REPRESENTATION OF SPEECH-LIKE SOUNDS IN THE DISCHARGE PATTERNS OF AUDITORY-NERVE FIBERS

by

BERTRAND DELGUTTE

Ingenieur diplôme de l' Ecole Polytechnique
(1974)

S.M., Massachusetts Institute of Technology
(1976)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS OF THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

August 1981

Signature of Author .....
Department of Electrical Engineering and
Computer Science, August 7, 1981

Certified by .....
Kenneth N. Stevens
Thesis Supervisor

Accepted by .....
Arthur C. Smith
Chairman, Department Committee on Graduate Students

REPRESENTATION OF SPEECH-LIKE SOUNDS IN THE DISCHARGE PATTERNS OF AUDITORY-NERVE FIBERS

by

BERTRAND DELGUTTE

Submitted to the Department of Electrical Engineering and Computer Science on August 7, 1981 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Electrical Engineering.

ABSTRACT

This thesis is an experimental study of how acoustic information used by humans to distinguish between speech sounds is coded at the level of the auditory-nerve. The experiments consisted in recording responses of single auditory-nerve fibers in anesthetized cats to computer-generated stimuli having properties important for phonetic distinctions. The stimuli were presented at two sound levels typical for conversational speech.

Part of the thesis addresses the question of whether it is possible to define a response measure from auditory-nerve fiber discharge patterns that provides a representation of the phonetically-important features of the short-time spectrum of speech over a wide range of stimulus levels and signal-to-noise ratios. Speech sounds can be categorized into "sonorants", characterized by a well defined formant pattern in the low frequencies and a periodicity at the fundamental frequency, and "obstruents", which have most of their energy above 1-2 kHz and a noise-like waveform.

Steady-state, two-formant vowels were used as models for sonorants. In response to these stimuli, the discharges of the majority of auditory-nerve fibers are synchronized to harmonics near one of the formant frequencies. This pattern contrasts with responses to broad-band noise stimuli, for which discharges are primarily synchronized to components near the fiber characteristic frequency (CF). By processing the response patterns with a bank of filters that extract response components around the CF of each fiber, and plotting the rectified and lowpass-filtered outputs of the filters against center frequency, a profile is formed in which local maxima provide information about the formant frequencies and the fundamental frequency of the vowel stimuli. The filters, which are compatible with the psychophysical concept of critical bands, can be realized by known synaptic operations in the central nervous system such as delays and summation. These results are valid at the two levels and in lowpass background noise typical for conversations.

Steady-state, voiceless fricative stimuli were used as models of obstruent sounds. The profile of average discharge rate against CF provides information about the frequency regions in which these stimuli have most of their energy. This information would suffice to make phonetic distinctions among fricatives produced with different places of articulation. In contrast, the response measures obtained by filtering the response patterns around CF fail to provide a representation of the phonetically-important information for fricatives. Because average discharge rate is not an adequate measure for vowels at high stimulus levels (Sachs and Young, J. Acoust. Soc. Am. 66, 858-875, 1979), these results imply that two different response measures would be required for the representation of phonetic information in sonorants and obstruents.

In contrast to vowels and fricatives, most speech sounds are not steady-state, and show rapid changes in amplitude and spectral characteristics. To study the effects of these rapid changes on the representation of spectral information, /da/-like formant trajectories were preceded by a context whose characteristics were manipulated to produce stimuli sounding like /da/, /ada/, /na/, /$\breve{s}$a/, /sa/, and other syllables. For stimuli like /da/ and /ada/, in which the formant transitions are preceded by a silence, fibers over a wide range of CF's responded with a peak in discharge rate during the formant transitions. For stimuli with a low-frequency context as in /na/, the discharge rates of low-CF fibers were reduced considerably during the formant transitions, whereas for stimuli with a high-frequency context as in /$\breve{s}$a/ and /sa/, the discharge rates of high-CF fibers were decreased, even though the formant frequencies were the same in all cases. Thus, the positions along the CF dimension of the peaks in discharge rate at consonantal release provide information for distinctions between stop, nasal and fricative consonants. In a separate experiment with /$\breve{s}$/ and /$\breve{c}$/ stimuli differing in the duration of their rise in amplitude at the onset, the characteristics of the peaks in discharge rate at the onset was shown to provide sufficient information to distinguish between the two stimuli. The peaks in discharge rate that occur in specific CF regions when the speech signal shows a rapid change in amplitude or spectral characteristics, peaks could be used by the central processor as markers to select regions of the spatio-temporal pattern of auditory-nerve discharges for further processing.

Thesis Supervisor: Kenneth N. Stevens

Title: Lebel Professor of Electrical and Bioengineering

ACKNOWLEDGEMENT

I wish to thank Dr. N.Y.S. Kiang for his role in initiating this work and guiding it to completion. He showed unrelenting faith, trust and good humor, and was always accessible, understanding and encouraging when needed. Though he let me determine the topics of investigation and the conduct of the research, he knew how to direct the work towards deeper questions by thought-provoking discussions and by establishing novel connections.

I express deep-felt gratitude to my thesis supervisor and faculty advisor Professor K.N. Stevens for his interest, guidance and support throughout the seven years I spent at MIT. I particularly appreciate his faith in encouraging this new line of investigation. His manner of looking at speech is the foundation of the thesis, and his ideas often directly influenced the design of the experiments.

I thank the thesis committee for their care in reading the manuscript and their valuable comments. Contact with these people was one of the most profitable experiences of my career at MIT. Professor H.S. Colburn made probing criticisms, and revealed the implications of burgeoning ideas. Dr. D.H. Klatt provided much advice about speech synthesis, and formulated pointed criticisms that forced me to sharpen certain conclusions. Professor W.T. Peake never failed to

detect confused thinking clouded by jargon, and made many suggestions that greatly improved the writing. Professor V.W. Zue, who had the difficult role of the "outsider", was valuable in pointing out points that needed clarification.

Stimulating discussions with S. Seneff led to some of the most interesting results of this thesis. Dr. D. Isenberg helped with the synthesis of the stimuli and the design of some of the experiments.

I owe much to the people of Eaton-Peabody Laboratory who created a pleasant working atmosphere. P.M. McGaffigan and L.W. Dodds drew the difficult figures without regard to time and effort, and readers will undoubtedly look at these drawings for their esthetic value as well as their information content. M.M. Jazak, P.M. McGaffigan, J.W. Larrabee, E.M. Marr, L.P. Miller, and J.M. Rho overcame their legitimate reluctance at performing surgeries on our furry little friends. I thank the engineering staff, particularly F.J. Stefanov and D.A. Steffens for their general helpfulness and specific involvement in this project. During the early stages of this work, Dr. M.C. Liberman, Dr. B. Wang and K.E. Michel taught me much about recording from auditory-nerve fibers.

This work was supported by NIH grants NS04332 and NS13126, by a C.J. Lebel Fellowship, and by a Grass Instruments Corporation Fellowship.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES OF CHAPTER I

FIGURES OF CHAPTER II

FIGURES OF CHAPTER III

FIGURES OF CHAPTER IV

INTRODUCTION

The processing of speech sounds by the ear is the first stage in the decoding of acoustic stimuli during speech communication. The discharge patterns of auditory-nerve fibers constitute the output of the ear. On the basis of the information present in these patterns, the brain can identify the linguistic message intended by the speaker. In recent decades a great deal of progress has been made in both auditory physiology and speech science. It has become almost routine to record the electrophysiological activity of auditory-nerve fibers from several species of mammals in response to acoustic stimuli. Such experiments with simple stimuli have led to theoretical concepts that might be useful in interpreting responses to complex stimuli such as speech. With the advances in techniques to analyze and synthesize speech, the acoustic characteristics that are used to distinguish between speech sounds are beginning to be understood. Thus the time seems appropriate to combine our knowledge of speech and auditory-nerve responses in studying how speech-like sounds are represented in the discharge patterns of auditory-nerve fibers. Experiments on the responses of auditory-nerve fibers to speech and speech-like stimuli have been reported, (Kiang and Moxon, 1972,1974; Hashimoto et al., 1975; Sachs and Young, 1979,1980; Young and

Sachs, 1979), but a systematic description of the coding of the acoustic characteristics that are used in language for phonetic distinctions is still lacking.

For students of language, a better understanding of speech processing by the auditory system could provide clues as to how classes of sounds that are used for phonetic distinctions are selected among the broader set of sounds that can be produced by the articulators, and lead to the formulation of constraints on how these sounds can be combined in speech. For auditory physiologists, the responses of auditory-nerve fibers to speech stimuli provide a test for functional descriptions of the processing that occurs in the ear. For auditory psychophysicists, descriptions of responses to speech-like sounds might serve as the basis of a search for models of auditory behavior that would be consistent with physiological mechanisms.

Studies of speech coding in the auditory nerve may also have practical applications in communication engineering and the treatment of communication disorders. Normal human subjects are able to communicate effectively in conditions of low signal-to-noise ratio or severe degradation of the signal. This is not the case for certain hearing-impaired subjects. Many current analysis/synthesis or speech recognition systems do not perform well in noise. If the good performance of normal listeners under signal degradation is at least partially determined by the type of processing that occurs in

the ear, one might hope that a better knowledge of this processing could be used in designing better hearing aids and communication systems. Another potential application is the design of cochlear prostheses in which one would try to mimic the normal pattern of activation of the auditory nerve more closely.

The experiments reported in this thesis consisted in recording responses of single fibers in the cat's auditory nerve to computer-generated stimuli having some of the acoustic characteristics that are important for phonetic distinctions. Because techniques to record and process auditory-nerve data have become routine, the main methodological consideration was the choice of the stimuli. The choice was guided by knowledge of the acoustic characteristics that are important for phonetic distinctions, and by expectations as to how the response properties of auditory-nerve fibers might interact with the characteristics of the stimulus to enhance phonetically-important information. In order to get a broad picture of speech coding in the auditory nerve, stimuli representing a wide range of acoustic characteristics have been used.

The short-time spectra of speech sounds fall into two major categories: Vowel-like or "sonorant" sounds have nearly periodic waveforms and prominent spectral peaks in the frequency region below 3 kHz, whereas noise-like or "obstruent" sounds have irregular waveforms and most of their

energy above 1-2 kHz. The manner in which the spectra of speech sounds are coded in the auditory nerve is studied in Chapter I for sonorants and Chapter II for obstruents, using simple steady-state stimuli. However, most sounds of speech are not even approximately steady-state, and for these sounds, portions of the speech waveform in the vicinity of rapid spectral changes seem to contain most of the phonetic information. In Chapter III, some phonetic distinctions involving rapid changes in amplitude and spectral characteristics are treated. In these three chapters, several response measures or processing schemes are proposed for extracting phonetically-important features from auditory-nerve data. As a test of the adequacy of these measures, and because speech communication typically occurs in noise, some results with speech-like stimuli presented in background noise are reported in Chapter IV.

Because these experiments were based on a broad set of stimuli, they had to be limited in other respects. First, recordings were restricted to the most sensitive population of auditory-nerve fibers, those with spontaneous discharge rates greater than about 18 spikes/s (Liberman, 1978). Second, stimuli were presented at only two levels which are typical for conversational speech. Limitations of this thesis are discussed in more detail at the end of Chapter I.

NOTES

The four chapters of the thesis are organized as self-contained papers that are intended to be published. This organization leads to minor repetitions, particularly in the description of the methods.

The results reported in Chapter I and Chapter IV are based on a set of nine vowel stimuli that are designated by phonetic symbols in the figures. For convenience, the phonetic symbols for some of these stimuli have been replaced by typewriter codes in the text:

/ɛ/ ---> /eh/

/ɑ/ ---> /a/

/ɔ/ ---> /oh/

/ə/ ---> /ax/

CHAPTER I

CODING OF VOWEL-LIKE SOUNDS IN THE AUDITORY NERVE

INTRODUCTION

Distinctions between speech sounds are based on a number of acoustic properties varying along different dimensions of the speech signal (Liberman, 1957; Fant, 1960,1973; Stevens and House, 1972; Stevens, 1980). It is useful to consider each speech sound as being characterized by the set of acoustic properties that distinguish it from the other speech sounds of the language (Jakobson et al., 1952; Chomsky and Halle, 1968; Ladefoged, 1971). It appears that many of these distinctive acoustic properties are shared by many speech sounds, so that a relatively small number of properties can establish a large number of phonetic distinctions in many languages. Classes of speech sounds that share one or several distinctive properties constitute natural categories such as vowels, stop consonants, and nasal sounds. How this general organization of speech sounds is represented in the auditory nervous system, even at its most peripheral stages, is unknown.

Some studies of responses of auditory-nerve fibers to speech and speech-like stimuli have been reported (Kiang and Moxon, 1972,1974; Hashimoto et al., 1975; Kiang et al.,

1979; Sachs and Young, 1979,1980; Young and Sachs, 1979; Delgutte, 1980). These studies suggest that concepts describing responses to simple stimuli are useful in interpreting responses to speech-like stimuli. The present paper is the first report of a series of electrophysiological experiments aimed at describing systematically how acoustic properties important for phonetic distinctions are represented in the discharge patterns of auditory-nerve fibers. The organization of speech sounds as complexes of distinctive properties is used to design simplified speech-like stimuli having only a few of the acoustic characteristics important for phonetic distinctions. Because each distinctive property is shared by many speech sounds, experiments with a small number of speech-like stimuli may be relevant to a whole class of speech sounds.

Speech sounds can be divided into "sonorant" sounds, characterized by a prominent formant structure in the low-frequencies and a quasi-periodic waveform, and "obstruent" sounds, that have mostly high-frequency energy and a noise-like waveform. Sonorants include vowels, glides and nasal consonants, whereas obstruents include stop and fricative consonants. This paper is concerned with simple sonorants, steady-state vowels. The second paper in this series is about simple obstruents, steady-state, voiceless fricative consonants. Whereas vowels and fricatives can have nearly steady-state spectral characteristics, many consonants

are characterized by rapid changes in amplitude and spectral characteristics. These regions of rapid change are studied in the third paper. The fourth paper describes how background noise degrades the representation of certain phonetically-important acoustic properties in the discharge patterns of auditory-nerve fibers.

Vowels are characterized by a periodicity at the fundamental frequency, and a spectrum having most of its energy near the frequencies of the first two or three formants. Because of constraints imposed by the speech production mechanism, the spectrum envelope of vowels is almost entirely specified by the formant frequencies (Fant, 1960,1973). Perceptual data suggest that phonetic distinctions among vowels are primarily based on the formant pattern (Peterson and Barney, 1952; Fant, 1973; Pols et al., 1969; Carlson and Granstrom, 1980). For sonorant sounds in general, the time-variations in formant frequencies are of great perceptual importance (Cooper et al., 1952; Liberman, 1957).

The coding of the formant pattern in the auditory-nerve has been investigated in considerable detail (Sachs and Young, 1979,1980; Young and Sachs, 1979). The study, based on steady-state vowels modelled after the American English vowels /I/, /eh/, /a/ and /u/, showed that, at stimulus levels typical for conversations, the profile of average discharge rate against fiber characteristic frequency

does not show clear peaks at the formant frequencies. In contrast, fine time patterns of discharge provide considerable information about the formant frequencies over a wide range of stimulus levels and signal-to-noise ratios.

The present study of the coding of vowels is based on a larger set of steady-state, two-formant vowels that were chosen to cover the range of formant patterns that is found in language. Two-formant vowels were chosen because they can be characterized by essentially two parameters. The nine stimuli are similar to the "cardinal vowels" that are used by phoneticians as reference sounds in identifying vowel-like sounds in unknown languages (Jones, 1956). This choice of stimuli should facilitate correlation between patterns of auditory-nerve activity and dimensions that are important for phonetic distinctions.

Though this paper is primarily concerned with the representation of the formant pattern, some attention is given to the fundamental frequency. This parameter, which is heard by listeners as voice pitch, is important for the voiced/voiceless distinction and for its prosodic functions. In tone languages, such as Chinese, variations in fundamental frequency as well as formant frequencies affect word identity. Because of the importance of formant frequencies and fundamental frequency in speech perception, a major goal of this paper is to define speech-processing schemes by which these parameters could be extracted from the discharge

patterns of auditory-nerve fibers over a wide range of stimulus levels and signal-to-noise ratios (Young and Sachs, 1979). Such processing schemes should be consistent with psychophysical data and current knowledge of central auditory processing.

I METHODS

A. Stimuli

The two-formant vowel stimuli are generated by a computer program that simulates the summed outputs of two bandpass filters excited by a periodic pulse train. The repetition rate of the pulses, which represents the fundamental frequency of voice, is 125 Hz for all stimuli. The bandpass filters model the resonances of the vocal tract associated with the formant frequencies F1 and F2. Details of stimulus generation are given in Appendix A.

Figure 1 shows the waveforms and spectra of the 9 stimuli that were used in these experiments. Although in all cases the spectra have clear peaks near the formant frequencies, the height of the components in the vicinity of these peaks depends on the relationship between the formant frequency and the fundamental frequency. When a formant frequency coincides with a harmonic of the 125-Hz fundamental this harmonic is considerably more intense than the adjacent harmonics. This occurs, for instance, for both formants of

/ax/ which correspond to harmonics 4 and 12. However, in general, the formant frequencies do not coincide with harmonics of the fundamental frequency. For instance, the 700-Hz first formant of /a/ is nearly halfway between the 625-Hz fifth harmonic and the 750-Hz sixth harmonic, so that the two harmonics have nearly equal amplitudes.

Figure 2 shows the positions of the 9 vowel stimuli in the F1-F2 plane. Because the stimuli have only two formants, the second formant frequencies were set to the values that are obtained in matching experiments with natural vowels (Carlson et al, 1975; Bladon and Fant, 1978). These matched values are usually close to the second formant of natural vowels, except for vowels with a high F2 such as /i/ and /e/ for which they are closer to F3. In Fig. 2, the vowel stimuli, except /ax/, are distributed along a quadrilateral where "peripheral" vowels in many languages tend to be found. The vowel /ax/, located roughly at the center of this quadrilateral, is often called "neutral vowel". This vowel, together with /i/, /ae/ and /u/, which have the most extreme formant frequencies, will be used in some of the figures to illustrate representative results.

Dimensions that are traditionally used in phonology to describe contrasts between vowels can be related to the diagram of Fig. 2. These dimensions will be useful in describing responses of auditory-nerve fibers. The vowels /i/, /ɨ/ and /u/, which have a low first formant, are called

"close" vowels because of the position of the tongue during their production. The contrasting "open" vowels /ae/ and /a/ are characterized by a high F1. There is also a contrast between the "back" vowels /u/, /oh/ and /a/, which have a low second formant, and the "front" vowels /i/, /e/, /eh/ and /ae/, which are characterized by a high F2. Besides these traditional phonological classifications, some investigators have argued for the perceptual importance of the "spread" dimension defined by the ratio F2/F1 (Fant, 1973; Chistovich et al., 1979). Figure 2 shows that /i/, /e/ and /ɨ/ are the most spread vowels, whereas /a/, /ae/ and /oh/ are the least spread.

B. Experimental procedures

The preparation of the animals and techniques to record from single auditory-nerve fibers are essentially as described in Kiang et al.(1965). Healthy adult cats weighting 1.6 to 3.5 kg were anesthetized with diallyl barbituric acid in urethane solution. A cannula was inserted into the trachea, the cartilaginous external auditory meatus was cut near the tympanic membrane, and the bulla cavity was opened to expose the round window. The posterior fossa of the cranial cavity was opened dorsally and the cerebellum retracted medially to allow access to the auditory nerve. The animal was placed in a sound-proof, vibration-insulated, electrically-shielded chamber (Ver et al., 1974).

An acoustic cavity containing the sound transducer (B&K 1" condenser microphone) and a calibrated probe microphone was sealed into the meatus. Transfer characteristics of the acoustic system in individual animals were always flat within $\pm$4 dB for frequencies below 5 kHz, and the maximum sound pressure in that frequency region was about 115 dB SPL. Sound levels of the vowel stimuli were set relative to the average of the transfer ratio over the frequency range 0.1-5 kHz. Figure 3 shows the mean transfer function for the 27 animals used in the present study. For frequencies below 8 kHz, the mean transfer function is nearly flat, and the standard deviation of inter-animal variations does not exceed 2 dB.

The state of the preparation was assessed by measuring the threshold of the click-evoked N1 potential recorded near the round window. The click attenuation level at the N1 threshold was between 70 and 85 dB. Single-fiber responses were recorded with micropipettes filled with 2M KCl and with impedances between 20 and 30 Mohm. For each fiber, the characteristic frequency (CF) and threshold at CF were estimated from a tuning curve measured using the algorithm described by Kiang et al. (1970) and Liberman (1978). To obtain a relatively homogeneous population, fibers whose threshold was more than 15-20 dB above Liberman's (1978) "best threshold curve" were discarded. In effect, this fiber selection procedure restricts recordings to the most sensitive

category of auditory-nerve fibers, i.e. those with spontaneous discharge rates greater than 18 spikes/s (Liberman, 1978). Figure 4 shows a scatter plot of threshold at CF against CF for all the fibers that were used in the present study. The threshold spread in any CF region exceeds that found for the most sensitive population of fibers in individual animals by about 10 dB (Liberman, 1978).

