The resulting, averaged correlation coefficients are used to identify the stimuli. This somewhat arduous process was followed so that a single r could be used in order to identify the stimuli.

2. Method

In a 2IFC paradigm, listeners discriminated between diotically presented signals, each composed of two 100-Hz-wide noise bands. The low-frequency (f_L) noise band was centered at 2500 Hz and the high-frequency ($f_L + \Delta f$) band was centered at 2750 Hz. The change in envelope correlation was the independent variable, and percent correct discrimination was the dependent variable.

One of the listeners, GR, had participated in experiments 1 and 2 prior to this experiment. The others were naive concerning this type of experiment, but had participated in other auditory experiments. As GN was unable to complete the experiment, VF was recruited midway through the experiment. All naive listeners began with the correlated versus uncorrelated condition, and data collection began after approximately 150 practice trials. Conditions were occasionally repeated, and there was no evidence of improvement throughout the experiment.

B. Results and discussion

Figure 6 shows the percent of correct discriminations as a function of r for the condition in which one interval contained bands of noise whose envelopes were independent and the other interval bands with envelopes that were partially correlated. Figure 7 presents the data for the fully versus partially correlated condition. Extrapolating from the data of Fig. 6, we see that a change in correlation of approximately 0.6 is expected to be needed in order to discriminate, with 75% accuracy, those noise bands with partially correlated envelopes from those with uncorrelated envelopes. In contrast, a change in correlation of only 0.15 is expected to be required in order to discriminate, with 75% accuracy, between noise bands of partially versus fully correlated envelopes.

The difference in correlation needed to detect a change from fully correlated ($\Delta r \approx 0.15$) as compared to the change needed to detect a change from zero correlation ($\Delta r \approx 0.6$) does not imply unequal sensitivity. Consider the theoretical probability density function for correlation coefficients shown in the top portion of Fig. 8. These distributions assume that the underlying population is bivariate normal (Bickel and Doksum, 1977). The histograms observed for our stimuli are also indicated. When r is near zero, the sampled distribution is symmetric and roughly bell shaped. As r increases, the distributions become increasingly skewed since values of r larger than 1 cannot occur. Finally, when the population correlation is 1, there is no variance since any draw will result in a correlation of 1.

In order to consider the auditory system's sensitivity to changes in correlation, a random variable whose variance is independent of its mean is desirable. The Fisher's r to z trans-

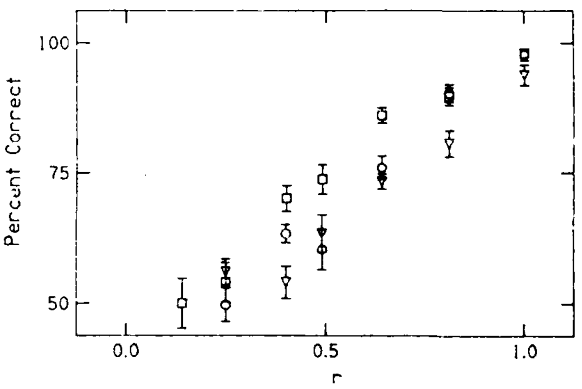


FIG. 6. Percent correct is plotted as a function of the correlation of the partially correlated signal. The envelopes of the standard noise bands were independent. Data for GN (▽), GR (□), and MS (○) are shown.

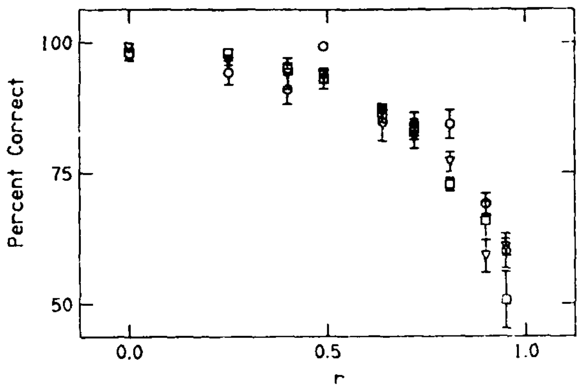


FIG. 7. Percent correct is plotted as a function of the correlation of the partially correlated signal. The envelopes of the standard noise bands were fully correlated. Data for VF (▽), GR (□), and MS (○) are shown.

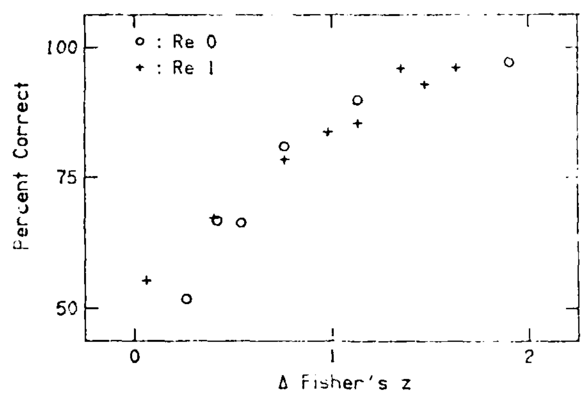


FIG. 9. Average percent correct is plotted as a function of Δ Fisher's z rather than r. (+) represents a reference correlation of one and (○) represents a reference correlation of zero.

form, indicated in Fig. 8, achieves this goal. Fisher's z is a normally distributed random variable, with a variance that is independent of its mean, z. (Simulations indicate that the transform may be reasonably applied to noise envelopes, even though the underlying distributions are Rayleigh rather than normal.) Following the signal detection approach, we assume that the observer's performance is dictated by changes in d'. It follows that detecting equal changes in z should yield equal discriminability, regardless of the standard. Thus one may expect that transforming the data of Figs. 6 and 7 from r to Δz should lead to more similar curves.

Figure 9 presents the data of Figs. 6 and 7 on a single abscissa, Δ Fisher's z. This figure was derived in the following manner. First, r's were transformed to z's. For the case in which the reference signal had uncorrelated noise bands, Δz is simply the z associated with the partially correlated signal (i.e., the z associated with an r of 0 is zero, so $\Delta z = z - 0 = z$). Next, consider the case in which the reference signal had perfectly correlated envelopes. In theory, an r of 1 is transformed to a z of infinity. Here, a noninfinite z was chosen so that the percent correct identifications were best predicted (according to the least squares criteria). The z so determined had a value of 1.9 and accounted for some 98% of the data's variance. Thus the Δz associated with the partially correlated envelopes was taken to be $1.9 - z$. Note that only the z associated with an r of 1 was altered (changes in other nonzero values were so small as to be insignificant). Mapping an r of 1 to a z of 1.9 may be viewed as a measure of the internal jitter or noise associated with the auditory and/or decision making system.

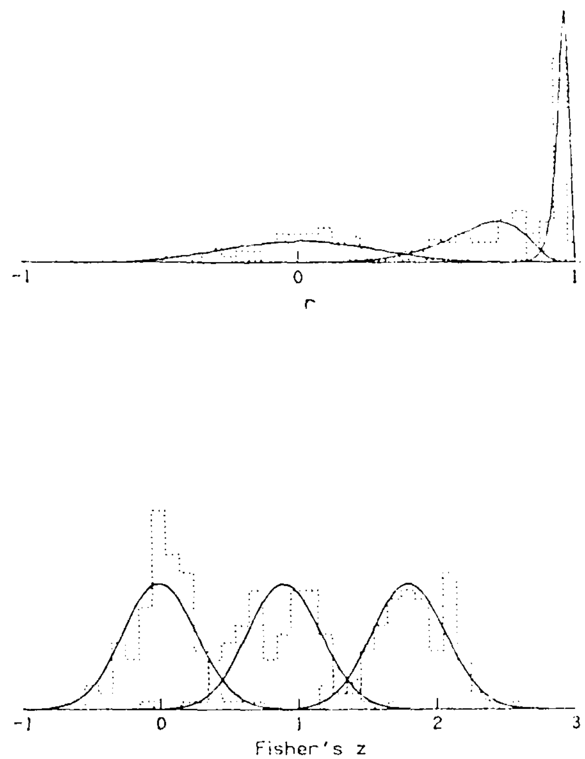


FIG. 8. The distribution of correlation coefficients is indicated on the r axis. The transformed distributions are indicated on the z axis. Solid lines represent theoretical distributions, and dashed lines are the observed histograms.

