AUDITORY PROCESSING OF COMPLEX SOUNDS

Edited by
William A. Yost and Charles S. Watson

PSYCHOLOGY LIBRARY EDITIONS:
COGNITIVE SCIENCE

ROUTLEDGE

PSYCHOLOGY LIBRARY EDITIONS:
COGNITIVE SCIENCE

Volume 27

AUDITORY PROCESSING OF COMPLEX SOUNDS

This page intentionally left blank

AUDITORY PROCESSING OF COMPLEX SOUNDS

Edited by
WILLIAM A. YOST AND CHARLES S. WATSON

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published in 1987 by Lawrence Erlbaum Associates, Inc.

This edition first published in 2017
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
711 Third Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN: 978-1-138-19163-1 (Set)
ISBN: 978-1-315-54401-4 (Set) (ebk)
ISBN: 978-1-138-65575-1 (Volume 27) (hbk)
ISBN: 978-1-138-65576-8 (Volume 27) (pbk)
ISBN: 978-1-315-62234-7 (Volume 27) (ebk)

Publisher's Note
The publisher has gone to great lengths to ensure the quality of this reprint but points out that some imperfections in the original copies may be apparent.

Disclaimer
The publisher has made every effort to trace copyright holders and would welcome correspondence from those they have been unable to trace.

AUDITORY PROCESSING OF COMPLEX SOUNDS

Edited by

Willliam A. Yost, PhD
Parmly Hearing Institute
Loyola University, Chicago, Illinois

Charles S. Watson, PhD
Department of Speech and Hearing Sciences
Indiana University, Bloomington, Indiana

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
1987 Hillsdale, New Jersey London

Lawrence Erlbaum Associates, Inc., Publishers
365 Broadway
Hillsdale, New Jersey 07642

Library of Congress Cataloging-in-Publication Data

Auditory processing of complex sounds.

"With the support of the AFOSR the Sarasota Workshop on Auditory Processing of Complex Sounds was held in April 1986"—Pref.

Includes bibliographies and indexes.

1. Auditory perception—Congresses. 2. Psychoacoustics—Congresses. I. Yost, William A. II. Watson, Charles S. III. United States. Air Force. Office of Scientific Research. IV. Sarasota Workshop on Auditory Processing of Complex Sounds (1986)

QP465.A93 A93 1986 152.1'5 86-29135

ISBN 0-89859-981-4

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

TABLE OF CONTENTS

SECTION V. SPEECH PERCEPTION:
Speech versus Non-Speech Perception and a New Model

SECTION VI. PERCEPTUAL ORGANIZATION OF COMPLEX SOUNDS:
Informational Masking, Stimulus Uncertainty, Learning, Attention, Memory, and Stream Segregation

This page intentionally left blank

CONTRIBUTORS:

Paul J. Abbas
Dept. of Speech Pathology-Audiology
University of Iowa
Iowa City IA 52242

P.E. Barta
Dept. of Biomedical Engineering
Johns Hopkins University School of Medicine
Baltimore MD 21205

Leslie R. Bernstein
Dept. of Psychology
University of Florida
Gainesville FL 32610

Jens Blauert
Ruhr-Universitat,
Lehrstuhl AEA
Postfach 102148
4630 Bochum 1
Fed. Republic of Germany

Brian P. Callaghan
Boys Town National Institute for Communication Disorders
555 N. 30th St.
Omaha NE 68131

Robert P. Carlyon
MRC Institute of Hearing Research
University Park
Nottingham England NG7 2RD

Nelson Cowan
Dept. of Psychology
210 McAlester Hall
University of Missouri
Columbia MO 65211

Pierre Divenyi
Speech & Hearing Facility
VA Medical Center
Martinez CA 94553

Raymond H. Dye, Jr.
Parmly Hearing Institute
Loyola University
6525 N. Sheridan Rd.
Chicago IL 60626