The 184-ms vowel stimuli were presented repeatedly at a rate of 100/min for the computation of post-stimulus time (PST) histograms (Gerstein and Kiang, 1960). The histograms were computed with a bin width of 0.05 ms on the basis of 250 to 500 presentations of the stimulus. This bin width is adequate for accurate estimation of the frequency components of the instantaneous discharge rate up to roughly 4 kHz (Johnson, 1978). Whenever possible, PST histograms in response to vowel stimuli were obtained for sound levels of 60 and 75 dB SPL, which are typical for conversational speech. Histograms were computed only for characteristic frequencies up to 4-5 kHz for the 60-dB vowels, and up to 8-10 kHz for the 75-dB vowels, because responses were hard to detect for higher CF's.

C. Processing of Spike Data

Period histograms were computed from the PST histograms by adding the histogram waveforms for each period of the sound burst. The first two and last period of each

burst were discarded because the response pattern in the vicinity of the transients could be significantly different from the pattern in the central portion of the burst. Discrete Fourier transforms of the period histograms were used to estimate the synchronization indices for each harmonic of the 125-Hz fundamental up to 5 kHz (Johnson, 1978; Young and Sachs, 1979). The synchronization index, which varies between 0 and 1, is a measure of how well fiber discharges are synchronized to a particular frequency component. It is defined as the magnitude of the Fourier component at one frequency divided by the DC component, which is the average discharge rate. Plots of synchronization index against harmonic frequency (as in Fig. 5) will be called "normalized harmonic spectra". Autocorrelation functions of period histograms were computed with the assumption that the histograms are periodic. The resulting functions have the same 8-ms period as the stimulus and have an absolute maximum corresponding to the mean square discharge rate at time zero.

To obtain data at regular intervals along the log CF dimension, it was found useful to average the spectra of period histograms for all fibers whose CF lies within a narrow band of frequencies. Specifically, such "band-average" spectra were computed at intervals of 0.25 octave using a trapezoidal weighting window with a central width of 0.25 octave and a total width of 0.55 octave (effective width 0.4 octave). The weighted number of units used in the computation of each band-average spectrum was about 3 on the average.

The band-averaging procedure is valid only if the spectra of fibers within each CF band are sufficiently similar. Figure 5 compares response spectra for 3 units in each of 5 CF bands with the corresponding band-average spectra. Whenever possible, units in the first column are near the low-frequency limit of each band, units in the second column near the center, and units in the third column near the high-frequency end. For the bands centered at 0.23, 0.46, and 1.83 kHz, the spectra of all three units share the same largest component, so that the band-average spectrum is also dominated by that component. This pattern is typical when the band center-frequency is close to one of the formant frequencies. For the band centered at 3.67 kHz, the band-average spectrum also resembles the spectra of the individual units, except for the unit in column 2. The spectra for the units in the band centered at 0.92 kHz show multiple peaks whose locations vary between units. The band-average spectrum has intense components over a broad frequency region where spectral peaks of the individual units are found. Such relatively large variations between units are usually restricted to narrow CF regions between the formant frequencies. Thus, in the majority of cases, band-averaging preserves the major features of the spectra of individual units.

II. RESULTS

A. Spatial distribution of spectral components for vowel stimuli

Figure 6 shows period histograms obtained for units with 7 different CF's in response to the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL. For each vowel, the response pattern changes considerably as a function of CF, so that a description of responses for the entire array of auditory-nerve fibers is required. For each CF, the response patterns to the three vowels are usually quite different, so that responses of individual units contain information about the identity of the stimulus. Figure 7 shows the normalized harmonic spectra of the period histograms shown in Fig. 6. Together, these two figures illustrate how fiber responses vary as a function of CF and of the formant frequencies of the vowel stimuli.

The 0.25-kHz first formant of /i/ is close to the CF of the 0.23-kHz unit. The largest component in the response spectrum of this unit is the first formant frequency, and the period histogram consists of two peaks separated by intervals of 1/F1. In the spectrum of the 0.35-kHz unit, the first formant is also the largest component, but the components at 2F1=0.5 kHz and 3F1=0.75 kHz are also large. Correspondingly, the period histogram shows two pairs of peaks separated by intervals of about 1/F1. For the other two vowels, frequency

components near F1 are also the largest in the response spectra of the units whose CF is within one octave of the first formant (the 0.65 and 1.1-kHz units for /ae/, and the 0.23 and 0.39-kHz units for /u/). However, unlike the first formant of /i/ which coincides with harmonic 2, the 0.3-kHz first formant of /u/ is approximately halfway between the second and third harmonic. The largest component in the spectrum of the 0.23-kHz unit for /u/ is the second harmonic, whereas the third harmonic component is the largest for the 0.39-kHz unit. Correspondingly, the number of peaks in the period histogram is two for the 0.23-kHz unit and three for the 0.39-kHz unit. A similar result is obtained for /ae/ in which the 0.8-kHz first formant is between harmonics 6 and 7: the period histogram of the 0.65-kHz unit has 6 peaks, whereas the histogram of the 1.1-kHz unit has 7 peaks.

The largest components in the response spectra of low-CF units are not always harmonics near F1. The CF of the 0.23-kHz unit is nearly two octaves below the first formant of /ae/. The period histogram for this unit shows two peaks at intervals of about 1/CF, and the the largest spectral component is harmonic 2 which is the closest to CF. The response spectrum of the 0.39-kHz unit for /ae/ also shows intense low-frequency components as well as components near F1.

Units whose CF is close to the second formant of a vowel stimulus (the 3.1-kHz unit for /i/, the 1.8-kHz unit for

/ae/, and the 0.65 and 1.1-kHz for /u/) have intense response components near F2, and their period histograms show peaks separated by intervals of about 1/F2. For the /i/ stimulus, the response spectrum of the 3.1-kHz unit has a peak at the fundamental frequency in addition to the peak near F2. Because the first and second formants of /i/ are separated by a wide frequency range, the largest response components in the spectra of the 0.65, 1.1 and 1.8-kHz units are not close to either F1 or F2. The spectra of the 1.1 and 1.8-kHz units show a broad peak near the CF and a peak at the fundamental. Correspondingly the period histograms show peaks at intervals of 1/CF during the first half of the period. The period histogram of the 0.65-kHz unit has peaks at irregular intervals and its spectrum shows many prominent components including those near CF.

Units whose CF is more than about one octave above the second formant tend to have response spectra with broad peaks at several frequencies. The spectra of the 3.1 and 5.2-kHz units show peaks near both F1 and F2 in response to the /ae/ stimulus, with the F1 peak being somewhat larger. In contrast, components near F2 are the largest for the 1.8 and 5.2-kHz units in response to /u/, while the fundamental component is the largest for the 5.2-kHz unit in response to /i/.

In summary, the largest spectral components in the response patterns of auditory-nerve fibers for vowel stimuli

are almost always harmonics close to one of the formant frequencies, the fundamental frequency or the fiber CF. The position of the fiber CF relative to the formant frequencies appears to be a good predictor of the main features of the response pattern.

Figure 8 shows pseudo-perspective displays of the distribution of spectral response components across the array of auditory-nerve fibers for the nine vowel stimuli presented at 75 dB SPL. For most vowels, the strongest activity is distributed along the horizontal dashed lines corresponding to the formant frequencies, indicating that the response patterns of auditory-nerve fibers are dominated by frequency components near the formant frequencies through most of the CF range. Prominent activity distributed along the f=CF line occurs only in the CF region between F1 and F2 for the vowels /i/, /e/ and /ɨ/, and in the low-CF region for /ae/, /a/, /eh/ and /oh/.

The array of auditory-nerve fibers can be divided into five CF regions within which the response spectra of Fig. 8 have many features in common. The five regions consist of a low-CF region below F1, a region centered at CF=F1, an intermediate region between F1 and F2, a region centered at CF=F2, and a high-CF region above F2. The characteristics of these CF regions can be related to the phonological classifications of vowels introduced in Sec. IA.

For all vowels, there is a region around the first formant where the harmonics closest to F1 dominate the response patterns. Distortion components of F1 are also large, and in particular, for the spread vowels /i/, /e/ and /ɨ/, there is a band of CF around the 2F1 place in which 2F1 can be the largest component. The CF region of dominance by F1 and its harmonics extends over about one octave below F1, and, for vowels in which F2 is at least 1.5 octave above F1 (/i/, /e/, /ɨ/, /eh/ and /ax/), about 1 to 1.25 octave beyond F1.

The region around CF=F1 is flanked on the low-CF side by another region in which the largest response components are the harmonics closest to the CF. For vowels with a high F1, such as /ae/ and /a/, this region extends up to about 0.4 kHz, whereas it is not found for vowels with a low F1 such as /i/, /u/, and /ɨ/, at least for CF's above 0.16 kHz. Thus the open/close dimension of phonology correlates both with the position of the F1-dominated region along the CF dimension and with the extent of the low-frequency CF-dominated region.

For all vowels, there is a CF region around F2 where harmonics near the second formant are the largest response components for frequencies above the fundamental. The fundamental response component can be larger than the F2 component over part of the region for vowels with a high F2 such as /i/, /e/, /eh/, /ae/ and /ɨ/. For /e/, /eh/, /ae/,

and /ax/ a considerable F1 component can also be found in that region, as well as distortion components such as F2±F1. For vowels whose second formant is more than 1.5 octave above F1 (except /i/), the region of F2 dominance extends over about 1/2 octave below F2. Above F2, the extent of this region seems to decrease with the second formant frequency, varying from more than one octave for the back vowels /u/ and /oh/ to less than 1/2 octave for /i/. Thus the front/back distinction of phonology correlates with the position and extent of the F2-dominated region along the CF dimension.

For vowels in which the two formant frequencies are more than 2 octaves apart (/i/,/e/, and /ɨ/), there is a CF region in which the response spectrum shows a broad peak near CF and a prominent fundamental component. For vowels in which F1 and F2 are closer together, there is a direct passage from the region dominated by F1 to the region dominated by F2 without an intermediate region. For these vowels, the boundary between the two regions is usually about 2/3 of the distance between F1 and F2, in logarithmic units. Thus, the spread dimension of phonology correlates with the extent of the intermediate region and of the regions dominated by F1 and F2.

Finally, there is a CF region above F2 in which the high-CF fibers respond to the intense, low-frequency components of the vowel stimuli. Response spectra in that region show broad peaks at the formant frequencies and the

to the voice pitch. An intense fundamental usually indicates that the envelope of the period histogram shows prominent fluctuations at the fundamental frequency, whereas a weak fundamental indicates a flatter envelope (Fig. 6 and 7). Figure 10 shows the distribution of the fundamental component across CF for the vowels /i/, /ae/ and /u/ presented at 60 and 75 dB SPL. At both stimulus levels, the fundamental component is large in the high-CF region for all vowels, in the CF region between F1 and F2 for /i/, and near 0.3-0.5 kHz for /ae/. More generally, large fundamental components are found in the CF region below F1 for open vowels, and between F1 and F2 for spread vowels. In the few cases in which a unit with a CF near the fundamental frequency was recorded from, the fundamental component was always large. In contrast, the fundamental component is small near the formant frequencies, except the second formant of /i/. There is also a tendency for the fundamental response component to be smaller at the higher stimulus level.

In summary, dimensions of vowel stimuli that are important for phonetic distinctions are represented in the auditory nerve as variations in the extent and position of different CF regions over which response characteristics are relatively homogeneous.

B. Possible speech processing schemes

Because psychophysical data show that the formant frequencies and fundamental frequency play a major role in speech perception, it is important to define processing schemes by which formant frequencies and fundamental frequency could be estimated from auditory-nerve responses. In this section two approaches to the estimation of speech parameters are introduced. In the first approach, these parameters are extracted from the response patterns exclusively, while in the second approach, more akin to that of Young and Sachs (1979), the value of the fiber characteristic frequency as well as the response pattern enters into the computation of the response measure. Both approaches are based on fine time patterns of discharge and preserve the place dimension defined by the characteristic frequency parameter.

1. Prominent spectral components

The first processing scheme to be discussed is based on the observation that the major zones of activity in Fig. 8 and 9 are located along the horizontal lines at the formant frequencies. Figure 11A shows the largest component in the response spectrum of individual fibers plotted against CF for the /i/, /ae/, /u/ and /ax/ stimuli presented at 75 dB SPL. Figure 11B shows the same response measure for the band-average spectra. For all vowels except /i/, the largest spectral component is close to one of the formant frequencies

over nearly the entire range of CF, and jumps from F1 to F2 without passing through intermediate frequencies. For the /i/ stimulus, the largest spectral component coincides with F2 only over a narrow strip, and is near CF for many fibers. With the exception of the second formant of /i/, the formant frequencies (more precisely the harmonics of the 125-Hz fundamental closest to the formant frequencies) could in principle be estimated by constructing a histogram of the distribution of the largest spectral component for the entire range of CF's, and picking the two largest modes in the histogram.

The last two rows of Fig. 11 show another response measure, which behaves similarly to the largest response component, but is based on the autocorrelation function of period histograms instead of the spectrum. Figure 12 shows the autocorrelation functions for narrow CF bands in response to the vowel /ax/ presented at 75 dB SPL. For bands with CF's near the first formant frequency, and in the high-CF region, the autocorrelation functions show peaks separated by intervals of about 1/F1, while near the place of the second formant the intervals between successive peaks are about 1/F2. Figure 11C shows the Reciprocal of the Mode of the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function for individual fibers plotted against CF in response to /i/, /ae/, /u/ and /ax/ presented at

75 dB SPL.(1) Fig. 11D shows the same measure for the band-average autocorrelation functions. In general, the RMISP has a CF dependence similar to the largest spectral component, though it is somewhat more "noisy". One difference is that, because only intervals between successive peaks are used in the computation of the RMISP, greater prominence is given to the high-frequency components of the response. In particular, values distributed along the F2 line are found over a wider CF range for /i/, /u/, and /ax/. Another difference occurs when a low-frequency formant is about halfway between two harmonics of the 125-Hz fundamental frequency. The largest spectral component in Fig. 11B is necessarily one of the two harmonics flanking the formant frequency. For instance, for the /u/ stimulus, it is either 0.25 or 0.375 kHz instead of F1=0.3 kHz, and either 0.625 or 0.75 kHz instead of F2=0.7 kHz. In contrast, the RMISP of the autocorrelation function in Fig. 11D more nearly coincides with the formant frequency over some range of CF's.

2. Average localized synchronized measures

The second approach to formant-frequency estimation is based on the observation that the response amplitude along the f=CF line in the perspective displays of Fig. 8 and 9 varies considerably with CF, and is almost always large at the

(1) Alternatively, a processing scheme could be based on the interval between time zero and the first peak of the autocorrelation function, which is usually close to the mode of the distribution of intervals between successive peaks.

places of the formant frequencies. The idea of a response measure based on response components along the f=CF line was introduced by Young and Sachs (1979) when they proposed the Average Localized Synchronized Rate (ALSR). Computation of the ALSR corresponds to analyzing fiber response patterns with a bandpass filter whose center frequency is close to CF. Plots of ALSR against f_c show clear peaks near the formant frequencies over a broad range of stimulus levels. In the original formulation of Young and Sachs (1979), the ALSR was sampled at the harmonics of the fundamental frequency, and the filter bandwidths were sufficiently narrow to resolve individual harmonics. In the present study, the ALSR was computed at regularly spaced samples along the log f_c dimension, and the filter characteristics were chosen in a manner more consistent with the psychophysical concept of critical bands. Figure 13 shows the transfer functions of four filters that were selected to evaluate the effects of filter characteristics on this response measure. Further departures from the formulation of Young and Sachs (1979) include computing the measure from the power spectra instead of the magnitude spectra of period histograms, and normalization by the mean square discharge rate. Because the power spectrum of the period histogram is approximately equal to the magnitude spectrum of the interval histogram (Johnson, 1978), the use of the power spectrum should make the measure more similar to the "average localized interval rate", a measure based on interval histograms which was later proposed

by Sachs and Young (1980). The normalization by the mean square discharge rate, which has little effect for narrow bandpass filtering schemes, improves the behavior of the measure in certain conditions. Due to this normalization, the response measures obtained by local filtering of fiber responses will be renamed Average Localized Synchronized Measures (ALSM).

The first filtering scheme that was used to compute ALSM's is based on a 1/6-octave Gaussian bandpass filter whose transfer function is illustrated in Fig. 13A. Figure 14A shows plots of ALSM against the center frequency of this narrow bandpass filter for the /i/, /ae/, /u/ and /ax/ stimuli presented at 75 dB SPL. The profile shows clear peaks at the formant frequencies above 1 kHz (F2 of /i/, /ae/ and /ax/). For lower-frequency formants, there is a peak at the formant frequency if it coincides with one of the harmonics (F1 of /i/ and /ax/), while there are two peaks flanking the formant frequency if it falls between two harmonics (F1 of /ae/ and /u/, F2 of /u/). For /ae/ and /ax/, there are also large peaks at the low-frequency harmonics (second and third harmonics for /ae/, second harmonic for /ax/). Such low-frequency peaks are found for all the relatively open vowels (/ae/, /a/, /eh/, /oh/ and /ax/), and are more prominent at 60 dB SPL. For /i/, and the other spread vowels /e/ and /ɨ/, there are large peaks at 2F1 and in the broad CF-dominated region between F1 and F2. These non-formant

peaks can lead to incorrect results if the formant frequencies are estimated by picking the two largest peaks in the profile of ALSM against f_c. The two peaks flanking low-frequency formants, however, might be merged by computing the ALSM with a broader filtering scheme.

Figure 13B shows the transfer function of a Gaussian filter with a 2/3-octave bandwidth that was used as an alternative to the 1/6-octave filter. Figure 14B shows profiles of ALSM against f_c for the broad-band filtering scheme. Unlike for the narrow-band filtering, there are peaks centered at the first formant of /ae/ and the second formant of /u/. However the peak at the second formant of /i/ is less prominent, and the amplitude of the non-formant peak at 2 kHz is increased. This loss of second-formant information is typical for spread vowels processed by broad-band schemes. A combination of a broad-band scheme in the low-frequency region and a narrow-band scheme in the high-frequencies would be preferable for formant-frequency estimation.

Instead of being based on spectra of period histograms, ALSM's can also be derived from autocorrelation functions. The curved dashed line in Fig. 12 shows that the amplitude of the autocorrelation function for times close to 1/CF varies greatly with CF, and in particular is large near the places of the formant frequencies. Sampling the autocorrelation at $t=1/f_c$ is mathematically equivalent to computing the ALSM with the cosinusoidal comb filter whose

transfer function is shown Fig. 13D. Compared with the bandpass schemes of Fig. 13A and 13B, comb filters extract frequency components near the harmonics of f_c and components near DC in addition to the components near f_c. The inset in Fig. 13D shows a block-diagram of a simple implementation of the cosinusoidal comb filter involving only one delay, multiplication and lowpass filter. Figure 14D shows plots of ALSM against f_c for the cosine comb filtering scheme of Fig. 13D. For all stimuli, there are clear peaks at the formant frequencies except for the first formant of /u/, where peaks occur at the harmonics flanking the formant frequency. The non-formant peaks between F1 and F2 for /i/ are less prominent than for the bandpass filtering schemes. However, the peaks at the low-frequency harmonics for the open vowels /ae/ and /ax/ remain large, and for all vowels the ALSM is large at high frequencies as $1/f_c$ approaches time zero where the autocorrelation functions have their maximum.

To estimate the formant frequencies, it is desirable to smooth the spectrum in order to eliminate the fine spectral structure due to voice excitation. In contrast, to estimate the fundamental frequency it is necessary to use a narrow-band spectral representation in which individual harmonics of the fundamental are well resolved. This is the case for the narrow-band filtering scheme of Fig. 14A, as there are clear peaks at many of the harmonics below 1 kHz. Specifically, for /i/ there are peaks at harmonics 2, 4, 5 and 7, for /ae/ at

harmonics 2, 3, 4, 6 and 7, for /u/ at harmonics 2, 3, 4, 5 and 6, and for /ax/ at harmonics 2, 4 and 8. With the possible exception of /ax/, these harmonic sequences would in principle suffice to estimate the fundamental frequency. Narrow-band comb filtering can be achieved by computing a weighted sum of samples of the autocorrelation function at times of $1/f_c$, $2/f_c$, $3/f_c$, etc. Roughly speaking, the more multiples of $1/f_c$ that are taken, the narrower the filtering. Figure 13C shows the transfer function of a possible narrow comb filter, with a bandwidth of 1/6 octave. The inset of Fig. 13C shows a simple realization of the narrow-comb filter, in which the series of delays at multiples of $1/f_c$ is replaced by a single delay placed in a feedback loop.(1) Profiles of ALSM against f_c for this narrow-comb filtering scheme are shown in Fig. 14C. There are peaks at essentially the same low-frequency harmonics as for the narrow-band scheme of Fig. 14A, so that this scheme would be useful in fundamental-frequency estimation. However, small peaks at non-harmonic frequencies are occasionally introduced (e.g. at 0.19 kHz for /u/ and at 0.33 kHz for /ax/). This scheme could also be useful for formant-frequency estimation since there are peaks at the positions of high-frequency formants and harmonics close to low-frequency formants are prominently represented.