Figure 9 is based on the averaged data of subjects GR and MS (the only two that participated in both portions of this experiment). When performance is plotted as a function of Δz, the curves are essentially the same. Thus sensitivity to changes in envelope correlation appear to be independent of the reference correlation, provided the data are presented in terms of Δz rather than r.

In the Appendix, the binaural data of Pollack and Trittipoe (1959) and Gabriel and Colburn (1981) are considered. In their studies, subjects indicated differences in the correlation of noises presented to the two ears. The Appendix presents these data in terms of changes in Fisher's z rather than changes in correlation. In general, the binaural data do not conform to the "equal changes in Fisher's z leads to equal discriminability" rule to the extent that the current data do.

V. GENERAL DISCUSSION

A. Comparisons with CMR

Having established that changes in envelope correlation are discriminable, we may address the apparent contradiction between the data of Schubert and Nixon (1970) and the CMR data of Hall *et al.* (1984) and others. It seems that Schubert and Nixon failed to demonstrate discriminability because they tested only low-frequency regions (350 Hz). We have shown that changes in envelope correlation are indeed detectable, but it remains to be shown that envelope correlations *per se* are responsible for the reduction in masking observed in the CMR paradigm.

These experiments are not sufficient to resolve this issue. The data of experiment 3 may, however, be used in order to examine the plausibility of such an argument. Based on Green *et al.* (1964), a 2500-Hz signal in noise should be detected when an E/N_0 of approximately 12 dB is used. Table II presents both Fisher's *z* scores and correlation coefficients for the case in which a tone is added to one of two noise bands whose envelopes are otherwise identical. The tone was added to a 100-Hz-wide band of noise centered at 2500 Hz. Several values of E/N_0 were considered, changes in values having been achieved by changing the level of the added 2500-Hz tone. The *z* scores are based on 100 computer simulations. Also indicated are the observed standard deviations and the expected percent correct for each E/N_0 considered. These predictions are based on the data of Fig. 7.

As is clear from Table II, an E/N_0 between 0 and 5 dB should be sufficient to detect the presence of a 2500-Hz signal if detection is based on changes in envelope correlation. The estimated signal-to-noise ratio is about 10 dB below the expected threshold for the detection of 2500-Hz tone in noise. This drop in E/N_0, about 10 dB, is in line with the change in threshold of about 5 to 10 dB expected based on the CMR experiments of Cohen and Schubert (1987), who used somewhat similar stimulus parameters.

These rough estimates indicate that it is reasonable to argue that envelope correlation and CMR experiments reflect similar processing, namely, the discriminability of changes in envelope correlation. The generality of this argument is, however, limited. At least two recent studies (Hall *et al.*, 1987; Schooneveldt and Moore, 1987) have demonstrated that CMR may be obtained at low frequencies. In contrast, the current data, and those of Schubert and Nixon (1970), establish that changes in envelope correlations are not detectable at low frequencies. This is one area in which envelope correlation discrimination and CMR are at odds. Whether this difference reflects a difference between CMR and envelope correlation for noise bands centered at both high and low frequencies remains to be determined.

TABLE II. Estimates of the envelope correlation between two bands of noise whose envelopes would be identical, except that a tone is added to one of the bands. Several values of E/N_0 were simulated, and the resulting Fisher's *z* scores (*z*), observed standard deviation of *z* scores (s_z), and correlations (*r*), are shown. The expected performance levels [*E* (% correct)], derived from Fig. 7, are also indicated.

E/N_0	*z*	s_z	*r*	*E* (% correct)
15	0.36	0.42	0.34	94
10	0.60	0.31	0.54	92
5	[illegible]	0.32	0.77	80
0	1.59	0.23	0.90	65
−5	[illegible]	0.22	0.97	58

VI. CONCLUSIONS

These experiments lead to four conclusions.

(1) There is a frequency region (f_L,Δf) for which changes in envelope correlations can be detected and, in this region, performance depends on Δf. In general, as the separation between the bands of noise is increased, performance levels drop, approaching chance as the frequency separation approaches 1 oct.

(2) The discrimination of noise envelopes is not derived solely from combination tones introduced by nonlinearities in the auditory system.

(3) The discrimination of changes in monaural envelope correlation appears to be independent of the standard or reference correlation.

(4) If CMR is considered in terms of the changes in envelope correlation that are concomitant with the addition of a tone to one of two noise bands whose envelopes are otherwise identical, the current data predict reasonable values for the release in masking measured for bands of noise centered at high frequencies. In contrast, CMR may be obtained using low-frequency bands of noise, but envelope correlations do not appear to be extracted at low frequencies.

ACKNOWLEDGMENTS

This work was supported, in part, by AFOSR and, in part, by an NIH post-doctoral fellowship. Portions of this article were presented at the 112th meeting of the Acoustical Society of America in Anaheim. I thank Dr. David M. Green, Dr. Leslie R. Bernstein, and Dr. Tim Forrest for helpful comments throughout this project as well as on earlier drafts of the article. I also thank J. W. Hall III and an anonymous reviewer for comments.

APPENDIX

Here, the binaural correlation data of Pollack and Trittipoe (1959) and Gabriel and Colburn (1981) will be considered. In their experiments, Pollack and Trittipoe (1959) asked listeners to discriminate changes in the interaural correlation of wideband noise. For the case in which the reference correlation was 0, a change in correlation[1] of 0.4 was needed for the subjects to perform at a level of 75% correct. When the reference correlation was 1, a change in correlation of 0.04 was needed for the subjects to perform at a level of 75% correct. Although the authors mention calculations concerning changes in Fisher's *z*, the extent of evaluation is not indicated. Figure A1 presents their data against the measure of Δ Fisher's *z*. Although their study indicates superior sensitivity (an interaural correlation of 1 was mapped to a *z* of 2.4), the function relating percent correct to changes in Fisher's *z* is quite similar to ours (Fig. 9). In general, equal changes in Fisher's *z* lead to equal discriminability. It should be noted that, for these binaural studies, the reported correlation coefficients are in terms of whole-waveform correlations rather than envelope correlations, and that waveform correlations are larger than correlations based on envelopes

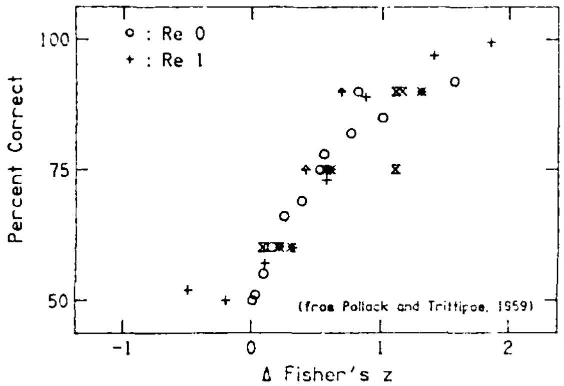


FIG. A1. As Fig. 9 except that the data are taken from the dichotic experiment of Pollack and Trittipoe (1959). Again, (+) indicates a reference correlation of 1 and (○) indicates a reference correlation of 0. Several other reference correlations are included, 0.19 (Δ), 0.5 (▽), 0.76(*), and 0.905 (x̄). The z's are based on whole waveform correlations,[1] not the correlations between envelopes.

alone (except at $r = 0$ and $r = 1$, where whole waveform and envelope correlations are the same).

Figure A2 plots data presented by Gabriel and Colburn (1981; subject DO). Again, percent correct is plotted as a function of Δ Fisher's z. Gabriel and Colburn used three stimulus conditions: noise bands centered at 500 Hz that had bandwidths of either 3 or 115 Hz and a 4.7-kHz low-pass noise. The reference correlations were either 0 or 1. The top graph of Fig. A2 is for the condition in which the reference correlation was 1, and the bottom graph the condition in which the reference correlation was 0. When the reference correlation was 1, changes in *correlation* of 0.006, 0.015, and 0.04 yielded 75% correct discriminations for the stimulus conditions of 3 Hz, 115 Hz, and 4.7 kHz, respectively. For the same conditions, changes in Fisher's z of 0.9, 0.4, and 0.3, respectively, were needed. When the reference correlation was 0, changes in *correlation* of 0.7, 0.4, and 0.28, respectively, yielded 75% correct discriminations. In these conditions, changes in Fisher's z of 0.95, 0.8, and 0.6, respectively, were needed.