Deborah A. Fantini
Dept. of Psychology
University of Minnesota
75 E. River Rd.
Minneapolis MN 54455

Glen R. Farley
Boys Town National Institute for Communication Disorders
555 N. 30th St.
Omaha NE 68131

Lawrence L. Feth
University of Kansas
Depts. of Speech, Language, Hearing
3129 Tomahawk Drive
Lawrence KS 66045

C. Craig Formby
University of Florida
Depts. of Communicative Disorders/ Neurology
Box J-174/ J.H. Miller Health Center
Gainesville FL 32610

Robert H. Gilkey
Central Institute for the Deaf
818 S. Euclid
St. Louis MO 63110

David M. Green
Dept. of Psychology
University of Florida
Gainesville FL 32605

Steven Greenberg
Dept. of Neurophysiology
University of Wisconsin
Madison WI 53706

Joseph W. Hall
School of Medicine
Division of
Otolaryngology
Head and Neck Surgery
The University of North
Carolina
Chapel Hill NC 27514

Patrick J. Harder
Parmly Hearing Institute
Loyola University
6525 N. Sheridan Rd.
Chicago IL 60626

William M. Hartmann
Physics Dept.
Michigan State University
East Lansing MI 48824

J.W. Horst
Institute of Audiology
University Hospital
P.O. Box 30.001, Groningen
The Netherlands

Tammo Houtgast
Institute for Perception
Soesterberg, Kampweg 5
Postbus 23
The Netherlands

Stewart H. Hulse
Dept. of Psychology
Johns Hopkins University
Baltimore MD 21218

Eric Javel
Boys Town National
Institute for Communication
Disorders
555 N. 30th St.
Omaha NE 68131

Gerald Kidd, Jr.
Boston University
Dept. of Communication
Disorders
48 Cummington St.
Boston MA 02215

Michael Kubovy
Dept. of Psychology
Rutgers State University of
New Jersey
Psychology Building/Busch
Campus
New Brunswick NJ 08903

Marjorie R. Leek
Dept. of Communication
Disorders
University of Minnesota
Minneaplois MN 55455

James D. Miller
Central Institute for the
Deaf
818 S. Euclid
St. Louis MO 63110

Brian C.J. Moore
University of Cambridge
Psychological Lab
Downing St.
Cambridge CB2 3EB
England

Donna L. Neff
Boys Town National
Institute for Communication
Disorders
555 N. 30th St.
Omaha NE 68131

Roy D. Patterson
MRC Applied Psychology
Dept.
15 Chaucer Rd.
Cambridge
England CB2 2EF

David B. Pisoni
Dept. of Psychology
Indiana University
Bloomington IN 47405

William S. Rhode
Dept. of Neurophysiology
University of Wisconsin
Madison WI 53706

Virginia Richards
Dept. Psychology
University of Florida
Gainesville FL 32610

Donald A. Robin
Dept. of Speech Pathology-Audiology
University of Iowa
Iowa City IA 52242

Fred L. Royer
Dept. of Speech Pathology-Audiology
University of Iowa
Iowa City IA 52242

Murray B. Sachs
Dept. of Biomedical Engineering
Johns Hopkins School of Medicine
Baltimore MD 21209

Robert D. Sorkin
Psychological Sciences Building
Pierce Hall
Indiana University
West Lafayette IN 47907

Liza J. Stover
University of Kansas
Depts. of Speech, Language, Hearing
3129 Tomahawk Drive
Lawrence KS 66045

Enrst Terhardt
Tech. Univ Muenchen
Arcisstr 21, D-8000
Muenchen 2
Fed. Republic of Germany

Neal F. Viemeister
Dept. of Psychology
University of Minnesota
75 E. River Rd.
Minneapolis MN 54455

Raymond L. Winslow
Institute for Biomedical Computing
Washington University
St. Louis MO 63110

Charles S. Watson
Dept. of Speech and Hearing Sciences
Indiana University
Bloomington IN 47405