(1) Another scheme, in which the output of the feedback loop $y_i(t,f_c)$ is rectified and lowpass filtered, gives very similar results.

In summary, several schemes have been proposed for the estimation of speech parameters from auditory-nerve responses. The largest spectral component and the RMISP of the autocorrelation function can detect the spectral peaks associated with the formant frequencies of all the vowel stimuli, with the possible exception of the second formant of /i/. The use of these measures for estimating fundamental frequency has not been investigated because the value of the fundamental enters directly into the computation of harmonic spectra and autocorrelation functions of period histograms. It seems likely, however, that in the CF regions where the fundamental response component is large (Fig. 10), the fundamental frequency could be estimated from the largest component in the low-frequency part of the short-time spectra of PST histograms, or from the reciprocal of the time of the largest peak of the short-time autocorrelation function of PST histograms.

The usefulness of the average localized synchronized measures depends on the type of filtering. Narrow-band or narrow-comb filtering schemes are necessary for fundamental-frequency estimation, and are also useful for formant frequencies. In contrast, broad-band and broad-comb filtering schemes are useful only for formant frequency estimation, and the broad-band scheme is inappropriate for detecting the second formant of spread vowels. The broad comb scheme might seem more accurate than the narrow-band and

narrow-comb schemes for estimating low-frequency formants because the peaks at the harmonics flanking the formant frequency that are found for narrow-band schemes are merged into one peak at the formant frequency. However, on grounds of parsimony, it seems preferable to use the same narrow-band or narrow-comb processing scheme for estimating formant frequencies and fundamental frequency, and following this scheme by a later stage of spectral smoothing to improve the estimate of low-frequency formants.

IV DISCUSSION

A. Limitations of the study

The goal of this study was to describe how acoustic properties that are used for phonetic distinctions among vowels are coded at the level of the auditory-nerve. Its validity is limited by the simplifications in the stimuli, the use of cat as experimental animal, and the experimental procedures.

The stimuli are idealized in that they have only two formants, and their fundamental and formant frequencies are steady-state. The effect of higher-order formants should be minimal for low-CF fibers because auditory-nerve fibers do not typically respond to frequency components more than one octave above their CF. However, differences in the responses to two-formant and full vowels would be expected in the CF region

above the second formant. For instance, Young and Sachs (1979), using four-formant stimuli, found fibers whose response patterns had significant components at the third formant frequency. Some of these differences must be perceptually relevant since listeners can distinguish two-formant vowels from full vowels. However, vowel identity is not, in general, affected by these differences, provided F2 of the two-formant vowel is set appropriately (Carlson et al., 1975). Responses of auditory-nerve fibers to two-formant and full vowels need to be compared systematically in order to find features of the response patterns that would reflect this perceptual invariance.

Possible differences in response between steady-state and non steady-state vowels are of considerable importance because portions of vowel waveforms during which formant frequencies are changing can contain essential information for vowel identification (Lindblom and Studdert-Kennedy, 1967; Strange et al., 1976). Formant frequencies of speech usually change slowly enough that they would remain within the bandwidths of the filtering elements in the cochlea during time intervals of about 10 ms. Thus, over each 10-ms interval, responses of auditory nerve to speech stimuli with changing spectral characteristics might be qualitatively similar to responses to a steady-state vowel. Preliminary results by Voigt et al. (1981) and Sinex and Geisler (1981) suggest that response measures that provide a

representation of the spectrum of steady-state vowels may also be useful for representing the short-term spectrum of stimuli with changing formant frequencies.

The use of cat as experimental animal raises questions about the applicability of our data to the coding of speech in the human auditory nerve. The general plan of the cochlea and auditory nerve shows great similarities between the two species (Schuknecht, 1974). Basic response properties of auditory-nerve fibers appear to be similar in many species of mammals, including primates (Evans, 1975). The behavior of the compound action potential (CAP) of the auditory nerve evoked by click and tone-burst stimuli follows the same trends in humans and other mammals (Kiang and Peake, 1960; Stephens et al., 1975; Montandon et al., 1975). Contribution to the CAP from different regions of the cochlea, estimated by masking with bands of noise with sharp cutoffs, are also similar in humans and other species of mammals (Teas et al., 1962; Eggermont, 1976). However, there is a clear difference between species in the range of frequencies to which the ear is responsive (Miller et al., 1963). In cats, the behavioral threshold to tones roughly follows the thresholds at CF of the most sensitive auditory-nerve fibers (Kiang et al., 1965). If a similar relationship holds in humans, one would expect human fiber thresholds to be considerably higher than cat thresholds for frequencies above 8-10 kHz. However, at the stimulus levels that were used in this study, responses of high-CF

fibers to the low-frequency vowel stimuli were minimal. Along similar lines of reasoning, one has to be careful about responses of very low CF fibers, for which there may be considerable differences in tuning characteristics between cats and humans.

It is possible that the barbiturate anesthesia causes significant differences between the auditory-nerve activity in our preparation and what would be observed in an awake animal. Pilot studies on unanesthetized, decerebrate animals do not show major changes in auditory-nerve activity (Bauer,unpublished). The CAP waveform does not vary much with the state of the animal (Peake et al., 1962). On the other hand, barbiturate anesthesia is known to suppress the activity of middle-ear muscles and is thought to affect the efferent olivo-cochlear bundle (OCB). However, the middle-ear muscles would not normally contract at the stimulus levels (<80 dB SPL) used in this study (Møller, 1965). The efferent system may be active during normal communication, but it seems that the auditory behavior of animals in which the OCB has been cut is not grossly affected, so that many of the important cues presumably do not depend on the state of the OCB (Dewson, 1968; Trahiotis and Elliot, 1970).

Another limitation of the study stems from fiber selection. If, as Kiang (in press) suggests, micropipettes such as those used in this study do not record from unmyelinated fibers, nothing is known about stimulus coding in

unmyelinated auditory-nerve fibers, which constitute about 5-10 % of the entire population (Arnesen and Osen, 1978). A further limitation is that recordings were restricted to the most sensitive population of presumably myelinated fibers, those with spontaneous discharge rates greater than 18 spikes/s (Liberman, 1978). The less sensitive populations of fibers, which constitute about 35 % of the myelinated fibers, have lower spontaneous rates, narrower tuning curves, slightly larger synchronization indices, a wider dynamic range, and show more prominent two-tone suppression (Liberman, 1978; Schalk and Sachs, 1980; Johnson, 1980). One would also have to study the response of these populations in order to get a complete description of speech coding in the auditory-nerve, particularly if these fibers have specialized connections in the cochlear nucleus.

Thus, in spite of considerable limitations that would require further investigation, it is felt that the present study provides a valid description of the main features of the responses of auditory-nerve fibers to vowel stimuli.

B. Comparison with previous work

Our data on responses to vowels are generally consistent with those of Sachs and Young (1979) and Young and Sachs (1979). However, there are differences in the relative amplitudes of the response components associated with F1 and

F2. In our data, the first formant was never found to dominate the response patterns of fibers whose CF is near F2, whereas that situation arises at high stimulus levels in the data of Young and Sachs (1979). The second formant can be the largest component in the high-CF region (e.g. for /u/ and /oh/) in our data, whereas F1 is always dominant in that region in the data of Young and Sachs. Peaks of ALSM's at 2F1 and 3F1 are usually lower than the F2 peak in our data, whereas the opposite seems to be the rule at high levels in the data of Young and Sachs. These differences may be due to differences in stimuli. The amplitude of the second formant relative to the first formant is 3 to 10 dB higher in our stimuli than in those of Young and Sachs (1979). The lack of possible suppression of the second formant by higher formants in our data may also be a factor. Another difference in stimuli is that the formant frequencies always coincide with harmonics of the fundamental frequency in the stimuli of Sachs and Young (1979). In our stimuli, when F1 coincides with a harmonic, the peak at 2F1 in plots of ALSM tends to be larger than for a similar vowel in which F1 is about halfway between two harmonics. Similarly, the data of Voigt et al. (1981) with whispered vowels show relatively small components at harmonics of F1.

Another finding which does not appear in the data of Young and Sachs (1979) is the weak synchronization of discharges with the second formant of /i/. This finding is

probably due to the high frequency of this formant (F2=3.2 kHz) for which weak synchrony is expected (Johnson, 1980), and, possibly, to the large bandwidth (BW2=250 Hz), which may result in less suppression of weak spectral components by the intense components near F2. This weak synchrony at F2 was not observed by Young and Sachs (1979) because the F2 of a two-formant /i/ is considerably higher than the F2 of a full formant vowel. However, because a broad spectral maximum in the 3-kHz region is essential for the identification of a sound as /i/, this weak synchrony is a problem for speech-processing schemes based on fine time patterns of discharge regardless of the number of formants in the stimulus. It is possible, however, that the resonance of the ear canal in the human (Mehrgardt and Mellert, 1977) would enhance the representation of components near 3 kHz in fiber response patterns.

C. Comparison with psychoacoustic and phonological data

The presence of peaks in the low-frequency (<3 kHz) region of the spectrum of a sound stimulus is an essential cue for listeners' identification of this sound as speech (Remez, 1979). Our data and those of Sachs and Young (1979) show that the response patterns of the majority of auditory-nerve fibers are dominated by components near the formant frequencies. In contrast, responses to broad-band noise stimuli are dominated by components near the fiber CF (De Boer and Kuyper, 1968; Ruggero, 1973; Møller, 1977; Evans, 1977; De Boer and De

Jongh, 1978). This contrast between a flat distribution of response components for broadband noise, and a distribution centered around the formant frequencies for vowels may be an important cue for the central processor in detecting speech sounds among environmental sounds.

The identity of vowels and sonorant sounds is determined primarily by the frequencies of the local spectral maxima associated with the formant frequencies. In general, it is not sensitive to the relative amplitudes of the formant peaks and to the general tilt of the spectrum (Carlson and Granstrom, 1980; Chistovich et al., 1979). In spite of the differences between our stimuli and those of Sachs and Young (1979), the responses of the majority of auditory nerve fibers are dominated by the formant frequencies in both sets of data. However, at high stimulus levels, the largest response component near the place of the second formant is often F1 in the data of Young and Sachs (1979) instead of F2 as in our data. This degradation in second-formant information might be compensated for by emphasizing high-frequency components in the response spectra prior to detection of the largest component. The RMISP of the autocorrelation function, which also emphasizes high-frequency component, might also be more stable to variations in formant amplitudes and spectral tilt. Clearly, systematic investigations of the effect of these parameters on auditory-nerve responses are needed.

Deletion of several non-formant harmonics from an /ae/-like stimulus has little perceptual effect when the harmonics are between F1 and F2, or in the high-frequencies, but is easily heard when the harmonics are below F1 (Carlson et al., 1979). Our study shows that fibers whose CF is close to a low-frequency harmonic of /ae/ show large response components at that harmonic. In contrast, at the stimulus levels that were used, there is usually little synchronization to harmonics above F2, unless they coincide with distortion or intermodulation products of the formant frequencies. This is also the case for harmonics between F1 and F2 when the formant frequencies are less than 1.5 octave apart. For spread vowels, like /i/, /e/ or /ɪ/, large response components at harmonics between F1 and F2 are found for fibers whose CF is between F1 and F2. No psychophysical data on the effects of deletion of harmonics are available for such vowels.

Though psychophysical data show that vowel identification depends primarily on the frequencies of the local spectral maxima associated with formant frequencies, the properties of the stimulus that the auditory system uses in estimating these frequencies are unclear. Some results (Chistovich, 1971) can be interpreted as if the auditory system detects the most intense harmonic near the formant frequency. Other results (Carlson et al., 1975) suggest that the auditory system computes a weighted average of the two or three most intense harmonics. This average would more nearly

coincide with the "true" formant frequency (i.e. the resonant frequency of the vocal tract) than with the most intense harmonic. In our data, when a harmonic is sufficiently close to a formant frequency so that its amplitude exceeds that of its neighbors by about 6 dB, the components synchronized to the neighboring harmonics are considerably suppressed, and all the formant estimation schemes yield the frequency of the largest harmonic. When the two largest harmonics in the stimulus have amplitudes within 6 dB, considerable response components are found at both harmonics, and some of the processing schemes (specifically the RMISP of the autocorrelation and the broad-band and broad-comb ALSM's) can yield an estimate close to the formant frequency. These results may differ for female and children voices because the separation between harmonics is greater. Responses of auditory-nerve fibers to speech-like stimuli in which the positions of the harmonics relative to the formant frequencies are systematically manipulated need to be studied.

When the formant frequencies of a two-formant vowel are less than about one octave apart, listeners can reliably set the frequency of a single-formant stimulus that provides a best match to the two-formant stimulus (Chistovich et al., 1979). The matched frequency is intermediate between the formant frequencies of the two-formant stimulus, and depends on the frequencies and amplitudes of the two formants. Correlates of this "center of gravity" effect are not directly

apparent in our data since, even for the vowels /a/ and /oh/ whose formant frequencies are within one octave, response patterns show intense components at harmonics near the formant frequencies rather than at some intermediate frequency.

Phonetic descriptions of vowels are traditionally organized in terms of degree of vocal-tract opening (tongue height) and place of major constriction. These dimensions are also useful in organizing phonological data on classes of speech sounds that behave similarly with respect to historic changes, dependence on phonetic context, or dialect variations. Our results show that these phonological dimensions have clear correlates at the level of the auditory nerve as variations in the position and extent of different CF regions within which certain response properties are homogeneous. Specifically, the open-close dimension is correlated with the position of a CF region around the F1 place over which discharges are primarily synchronized to F1 and its harmonics, and with the extent of a low-CF region in which harmonics near CF are the largest components. The front-back dimension is correlated with the position and extent of a region centered at the F2 place over which F2 is the largest component, and with the extent of a high-CF region in which response spectra have broad peaks at F1, F2 and the fundamental frequency. The spread dimension, proposed by Fant (1973), is correlated with the extent of an intermediate region between F1 and F2 over which large response components

near CF and F0 are found, and with the extent of the regions centered at F1 and F2. These correlations between phonological dimensions and properties of auditory-nerve activity are of a continuous nature, and there is no basis in our data for dividing, for instance, the tongue-height dimension into well defined "open" and "close" vowel categories.

Because steady-state vowels are complex periodic tones, our results are relevant to theories of pitch perception. Two types of theories that have been proposed for the low pitch of complex tones are "place" and "periodicity" theories. Place theories require that there is a response measure derived from auditory-nerve data that, when plotted against characteristic frequency, shows peaks at the places of the low-frequency harmonics of the fundamental. Periodicity theories usually assume that the fundamental frequency is estimated from periodicities in the envelope of fiber response patterns, or from peaks at the fundamental period in autocorrelation functions of response patterns. A variety of psychophysical data seem to be more easily explained by place theories (Plomp, 1976; Goldstein, 1978). The data of Sachs and Young (1979) suggest that the profile of average discharge rate against CF does not show peaks at low-frequency harmonics of the fundamental, so that this measure would not be appropriate for place theories. However, for all vowels except possibly /ax/, the narrow-band and narrow-comb ALSM's

show clear peaks at the frequencies of enough harmonics so that the fundamental frequency could in principle be estimated. A critical factor in the success of these schemes is the narrow bandwidth of the filter, which was set at 1/6 octave because it is a lower bound on critical bandwidth estimates (Sharf, 1970). According to this interpretation, critical bands would be primarily a central phenomenon since the narrow filters operate on auditory-nerve fiber discharge patterns, which are the outputs of peripheral processing. However, the strength of the representation of low-frequency harmonics in the ALSM depends on the spectrum envelope and on stimulus level, so that that one would have to be cautious in generalizing place pitch theories from one type of stimulus to another.

In cases when the place representation is weak or degraded, the fundamental frequency could be estimated from periodicity cues. For responses of auditory-nerve fibers to vowel stimuli, such cues are present in the CF regions where the fundamental response component is large. The fundamental response component is probably generated by rectification of the periodic fluctuations in the envelope of the stimulus waveform rather than from the relatively weak fundamental component in the stimulus (Hashimoto et al., 1975; Delgutte, 1980). Consistent with this interpretation, the amplitude of the fundamental response component tends to decrease with stimulus level, as would be expected from the saturation of

rate-intensity functions (Smith and Brachman, 1980b). When periodic fluctuations at the fundamental frequency in the envelope of a vowel stimulus are removed by randomizing the phases of the harmonics, listeners hear the stimuli as being "harsh and aperiodic", but the low pitch sensation does not disappear (Carlson et al., 1979). Thus, it seems that some aspect of vowel quality is related to the intense fundamental component of auditory-nerve fiber responses, but it is not clear that this aspect is the same as low pitch.

D. Speech-processing schemes and engineering approaches

Many of the speech processing schemes that have been proposed in this study resemble algorithms that have been used to represent the speech signal in communication engineering. The comparison between the narrow-band ALSM's, which are preferable for fundamental-frequency estimation, and broad-band ALSM's, which provide a better representation of formant frequencies, is reminiscent of narrow-band and broad-band spectrograms. The broad-comb ALSM, obtained by evaluating the autocorrelation of response patterns at 1/CF, is similar to the autocorrelation vocoder (Schroeder, 1962), with the important difference that computation of the ALSM is preceded by peripheral auditory processing. The processing scheme based on the RMISP involves reducing the signal to a series of pulses coincident with local maxima and looking at the frequency distribution of the intervals between these pulses. This approach resembles those of Baker (1975) and

Carlson et al. (1975), except that the pulses are coincident with peaks of the autocorrelation function instead of zero-crossings of the waveform.

One technique which has not been used previously for formant-frequency estimation is comb filtering. Comb filtering is well adapted to the special properties of auditory-nerve fiber responses. Due to the nonlinear processing that occurs in the cochlea, the response to a single tone stimulus consists of a DC component, a component at the stimulus frequency, and many harmonic distortion components. A comb filter whose "center frequency" is set at the stimulus frequency extracts all these response components, whereas a bandpass filter extracts only the fundamental component. For this reason, peaks of the ALSM at 2F1 tend to be less prominent for comb-filtering schemes than for bandpass schemes (Fig. 14). One disadvantage of comb filtering is that the filter output is large when the center frequency is set at a subharmonic of the stimulus frequency. However, in the case of ALSM computation, this is rarely a problem since peripheral filtering typically prevents auditory-nerve fibers from responding to components that are more than one octave above CF. This example illustrates how, because response signals of auditory-nerve fibers have different properties from speech signals, new ways of processing speech may develop from physiological studies.

E. Realism of the speech-processing schemes

The speech-processing schemes that have been proposed can be considered as models of the central processor. It is important to evaluate the compatibility of such processing schemes with anatomical, physiological and behavioral knowledge on the auditory system. A relevant anatomical finding is that all auditory centers in the brain up to the level of the auditory cortex include at least one cochleotopically-organized population of cells (Clopton et al., 1974). Consistent with this fact, all response measures have been plotted against the logarithm of the characteristic-frequency, which seems to be linearly related to the locus of innervation of auditory-nerve fibers along the cochlea (Kiang, in press).

Another common property of the processing schemes is that they do not depend on the phases of the response components since they are based on the power spectrum or the autocorrelation function of period histograms. Thus, if the magnitude of the response spectrum is not strongly affected by changes in the phase spectrum of the stimulus, the processing schemes fit with the psychophysical finding that human subjects are not usually very sensitive to the phase spectrum. The limited available evidence for two-tone stimuli suggests that the amplitudes of response components are not very sensitive to changes in stimulus phase (Brugge et al., 1969; Littlefield, 1973; Arthur, 1976). Investigations of the

effects of phase manipulations on speech perception and on responses of auditory-nerve fibers are needed.

In the computation of the harmonic spectra and autocorrelation functions of period histograms, we assumed knowledge of the fundamental frequency. Because estimation of the fundamental frequency of speech is a difficult problem, this assumption is not desirable. In principle, the proposed processing schemes could also be based on the short-time spectrum or short-time autocorrelation of fiber discharge patterns (Flanagan, 1972). In Chapter IV, we will show that the speech processing schemes are also effective when they are based on short-time analysis of PST histograms with a 10-20 ms window. Such windows provide bandwidths that are small enough to resolve the harmonics of the fundamental for male voices.

Many of the speech-processing schemes are based on the autocorrelation functions of period histograms. Licklider (1951) suggested that samples of the short-time autocorrelation of fiber response patterns could be computed from known synaptic mechanisms such as delay and summation. Following the ideas of Jeffress (1948) on crosscorrelation, Licklider (1951) pointed out that evaluating the autocorrelation function for a population of delays would be a means to transform a temporal code into a spatial code, and proposed that the auditory system creates a two-dimensional representation of acoustic stimuli along the characteristic frequency and characteristic delay parameters. Figure 12

gives an example of such a representation for the vowel /ax/. Similar representations (based on crosscorrelation between the outputs of the two ears) have been useful for interpreting binaural psychophysical phenomena (Colburn and Durlach, 1978). There has been no physiological evidence for these ideas, though neurons similar to the collicular units found by Kallert et al. (1970), which showed harmonically-related peaks in their frequency response, would be expected on the basis of autocorrelation analysis. In any case, the autocorrelation function is a possible alternative to the interval histogram proposed by Goldstein and Srulovicz (1977) and Sachs and Young (1980).