The reader should recognize the unusual finding that, when the reference correlation was 1, increased bandwidth led to poorer performance. The fact that the change in Fisher's z is monotonic with bandwidth does not reflect a quick fix of this situation. Rather, it is a manifestation of the model's failure to adequately describe the data: Contrary to the theoretical expectation of equal variance, the estimated distributions in the 115-Hz and 4.7-kHz conditions had variances that depended on the reference signal's correlation. In Fig. A2, the difference in variance is manifested as a change in slope, the slope being shallower (larger variance) when the reference correlation was 1. Only in the 3-Hz condition do equal changes in Fisher's z lead to equal discriminability.

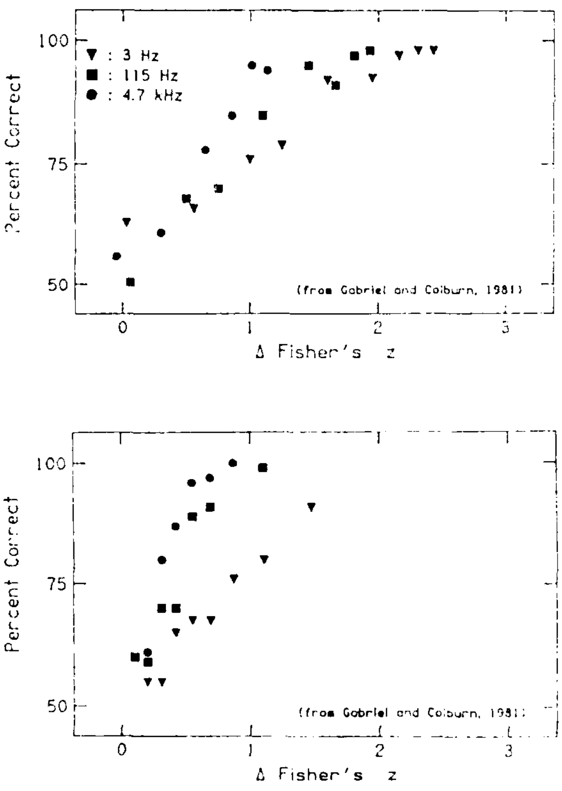


FIG. A2. Taken from Gabriel and Colburn (1981, Fig. 3). Percent correct discriminations is plotted as a function of Δz. Two reference correlations were used, 1 (top panel) and 0 (bottom panel). Three dichotically presented stimuli were used, two narrow-band noise signals centered at 500 Hz, bandwidths of 3 (▼) and 115 (■) Hz, and a 4.7-kHz low-pass noise (●).

[1]The correlations reported by Pollack and Trittipoe were actually squared correlations (see Jeffress and Robinson, 1962).

Bernstein, L. R., and Trahiotis, C. (1985). "Lateralization of sinusoidally amplitude-modulated tones: Effects of spectral locus and temporal variation," J. Acoust. Soc. Am. 78, 514–523.

Bickel, J., and Doksum, K. A. (1977). *Mathematical Statistics: Basic Ideas and Selected Topics* (Holden-Day, San Francisco, CA).

Cohen, M. F., and Schubert, E. D. (1987). "Influence of place synchrony on detection of sinusoids," J. Acoust. Soc. Am. 81, 452–458.

Gabriel, K. J., and Colburn, H. S. (1981). "Interaural correlation discrimination: I. Bandwidth and level dependence," J. Acoust. Soc. Am. 69, 1394–1401.

Goldstein, J. L. (1965). "An investigation of monaural phase perception," Doctoral dissertation, The University of Rochester, Rochester, NY.

Green, D. M., McKey, M. J., and Licklider, J. C. R. (1964). "Detection of a pulsed sinusoid as a function of frequency," in *Signal Detection and Recognition by Human Observers*, edited by J. Swets (Wiley, New York).

Hall, J. W., III, Haggard, M. P., and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am. 76, [illegible]–56.

Hall, J. W., III, Haggard, M. P., and Harvey, D. G. (1984). "Release from masking through ipsilateral and contralateral comodulation of a flanking band," J. Acoust. Soc. Am. Suppl. 1 76, S76.

Hall, J. W., III, Grose, J. H., and Haggard, M. P. (1987). "Comodulation masking release for complex signals," submitted to J. Acoust. Soc. Am.

Henning, G. B. (1980). "Some observations on the lateralization of complex waveforms," J. Acoust. Soc. Am. 68, 446–454.

Henning, G. B., and Ashton, J. (1981). "The effect of carrier and modulation frequency on lateralization based on interaural phase and interaural group delay," Hear. Res. 4, 185–194.

Jeffress, L. A., and Robinson, D. E. (1962). "Formulas for the coefficient of interaural correlation of noise," J. Acoust. Soc. Am. 34, 1658–[illegible].

Licklider, J. C. R., and Dzendolet, E. (1948). "Oscillographic scatterplots illustrating various degrees of correlation," Science 107, 121–124.

McFadden, D. (1986). "Comodulation masking release: Effects of varying

level, duration, and time delay of the cue band," J. Acoust. Soc. Am. 80, 1658–1667.
McNemar, Q. (1969). *Psychological Statistics* (Wiley, New York).
Pollack, I., and Trittipoe, W. J. (**1959**). "Binaural listening and interaural cross correlation," J. Acoust. Soc. Am. **31**, 1250–1252.
Schooneveldt, G. P., and Moore, B. C. J. (**1987**). "Comodulation masking release (CMR): Effects of signal frequency, flanking-band frequency, masker bandwidth, flanking-band level, and monotic versus dichotic presentation of the flanking band," J. Acoust. Soc. Am. (in press).
Schubert, E. D., and Nixon, J. C. (**1970**). ONR Tech. Rep. NR 140, 253.

Components of monaural envelope correlation perception

Virginia M. Richards
Psychoacoustics Lab
Department of Psychology
Univ. of Florida, Gainesville, FL 32611

In Press: Hearing Research.

ABSTRACT

The ability of listeners to discriminate between simultaneously presented bands of noise whose envelopes were either the same or statistically independent was determined. Bands of 100-Hz wide noise were employed which had low and high center frequencies of (2500,2750), (2500,3000), (2500,3500) and (4000,4400) Hz. Average discriminations were above 90 percent correct except for the (2500, 3500) Hz condition, which yielded an average of 77 percent correct. Next, a factorial stimulus design was employed in order to determine the relative importance of envelope and power spectrum cues. The results indicate that in the absence of power spectrum cues, bands with the same envelopes could be discriminated from bands with statistically independent envelopes. When the envelopes were always the same, listeners were able to discriminate between power spectra that were either the same or different. In contrast, when the envelopes were always different, listeners were unable to discriminate between the same and different power spectra.

Key Words: Comodulation masking release, Envelope, Power spectrum

INTRODUCTION

Hall, Haggard and Fernandes (1984) showed that the detectability of a tone in a band of noise is improved when synchronous bands of noise, rather than independent bands of noise, are presented at frequencies removed from the critical band containing the signal. This reduction in masking, termed comodulation masking release or CMR (see also McFadden, 1986; Cohen and Schubert, 1987a), indicates that the auditory system is sensitive to temporal synchrony. Recent experiments (Cohen and Schubert, 1987b; McFadden, 1987) have demonstrated that temporal synchrony affects the detectability of not just tones, but the detectability of bands of noise as well.

In an experiment motivated by the CMR paradigm, Richards (1987) asked subjects to discriminate between simultaneously presented 100 Hz wide bands of noise whose envelopes were either the same or statistically independent. Discriminations were not possible unless the center frequency of the bands of noise exceeded 1000 Hz. For center frequencies above 1000 Hz, discriminations were above chance when the two bands of noise were separated by less than an octave. Richards argued that for high frequencies, the auditory system is able to compare information contained in the envelopes of diotically presented bands of noise.