William A. Yost
Parmly Hearing Institute
Loyola University
6525 N. Sheridan Rd.
Chicago IL 60626

This page intentionally left blank

Preface

This book is the result of a workshop on the processing of complex sounds held in Sarasota Florida in April 1986. The workshop was supported by the Air Force Office of Scientific Research (AFOSR), Life Sciences, and was chaired by the editors of this book. A series of recent events led to the workshop and to publication of this book. In 1982, Dr. John Tangney of AFOSR asked the Committee on Hearing, Bioacoustics and Biomechanics (CHABA) of the National Academy of Sciences to survey recent developments and trends in the biological and behavioral study of the auditory system. The result of the request from AFOSR was a 1983 Symposium on Basic Research in Hearing organized by CHABA and sponsored by the AFOSR (Dolan and Yost, J. Acoust. Soc. Am. 78, No.1 Part 2, 1985). After reviewing the proceedings of the CHABA Symposium and considering its program goals, AFOSR began a program of support for research on complex auditory perception. The support by the AFOSR, the discussions at the CHABA Symposium, and the increase volume of research on the topic of auditory processing of complex sounds stimulated us to organize a meeting on this topic. With the support of the AFOSR the Sarasota Workshop on Auditory Processing of Complex Sounds was held in April, 1986.

Although the chapters in this book include most of the topics presented at the workshop, the book is not a "proceedings." We did not organize the workshop with the intent of publishing a book. The topics were chosen from the many excellent submitted papers in order to sample as diverse a cross-section of research as possible and yet provide continuity to the three day meeting. The quality and quantity of abstracts submitted for inclusion in the workshop and the enthusiastic and insightful discussions at the meeting convinced us and the participants that a timely publication devoted to these topics would be a useful contribution. Therefore, following the workshop the authors prepared chapters in camera-ready form in order to produce a book in a short period of time. The chapters are not just transcriptions of the presentations given at the workshop, but were written as brief papers on the topic of the author's interest. Authors were encouraged to provide some background and to make sure the germinal references on their topic were included in the bibliography.

Assistance for this project has come from many sources. The Workshop and the book would not be possible without the foresight and dedicated support of John

Tangney as a Program Director in the Life Sciences Division of AFOSR. The staff at the Sarasota Sheraton Hotel provided a pleasant environment in which to meet. Lawrence Erlbaum Press has been very helpful in assisting us in getting the book out quickly. The staff of the Parmly Hearing Institute at Loyola University, especially Marilyn Larson, Beth Langer, Ned Avejic, and Scott Stubenvoll have been invaluable, as has the staff of the Department of Speech and Hearing Sciences at Indiana University, especially Janet Farmer. Most importantly we want to thank the authors of the chapters and the participants and observers at the Workshop. Without their stimulating ideas, high quality research, and good-natured acceptance of our deadlines and ultimatums, there would be no book.

We hope this book captures some of the excitement we felt at the workshop as we discussed this new era in the study of auditory processing. All of the important contributions that are being made to understanding auditory processing of complex sounds could not be included in this single volume. However, the chapters do touch base with many of the lines of research and theory on complex sound and its perception, and they should provide both food for thought and a broad introduction to the literature on a topic that we are sure will be studied intensely during the next couple of decades.

WAY
CSW

Introduction: Auditory Processing of Complex Sounds

William A. Yost
Parmly Hearing Institute
Loyola University
Chicago, Illinois

Charles S. Watson
Department of Speech and Hearing Sciences
Indiana University
Bloomington, Indiana

This Introduction provides a general overview of some of the major concepts that appear throughout the chapters of the book. It is not an exhaustive coverage of the subjects discussed in the individual chapters, although it does attempt to highlight some of the new directions taken by the authors in their study of auditory processing of complex sounds.