The processing scheme based on the largest spectral component requires a full spectral analysis (a filter bank) in every narrow band of CF. Similarly, the RMISP of the autocorrelation function requires the computation of the autocorrelation function for a population of time delays in each CF band in order to obtain an estimate of the distribution of intervals between peaks. In contrast, the ALSM's are more parsimonious processing schemes because only one filtering element is required in each band of CF. However, these schemes assume that knowledge of fiber CF's is "wired in", since the center frequency of the central filters has to coincide with fiber characteristic frequency. The ALSM's also have the advantage of being more easily related to the psychophysical concept of critical band than the schemes

based on the largest spectral component or the RMISP of the autocorrelation.

Among the different ALSM's, the comb-filtering schemes have the advantage over the bandpass schemes that specific mechanisms have been proposed for the implementation of the filters. These mechanisms, which are illustrated in the insets of Fig. 13, are based entirely on operations known to exist at synapses in the central nervous system. To permit estimation of the first three formant frequencies in speech, the delays would have to vary from about 0.3 to 5 ms. The largest delays would thus require long, thin fibers or several synaptic relays, and the system's ability to preserve precise synchrony of discharges can be questioned. For the broad comb filtering scheme, time jitter equal to 12 % of the delay had little effect on the plots of ALSM against center frequency.

Thus, on the grounds of effectiveness, parsimony, and compatibility with psychophysical and physiological data, the ALSM's based on narrow-comb filtering would seem to be the best speech-processing scheme. However, in the absence of systematic data on processing of complex stimuli by the auditory nervous system, arguments to favor one type of processing scheme over another are not compelling, so that evaluation of the different schemes should be continued. One evaluation procedure consists in studying the behavior of the speech processing schemes for a broader set of stimuli and under psychophysical conditions that leave speech

intelligibility relatively invariant. In forthcoming papers, we will test the proposed processing schemes for high-frequency sounds such as fricative consonants, and in the presence of moderate background noise.

APPENDIX A: STIMULUS GENERATION

The two-formant vowel synthesizer consists of a voicing source in cascade with two bandpass filters connected in parallel. The voicing source generates impulses at a rate equal to the fundamental frequency. The impulse train is sent to a lowpass filter $H_g(s)$ representing the effects of glottal source spectrum and radiation characteristic. The frequency response of this lowpass filter is given by:

$$H_g(s)=\frac{1}{s+2\pi BW_g} \qquad (A1)$$

where s is the complex radian frequency and BW_g=0.5 kHz. The frequency response of each bandpass filter $H_i(s)$ [i=1,2] is specified from the formant frequency F_i and the formant bandwidth BW_i as follows:

$$H_i(s)=\frac{2\pi F_i}{(s-s_i)(s-s_i^*)} \qquad (A2)$$

with $s_i=-\pi BW_i+j2\pi F_i$. The parallel configuration of the filters was chosen over a cascade configuration so that the relative amplitude of the two formants A2/A1 would be under control. To make the low-frequency components more realistic, the second bandpass filter is followed by a highpass filter (Klatt, 1980b). The frequency response of the highpass filter $H_c(s)$ is given by:

$$H_c(s)=\frac{s}{s+2\pi BW_c} . \qquad (A3)$$

with BW_c=1.5 kHz. The outputs of the two branches are added out of phase to avoid the generation of a spectral zero between the two formants frequencies. The formant parameters for the nine vowel stimuli are listed in Table A-I.

The synthesizer was implemented by the impulse-invariance method (Oppenheim and Schafer, 1975) using a sampling rate of 20 kHz. A period of the steady-state output waveform was replicated 23 times to generate a burst of sound with a duration of 184 ms. A 2-ms linear rise-fall time was then applied to the burst to suppress discontinuities at the onset and offset. The computations were made in floating-point arithmetic on a PDP 11/34 computer. The output of the synthesizer was quantized to form a sequence of 12-bit integers in such a way that the maximum absolute value of the signal corresponds to the maximum quantizer level. The 12-bit integer sequences were sent to a digital-to-analog converter and lowpass filtered at 10 kHz by a fifth-order elliptic filter. All stimuli were sent to the earphone so that positive voltages in Fig. 1 represent condensation. The largest distortion components in the acoustic stimuli did not exceed 25 dB SPL for the 75 dB vowels.

Table A-I Formant parameters for the two-formant vowels

STIMULI	F1 (kHz)	BW1 (Hz)	F2 (kHz)	BW2 (Hz)	A2/A1 (dB)
/i/	0.25	50	3.2	250	-5
/e/	0.40	60	2.4	120	-6
/eh /	0.60	60	2.0	120	-5
/ae/	0.80	70	1.8	120	-6
/a/	0.75	130	1.2	70	-6
/oh /	0.55	90	1.0	100	-7
/u/	0.30	60	0.7	110	-9
/ɨ/	0.30	60	1.6	150	-9
/ax /	0.50	80	1.5	100	-9

A2/A1 is the ratio of the magnitudes of the transfer function of the entire synthesizer at the formant frequencies.

FIGURE CAPTIONS

Fig. 1

(a) Waveforms of one period of each of the 9 vowel stimuli. These waveforms are generated by a computer program described in appendix A. The amplitude scale was computed in pressure units for the 60 dB SPL stimuli, with the assumption that the transfer function of the acoustic system has a flat magnitude. (b) Harmonic spectra of the vowel waveforms shown in (a). Dotted lines show the positions of the formant frequencies F1 and F2.

Fig. 2

Second formant frequency plotted against first formant frequency for the 9 vowel stimuli. Values of the formant parameters are listed in Table A-I.

Fig. 3

Mean sound pressure near the tympanic membrane for constant-amplitude electric signal into the earphone for 27 cats. Vertical markers correspond to $\pm$ one standard deviation. These curves were obtained by (1) subtracting the average of the transfer ratio over 0.1-5 kHz from the transfer ratio for each animal in order to eliminate irrelevant variations in overall level, (2) taking the mean and standard deviation at each frequency across the 27 animals, and (3)

adding a typical level (115 dB SPL) to the mean transfer ratio.

Fig. 4

Thresholds at CF plotted against characteristic frequency for 313 units from 27 animals. The solid line shows the "best threshold curve" obtained by Liberman (1978) for cats raised in a low-noise environment.

Fig. 5

Normalized harmonic spectra for individual units and band-average spectra in response to the vowel /i/ presented at 75 dB SPL. Response spectra for 3 units in each of 5 CF bands are shown in the first three columns, and the corresponding band-average spectra are shown in the last column. The center frequencies of the 0.55-octave bands are spaced one octave apart. The number of units involved in the computation of each band average is, from low to high frequencies: 3, 5, 7, 5, and 7. The CF of each unit is listed above the spectrum, and the center frequency of each band is listed above the band average. The positions of the formant frequencies F1 and F2 are shown by dotted lines, and the position of the fiber CF is marked by an arrow below the frequency axis. The DC component (equal to 1 for all spectra) is omitted from the plots for clarity.

Fig. 6

Normalized period histograms for auditory-nerve fibers with 7 different CF's in response to the /i/, /ae/, and /u/ stimuli presented at 75 dB SPL. Each period histogram is normalized by the maximum discharge rate. The approximate CF's of the fibers are shown at the left and are spaced 3/4 octave apart. Numbers identifying the animal and the fiber are listed above each histogram. The horizontal markers at the top correspond to time intervals of 1/F1 and 1/F2, and the markers above each histogram correspond to intervals of 1/CF. There is an arbitrary delay between the onset of the stimulus period and zero time of the period histogram. This delay (in ms) is listed as follows for each row in increasing order of CF: 3.7 (2.7 for /u/), 3.0 (3.5 for /u/, 2.0 for /i/), 2.4, 1.9, 1.7, 1.4, and 1.2.

Fig. 7

Normalized harmonic spectra of the period histograms shown in Fig. 6. Dotted lines show the positions of the formant frequencies F1 and F2, and an arrow marks the position of the fiber CF.

Fig. 8

Pseudo-perspective representation of normalized band-average power spectra for 0.55-octave CF bands in response to the 9 vowel stimuli presented at 75 dB SPL. The normalized power spectrum is the square of the normalized harmonic spectrum shown in Fig. 5 and 7. Each band-average power spectrum is

plotted with frequency along the oblique axis, and amplitude along the vertical axis. Spectrum points with an amplitude lower than 0.05 are omitted for clarity. The center frequencies of the CF bands are sampled every quarter octave. Horizontal dashed lines show the positions of the fundamental frequency F0 and the formant frequencies F1 and F2 along the frequency axis. Oblique dashed lines mark the places of the formant frequencies along the CF dimension. The curved dashed line is the locus of points for which frequency is equal to CF.

Fig. 9

Same as Fig. 8 for the 60-dB vowel stimuli.

Fig. 10

Synchronization index at the fundamental frequency plotted against band center frequency for 0.55-octave CF bands in response to the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL. The CF bands are the same as in Fig. 8 and 9. The places of the formant frequencies F1 and F2 are marked by dashed lines.

Fig. 11

A. Frequency of the maximum of the harmonic response spectrum plotted against characteristic frequency for the /i/, /ae/, /u/, and /ax/ stimuli presented at 75 dB SPL. Open circles show the frequency of the maximum for all harmonics except the fundamental for one auditory-nerve fiber. If the fundamental

component is more intense than this maximum, a filled triangle is plotted at the ordinate of the fundamental. Dashed vertical lines mark the places of the formant frequencies along the CF dimension, and dashed horizontal lines show the positions of the formant frequencies along the frequency axis.
B. Same as A for the band-average spectra. The CF bands are the same as in Fig. 8.
C. Reciprocal of the Mode of the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function of the period histogram plotted against CF for the /i/, /ae/, /u/ and /ax/ stimuli presented at 75 dB SPL. To measure the RMISP, local maxima (peaks) were first estimated for the autocorrelation function of the period histogram for each fiber. A histogram of intervals between successive peaks was then computed, using a 0.05 ms bin width, and weighting each interval by the mean amplitude of the two peaks. The RMISP is defined as the reciprocal of the center of gravity of the histogram points within $\pm$12 % of the mode. Taking the center of gravity improves the accuracy of the estimate for intervals not much larger than the bin width. In the figure, vertical dashed lines show the places of the formant frequencies, and horizontal dashed lines are drawn at the ordinates corresponding to intervals of 1/F1 and 1/F2.
D. Same as C for band-average autocorrelation functions. The CF bands are the same as in Fig. 8.

Fig. 12

Pseudo-perspective representation of normalized band-average autocorrelation functions for 0.55-octave CF bands in response to the vowel /ax/ presented at 75 dB SPL. The band-average autocorrelation functions are computed from the autocorrelation functions of period histograms for individual units using the same method as for band-average spectra. The CF bands are the same as in Fig. 8. Each band-average autocorrelation function is normalized by the mean square discharge rate, and plotted with time along the oblique axis and normalized amplitude along the vertical axis. Oblique dashed lines mark the places of the formant frequencies along the CF dimension. The curved dashed line is the locus of points for which time is equal to 1/CF. The oblique markers on the right of the figure show time intervals of 1/F1 and 1/F2.

Fig. 13

Different filtering schemes used for the computation of the Average Localized Response Measures (ALSM). Each plot shows the transfer function of a filter operating on the period histograms of auditory-nerve fibers whose CF is within $\pm 1/4$ octave of the center frequency of the filter f_c. If the result of the filtering operation for fiber i with characteristic frequency CF_i is denoted $A_i(f_c)$, each ALSM is defined by:

$$ALSM(f_c) = \sum_i W_{f_c}(CF_i)\ A_i(f_c) \ / \sum_i W_{f_c}(CF_i) \qquad (13\text{-}1)$$

where $W_{f_c}(CF)$ is a trapezoidal weighting function centered at $CF=f_c$ with a central width of 1/6 octave and a total width of 1/2 octave. Each ALSM is computed for values of f_c from 0.1 to 5 kHz, in steps of 1/12 octave.

A. Gaussian bandpass filter with a bandwidth of 1/6 octave. The filter transfer function is given by:

$$H_{f_c}(f)=\exp-\pi[(f-f_c)/b_c]^2 \qquad (13\text{-}2)$$

where $b_c=0.116\ f_c$ is the filter bandwidth. The result of the filtering operation for fiber i is computed from:

$$A_i(f_c)=\sum_{0<f_k<5} H_{f_c}(f_k)\ P_i(f_k) \qquad 13\text{-}3$$

where f_k is the frequency of each harmonic in kHz, and $P_i(f)$ is the power spectrum of the period histogram normalized by the mean square discharge rate.

B. Gaussian bandpass filter with a bandwidth of 2/3 octave. The implementation is the same as in A except $b_c=0.47\ f_c$.

C. Comb filter with a bandwidth of 1/6 octave. The inset shows a block diagram illustrating the implementation of the filter. The output $A_i(t,f_c)$ is the short-time crosscorrelation between the period histogram $r_i(t)$ for fiber i and $y_i(t)$, which is obtained by filtering $r_i(t)$ by the delayed-feedback mechanism illustrated in the lower part of the inset. The feedback gain a_{f_c} controls the bandwidth of the comb filter and is set to 0.7 for a 1/6-octave filter. The lowpass filter was chosen as in D, and the final result is normalized by the mean square discharge rate.

D. Comb filter with cosinusoidal transfer function. The inset shows a block diagram illustrating the implementation of the

filter. The output $A_i(t,f_c)$ for fiber i is the short-time autocorrelation of the period histogram $r_i(t)$ evaluated at $1/f_c$. The impulse response of the lowpass filter was chosen to be rectangular with a duration equal to the 8-ms fundamental period of the vowel stimuli, so that the short-time autocorrelation would be equal to the periodic autocorrelation. The final result of the filtering is normalized by the mean square discharge rate.

Fig. 14

Four Average Localized Response Measures plotted against center frequency of the analyzing filter for the /i/, /ae/, /u/ and /ax/ stimuli presented at 75 dB SPL. For each measure, the place of the formant frequencies along the f_c dimension is marked by dashed lines.

A. ALSM obtained using the 1/6-octave Gaussian bandpass filter of Fig. 13A.

B. ALSM obtained using the 2/3-octave Gaussian bandpass filter of Fig. 13B.

C. ALSM obtained using the 1/6-octave comb filter of Fig. 13C.

D. ALSM obtained using the cosine comb filter of Fig. 13D.

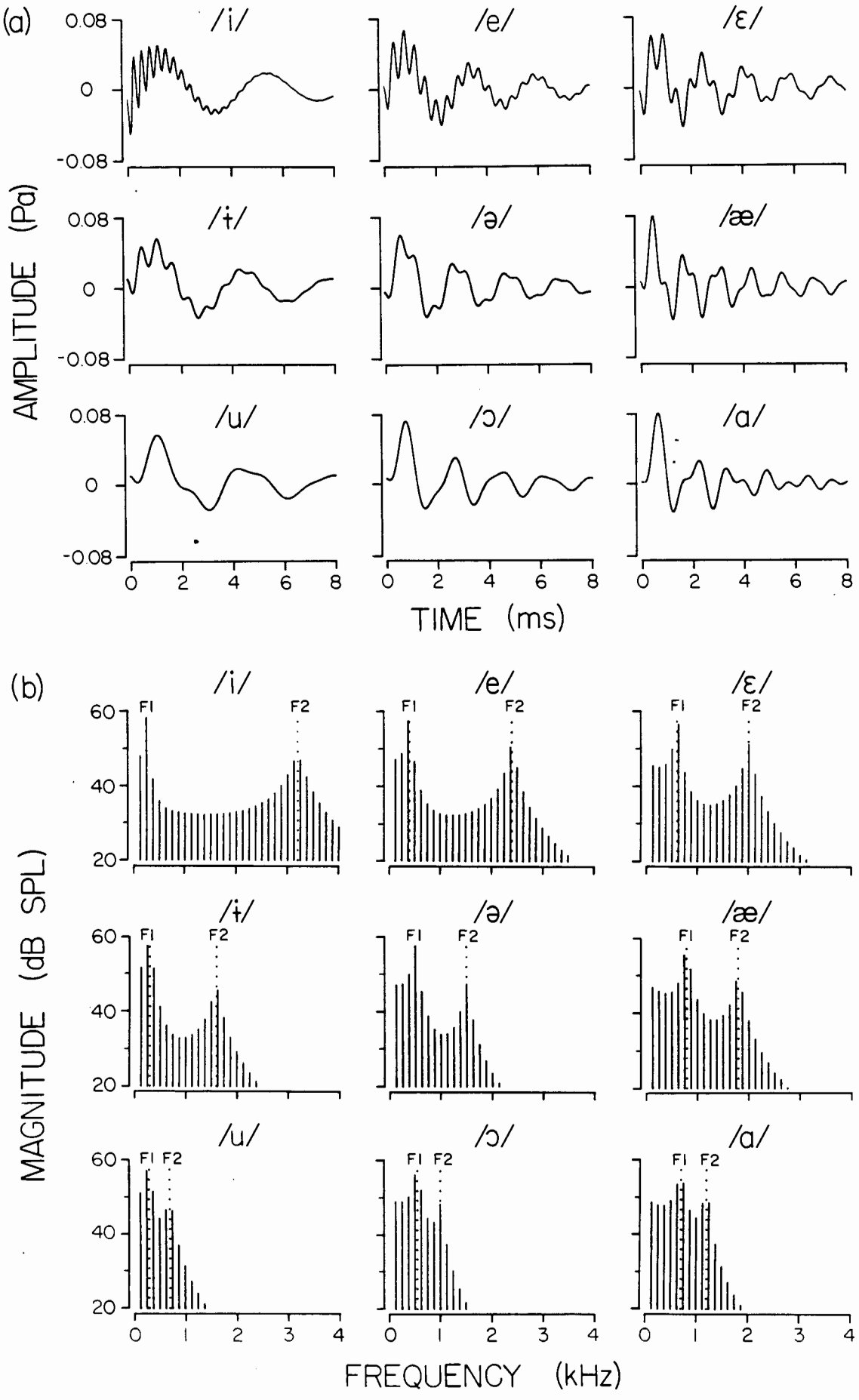
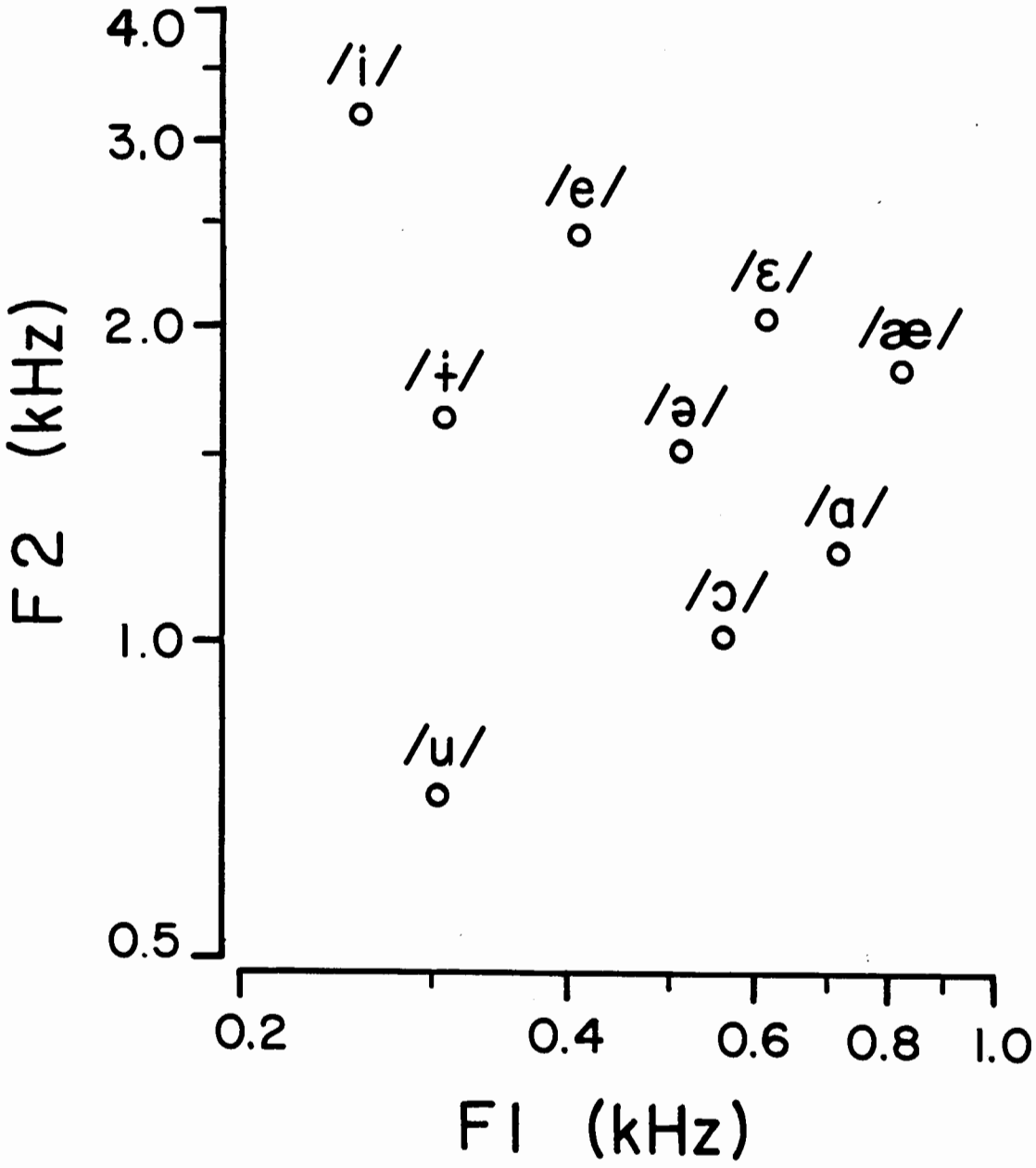