This conclusion does not take into account the possible cues derived from the power spectra of Richards' stimuli. Figure 1 presents Richards' stimuli (from Richards, 1987). When the two bands of noise had the same envelopes (left panel), the power spectra were also the same (save for a shift of frequency). Likewise, when the bands had statistically independent envelopes (right panel), the power spectra had phases and amplitudes that were independent. Thus, the correlated vs uncorrelated envelope cue co-varied with a similar vs dissimilar spectral cue. The same argument may be made concerning the CMR experiments; spectral cues are not considered.

-------- Figure 1 --------

There are at least two ways in which spectral cues might be have been incorporated in Richards' experiment. First, drawing from the work of Green and his colleagues (Green, 1988), one might hypothesize the existence of a 'mini-profile analyzer'. Such a mechanism would compare the shape of the spectra of the simultaneously presented bands of noise. It seems unlikely that such a mechanism was available in Richards' experiment since the 100 Hz wide bands of noise were narrow relative to the high-frequency critical bands. A second, seemingly more likely strategy would be to compare the total energy of the two bands of noise. When the two bands of noise had identical envelopes, the total energy in each spectral region was equal. In contrast, when the bands of noise were independent, the total energy of the two bands of noise differed. Thus, a gross, simultaneous energy comparison would afford an opportunity to discriminate between correlated and uncorrelated bands of noise.

The current experiment was designed to determine the relative importance of the envelope and power spectrum cues. To this end, a factorial stimulus design was employed. The four stimuli are indicated in shorthand along the top and side of Table 1. Each waveform is the sum of two bands of noise. 'E' indicates that the two noise bands had the same envelopes, and 'S' indicates that the two noise bands had the same power spectra (save for the shift in center frequency). A '$\bar{\ }$' indicates the complement: '$\bar{E}$' indicates that the two noise bands had different envelopes and '$\bar{S}$' indicates that the two noise bands had different power spectra. Using this notation, the four possible stimuli are:

ES: The two bands of noise had the same envelopes and the same power spectra.

$E\bar{S}$: The two bands of noise had the same envelopes, but different power spectra.

$\bar{E}S$: The two bands of noise had different envelopes but the same power spectra.

$\overline{ES}$: The two bands of noise had different envelopes and different power spectra.

In a 2IFC paradigm, listeners were asked to discriminate between two of the stimuli described above. All possible stimulus comparisons were tested, yielding six experimental conditions. These are represented by the cells of Table 1. For example, discriminating between ES and $\overline{ES}$ waveforms (lower left hand corner) is the condition examined by Richards; in one interval of the 2IFC presentation the bands of noise were the same (ES), and in the other interval the two bands of noise were statistically independent ($\overline{ES}$). Note that the ES and $\overline{ES}$ conditions are analogous to the 'correlated' (sometimes called 'coherent') and 'uncorrelated' (sometimes called 'independent') cue conditions often used in CMR experiments (McFadden, 1986; Cohen and Schubert, 1987).

-------- Table 1 --------

I. <u>Stimuli</u>

The stimuli will be referred to as described above: ES, $E\overline{S}$, $\overline{E}S$, and $\overline{ES}$. Each was the sum of two bands of noise, which were computer-generated in accordance with the following equations.

1. ES

The ES bands of noise are as presented in the left panel of Figure 1. Each of the two bands of noise is the sum of several cosines:

$$w_1(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \qquad (1a)$$

$$w_2(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f + \delta \cdot i) \cdot t + \phi_i), \qquad (1b)$$

where δ is the frequency separation of the tones added to generate the noise and Δf is the difference in center frequency between the two bands of noise, $w_1(t)$ and $w_2(t)$. Here δ was 10 Hz and m was 5, yielding a band of noise nominally 100 Hz in width. The amplitudes (a_i) were chosen at random from Rayleigh-distributed values, and the phases (ϕ_i) were chosen from values uniformly distributed between zero and 2π. Again, ES bands have the feature that the envelopes and power spectra are the same.

2. $E\overline{S}$

The $E\overline{S}$ bands of noise have the same envelopes but different power spectra. In practice, the $E\overline{S}$ stimuli were created using the same stimulus parameters described in Eqns. (1), but the amplitudes were reversed and the phases were reversed and multiplied by -1.

$$w_1(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \qquad (2a)$$

$$w_2(t) = \sum_{i=-m}^{m} a_{-i} \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f + \delta \cdot i) \cdot t - \phi_{-i}). \qquad (2b)$$

It should be noted that the power spectra are not statistically independent, but uncorrelated (one is the mirror image of the other). The appendix shows that these two noise bands have the same envelopes.

3. $\overline{E}S$

For the $\overline{E}S$ bands, the amplitudes of the tones that composed the two bands of noise were the same, but the phases were chosen independently:

$$w_1(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \qquad (3a)$$

$$w_2(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f + \delta \cdot i) \cdot t + \theta_i), \qquad (3b)$$

where ϕ_i and θ_i were chosen independently from values uniformly distributed from zero to 2π. The effect of phase randomization is to generate noise bands with similar power spectra, but different envelopes.

4. $\overline{ES}$

The $\overline{ES}$ noise bands were generated in a manner similar to Eqns. (1), (2), and (3), except that the amplitudes and phases were chosen independently for the two noise bands. The construction will be detailed in the procedure section. An example of an $\overline{ES}$ waveform is shown in the right panel of Fig. 1.

II. <u>Procedure</u>

In a 2IFC paradigm, listeners attempted to discriminate between two signals that differed in one or more aspects (i.e., envelopes and/or power spectra). For example, listeners might indicate which of two intervals contained the two bands of noise whose envelopes were the same. The percentage of correct discriminations was the dependent variable.

The six conditions tested are represented by the cells of Table 1. The first column of Table 1 presents those conditions in which listeners discriminated between the ES bands, which shared both envelopes and spectra, and bands that differed in one or more aspects: the $E\overline{S}$ bands had different spectra, the $\overline{E}S$ bands had different envelopes, and the $\overline{ES}$ bands had different envelopes and different spectra. The bottom row of Table 1 presents those conditions in which listeners discriminated between independent bands of noise ($\overline{ES}$) and bands that shared at least one aspect; envelopes, spectra or both. For example, in the [$E\overline{S}$,$\overline{ES}$] condition, the $E\overline{S}$ bands were presented in one interval, and the $\overline{ES}$ bands were presented in the other interval. Listeners were asked to indicate which signal was composed of bands that had identical envelopes ($E\overline{S}$).

The cells of Table 1 present the center frequencies ($f_L, f_L+\Delta f$) that were used. All comparisons included bands of noise centered at 2500 and 2750 Hz, which were chosen because subjects were able to discriminate between the ES and $\overline{ES}$ waveforms on an average of 98 percent of the trials. It was reasoned that the use of the 2500 and 2750 Hz center frequencies would allow ample opportunity for <u>lower</u> scores to be observed. As presented in Table 1, four pairs of center frequencies were examined in the [ES,$\overline{ES}$] and [ES,$E\overline{S}$] test conditions.

The different conditions were completed in random order, except that all listeners completed the [ES,$\overline{ES}$] conditions first. For each of the six experimental conditions represented in Table 1, the row and column stimuli had equal a-priori probability of being in the first interval.

The stimuli were generated on an IBM-PC microcomputer. Depending on the stimulus condition, one of two presentation algorithms was followed. For the conditions presented along the bottom row of Table 1, thirty-two low- and high-frequency noise band pairs were generated and stored. The column stimulus was the sum of the 'paired' low- and high-frequency bands, and the $\overline{ES}$ stimulus was the sum of independently chosen low- and high-frequency noise bands. For the remaining conditions ([ES,$E\overline{S}$], [ES,$\overline{E}S$], and [$E\overline{S}$,$\overline{E}S$]), thirty-two low- and high-frequency pairs were generated and stored, sixteen pairs for the row stimulus, and sixteen pairs for the column stimulus. The stimuli were chosen randomly from among each of the sixteen pairs.