This workshop brought together investigators with a remarkable diversity of approaches to the general problem of how humans (and nonhumans) process (or "hear," or "perceive") complex sounds. The only common denominator at the onset was that each had responded to an announcement (mailed or published in a journal), asking for contributed papers for a "workshop on complex sound perception." Surprisingly, this yielded a range of topics, research paradigms, and theoretical perspectives with some well-defined themes.

We anticipated that "complexity" would mean different things to different people, but the range of meanings that can be inferred from these twenty-eight papers is actually relatively small. In general, "simple sounds" are considered to be the individual pure tones or noise bursts that have served as the stimuli in most studies of the auditory system since Helmholtz. "Complex stimuli" mean those that vary systemically in either their spectrum, or in time, or both. While most of the contributors created complex stimuli to test particular hypotheses about auditory processing, a few dealt with natural or environmental sounds, speech, birdsongs, or music.

Many of the authors avoided the need to discuss physical criteria for stimulus "complexity," and instead opted for distinctions based on mechanisms of processing.

"Simple processing" in the spectral domain was equated by most authors with a critical band (CB) model, and in the temporal domain with the time constant of a simple temporal integrator. "Complex processing" was shown to require a considerable variety of mechanisms beyond these traditional workhorses of auditory theory, including spectral-shape and temporal-pattern detectors, and even more elaborate analyzers (hardware, software, or both) whose operation in many cases requires knowledge of the sources of complex sounds.

In general, the contributions can be divided into: (1) spectral processing, (2) temporal processing, (3) pitch, (4) speech, (5) physiological processing, and (6) perceptual organization; including "object" or event perception and central mechanisms. These a posteriori categories cannot, of course, capture the scope of numerous papers that treated more than one of these topics as, for example, several that dealt with stimuli varying both in spectrum and in time. The papers have been grouped into these six categories, but the reader is warned not to expect discussions of spectral processing to be confined to papers in the section bearing that name, and so forth.

The chapters that deal with spectral aspects of complex processing generally agree, as observed above, that considerably more elaborate frequency analysis can be demonstrated in psychoacoustic experiments than is predictable from a "bare-boned" critical-band filter bank. It should be stressed that none of these "failures of critical band theory" in fact provide evidence against the CB as an initial stage in frequency analysis. Several lines of investigation, however, demonstrate that when it is to the advantage of the listener to do so, he or she can simultaneously process energy arriving in several critical bands. That ability is demonstrated in two types of experiments. In one, a broad-band spectral array itself is treated as the meaningful event (i.e., a "signal"), rather than just one part of the spectrum (that associated with the output of a single auditory filter). Studies of "profile analysis" or spectral shape discrimination and its derivatives are examples of this approach. In the other, it is shown that temporal correlations among the noise levels across critical bands can reduce the masking efficiency of a critical-band masker (co-modulation release from masking or CMR). In both cases, mechanisms are implied which are simultaneously sensitive to the relative levels in each of a number of adjacent auditory channels. Common sense would have predicted at least one of these findings; vowel identification obviously requires recognition of spectral shape.

Some of the chapters discuss the nature of the physiological code that might subserve spectral pattern processing. The consensus seems to be that rate codes and temporal codes are both used by the central nervous system to process complex spectral patterns. These lines of research (both psychophysical and physiological) promise to establish the limits within which such spectral-shape- or profile-based recognition can operate.

Many sounds of everyday life may be described as temporal sequence of stimuli. If very similar (highly correlated) sounds occur in close temporal proximity, then under many circumstances, the auditory system is most sensitive to the first arriving information rather than to the pattern of the events. Studies of the precedence effect have provided insights into the mechanisms that govern the influence of the first acoustic wavefront. When the sequence of sounds is made up of different or uncorrelated acoustic events, the temporal pattern may lead to a variety of perceptions. Often times one part of a temporal pattern may be "heard out" from the background of the rest of the sound. In many contexts the last acoustic events are the most salient. The analogy to the foreground/background concepts of stream segregation (as derived from Gestalt Theory) is one theoretical approach to describe the dominance or saliency of certain aspects of a complex temporal pattern. Several computational schemes also provide insights into how to model discrimination among different sequences of sound. A variety of lines of research show the major role played by temporal modulation in our perception of complex sounds. The abundance of useful information available in the temporal code of the auditory nerve provides a physiological argument favoring temporal modulation as a variable around which many perceptions of complex sounds appear to be organized.