(a)
/i/
/e/
/ɛ/
/ɨ/
/ə/
/æ/
/u/
/ɔ/
/ɑ/
0.08
0
-0.08
AMPLITUDE (Pa)
0 2 4 6 8
TIME (ms)
(b)
F1
F2
60
40
20
MAGNITUDE (dB SPL)
0 1 2 3 4
FREQUENCY (kHz)

Fig. 1

Fig. 2



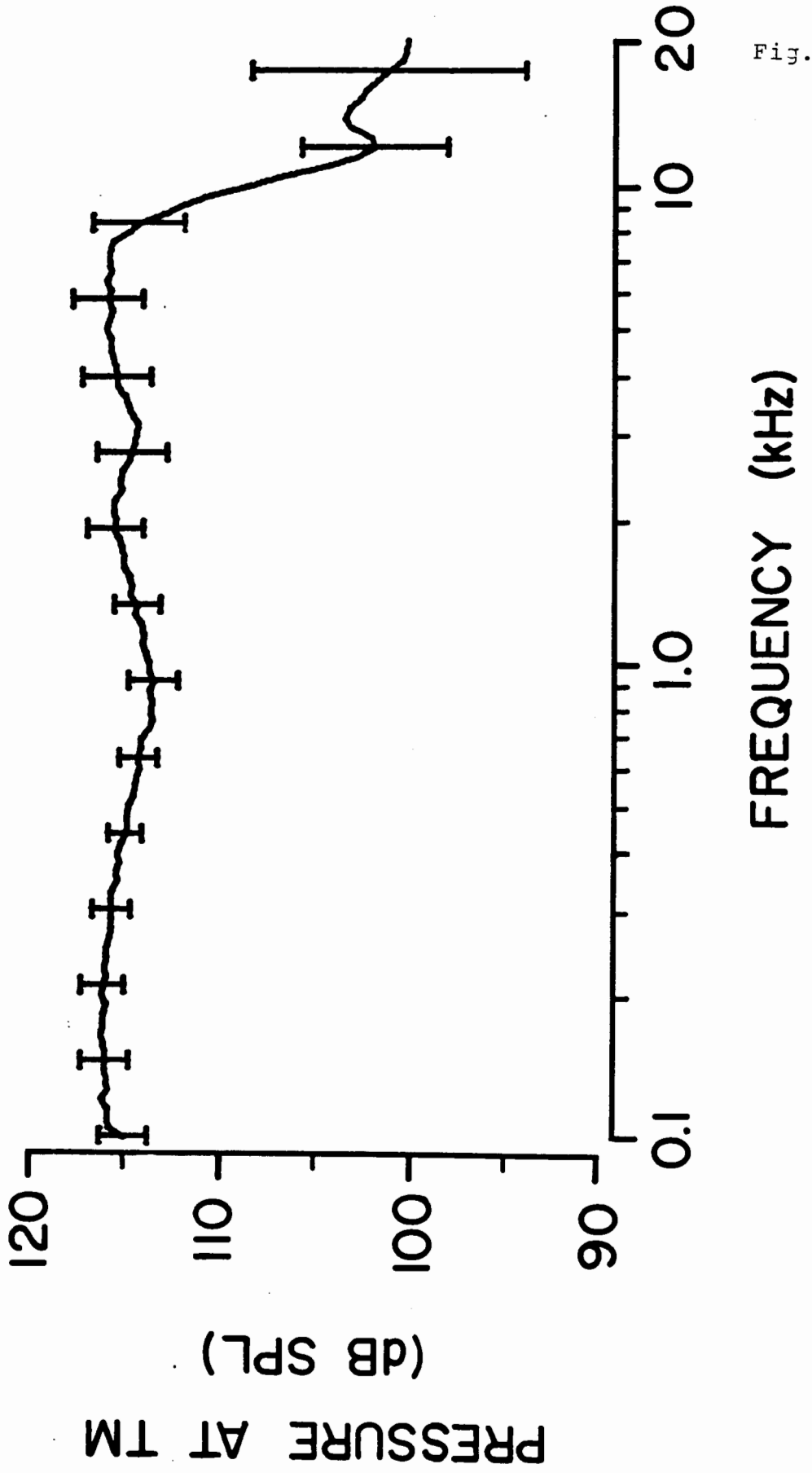


Fig. 3

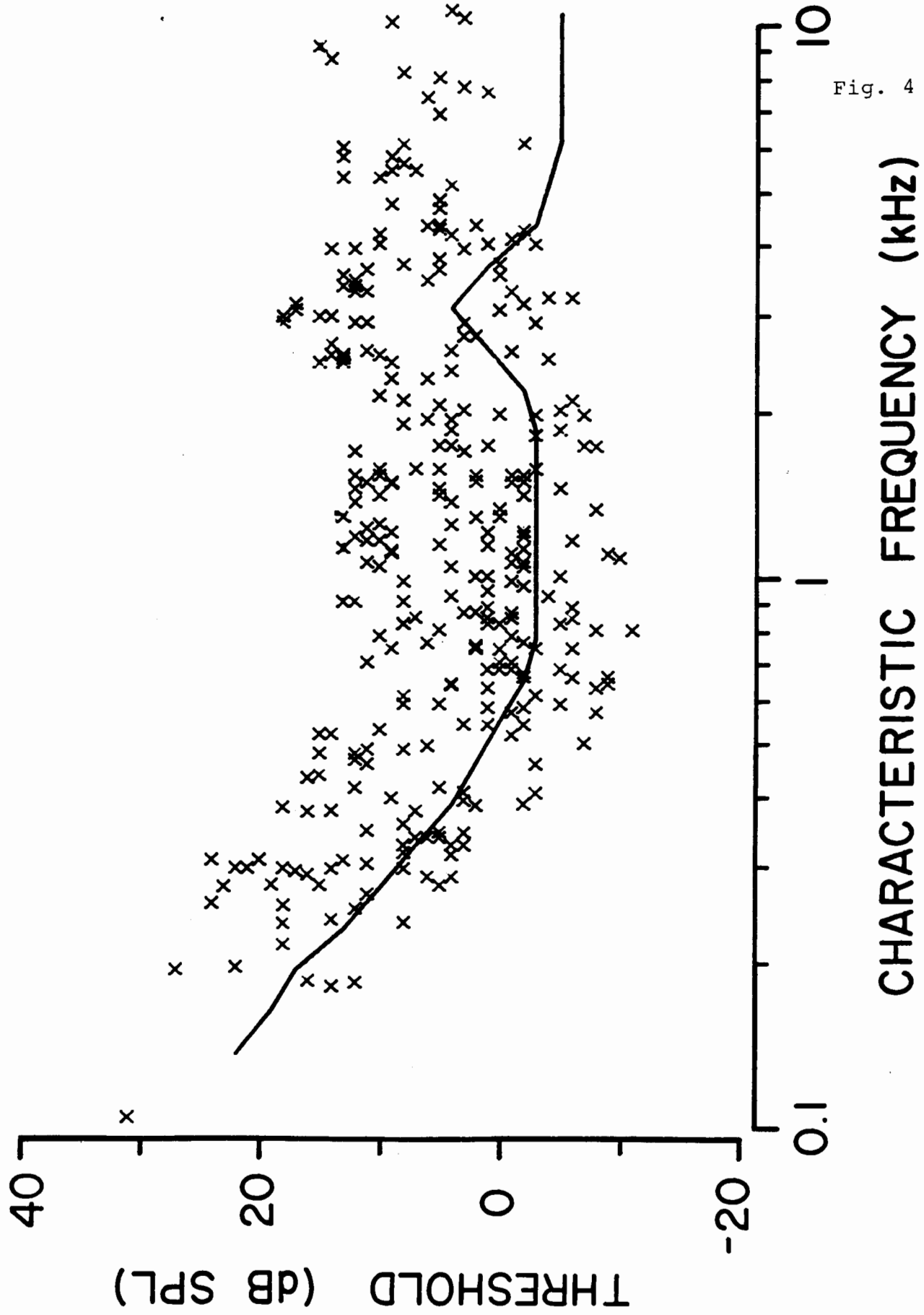
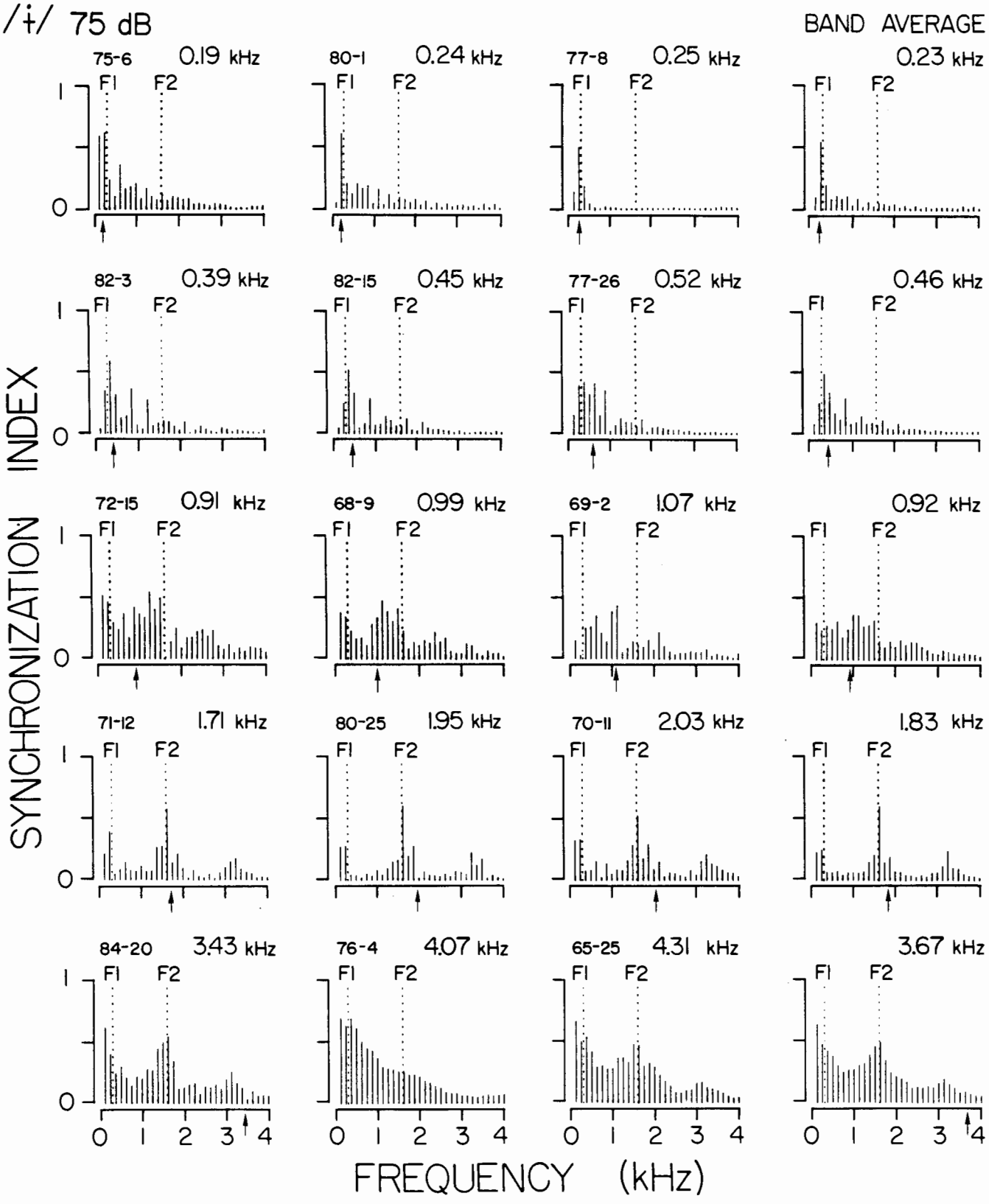

THRESHOLD (dB SPL)
40
20
0
-20
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 4

Fig. 5


/i/ 75 dB
BAND AVERAGE
75-6 0.19 kHz
80-1 0.24 kHz
77-8 0.25 kHz
0.23 kHz
82-3 0.39 kHz
82-15 0.45 kHz
77-26 0.52 kHz
0.46 kHz
72-15 0.91 kHz
68-9 0.99 kHz
69-2 1.07 kHz
0.92 kHz
71-12 1.71 kHz
80-25 1.95 kHz
70-11 2.03 kHz
1.83 kHz
84-20 3.43 kHz
76-4 4.07 kHz
65-25 4.31 kHz
3.67 kHz
F1
F2
SYNCHRONIZATION INDEX
1
0
0 1 2 3 4
FREQUENCY (kHz)

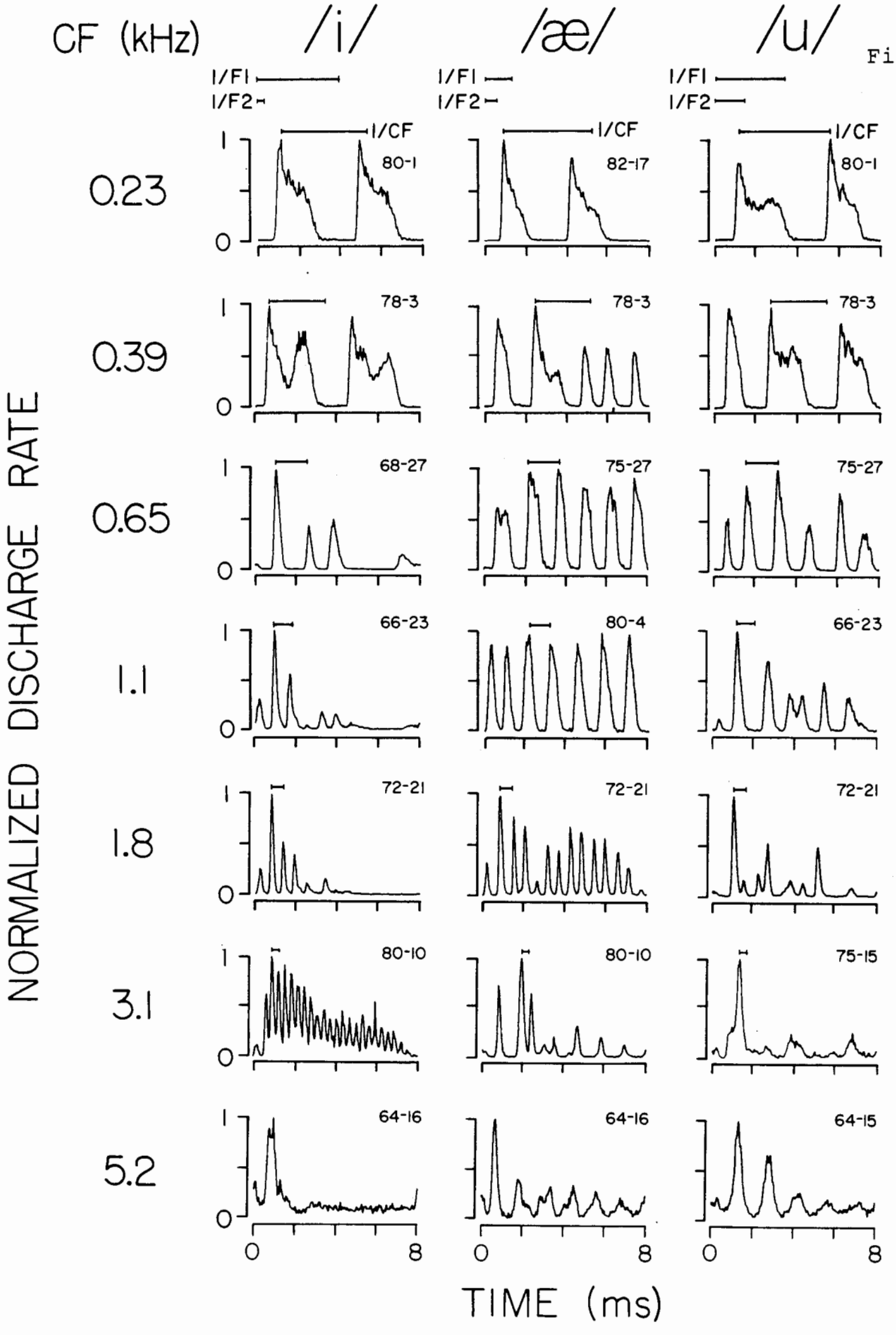

CF (kHz)
/i/
/æ/
/u/
1/F1
1/F2
1/CF
0.23
0.39
0.65
1.1
1.8
3.1
5.2
80-1
82-17
78-3
68-27
75-27
66-23
80-4
72-21
80-10
75-15
64-16
64-15
NORMALIZED DISCHARGE RATE
TIME (ms)
0
8

Fig. 6

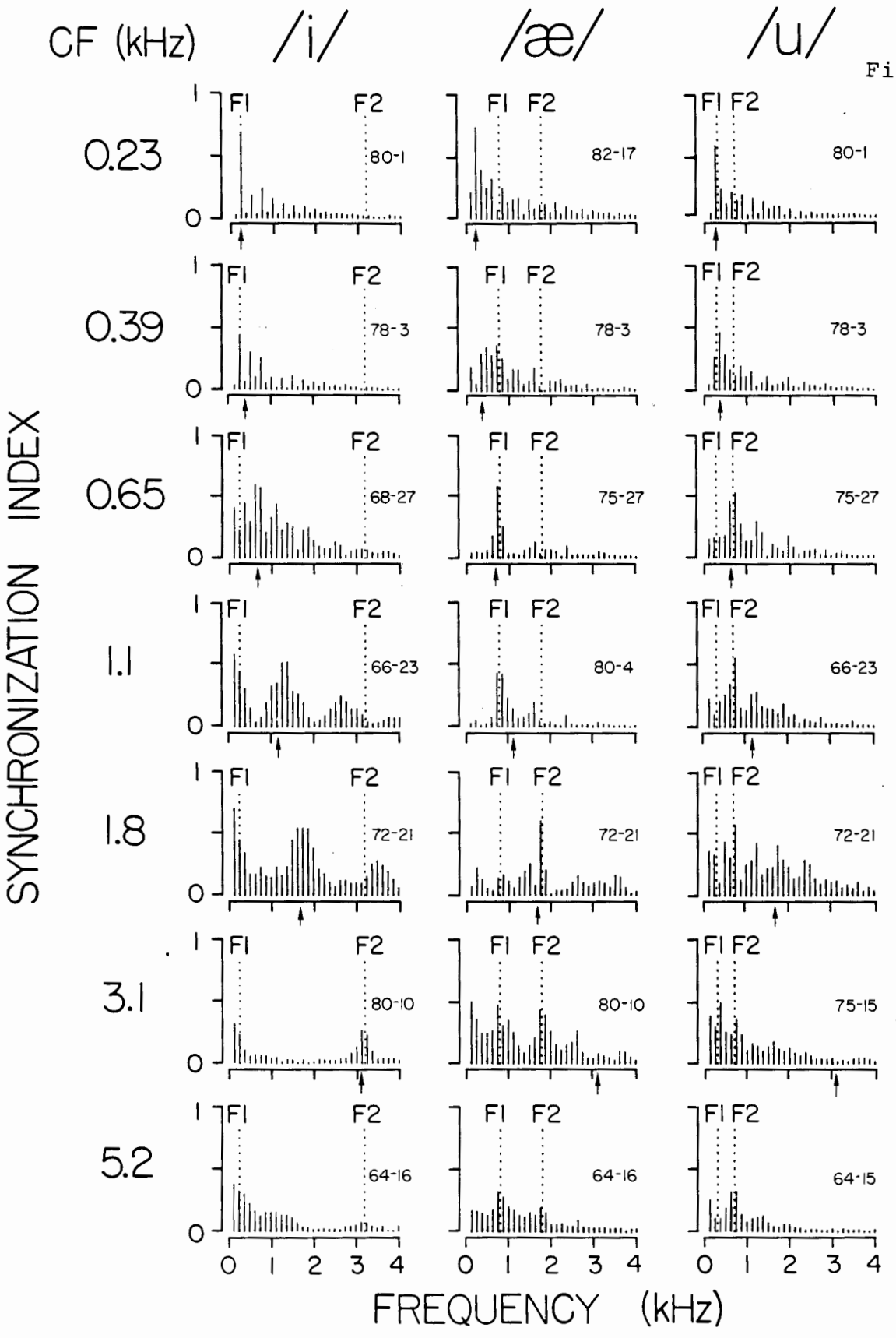

CF (kHz)
/i/
/æ/
/u/
0.23
0.39
0.65
1.1
1.8
3.1
5.2
F1
F2
80-1
82-17
80-1
78-3
78-3
78-3
68-27
75-27
75-27
66-23
80-4
66-23
72-21
72-21
72-21
80-10
80-10
75-15
64-16
64-16
64-15
1
0
0 1 2 3 4
SYNCHRONIZATION INDEX
FREQUENCY (kHz)