All waveforms were played through a 12-bit D/A at a sampling rate of 14.3 kHz and low-pass filtered (Kemo VBF/23) at 6 kHz. The signal duration was 100 ms, plus 5-ms cosine-squared onset/offset ramps. The two listening intervals were separated by 300 ms, and both intervals were visually indicated on a display screen. Following the listener's response, visual feedback was displayed for 240 ms.

The stimuli were presented diotically via Sennheiser HD 414 SL earphones at an average total level of 75 dB SPL. Subjects listened in a sound-treated room. Conditions were tested in 50-trial blocks, six blocks at a time. Thus, each data point is based on 300 trials. Listeners typically completed 18 to 24 blocks a day. Each condition was completed before practice for the next condition was begun.

Listeners were undergraduate students paid to participate, except GR, who is the author. All had previously participated in similar experiments, and little practice was completed prior to data collection. When a novel condition was introduced, listeners typically required only 50 to 100 trials in order to 'learn' the task. In order to assess possible effects of practice, approximately 1/3 of the conditions were repeated. Of those repeated, about 10% led to significantly different averages. For the repeated conditions, the average of only the last 300 trials are reported.

III. Results

Table 2 lists the average percent correct discriminations for each of the conditions tested. The averages are based on four subjects. The standard errors of the mean, based on averages of the four listeners and six estimates per listener, are indicated in parentheses. The presentation parallels the conditions of Table 1, the scores referring to the center frequencies indicated in Table 1.

---- Table 2 ----

Figure 2 presents the data for two of the experimental conditions, $[ES,\overline{ES}]$ (open bars) and $[E\overline{S},\overline{ES}]$ (filled bars). For each subject, a histogram plots the percent correct for each $(f_L, f_L+\Delta f)$ center frequency pair tested. The error bars indicate the standard errors of the mean.

---- Figure 2 ----

First consider the data obtained in the $[ES,\overline{ES}]$ condition (open bars). For all four subjects, increasing the center frequency of the higher noise band from 2750 to 3500 Hz (holding the lower frequency noise band fixed at 2500 Hz) yielded poorer performance. The extent of the drop in performance was subject dependent, with subjects VF and KK changing relatively little, and subjects GR and MS changing relatively more. In the (4000,4400) Hz condition, discrimination was nearly perfect. These data constitute a replication of the experiment presented by Richards (1987), with the same results.

Next consider the data in the $[E\overline{S},\overline{ES}]$ condition (filled bars), which is similar to the $[ES,\overline{ES}]$ condition, except that the two bands of noise always had different power spectra. Clearly, envelope similarity is detectable in the absence of co-varying spectral cues. For three of the four subjects (KK, GR, and MS), increasing the frequency separation of the two noise bands led to decreased discriminability. Reading from Fig. 2, for fifteen of the sixteen possible comparisons, performance levels in the $[ES,\overline{ES}]$ condition were superior to performance levels obtained in the $[E\overline{S},\overline{ES}]$ condition. On average, the difference was thirteen percentage points, but the effect was subject dependent. Subjects VF and KK were little affected (an average difference of 5.2 percent correct), while subjects GR and MS showed larger changes (an average difference of 20.5 percent correct).

The data presented in Figure 2 indicate that listeners were able to discriminate between bands of noise with either identical or statistically independent envelopes, even when spectral cues were absent. For two of the subjects, however, performance levels were superior when both spectral and envelope cues were available. The relatively little change observed for the other two subjects may reflect a performance ceiling in the $[ES,\overline{ES}]$ condition.

Figure 3 presents the data for the $[\overline{E}S,\overline{ES}]$, $[ES,E\overline{S}]$, $[ES,\overline{E}S]$, and the $[E\overline{S},\overline{E}S]$ conditions. The bands of noise were centered at 2500 and 2750 Hz. In the $[\overline{E}S,\overline{ES}]$ condition, none of the four listeners was able to reliably perform above chance levels (50% correct), even though three of the four subjects completed in excess of 1000 trials. That is, when the envelopes were always different, changes in power spectra were not detectable. Most of the efforts were limited to center frequencies of 2500 and 2750 Hz, but others were occasionally tested. In no instance did performance appear to rise above chance. These data indicate that simultaneous comparisons of either spectral shape or energy levels is not an important aspect of what has been termed envelope correlation detection (Richards, 1987).

Figure 3 shows that subjects were able to discriminate between the ES and $E\overline{S}$ waveforms (an average of 85 percent correct discriminations). That is, when both intervals contained noise bands with identical envelopes, changes in power spectra were detectable. This contrasts with the finding that the subjects were unable to discriminate between similar and dissimilar power spectra when the noise bands had dissimilar envelopes (the [$\overline{E}S$,$\overline{ES}$] condition described above).

---- Figure 3 ----

In the [ES,$\overline{E}S$] condition, the listeners discriminated between bands of noise whose envelopes were either identical or different, but the noise bands always had the same power spectra. The fact that discriminations were good (an average of 97 percent correct) supports the notion that similar and dissimilar envelopes are discriminable even when spectral and envelope cues do not co-vary.

Finally, subjects were able to discriminate, with an average accuracy of 84 percent correct, between the $E\bar{S}$ waveforms, which had identical envelopes and different power spectra, and the $\bar{E}S$ waveforms, which had similar spectra but different envelopes. The interpretation of this finding is not clear, as the role of spectral cues is unclear.

In summary, listeners were always able to discriminate between waveforms composed of bands of noise that had identical envelopes and waveforms composed of bands of noise that had independent envelopes. When the bands of noise had power spectra that were either the same or different, listeners were unable to discriminate the change unless the bands of noise had identical envelopes. Finally, subjects performed well in the $[E\bar{S},\bar{E}S]$ condition, in which both envelopes and power spectra changed. In that condition, it is not clear whether the discrimination was based on changes in envelopes, changes in power spectra, or both. It seems reasonable to assume that the discriminations were based on changes in envelope similarity, since changes in spectral similarity were not detectable unless the noise bands had the same envelopes.

IV. Discussion

The data presented above introduce two discrepancies, both involving the discriminability of changes in the power spectrum of the narrow bands of noise. The first involves the data of Figure 2: If removing power spectrum cues led to a reduction in the discriminability of envelope similarity ($[ES,\overline{ES}]$ *vs.* $[E\bar{S},\overline{ES}]$), an average difference of 13 percent), why were subjects unable to to detect (an average of 50 % correct) the change in power spectra between the $\bar{E}S$ and $\overline{ES}$ waveforms? Second, given that subjects discriminated between the ES and $E\bar{S}$ waveforms on an average of 85 percent of the trials, why were listeners unable to exceed chance performance in the $\bar{E}S$ and $\overline{ES}$ condition?

Before examining these inconsistencies at length, the reader should be made aware of potential problems in the experimental procedure. First, as has been observed in other CMR experiments (McFadden, 1986; Buus, 1985), there are considerable between-subject differences. Indeed, for the current experiment, a consistent ranking of the dependent variable as a function of the condition across the four listeners cannot be achieved.

Second, due to the large number of conditions tested, there is no guarantee that the subjects were comfortable with the discrimination response required for each of the conditions tested. Although listeners gave no indication of confusion, it is possible that other experimental procedures (for example, a same-different procedure) or further practice might have altered the magnitude of the performance levels. Certainly longer signal durations would have lead to superior performance levels (Richards, 1988), but whether the relative performance levels would have been altered remains to be determined. For these reasons, moderate changes in performance level deserve little emphasis. While such limitations argue for further experimentation, we do not feel that they are so crippling as to affect the basic observations noted above. The difference between discriminability ([ES,$E\bar{S}$]) and indiscriminability ([$\bar{E}S,\overline{ES}$]) is clear, and so will be the primary comparison of interest. The difference in performance levels between the [ES,$\overline{ES}$] and the [$E\bar{S},\bar{E}S$] conditions will receive little attention as the difference was subject dependent.