There are only so many words that can be used to describe a sound. One of the most common is "pitch." Although there is some disagreement about the precise definition of pitch, a variety of complex sounds are capable of producing sensations listeners refer to has having that perceptual attribute. Many authors consider pitch to be a major organizing feature for our perceptions of complex sounds. Models based only on auditory neural tuning or only on neural temporal periodicity, have failed to provide adequate descriptions of the pitch evoked by many complex sounds. Thus, the debate concerning whether complex pitch is spectrally or temporally based continues. Much of the research in this book suggests that the extraction of pitch from complex stimuli is not an "either-or" question. In both spectral shape processing and pitch processing neural tuning and temporal coding must be considered. In addition, although the auditory nerve contains a wealth of temporal and spectral information, central mechanisms are probably required to fully process the peripheral neural code in a manner adequate to account for complex pitch perception.

If a complex sound contains short term spectral changes then these might give rise to pitches which listeners could use in processing these sounds. The work on stream segregation, spectral shape discrimination, and tonal pattern recognition suggest the need to consider possible long-term and short-term spectral cues that may be used to detect, discriminate, or identify many complex sounds.

Most of the chapters generally conclude, not only that the peripheral mechanisms of auditory tuning and simple temporal integration are inadequate to explain the hearing of complex sounds, but also that some fairly elaborate central processing must be involved. A few papers explicitly deal with selective attention, short-term memory capacity, and other such cognitive constructs. It is clear that the "passive" auditory system is in fact very dynamic and can effectively be "programmed" to look like quite a variety of acoustic information processing devices. If we are to cope with such practical issues as auditory code learning (speech or non-speech), it is essential that we learn some of the primary limitations within which the central processor functions. How long can a sound be, if it is to be accurately recalled, or recognized later? How much of a complex sound must be processed "categorically," if any? Within what parameters must selective auditory attention function? Are there two auditory modes, one for speech and one for non-speech? Or, do we process very familiar sounds (e.g. speech in our native tongue) differently from novel sounds? Several papers made efforts to deal with these issues, but it is clear that a great deal remains to be done before we will understand the actual auditory processing that occurs at a cocktail party.

One fascinating line of thought carries on from the tradition of Gestalt Psychology. Certain organizing principles seem to be used when we hear a novel complex sound. Sometimes a portion of a total waveform "stands out", i.e. seems to be closer. That is an instance of auditory Gestalt perception. Those chapters concerned with Gestalt theory suggest a number of possible organizing rules for processing novel sounds. Certainly, frequency similarity is one of the potent determinants of forming a "figure" from the "background." It appears that musicians may be ahead of basic scientists in understanding some of these organizing axioms. Many of these concepts appear to be applicable whether we use speech and human communication, complex non-speech sounds, music, or an animal model, such as songbirds, as our tool for understanding auditory processing.

In general, the chapters in this book do two things. They provide many examples of why tuning and/or simple temporal integration are not sufficient mechanisms to account for the perception of complex sounds. They also review a variety of recent experimental findings that should be considered when models or theories of perceptual organization are proposed to describe auditory perception of complex sounds. These chapters provide a few answers and many, many questions. It is the latter, that indicates a rich future for the study of hearing.

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 <u>temporal</u> 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 each number of components (5, 11, 21, and 43) by randomly 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 temporal waveforms generally differed on each presentation. The amplitude spectra, however, were identical.