Fig. 7

Fig. 8

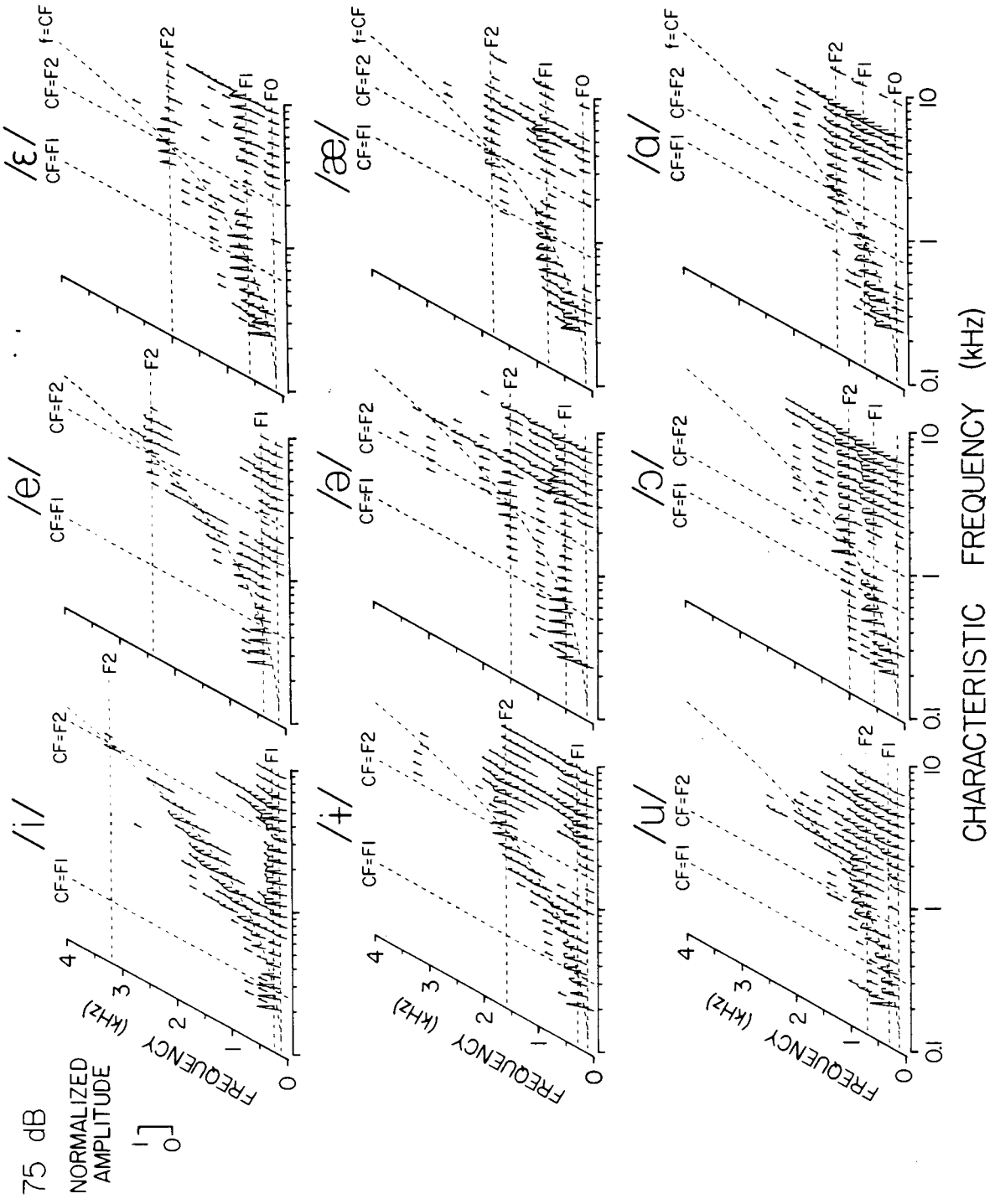
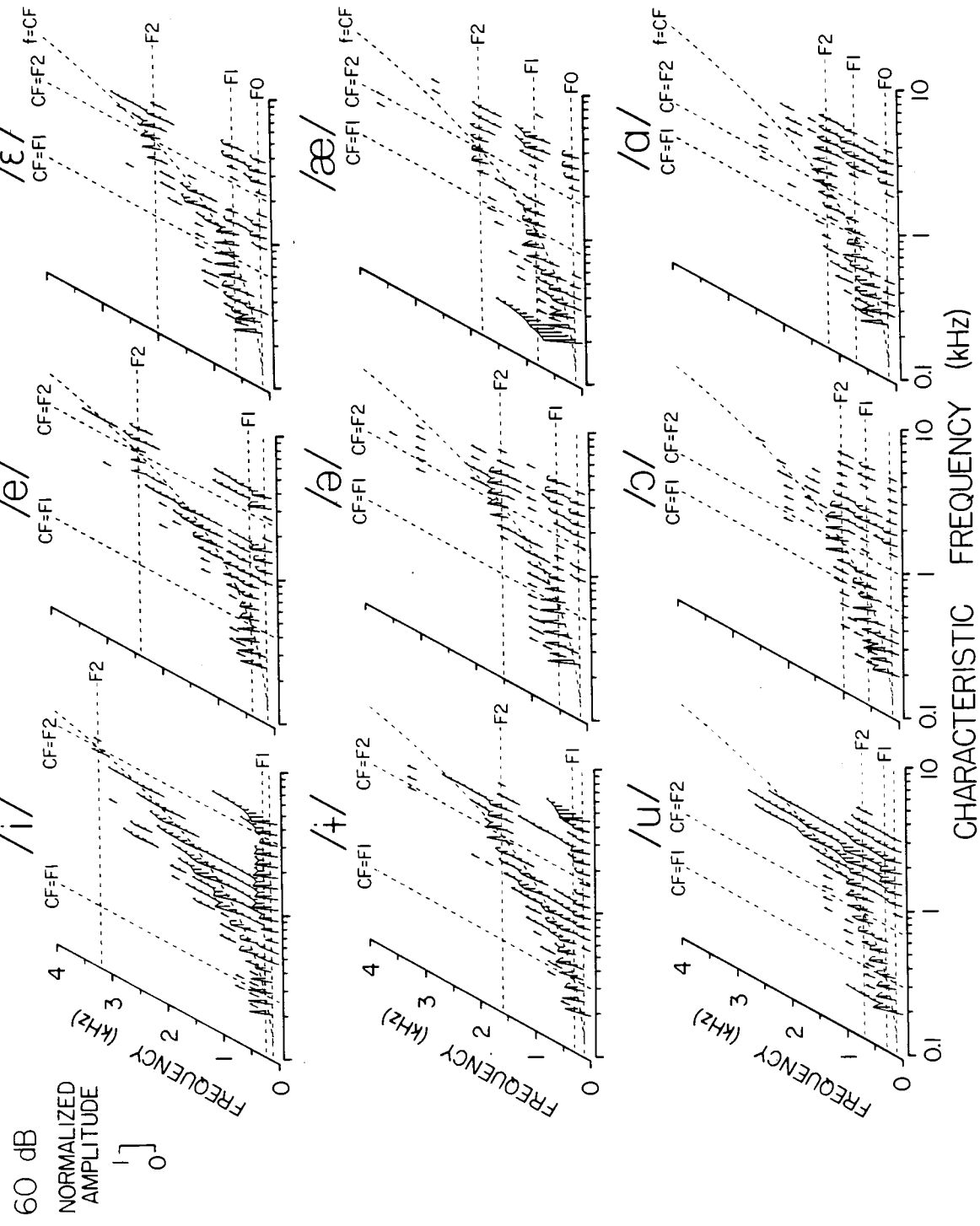

75 dB
NORMALIZED AMPLITUDE
1
0
/i/
/e/
/ɛ/
/ɨ/
/ə/
/æ/
/u/
/ɔ/
/ɑ/
CF=F1
CF=F2
f=CF
F0
F1
F2
FREQUENCY (kHz)
0
1
2
3
4
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 9


/i/
/e/
/ɛ/
/ɨ/
/ə/
/æ/
/u/
/ɔ/
/ɑ/
CF=F1
CF=F2
f=CF
F0
F1
F2
FREQUENCY (kHz)
0
1
2
3
4
0.1
1
10
60 dB
NORMALIZED AMPLITUDE
CHARACTERISTIC FREQUENCY (kHz)

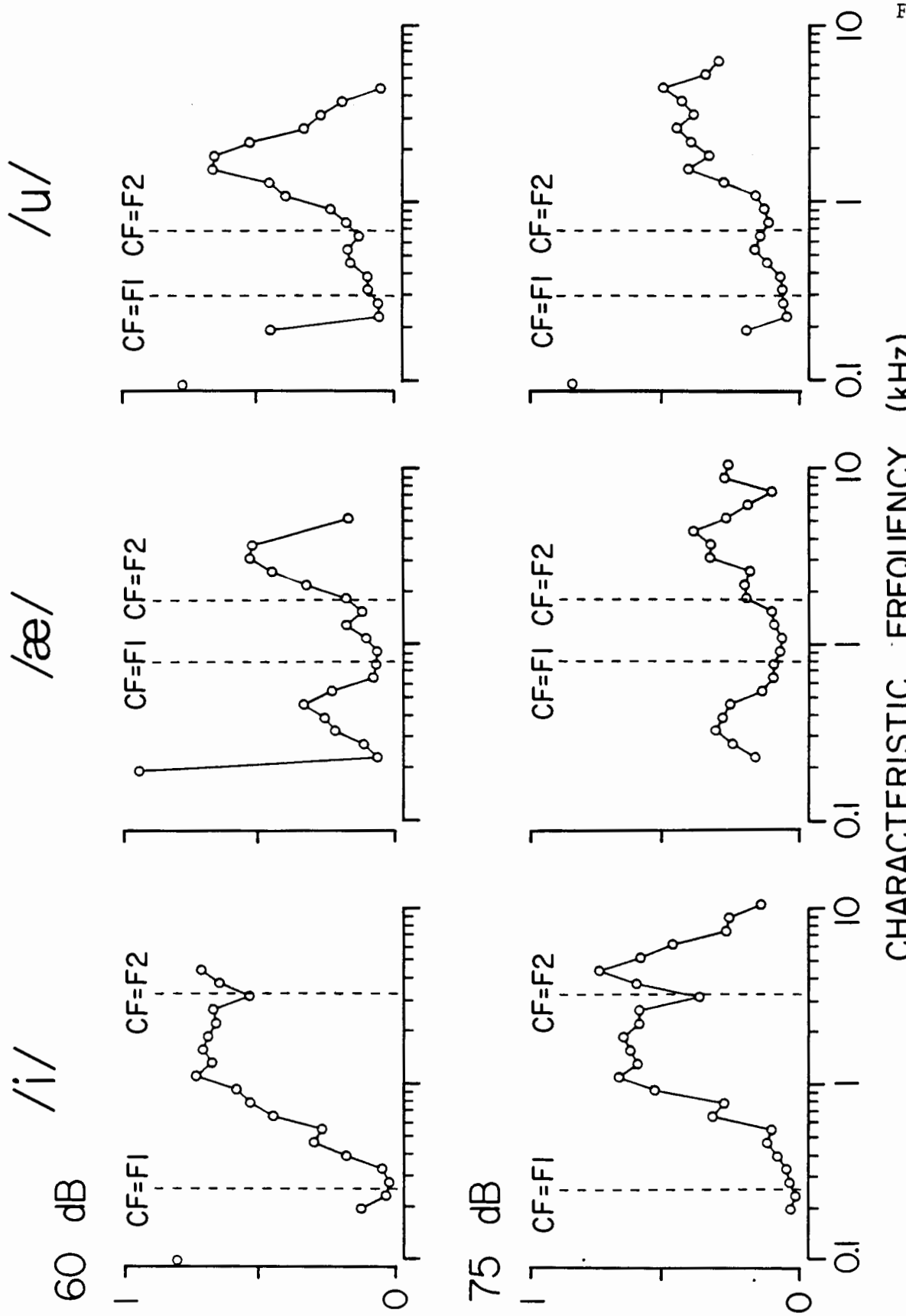

/i/
/æ/
/u/
60 dB
75 dB
CF=F1
CF=F2
SYNCHRONIZATION INDEX AT FO
1
0
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 10

Fig. 11

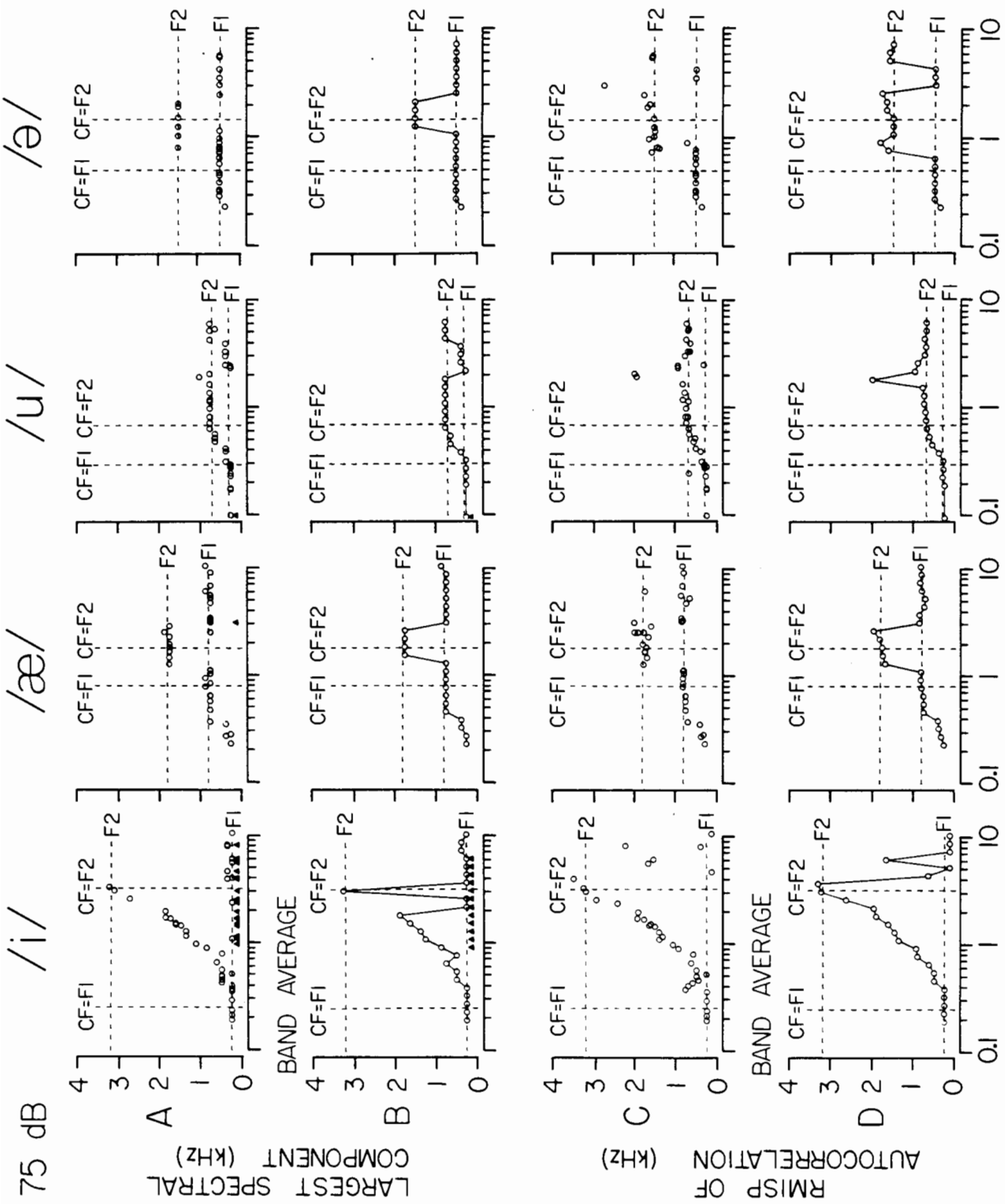

75 dB
/i/
/æ/
/u/
/ə/
CF=F1
CF=F2
F1
F2
A
B
C
D
BAND AVERAGE
LARGEST SPECTRAL COMPONENT (kHz)
RMISP OF AUTOCORRELATION (kHz)
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 12

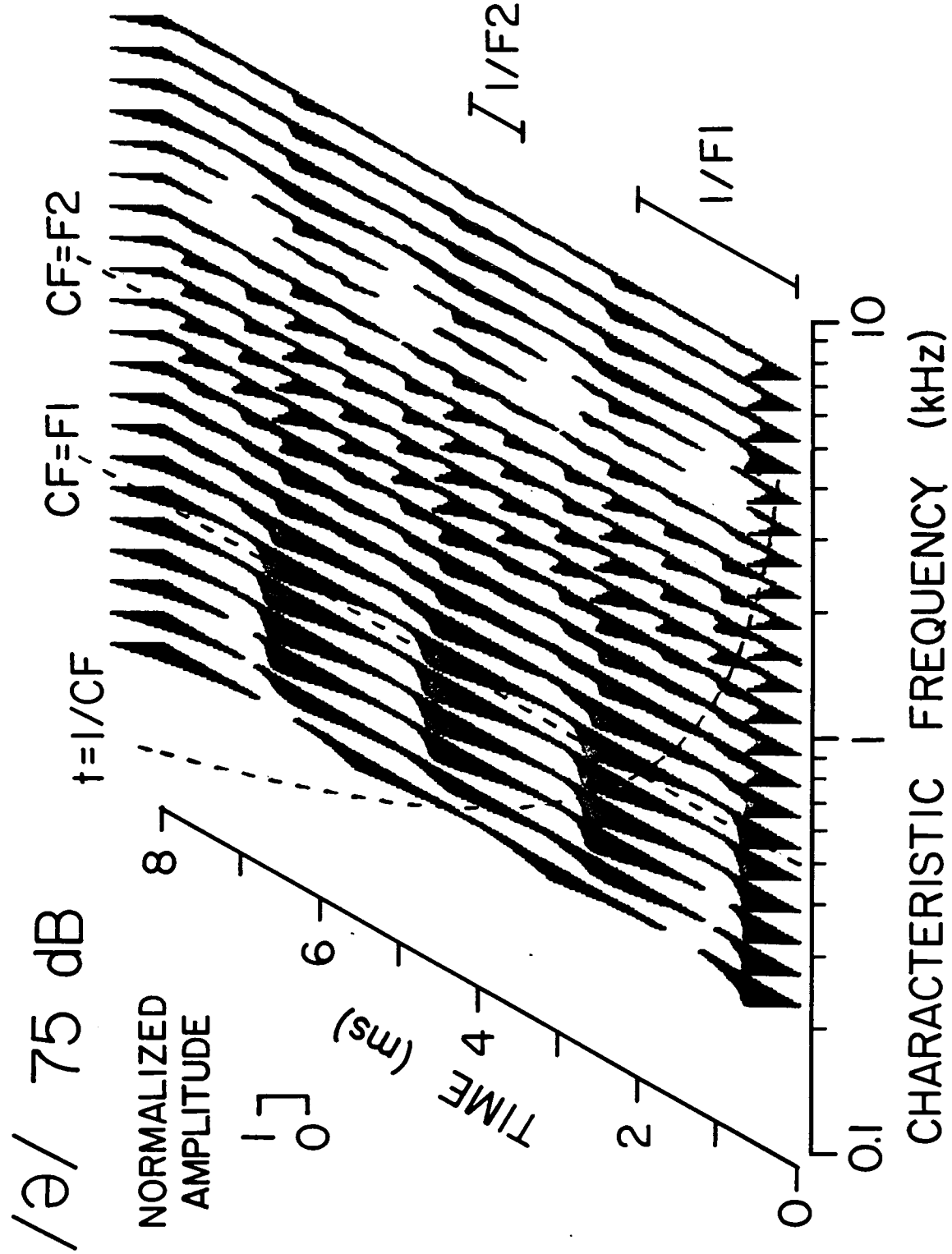

/ə/ 75 dB
NORMALIZED AMPLITUDE
1
0
TIME (ms)
0
2
4
6
8
t=1/CF
CF=F1
CF=F2
1/F2
1/F1
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 13