In order to address the change in performance levels between the [ES,$E\bar{S}$] and the [$\bar{E}S$, $\overline{ES}$] conditions, we shall examine an assumption that the two dimensions, envelope and spectral, provide a sufficient basis from which to describe the discriminations. The possibility that these dimensions are not sufficient is explored below, but no satisfactory explanation of the data is presented.

Are envelope and spectral considerations sufficient? To this point we have ignored changes in the 'fine structure' of the signals. Consider the bands of noise that made up the ES waveforms. The bands had the same envelopes and the same power spectrum (save for a shift in center frequency). Further, the phase function (the phase of the fine structure as a function of time; Davenport and Root, 1958) was the same. In contrast, the bands that made up the $E\bar{S}$ waveforms had identical envelopes, but the power spectrum and the phase functions were not the same. Might dynamic changes in the phase functions affect the discriminations?

A direct comparison of the phase functions seems an unlikely cue; the fine structure is not 'extractable' at the high frequencies used here. However, there is the possibility that changes in fine structure might be manifested in other ways.

Initially we considered McFadden's (1987; see also McFadden, 1975) observation that for CMR experiments, and by extension the current experiment, it is often unreasonable to assume that envelopes extracted from the low- and high-frequency regions bands match. The higher-frequency band is probably better described as a sum of the high- and low-frequency bands. This point is especially germane here since the experimental bands were relatively close, with the center frequencies being separated by no more than a critical band. For these reasons we considered the envelopes of the summed ES, $\bar{E}S$, $E\bar{S}$ and $\overline{ES}$ waveforms.

The bands of noise that comprise the ES, $\bar{E}S$, $E\bar{S}$ and $\overline{ES}$ waveforms were generated, the paired bands added, and the envelopes of the resulting waveforms extracted. The root-mean-squared (RMS) values for the extracted envelopes were then computed. The RMS values of the ES and $E\bar{S}$ waveforms did not differ significantly (for similar parameter conditions, the envelopes of the summed waveforms were indistinguishable). Nor were the RMS values for the $\bar{E}S$ and $\overline{ES}$ waveforms significantly different. The RMS values for the ES and $E\bar{S}$ waveforms were, however, significantly larger than the RMS values for the $\bar{E}S$ and $\overline{ES}$ waveforms. Thus, the change in the RMS values of the envelopes of the summed waveform cannot account for the difference in performance levels for the [ES,$E\bar{S}$] and the [$\bar{E}S$,$\overline{ES}$] conditions. The simulations do, however, indicate that the peak-to-valley-ratio, or the 'modulation depth', of the envelopes of the summed waveforms may contribute to the discrimination between simultaneously presented bands of noise that have either identical (ES and $E\bar{S}$) or statistically independent ($\bar{E}S$ and $\overline{ES}$) envelopes.

It is evident that changes in the phase functions do not alter the summed bands in a manner consistent with our results. We are currently investigating the extent to which other peripheral interactions may contribute to the discrimination between waveforms composed of correlated and independent bands of noise.

V. Summary

(1) The ability to discriminate between bands of noise whose envelopes were either the same or statistically independent was accomplished in the absence of spectral cues. If spectral cues co-varied with envelope cues, performance levels were typically superior to those obtained in the absence of co-varying cues.

(2) When the bands of noise had envelopes that were not the same, listeners were unable to discriminate between noise bands whose power spectra were either the same of different. When the noise bands had identical envelopes, listeners were able to indicate whether the bands of noise had the same power spectra.

Acknowledgment

This work was supported in part by AFOSR and in part by an NIH post-doctoral fellowship. I thank Drs. Leslie R. Bernstein, Tim Forrest, David M. Green and Dennis McFadden for helpful comments on earlier drafts of this paper.

Appendix

Here is shown that the waveforms described by Eqns. (2a) and (2b) have identical envelopes. To this end, the envelopes of each of the two waveforms will be derived.

The envelope of a narrow band of noise, A(t), may be written in terms of the original waveform, w(t), and its Hilbert Transform, $\hat{w}(t)$:

$$A^2(t) = w^2(t) + \hat{w}^2(t). \qquad \text{(A1)}$$

Because the Hilbert transform is linear, the Hilbert Transform of a sum of cosines is the sum of the Hilbert Transforms of the cosines. Further, because the Hilbert Transform of a cosine is a sine, the Hilbert Transform of the noise bands described by Eqn. (2a):

$$w_1(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \qquad \text{(2a)}$$

is

$$\hat{w}_1(t) = \sum_{i=-m}^{m} a_i \cdot \sin(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i).$$

Squaring and adding:

$$A_1{}^2(t) = w_1{}^2(t) + \hat{w}_1{}^2(t) = \left(\sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \right)^2$$

$$+ \left(\sum_{i=-m}^{m} a_i \cdot \sin(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \right)^2.$$

Expanding the right hand side,

$$A_1^2(t) = \sum_{i=-m}^{m} a_i^2 \cdot \cos^2(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i)$$

$$+ \sum_{i=-m}^{(m-1)} \sum_{j>i} a_i \cdot a_j \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \cdot \cos(2 \cdot \pi \cdot (f_L + \delta \cdot j) \cdot t + \phi_j)$$

$$+ \sum_{i=-m}^{m} a_i^2 \cdot \sin^2(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i)$$

$$+ \sum_{i=-m}^{(m-1)} \sum_{j>i} a_i \cdot a_j \cdot \sin(2 \cdot \pi \cdot (f_L + \delta \cdot i) \cdot t + \phi_i) \cdot \sin(2 \cdot \pi \cdot (f_L + \delta \cdot j) \cdot t + \phi_j).$$

Because $\cos^2\alpha + \sin^2\alpha = 1$, and because $\cos\alpha \cdot \cos\beta + \sin\alpha \cdot \sin\beta = \cos(\alpha - \beta)$, $A_1^2(t)$ may be written as

$$A_1^2(t) = \sum_{i=-m}^{m} a_i^2$$

$$+ \sum_{i=-m}^{(m-1)} \sum_{j>i} a_i \cdot a_j \cdot [\cos(2 \cdot \pi \cdot \delta (i-j) \cdot t + (\phi_i - \phi_j))]. \qquad (A2)$$

Next, we shall show that the complementary waveform, $w_2(t)$ (Eqn. (2b)), has the same envelope. Although the noise band's center frequency has no bearing on the shape of the envelope (provided the center frequency is large relative to the bandwidth), the difference in center frequency will be maintained in order to demonstrate that point.

$$w_2(t) = \sum_{i=-m}^{m} a_{-i} \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f + \delta \cdot i) \cdot t - \phi_{-i}). \qquad (2b)$$

Equation (2b) may be rewritten, replacing i with -i, and changing the order of addition,

$$w_2(t) = \sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f - \delta \cdot i) \cdot t - \phi_i). \qquad (A3)$$

After the change of variables, $w_2(t)$ resembles $w_1(t)$, except that the center frequency has been increased, the phase constant is subtracted rather than added, and the frequencies are decremented from high to low.

The Hilbert Transform of $w_2(t)$ is given by

$$\hat{w}_2(t) = \sum_{i=-m}^{m} a_i \cdot \sin(2 \cdot \pi \cdot (f_L + \Delta f - \delta \cdot i) \cdot t - \phi_i).$$

Squaring and adding the waveform and its transform yields

$$A_2^{\,2}(t) = w_2^{\,2}(t) + \hat{w}_2^{\,2}(t) = \left(\sum_{i=-m}^{m} a_i \cdot \cos(2 \cdot \pi \cdot (f_L + \Delta f - \delta \cdot i) \cdot t - \phi_i) \right)^2$$

$$+ \left(\sum_{i=-m}^{m} a_i \cdot \sin(2 \cdot \pi \cdot (f_L + \Delta f - \delta \cdot i) \cdot t - \phi_i) \right)^2.$$

As before, we expand the right hand side and combine like terms to obtain

$$A_2^{\,2}(t) = \sum_{i=-m}^{m} a_i^{\,2}$$

$$+ \sum_{i=-m}^{(m-1)} \sum_{j>i} a_i \cdot a_j \cdot [\cos(2 \cdot \pi \cdot \delta \cdot (j-i) \cdot t + (\phi_j - \phi_i))].$$

Since the cosine is an even function, this may be rewritten as

$$A_2^{\,2}(t) = \sum_{i=-m}^{m} a_i^{\,2}$$

$$+ \sum_{i=-m}^{(m-1)} \sum_{j>i} a_i \cdot a_j \cdot [\cos(2 \cdot \pi \cdot t \cdot \delta (i-j) + (\phi_i - \phi_j))]. \qquad (A4)$$

Thus, the waveforms represented by Eqns. (2a) and (2b) have identical envelopes ((A2)=(A4)).