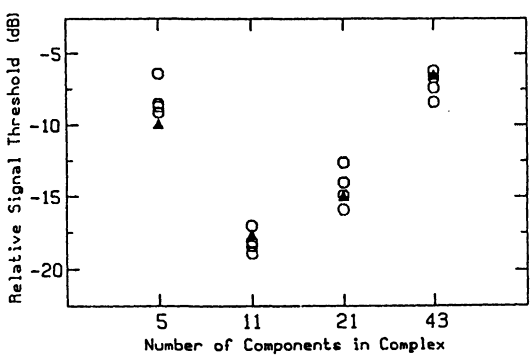


Figure 1. Signal threshold (dB) as a function of the number of components in the complex. Open circles: data obtained for each of the four phase-randomizations when the phase of each component was fixed throughout a block of trials ("fixed-phase" condition). Filled triangles: data from the "random-phase"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

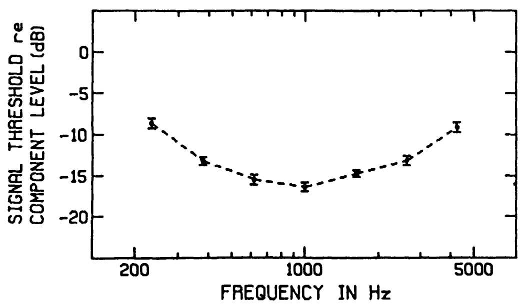
This question also bears on that of how the auditory system codes intensity. There are, at least, two different mechanisms that have been proposed as the basis for detecting changes in the intensity of sinusoidal components. One is what we will call the "rate" model. It assumes that changes in acoustic 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.



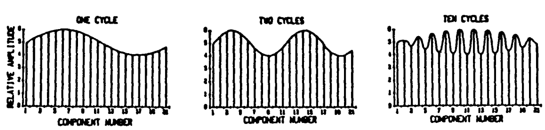
One other result from this recent study also deserves mention. The experiment described immediately above was repeated with one important exception. The median level of the standard was 60 rather than 40 dB SPL. This higher intensity level would be expected to produce firing rates at or close to saturation 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 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.

Auditory discrimination of complex sounds: The effects of amplitude perturbation on spectral shape discrimination.

Gerald Kidd, Jr.
Department of Communication Disorders, Boston University, 48 Cummington Street, Boston, Massachusetts 02215

This paper describes some of the conditions limiting performance in spectral shape discrimination experiments. Specifically examined is the relationship between detection of an alteration in a broadband reference spectrum, the effects of "random perturbation" in amplitude of the reference spectrum and the information available to the observer in sequential stimulus presentations. The experimental results help illustrate the complexities of devising a comprehensive model of "auditory profile analysis".

INTRODUCTION

The work described in this paper is based on a series of experiments performed during the past four years dealing with the perception of complex sounds. The author's collaborators on portions of this work were David M. Green, Thomas E. Hanna and Christine R. Mason.

From infancy, human listeners learn to identify a vast number of complex, nonspeech sounds based on their characteristic patterns of time-varying acoustic energy. Thus, we are able to detect, locate, attend to, and derive information from a multitude of sound sources in our environment.

As a result of the availability of electronic sound production equipment over one-half century ago, the study of auditory perception has focused on relating observer's perceptions to simple acoustic signals. Pure tones, and to some degree, bands of random noise, are readily characterized by the perceptual correlates of intensity, frequency and duration; that is, loudness, pitch and apparent duration. Further, it is relatively easy to vary the physical properties of these signals in a well-defined manner and expect that detectability or discriminability will also vary in an orderly and meaningful way.

An important goal of current research in audition is to relate the theories and models developed over the past half-century using simple acoustic signals to listener's auditory behavior in the complex, time-varying acoustic environment we live in. This is now a realistic goal because the technical tools have become available for generating complex sounds of arbitrary composition quickly and accurately. Just as the advent of modern electronics expanded and shaped auditory research early in the century, so has the availability of low-cost digital computers allowed the synthesis of virtually any physically realizable waveform. Thus, it is possible to devise a whole new class of experiments that would have been difficult, if not impossible, twenty years ago.