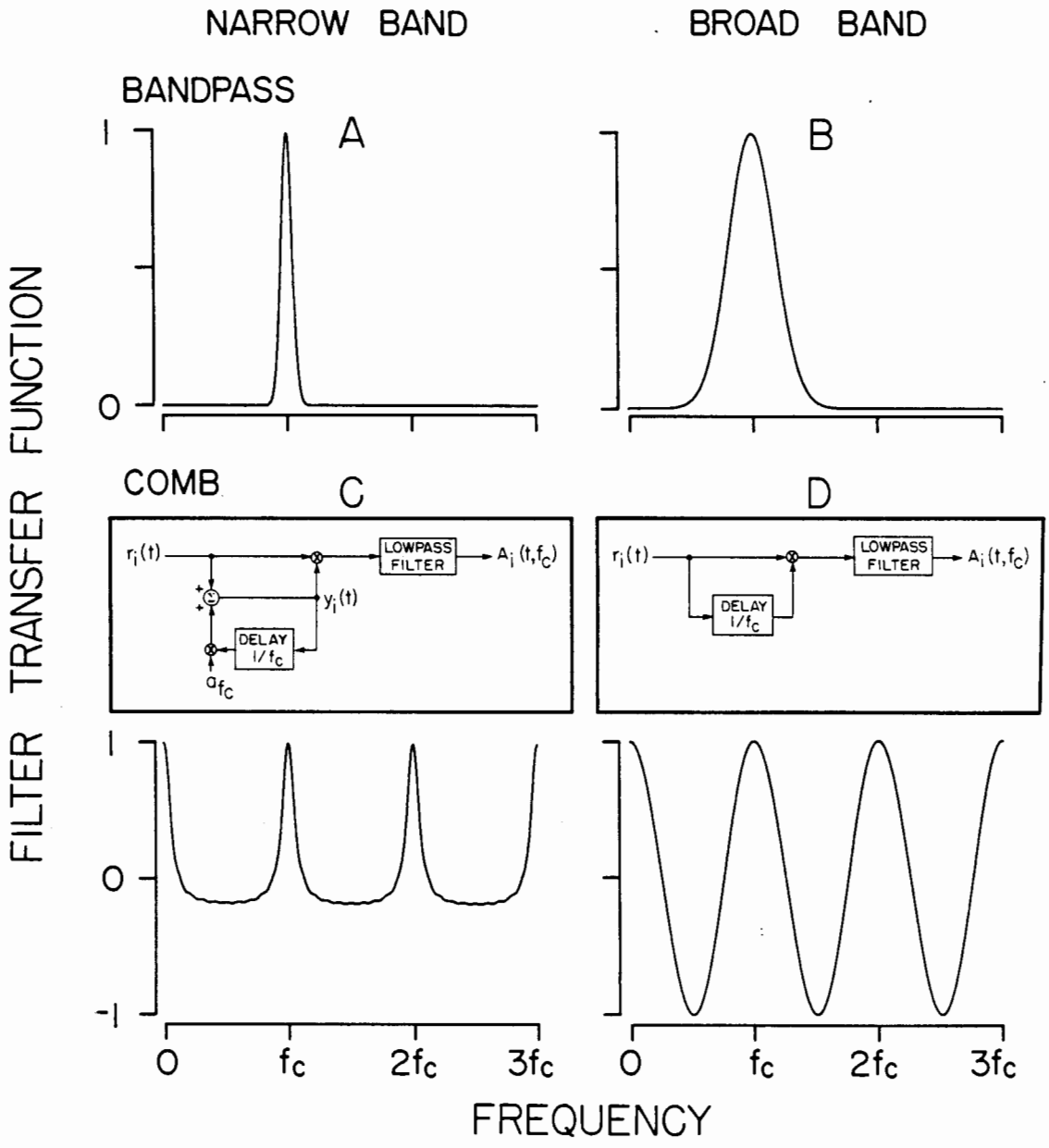
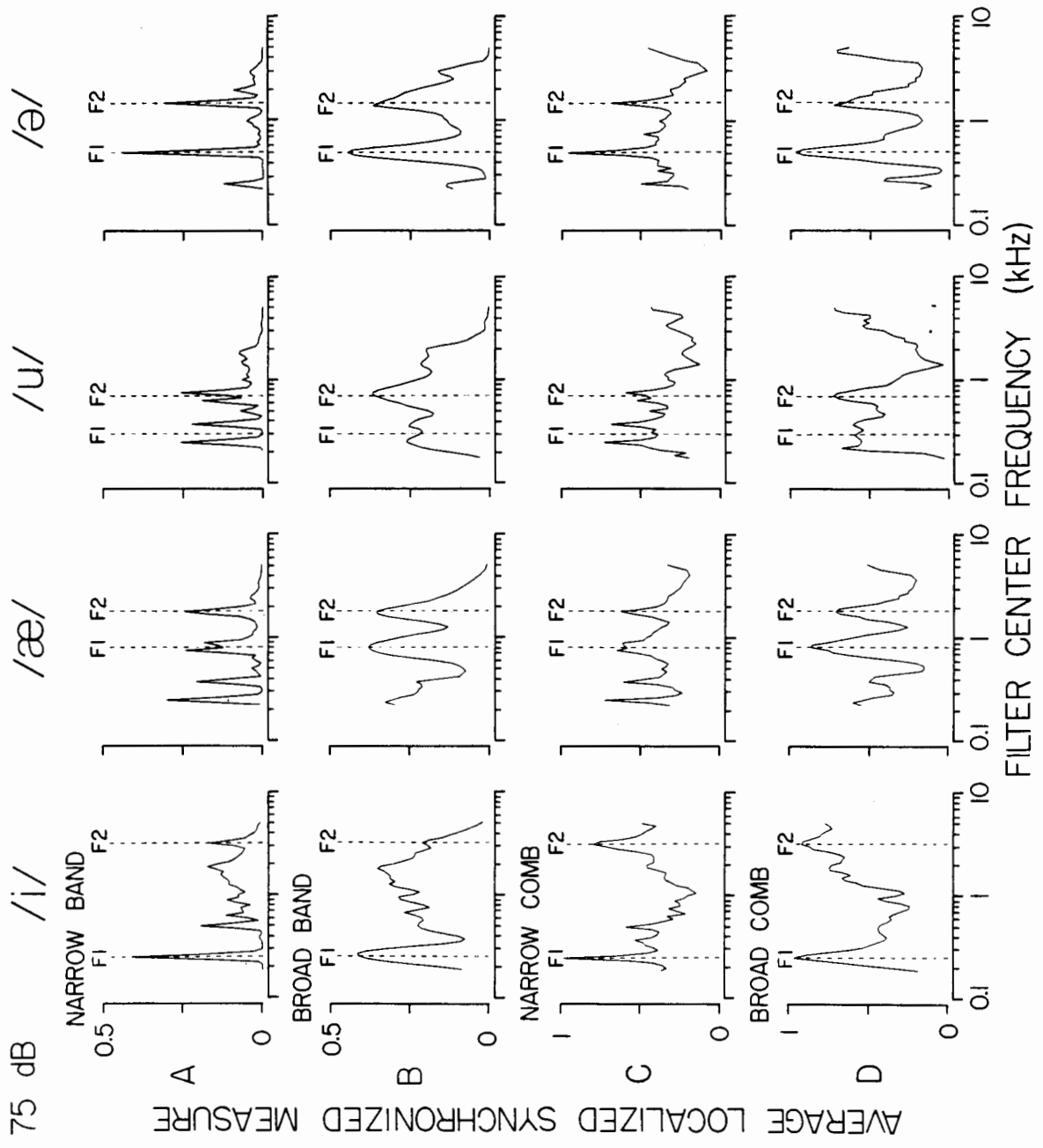

NARROW BAND
BROAD BAND
BANDPASS
A
B
COMB
C
D
FILTER TRANSFER FUNCTION
1
0
-1
$r_i(t)$
LOWPASS FILTER
$A_i(t,f_c)$
$y_i(t)$
DELAY $1/f_c$
q_{f_c}
0
f_c
$2f_c$
$3f_c$
FREQUENCY

Fig. 14


75 dB
/i/
/æ/
/u/
/ə/
NARROW BAND
BROAD BAND
NARROW COMB
BROAD COMB
F1
F2
A
B
C
D
0.5
1
0
0.1
10
AVERAGE LOCALIZED SYNCHRONIZED MEASURE
FILTER CENTER FREQUENCY (kHz)

CHAPTER II

CODING OF VOICELESS FRICATIVE CONSONANTS IN THE AUDITORY NERVE

INTRODUCTION

One of the basic distinctions of speech is that between vowel-like ("sonorant") sounds and noise-like ("obstruent") sounds. Sonorants, which include vowels, glides and nasal consonants, have a quasi-periodic waveform, considerable energy in the low frequencies, and a well-defined formant structure. Obstruents, which include fricative and stop consonants, have an irregular waveform, most of their energy above 1-2 kHz, and an "incomplete" formant pattern, in the sense that not all the resonances of the vocal tract are excited by the turbulence noise. The manner in which formant patterns for sonorants are coded in the auditory nerve has been studied in some detail using steady-state vowel stimuli (Sachs and Young, 1979,1980; Young and Sachs, 1979; Chapter I). In contrast, systematic studies of the response of the auditory nerve to obstruents are lacking, though there are reports of experiments with such stimuli (Kiang and Moxon, 1972,1974; Delgutte, 1980). The purpose of the present experiments was to describe how the spectra of obstruents are represented in the discharge patterns of auditory-nerve fibers. The study is based on four computer-generated noise bursts sounding like the voiceless fricatives /x/ (as in

German "Bach"), /x͡š/, /s/ and /f/. These stimuli are intended to cover roughly the range of voiceless obstruent spectra that are found in language, though, of course, they do not constitute a complete inventory of fricatives found in all languages.

Even though this study is based on a small number of stimuli, it may have more general implications because the spectra of voiceless fricatives resemble those of other obstruents. Specifically, the spectra of the release bursts for the stop consonants /k/ and /g/ are similar to the spectrum of /x/, the burst spectra of /t/ and /d/ are similar to the one of /s/, and, within limits, the spectra at the release of /p/ and /b/ resemble the one of /f/ (Halle et al., 1957; Zue, 1976). In addition, the steady-state spectra of the affricate /č͡š/ and the fricative /x͡š/ are similar, and the fricatives /θ/ and /f/ in isolation are hard to distinguish (Heinz and Stevens, 1961). The voiced fricatives /v/, /ž͡z/ and /z/ have about the same spectral characteristics as their voiceless correspondents /f/, /x͡š/ and /s/ respectively, except for the presence of considerable low-frequency energy.

Most of the energy of fricative sounds is located in specific bands of the frequency spectrum (Strevens, 1960). To a large extent, distinctions between fricatives are based on the positions of these bands (Harris, 1958; Heinz and Stevens, 1961). A simple manner in which this information could be represented in the auditory nerve is in the profile

of average discharge rate against characteristic frequency. According to this scheme, fibers with CF's in the frequency bands where the stimuli have most of their energy would respond with larger discharge rates than the other auditory-nerve fibers. In the case of vowel stimuli, the adequacy of this scheme is limited at high stimulus levels by the saturation of rate-intensity functions and suppression effects (Sachs and Young, 1979). In contrast, it is possible to define response measures based on fine time patterns of discharge that provide a good representation of the formant patterns of vowels over a wide range of stimulus levels (Young and Sachs, 1979; Chapter I). However, in the case of stimuli with intense high-frequency components such as fricative consonants, such schemes may be limited by the increasingly weak synchrony of discharges as frequency increases (Rose et al., 1967; Johnson, 1980). A goal of the present study is to evaluate the ability of average discharge rate and fine time patterns of discharge to represent the essential information for distinctions among voiceless fricative stimuli.

I METHODS

A. Stimuli

The fricative stimuli were generated by a computer program that simulates the summed outputs of one to three bandpass filters excited by broadband noise. The noise source, which models the turbulence generated at a

constriction of the vocal tract, is approximately white and Gaussian, and has the same waveform for all stimuli. The noise waveform lasts 200 ms, and has a rise-fall time of 75 ms. The bandpass filters represent the resonances of the vocal tract that are excited by the turbulence noise. They are second-order filters implemented by the impulse-invariance method (Oppenheim and Schafer, 1975) using a sampling rate of 20 kHz. The outputs of the bandpass filters are added with alternate phases, quantized to 12 bits, converted to analog signals, and lowpass filtered at 10 kHz.

Figure 1 shows the power spectra of the four fricative stimuli that were used in these experiments. Because in the production of fricative consonants only the resonances of the portion of the vocal tract that is located in front of the constriction are excited by the frication noise, the formants of fricatives can only be labeled by reference to a neighboring vowel (Fant, 1960). The excited formants are manifested by peaks at frequencies P1, P2 and P3 in the power spectra. The frequencies and bandwidths of the excited formants are listed in Table I. Natural fricatives differ from these computer-generated stimuli in that they often have a spectral zero below the frequency of the lowest excited formant (Heinz and Stevens, 1961).

The spectrum of the /x/ stimulus has a prominent peak at 2 kHz and an abrupt high-frequency cutoff near 4 kHz. The spectra of /$\overset{y}{s}$/ and /s/ are basically bandpass, with local

spectral prominences not exceeding 10 dB in amplitude within the passband. Some of these prominences may not be detectable by human listeners (Malme, 1959), so that the perceptually important features of the spectra are probably the low-frequency and high-frequency cutoffs. These cutoffs are located near 2 and 7 kHz for /ʃ/, and near 4 and 8 kHz for /s/. The spectrum of /f/ is nearly flat in the low frequency region, and rises gradually above 4 kHz to reach a broad maximum near 9 kHz.

B. Experimental procedures and data processing

The experimental results to be reported are based on recordings from single auditory-nerve fibers in anesthetized cats. The preparation of the animals, the stimulus generation system, and the recording procedures are as described in Chapter I. Recordings were restricted to the most sensitive population of auditory-nerve fibers, those with spontaneous discharge rates greater than 18 spikes/s (Liberman, 1978).

Stimulus levels were set by a three-step procedure. First, average transfer characteristics of the acoustic system over 20 experiments were obtained prior to this experimental series. Second, "standard" level settings were computed for each stimulus from their power spectra and the average transfer characteristics. Third, for each experiment, stimulus levels were set by adding to the standard settings a correction based on the mean magnitude of the transfer ratio

for that experiment over the frequency range 0.1-5 kHz. Thus the actual levels of the stimuli may vary somewhat from one experiment to another, though the standard deviation of the transfer ratios for the animals used in this series did not exceed 2 dB for frequencies below 10 kHz. The /x/, /ʃ/ and /s/ stimuli were presented at levels of 45 and 60 dB SPL, which are appropriate for speech in which the levels of the vowels would be about 60 and 75 dB SPL respectively. The /f/ stimulus was presented at levels that were 10 dB lower than those of the other stimuli, as would occur in real speech. These two stimulus levels will be referred to as "low" and "high" levels, respectively.

The 200-ms fricative stimuli were delivered 200 to 400 times at a rate of 100/min for the computation of post-stimulus time (PST) histograms (Gerstein and Kiang, 1960). A first set of PST histograms, computed with a bin width of 0.25 ms, was used to estimate average discharge rate in two different intervals. The "steady-state" rate was computed by summing the histogram bins within a trapezoidal time window beginning at the onset of the stimulus, and having a central duration of 50 ms and a total duration of 200 ms. The "onset" rate was measured using a window of value 1 from 0 to 10 ms, then decreasing linearly to reach 0 at 50 ms.

A second set of PST histograms, computed with a bin width of 0.025 ms for the central 100-ms segment of the stimuli, was used to estimate power spectra and

autocorrelation functions. This bin width is adequate for estimating frequency components of the instantaneous discharge rate up to about 8 kHz (Johnson, 1978). The DC component of the histogram was removed prior to the computation of power spectra and autocorrelation functions. Power spectra were estimated by (1) dividing the histogram into 16 overlapping 12.8-ms segments, each of which was weighted by a Kaiser window, and (2) averaging the magnitude square of the discrete Fourier transforms (DFT) of these segments (Oppenheim and Schafer, 1975). The resulting spectra have a frequency resolution of about 80 Hz. The autocorrelation functions were estimated for delays up to 12.8 ms by the DFT method (Oppenheim and Schafer, 1975).

In order to obtain data at regularly-spaced samples along the log CF dimension, power spectra from all fibers whose CF lies in a narrow band of frequency were averaged, using a trapezoidal weighting window with a central width of 0.25 octave and a total width of 0.55 octave (effective width 0.4 octave). The weighted number of fibers per band was about 3 on the average, and the CF bands were sampled every 0.25 octave. Overall, more than 100 fibers from 17 animals were involved in this band averaging. The legitimacy of this procedure has been discussed in Chapter I for responses to vowel stimuli. Band averaging was also applied to onset discharge rate and steady-state rate, using the same weighting windows as for the power spectra.

The power spectra and autocorrelation functions of PST histograms were used to compute the response measures that were found useful in Chapter I for extracting spectral parameters of vowel stimuli. One of these measures is the Reciprocal of the Mode of the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function. To compute this measure, a histogram of the frequency distribution of the intervals between successive peaks of the autocorrelation function was constructed, weighting each interval by height of the second peak, and the RMISP was set to the reciprocal of the center of gravity of the intervals within ± 12 % of the histogram mode. Another class of response measures are the Average Localized Synchronized Measures (ALSM), which are similar to the ALSR proposed by Young and Sachs (1979). These measures are obtained by filtering response patterns of auditory-nerve fibers by a filter whose center frequency f_c is near the fiber characteristic frequency. The ALSM's differ by the choice of the filter, comb or bandpass. For the ALSM based on a bandpass filtering scheme, the result A_j of the filtering operation for fiber j was computed from:

$$A_j = \sum_{0<f_k<10} P_j(f_k)\ H(f_k) \qquad (1)$$

where f_k is frequency in kHz, H(f) is the transfer function of a Gaussian filter, and $P_j(f_k)$ is the power spectrum for fiber j. The filter transfer function is given by:

$$H(f) = \exp -\pi[(f-f_c)/b_c]^2 \qquad (2)$$

where the bandwidth b_c is 0.116 f_c (1/6 octave) or 80 Hz, whichever is larger. For the ALSM based on a comb filtering scheme, the result of the filtering operation for each fiber was obtained by evaluating the autocorrelation function of the PST histogram at time $t=1/f_c$. For both filtering schemes, the result of the filtering operation for each fiber was normalized by the mean square discharge rate, and the ALSM was computed by averaging these normalized results for all fibers whose CF is in a narrow frequency band centered at f_c, using a trapezoidal weighting window with a central width of 1/6 octave and a total width of 1/2 octave. This computation was made for values of f_c ranging from 0.2 to 10 kHz, in 1/12-octave intervals.

II RESULTS

A. Average discharge rate

Figure 2 shows steady-state discharge rate plotted against characteristic frequency for the four fricative stimuli presented at the low level. For all stimuli, discharge rate is large for fibers whose CF's are in the frequency regions where the stimuli have most of their energy. It is close to spontaneous discharge rate for fibers with CF's more than 2-3 kHz above the highest formant frequency, and for those with CF's below 0.5 kHz. For /š/, /s/ and /f/, the CF regions with the largest discharge rates correspond well with the frequency regions where the stimuli have most of their

energy. In contrast, for /x/, the largest discharge rates are found below the place of the 2-kHz peak in the stimulus spectrum. The profiles for the other stimuli also show a local prominence near 1.2 kHz, so that the mismatch between the 2-kHz peak in the spectrum and the maximum of the rate profile for /x/ may be due to peculiarities in fiber sampling. In any case, the rate profiles for the four stimuli are clearly distinct at this stimulus level.

Figure 3 shows steady-state rate plotted against CF for the four stimuli presented at the high level. Compared to the low-level condition, discharge rate for /s/ and /f/ is increased throughout the range of CF's, so that the profiles preserve their characteristic shape. In contrast, for the /x/ and /$\overset{v}{s}$/ stimuli, there is little or no increase in rate at the places of the formant frequencies, while discharge rate in the surrounding regions increases considerably. Although the profiles for /x/ and /$\overset{v}{s}$/ both have plateaus extending over a wide CF region, they clearly differ in their high-CF cutoff.

Figure 4 shows that differences in the rate profiles for the high-level stimuli are enhanced when the onset discharge rate is used as a response measure. The onset rate is smaller than the steady-state rate in the CF region below 0.5 kHz and above the highest formant frequency. In contrast, onset rate is larger than steady-state rate near the places of the formant frequencies. For the low-level condition (not shown), large onset discharge rates are even more restricted

to CF regions near the formant frequencies, but the representation of the spectral peak of /f/ is weak, so that it is not clear that onset rate is a better measure than steady-state rate at that level.

Thus, at each stimulus level, the profile of either onset rate or steady-state rate against CF provides a rough indication of the frequency regions where the stimuli have most of their energy, though the precise positions of the formant frequencies are not clearly apparent. The changes in the profile with stimulus level are considerable, but, at least for this limited set of stimuli, in no case does the profile for one stimulus at the low level resemble the profile for another stimulus at the high level.