References

Buus, S. (1985). "Release from masking caused by envelope fluctuations," J. Acoust. Soc. Am. **78**, 1958-1965.

Cohen, M. F., and Schubert, E. D. (1987a). "Influence of place synchrony on detection of sinusoids," J. Acoust. Soc. Am. **81**, 452-458.

Cohen, M. F., and Schubert, E. D. (1987a). "The effect of cross-spectrum correlation on the detectability of a noise band," J. Acoust. Soc. Am. **81**, 721-723.

Davenport, W. P. Jr., and Root, W. L. (1958). An introduction to the Theory of Random Signals and Noise, McGraw-Hill Book Co. New York, New York.

Hall, J. W. III., Haggard, M. P. and Fernandes, M. A. (1984). "Detection in noise by spectro-temporal pattern analysis," J. Acoust. Soc. Am., **81**, 452-458.

Henning, G. B. (1980). "Some observations on the lateralization of complex waveforms," J. Acoust. Soc. Am. **68**, 446-454.

McFadden, D. (1975). "Beat-like interaction between periodic waveforms," J. Acoust. Soc. Am., **57**, 983.

McFadden, D. (1986). "Comodulation masking release: Effects of varying level, duration an time delay of the cue band," J. Acoust. Soc. Am., **80**, 1658-1667.

McFadden, D. (1987). "Comodulation detection differences using noise-band signals," J. Acoust. Soc. Am., **81**, 1519-1527.

Richards, V. M. (1986) "Monaural envelope correlation perception," J. Acoust. Soc. Am., **80**, S106(A).

Richards, V. M. (1987) "Monaural envelope correlation perception," J. Acoust. Soc. Am., in press.

Richards, V. M. (1988) "Monaural synchrony perception: Effects of duration, bandwidth and center frequency," AFOSR meeting, Chicago, Illinois.

Table Captions

Table 1. The experimental conditions are presented. Each cell indicates discrimination between the row and column stimuli. The center frequencies that were tested are indicated in each cell.

Table 2. The average percent correct is shown for each of the experimental conditions presented in Table 1. The standard errors of the mean, based on six estimates for all of the four listeners, are indicated in parentheses.

Figure Captions

Figure 1. Two stimuli are shown. The top panel presents the stimulus whose envelopes and power spectra are the same (ES). The bottom panel shows two statistically independent noise bands ($\overline{ES}$). The summed waveform, the amplitude spectra, and the waveforms of the individual noise bands are drawn.

Figure 2. Percent correct discriminations are indicated for two experimental conditions, [$ES,\overline{ES}$] (open bars) and [$E\bar{S},\overline{ES}$] (filled bars), and for each of the ($f_L, f_L+\Delta f$) center frequencies tested. The data for each listener are plotted separately. Error bars indicate the standard errors of the mean.

Figure 3. Percent correct discriminations are indicated for the [$\bar{E}S,\overline{ES}$], [$ES,E\bar{S}$], [$ES,\bar{E}S$], and [$E\bar{S},\bar{E}S$] conditions. The noise bands were centered at 2500 and 2750 Hz.

Table 1

Experimental Design

Center frequencies are indicated in parentheses

	ES	$E\bar{S}$	$\bar{E}S$	$\overline{ES}$
ES	***			
$E\bar{S}$	(2500,2750)	***		
$\bar{E}S$	(2500,2750)	(2500,2750)	***	
$\overline{ES}$	(2500,2750) (2500,3000) (2500,3500) (4000,4400)	(2500,2750) (2500,3000) (2500,3500) (4000,4400)	(2500,2750)	***

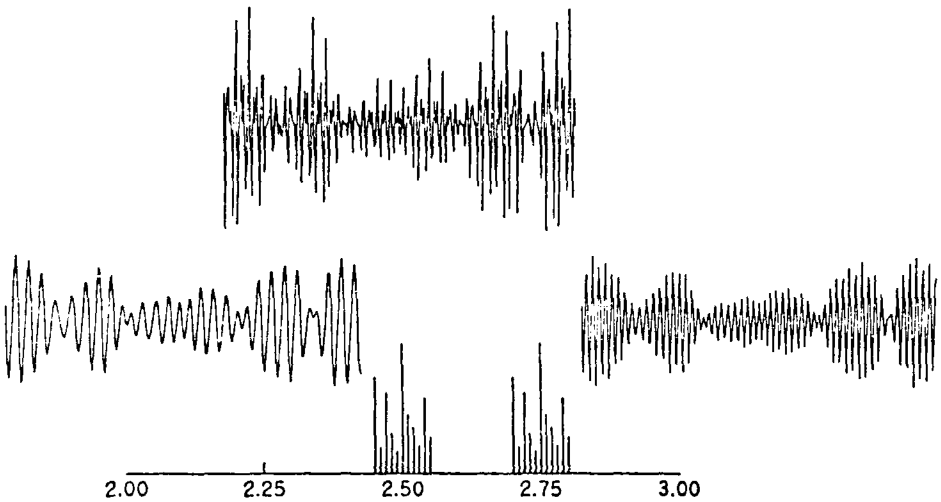
Table 2

Average Percent Correct

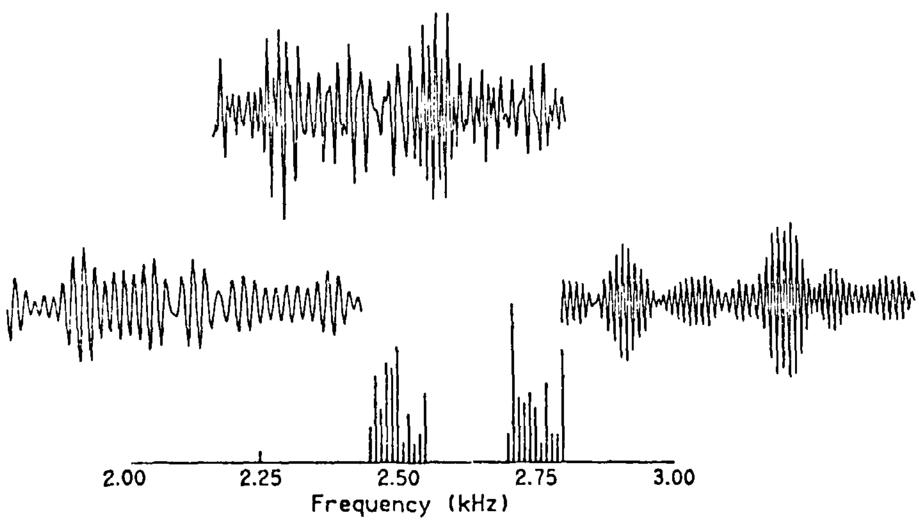
For the test conditions shown in Table 1

	ES	$E\bar{S}$	$\bar{E}S$	$\bar{E}\bar{S}$
ES	***			
$E\bar{S}$	85.2 (2.5)	***		
$\bar{E}S$	97.4 (0.8)	83.9 (2.3)	***	
$\bar{E}\bar{S}$	98.4 (0.4) 90.1 (1.2) 77.0 (2.7) 93.5 (0.6)	81.4 (2.2) 73.8 (2.6) 70.8 (3.6) 80.8 (2.8)	50.3 (1.4)	***