As the technical limitations on research in complex sound perception disappear, fundamental questions remain about what should be measured, how we should measure it and why it is important. Which aspects of complex acoustic signals have meaningful perceptual correlates? Which physical properties of sounds may be varied to produce orderly detection or discrimination functions that teach us something about perception of complex sounds in the "real world"?

EXPERIMENTAL RESULTS AND DISCUSSION

The experiments described in this paper were designed to measure the discriminability of complex sounds on the basis of spectral shape. Spectral shape discrimination, per se, received relatively little attention until a recent series of papers by Green and his colleagues from Harvard University (Spiegel et al., 1981; Spiegel and Green, 1982; Green et al., 1983; Green and Kidd, 1983; Green, 1983; Green et al., 1984; Mason et al., 1984; Green and Mason, 1985; Kidd, Mason and Green, 1986). The basic paradigm requires that the listener detect an alteration (such as an intensity increment or decrement) to a broadband spectrum. While such a measurement would seem to be a minor variant on the traditional tone-in-noise masking experiment, one important improvisation was added by Spiegel et al. (1981). They limited the usefulness of choosing the more intense sound of each pair by randomly "roving" the overall level of each sound from interval to interval of every 2AFC trial. They used the term "profile analysis" to describe the process whereby changes in spectral shape were discriminable. The first figure is a schematic of common reference and comparison stimuli that are used in spectral shape discrimination experiments. In this figure, two complex, multitone spectra are shown. The spectrum on the left is the reference spectrum and is composed of 21 equal-amplitude tones equally spaced in logarithmic frequency from 200 to 5000 Hz. The spectrum on the right is the comparison spectrum and is identical to that on the left, except that a signal has been added. The signal is a 1000 Hz sinusoid added in-phase to the 1000 Hz component of the reference spectrum. The result of the addition of the signal is an increment in the amplitude of a part of the reference spectrum. The listener may discriminate between the two sounds on the basis of a qualitative cue that reflects an alteration in the shape of the reference spectrum. To assure that discrimination is based on spectral shape, however, a within-trial random rove of the overall level of the sounds is employed. Thus, as shown in Figure 1, the nonsignal stimulus may have a greater overall SPL than the sound containing the signal limiting the usefulness of overall level as a cue in discrimination. All of the experimental results described in this paper were obtained using a 40-dB within-trial random rove of overall level. If the listener were to attend to the output of a single critical band centered on the 1000 Hz component, that level would contain little useful information. In this procedure, the level of any given sound is chosen randomly from a rectangular distribution of levels (cf. Mason et al., 1984).

The procedure of randomly roving the value of a reference sound to limit the information available for discrimination has been widely used in auditory psychophysics. For example, Harris (1952) used a "roving standard" frequency to measure the effects of

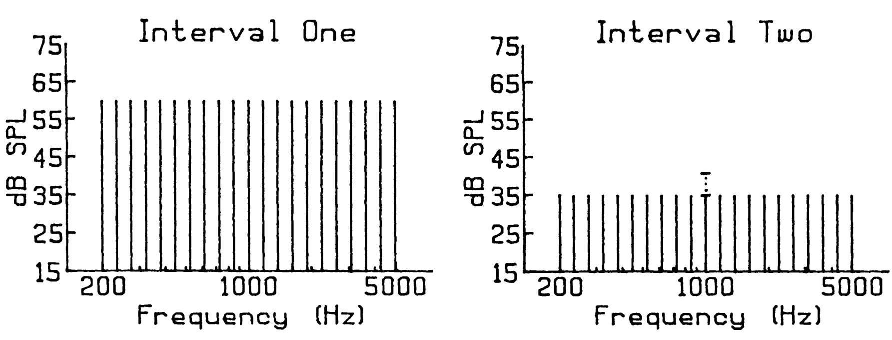


Figure 1. Multitone spectra used in spectral shape discrimination experiments.