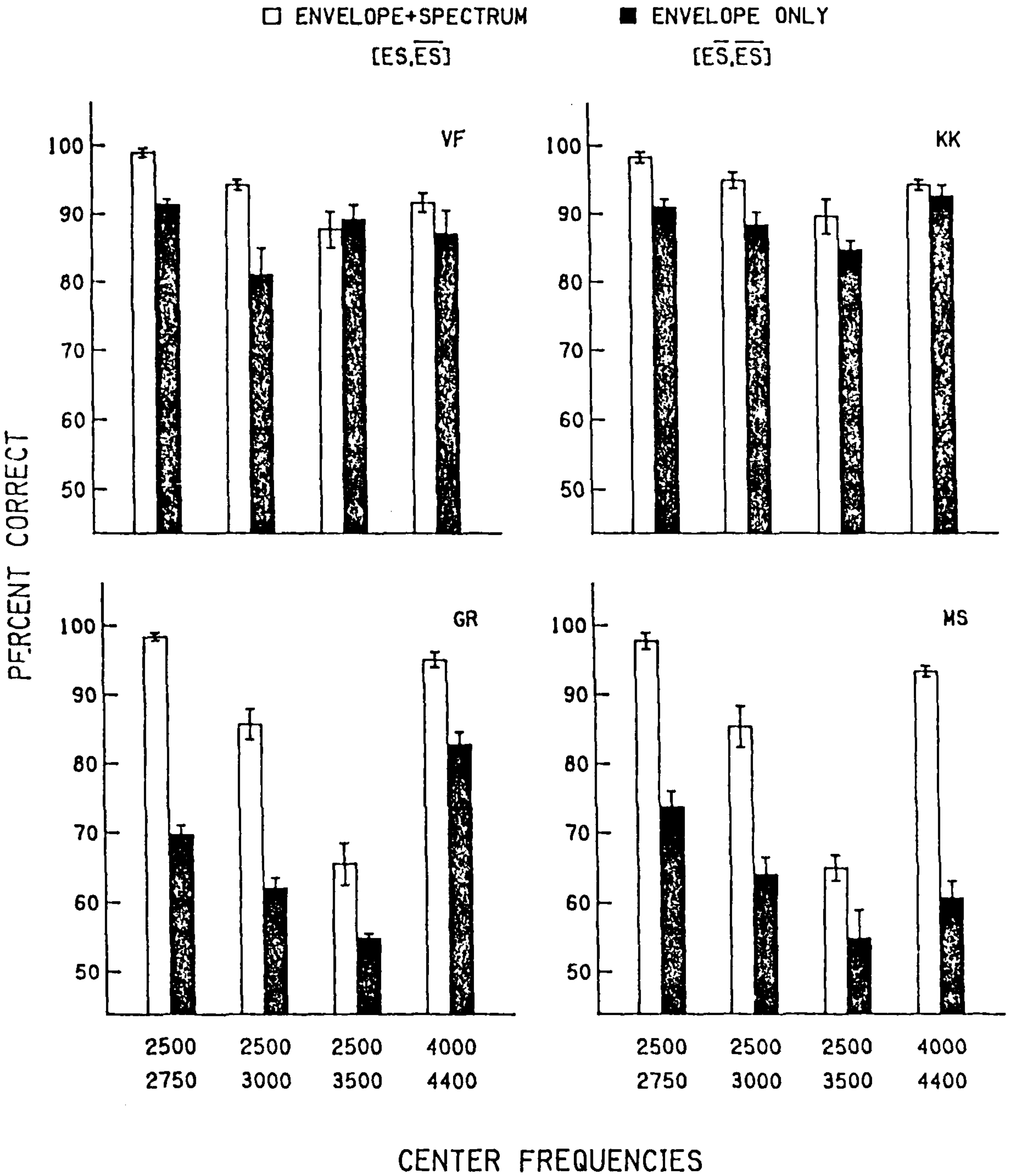
CORRELATED



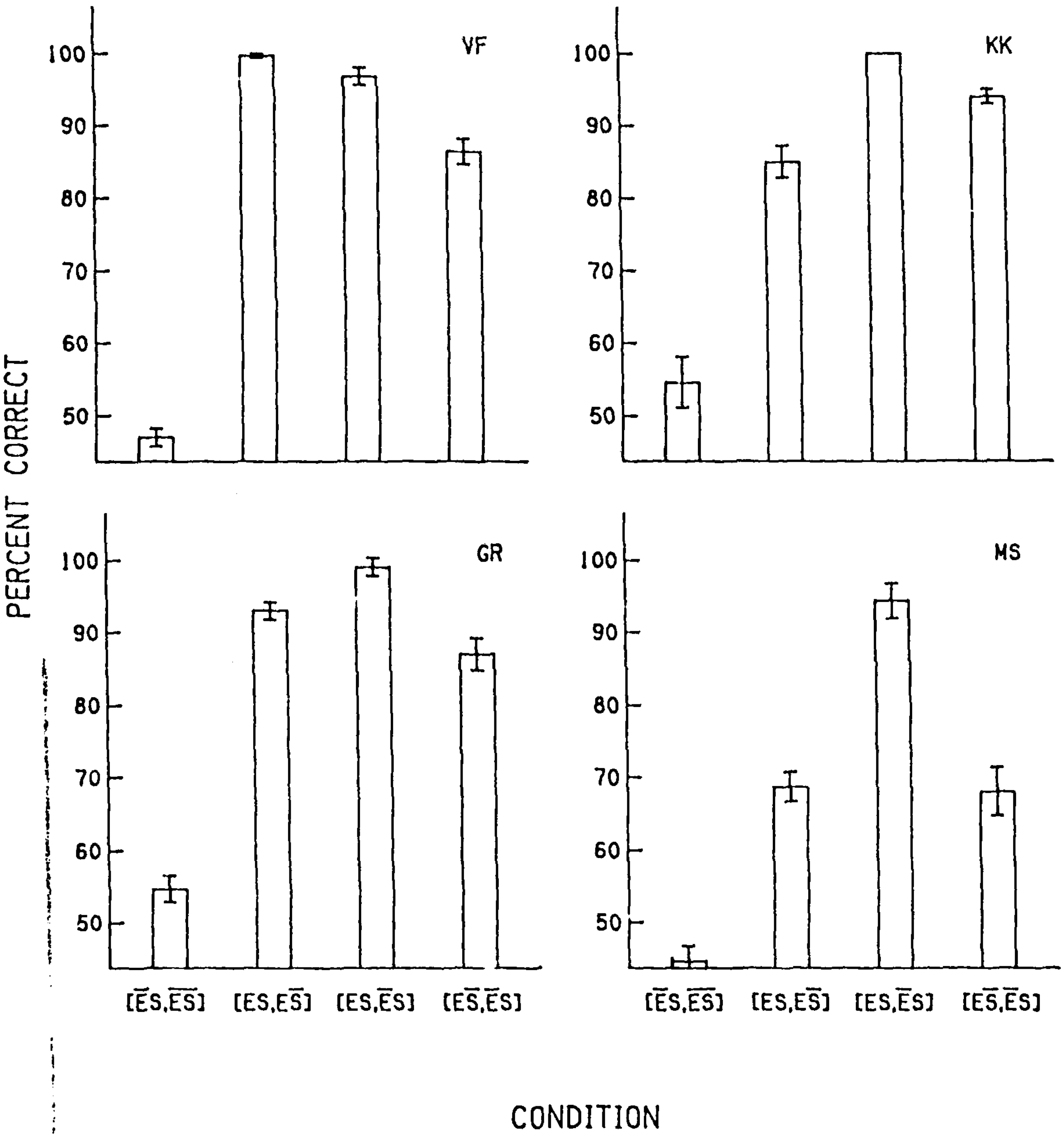
UNCORRELATED



1


□ ENVELOPE+SPECTRUM
[ES,$\overline{ES}$]
■ ENVELOPE ONLY
[$\overline{ES}$,$\overline{ES}$]
VF
KK
GR
MS
PERCENT CORRECT
100
90
80
70
60
50
2500 2750
2500 3000
2500 3500
4000 4400
CENTER FREQUENCIES

2



PERCENT CORRECT
VF
KK
GR
MS
100
90
80
70
60
50
[ĒS,ĒS] [ES,ES̄] [ES,ĒS] [ES̄,ĒS]
CONDITION