B. Fine time patterns of discharge

Because the fricative stimuli are bursts of pseudo-random noise that have the same waveform for each stimulus presentation, power spectra of PST histograms can be used to identify which frequency components fiber discharges are synchronized to. Figure 5 shows normalized power spectra of PST histograms for 3 fibers in response to the /ʃ/ and /f/ stimuli presented at the high level. The CF of the 0.71-kHz unit is far below the frequency range where the stimuli have most of their energy. The response spectra of this unit are similar for the two stimuli, with a peak near CF and a smaller peak in the low frequencies. The peak at CF, which is also

found in the response spectra of auditory-nerve fibers to broadband noise stimuli (De Boer and Kuyper, 1968; Møller, 1977; Evans, 1977), is consistent with the tuning that occurs in the cochlea. The CF of the 2.43-kHz unit is near the lowest formant of /ʃ/, but nearly two octaves below the formant of /f/. The response spectra for both stimuli have a peak near the CF and at low-frequencies, but the relative amplitude of the low-frequency peak is much greater for /f/ than for /ʃ/. The CF of the 10.5-kHz unit is close to the formant of /f/, and nearly one octave above the highest formant of /ʃ/. The spectra for both stimuli have significant components only in the low frequencies, but these components have a higher amplitude for the /ʃ/ stimulus than for the /f/ stimulus. The absence of a CF peak (not shown) is consistent with the weak synchrony of responses for high-frequency tones (Rose et al., 1967; Johnson, 1980). Thus, the gross features of the power spectra of PST histograms for fricative stimuli vary considerably with CF, so that a complete description of responses to fricatives must include these variations for the entire array of auditory-nerve fibers.

Figure 6 shows normalized band-average power spectra for narrow CF bands covering a wide range of CF's in response to the four fricative stimuli presented at the low level. For all stimuli, the activity is distributed primarily along the f=CF line and in the low frequencies. Prominent activity along the f=CF line is found only for fibers with CF's below

3-4 kHz. There are also response components at frequencies that are about twice the CF. These components are probably generated by nonlinear distortion in the cochlea. The activity in the low-frequency region tends to spread to higher frequencies as CF increases. These response components probably do not originate from the weak low-frequency components of the stimulus, which would be below the thresholds of high-CF fibers, but are more likely generated by nonlinear distortion in the cochlea. One would expect that, following some sharply-tuned stage of processing, the response variable would be a narrow-band noise whose envelope would have prominent components near the bandwidth of the filtering elements. Rectification of the envelope at a later stage of processing would generate low-frequency components such as those observed in the discharge patterns of auditory-nerve fibers. Consistent with this interpretation of the generation of low-frequency components, the spread of these components to the high frequencies as CF increases would be expected on the basis of the increase in tuning curve bandwidths (Kiang et al., 1965).

Though auditory-nerve activity tends to be distributed in the same regions of the diagram of Fig. 5 for all stimuli, the amplitudes of the response components vary between stimuli. For instance, for /x/, there is little activity for CF's above 5 kHz, while clear low-frequency components are still found at 11 kHz for /s/ and /f/. For the

/x/ stimulus, there is a band of CF's near 2 kHz in which the response spectra are dominated by components near the lower formant frequency, and a few fibers with CF's between 4 and 10 kHz also show response components near the 2-kHz formant frequency. Broad bands of CF's within which response spectra are dominated by components near a formant frequency are typical for vowel stimuli (Young and Sachs, 1979; Chapter I). This type of response is not found for the higher-frequency formants of /x/ and for the formants of the other stimuli.

Figure 7 shows the band-average power spectra in response to the four stimuli for the high-level condition. The response activity is generally distributed in the same areas as at the lower level. One difference is that low-frequency response components tend to be smaller in the CF regions where the stimuli have most of their energy. In these CF regions, the PST histograms are basically flat, so that the power spectra have only weak components besides DC (which is removed prior to the computation of the power spectra). The low amplitude of non-DC components is illustrated in Fig. 8, which shows the ratio of the square of the average discharge rate to the variance of the PST histogram plotted against CF for the four fricative stimuli presented at the high level. Large values of this ratio indicate that the discharge rate in each bin of the histogram deviates little from the average discharge rate. The ratio is largest for fibers with CF's near the formant frequencies that are above about 3 kHz. This

pattern holds at the lower level (not shown), but large ratios are also found for low-CF fibers because their discharge rate deviate little from spontaneous discharge rate.

In summary, for both stimulus levels, there are differences in fine time patterns of discharge between the responses to the four fricative stimuli. However, the largest response components in any CF region tend to be the same for all stimuli, and prominent response components at a formant frequency over a band of CF's are found only for the lower formant of /x/.

C. Speech processing schemes

The speech processing schemes that were found useful to estimate the formant frequencies of vowel stimuli were based on the observation that response components near the formant frequencies were in general more prominent than non-formant components. Because these findings do not in general apply to the fricative stimuli (Fig. 6 and 7), one would expect that these processing schemes would be less effective. Figure 9 shows one such response measure, the Reciprocal of the Mode of the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function, plotted against CF for the four fricative stimuli presented at the high level. This measure has about the same CF dependence as the largest component in the response spectrum, but gives more prominence to high-frequency

components. For vowel stimuli, the RMISP of the autocorrelation is close to one of the formant frequencies for the majority of fibers, so that profiles of RMISP against CF show prominent horizontal bands at the formant frequencies. In contrast, for the fricative stimuli, the RMISP generally coincides with fiber CF for frequencies below 3-4 kHz, and is not well defined for high-CF fibers since peaks of the autocorrelation function become hard to detect. The RMISP is equal to one of the formant frequencies only for a narrow CF band around the lower formant of /x/ and for scattered high-CF fibers.

Other processing schemes that were found useful for estimating formant frequencies of vowel stimuli are the Average Localized Synchronized Measures (ALSM), which are obtained by filtering the fiber response patterns around the CF. Figure 10 shows a plot of ALSM against center frequency of a narrow bandpass filter for the four fricative stimuli presented at the high level. For vowel stimuli, narrow-band ALSM's and related measures (Sachs and Young, 1979; Young and Sachs, 1979) show prominent peaks near the formant frequencies. In contrast, the ALSM for fricative stimuli shows peaks at the formant frequencies only for the lowest formants of /x/ and /$\overset{y}{s}$/. Furthermore, for /s/ and /f/, the ALSM shows a non-formant peak near 2.3 kHz that is as prominent as the formant peaks of /x/ and /$\overset{y}{s}$/. There are also low-frequency non-formant peaks, even though low-frequency

components in the stimulus spectrum are weak. Figure 11 shows plots of ALSM against center frequency for a comb-filtering scheme which, in some respects, was found preferable to the bandpass schemes for the vowel stimuli. In contrast to the bandpass ALSM of Fig. 10, the broad-comb ALSM of Fig. 11 is large for high center frequencies because the filter extracts the low-frequency response components as well as those near CF. Nevertheless, there are no clear peaks at the formant frequencies except for the low-frequency (< 3 kHz) formants of /x/ and /$\overset{v}{s}$/. The ALSM's of Fig. 10 and 11 are normalized by the mean square discharge rate. Because the variance of the PST histograms is lowest near the places of high-frequency formants, unnormalized ALSM's actually show dips near these formant frequencies.

Thus, though fine time patterns of discharge do contain information about the spectra of fricative stimuli, response measures that extract the spectral peaks associated with low-frequency formants of vowel stimuli are inadequate for estimating the prominent spectral components of fricative stimuli.

III DISCUSSION

A. Relation to studies of stimulus coding in the auditory nerve

One major result of this study is that the profiles of average discharge rate against CF are clearly distinct for the four fricative stimuli at both stimulus levels that were investigated. In general, the frequency regions over which the stimuli have most of their energy correspond well with the CF regions with the largest discharge rates. In contrast, at comparable speech levels, the rate profiles for vowel stimuli do not show peaks at the CF's corresponding to the formant frequencies (Sachs and Young, 1979). One factor explaining this difference is that the fricatives have lower intensities than vowels, so that saturation of the rate-intensity functions does not have as much effect. For instance, discharge rate at the place of the higher formant of /s/ is higher at 60 dB than at 45 dB (Fig. 2 and 3), demonstrating that saturation has not occurred in this case. Nevertheless, fibers at the place of the lowest formants of /x/ and /ʃ/ do not show an increase in rate for the 15-dB increase in level, while the rate of neighboring fibers increases, so that saturation does influence the shape of the rate profile in these cases. The effect of saturation should be smaller for fibers with spontaneous discharges rates below 18 spikes/s, which were excluded from this study. These fibers have higher thresholds (Liberman, 1978) and a wider dynamic range (Schalk

and Sachs, 1980). In addition, cat behavioral thresholds are considerably lower than human thresholds at high frequencies (Miller et al, 1963). If such a difference also applied to auditory-nerve fiber thresholds, one would expect the effect of saturation to be lower in the human.

Another factor explaining differences in rate profiles between vowels and fricatives is the absence of intense low-frequency components for fricatives. For vowel stimuli, the intense components near the first formant frequency suppress the response of fibers at the places of the higher formants, so that rate-intensity functions have a negative slope over a certain range of intensities (Sachs and Young, 1979). This phenomenon does not seem to occur in our fricative data, since there is no clear decrease in discharge rate at the high level. However, one would need to study fricative responses over a wider range of levels to draw final conclusions on the effect of suppression. There may also be suppression without nonmonotonic rate-intensity functions. Fibers with CF's above 3 kHz respond to intense low-frequency components (Kiang and Moxon, 1974), so that the profile of discharge rate against CF for vowel stimuli extends far into the high-CF region at high stimulus levels (Sachs and Young, 1979). In contrast, fibers with CF's more than 2/3 octave above the highest formants of fricatives do not respond much to components near the formant frequencies, so that the rate profiles for different fricatives are clearly distinguished by

their high-CF cutoff. The low-CF cutoff is not so clearly defined because fibers with CF's below the lowest formant of fricatives respond to the weak low-frequency components of the stimuli at high levels. The spectral zero below the lowest formant that often occurs in actual fricatives (Heinz and Stevens, 1961) might result in a clearer low-CF cutoff in the rate profile.

For the high-level condition, the frequency regions in which the fricative stimuli have most of their energy are more clearly apparent in the profiles of onset rate against CF than in the profiles for steady-state rate. This finding, which also applies to vowel stimuli (Young et al., 1981), is probably related to the wider dynamic range of auditory-nerve fibers at the onset of tone bursts (Smith and Brachman, 1980a), and to the lower stimulus level during the 50-ms time window used to measure onset rate, since the fricative stimuli have a 75-ms rise time.

Responses of auditory-nerve fibers to vowel and fricative stimuli also differ considerably in fine time patterns of discharge. Whereas spectra of PST histograms for vowel stimuli are usually dominated by frequency components near the formant frequencies (Young and Sachs, 1979; Chapter I), spectra for fricative stimuli have large components only near the fiber CF and in the low frequencies. In that respect, responses to fricative stimuli are similar to responses to broad-band noise, which also have their most

intense components at the CF (De Boer and Kuyper, 1968; Ruggero, 1973; Møller, 1977; Evans, 1977; De Boer and De Jongh, 1978). The upper frequency limit of the CF components in response to fricative stimuli is about 3-4 kHz, somewhat lower than the 5-6 kHz limit given by Johnson (1980) for detectable synchrony of discharges to tonal stimuli. We have suggested that the low-frequency response components are due to rectification of the envelope of narrow-band noise generated by cochlear filtering of the fricative stimuli. The decrease in low-frequency components with increasing intensities for fibers with CF's near the formant frequencies is consistent with this interpretation because a similar decrease with stimulus level occurs for response components synchronized to the modulating waveforms of amplitude-modulated tones (Smith and Brachman, 1980b; Delgutte, 1980), and for the fundamental component of responses to vowel stimuli (Chapter I). This reduction of envelope components at high stimulus levels has been interpreted as a consequence of the saturation of rate-intensity functions (Smith and Brachman, 1980b; Delgutte, 1980).

Whereas the CF and low-frequency response components are prominent in response to fricative stimuli, components at the formant frequencies are not, except for the lowest formant of /x/ and, possibly, /$\overset{y}{s}$/, which are the lowest formants for this set of stimuli. In chapter I, the 3.2-kHz second formant

of a two-formant /i/ vowel was found to have a much weaker representation in fine time patterns of discharge than the lower-frequency vowel formants. The decrease in the strength of discharge synchronization with increasing frequency (Johnson, 1980) predicts the relatively low amplitude of response components at the high-frequency formants, but does not account for the absence of a wide band of CF's over which the largest response component is near the formant frequency. For low-frequency formants, the spread of formant response components over a wide band of CF's is due to suppression of non-formant response components by the intense formant components (Young and Sachs, 1979). Apparently, suppression is not as strong for high-frequency formants, possibly because high-frequency formants have larger bandwidths than low-frequency formants (Fant, 1960).

B. Central processing of fricative stimuli

Phonetic distinctions among steady-state voiceless fricatives are based more on the general location of the bands of energy in the stimulus spectra than on the existence of spectral peaks associated with formant frequencies (Harris, 1958; Heinz and Stevens, 1961). Though the perceptual importance of spectral peaks within the passband of fricative stimuli has not been studied systematically, it is common practice in speech synthesis to generate the fricatives /$\overset{\vee}{s}$/, /s/, and /f/ as bandpass-filtered noise without formant peaks. The perception of /s/ or /$\overset{\vee}{s}$/ is elicited by a broad band of

noise having considerable energy near 4-5 kHz, and little energy in the low frequencies. Noise with a low-frequency cutoff near 2-3 kHz is heard as /$\overset{v}{s}$/, whereas noise with a cutoff near 3-4 kHz evokes a /s/ percept (Harris, 1958; Heinz and Stevens, 1961). The fricative /f/ or /θ/ is heard when there is weak, flat noise in the low-frequencies; a spectral peak above 7-8 kHz makes identification of these consonants more likely, but is not essential. Perception of /x/, however, seems to require a prominent spectral peak in the 1.5-2.5 kHz region. Identification of isolated fricatives is facilitated by the fact that the number of phonologically-distinct voiceless fricatives in any language rarely exceeds five, whereas vowel systems with more than 15 contrasts are not uncommon. Thus, a sufficient requirement for response measures based on auditory-nerve data is to provide a rough representation of the gross shape of the spectra of fricative stimuli.

The profile of average discharge rate against CF seems to satisfy this requirement for both stimulus levels that were investigated, provided the time window over which discharge rate is averaged is chosen appropriately. Fine time patterns of discharge also contain information about the spectra of fricative stimuli, but this information is either limited to certain frequency regions or strongly level-dependent. For instance, for the high-level stimuli, the ratio of the square of the mean discharge rate to the

variance of the instantaneous rate shows maxima for CF's near the high-frequency formants (Fig. 8), but this ratio is also large for spontaneous activity, so that it could not be used for identifying formants unequivocally. Similarly, the ALSM's and the RMISP of the autocorrelation function fail to identify correctly the frequency regions of maximum energy, except for the lowest formants of /x/ and possibly /ʧ/. For vowel stimuli at comparable speech levels, the situation is exactly the opposite, as average discharge rate fails to show peaks at the formant frequencies, while the measures based on fine time patterns of discharge provide a prominent representation of the formants (Sachs and Young, 1979; Young and Sachs, 1979; Chapter I). Thus, no single response measure that has been proposed seems to represent adequately the spectral features that are important for phonetic distinctions among both vowel and fricative stimuli.

Comparison between the profiles for onset rate and steady-state rate suggests that it would be advantageous for the central processor to average discharge rate over a time window whose position relative to the onset of the stimuli would be earlier for higher intensities. PST histograms in response to tone bursts with long rise times show a peak in discharge rate whose latency decreases as stimulus level increases (Smith and Brachman, 1980a). Because fricative stimuli have a gradual onset, the central processor might position the averaging window relative to the peak in

discharge rate. In continuous speech, short-term adaptation of fibers by previous speech segments would probably suppress many of these peaks in discharge rate (Smith, 1977; Harris and Dallos, 1979). However, because there is always a frequency region in which the intensity of fricatives exceeds that of vowels, one would still expect to find peaks in discharge rate at the onset of the fricative in at least some CF regions.

The profile of average discharge rate against CF would also be useful for distinctions between vowels and fricatives, and more generally between sonorants and obstruents. For vowel stimuli, the largest discharge rates are found near the place of the first formant, which is always below 1 kHz (Sachs and Young, 1979). In contrast, for none of the fricative stimuli is the maximum of the rate profile below 1 kHz.

Another aspect of distinctions between vowels and voiceless fricatives is that the waveforms of vowels show a low-frequency periodicity at the fundamental frequency of vocal-fold vibration, whereas voiceless fricatives have a noise-like waveform. In response to vowels, autocorrelation functions of PST histograms for auditory-nerve fibers over a wide range of CF's show a prominent peak at the fundamental period (Chapter I). In response to the fricative stimuli, autocorrelation functions for low-CF fibers consist of damped oscillations at the CF superimposed on a DC component (not shown). Autocorrelation functions for high-CF fibers show

only one major peak at time zero. In no case do the autocorrelation functions for voiceless fricatives have a prominent peak at a time that would be appropriate for the fundamental period. Cues to the voiced/voiceless distinction are also present in the fine structure the profiles of ALSM's against filter center frequency. For vowel stimuli, narrow-band ALSM's show clear peaks at the positions of many of the low-frequency harmonics of the fundamental (Chapter I). Though the narrow-band ALSM for fricative stimuli also shows peaks in the low-frequency region, these peaks are not as narrow as those found for vowels, and their frequencies are not harmonically related.

Phonetic distinctions among voiceless fricatives, particularly the weak fricatives /f/ and /θ/, are not entirely based on the steady-state power spectra, but also on dynamic cues such as directions of formant transitions into neighboring vowels (Harris, 1958; Heinz and Stevens, 1961; Mann and Repp, 1980), and intensity relative to neighboring vowels (Heinz and Stevens, 1961; McCasland, 1979; Gurlekian, 1981; Stevens, 1981). The representation of some of these dynamic cues in the auditory nerve is studied in the next chapter using consonant-vowel syllables.

Table I Formant parameters for the fricative stimuli

STIMULI	P1 (kHZ)	BW1 (Hz)	P2 (kHz)	BW2 (Hz)	P3 (kHz)	BW3 (Hz)
/x/	2.0	200	3.5	400		
/ʂ/	2.8	250	4.0	500	6.0	700
/s/	5.0	800	7.5	1000		
/f/	9.5	2000				

FIGURE CAPTIONS

Fig. 1

Power spectra of the four fricative stimuli. The spectra were computed from the waveforms of the computer-generated stimuli using the same method as described in Sec. IB for the power spectra of the PST histograms, and were corrected for the average transfer characteristics of the acoustic system over 27 animals. The vertical scale was computed in dB relative to $(20\ \mu Pa)^2/Hz$, assuming that the stimulus level is 60 dB SPL for /x/, /ʃ/ and /s/, and 50 dB SPL for /f/. Vertical dashed lines mark the positions of the formant frequencies P1, P2, and P3.

Fig. 2

Steady-state discharge rate plotted against CF for the four stimuli presented at the low level. The steady-state rate is measured as described in Sec. IB. Each circle represents discharge rate for one auditory-nerve fiber. Data from 17 animals are pooled, and the fibers are not necessarily the same for all stimuli. The continuous lines represent the band-average discharge rates for 0.55-octave CF bands sampled every quarter octave. The positions of the formant frequencies along the CF dimension are marked by dashed lines.

Fig. 3

Same as Fig. 2 for the high-level condition. The fibers are not necessarily the same as in Fig. 2.

Fig. 4

Same as Fig. 3 for the onset discharge rate. The onset rate is measured as described in Sec. IB.

Fig. 5

Normalized power spectra of PST histograms for three fibers in response to the /ʃ/ and /f/ stimuli presented at the high level. Vertical dashed lines mark the positions of the formant frequencies, and an arrow below indicates the position of the fiber CF. The spectra are normalized by the square of the mean discharge rate, so that the vertical scale has units of Hz^{-1}.

Fig. 6

Pseudo-perspective representation of normalized band-average power spectra for 0.55-octave CF bands in response to the four fricative stimuli presented at the low level. Each band-average power spectrum is normalized by the square of the mean discharge rate, and is plotted with frequency along the oblique axis, and magnitude along the vertical axis. Spectrum points with an amplitude lower than $0.4\ 10^{-3}\ Hz^{-1}$ are omitted for clarity. The center frequencies of the CF bands are sampled every quarter octave. Horizontal dashed lines mark the positions of the formant frequencies along the frequency

axis, and oblique dashed lines mark the places of the formant frequencies along the CF dimension. The curved dashed line is the locus of points for which frequency is equal to CF.

Fig. 7

Same as Fig. 6 for the high-level stimuli.

Fig. 8

Ratio of the square of the mean discharge rate to the variance of the PST histogram plotted against CF for the four stimuli presented at the high level. Each circle represents the ratio for one auditory-nerve fiber. The continuous lines represent the ratios for the band-average data. The positions of the formant frequencies along the CF dimension are indicated by dashed lines.

Fig. 9

Reciprocal of the Mode of Intervals between Peaks (RMISP) of the autocorrelation function of PST histograms plotted against CF for the four stimuli presented at the high level. Each circle represents the RMISP for one auditory-nerve fiber. The positions of the formant frequencies along the CF dimension are indicated by vertical dashed lines, and horizontal dashed lines are drawn at the ordinates corresponding to intervals of 1/P1 and 1/P2.

Fig. 10

Average Localized Synchronized Measure (ALSM) plotted against center frequency of a narrow Gaussian bandpass filter for the

four stimuli presented at the high level. The positions of the formant frequencies are indicated by dashed lines.

Fig. 11

Same as Fig. 10 for an ALSM based on a cosinusoidal comb filter.

Fig. 1

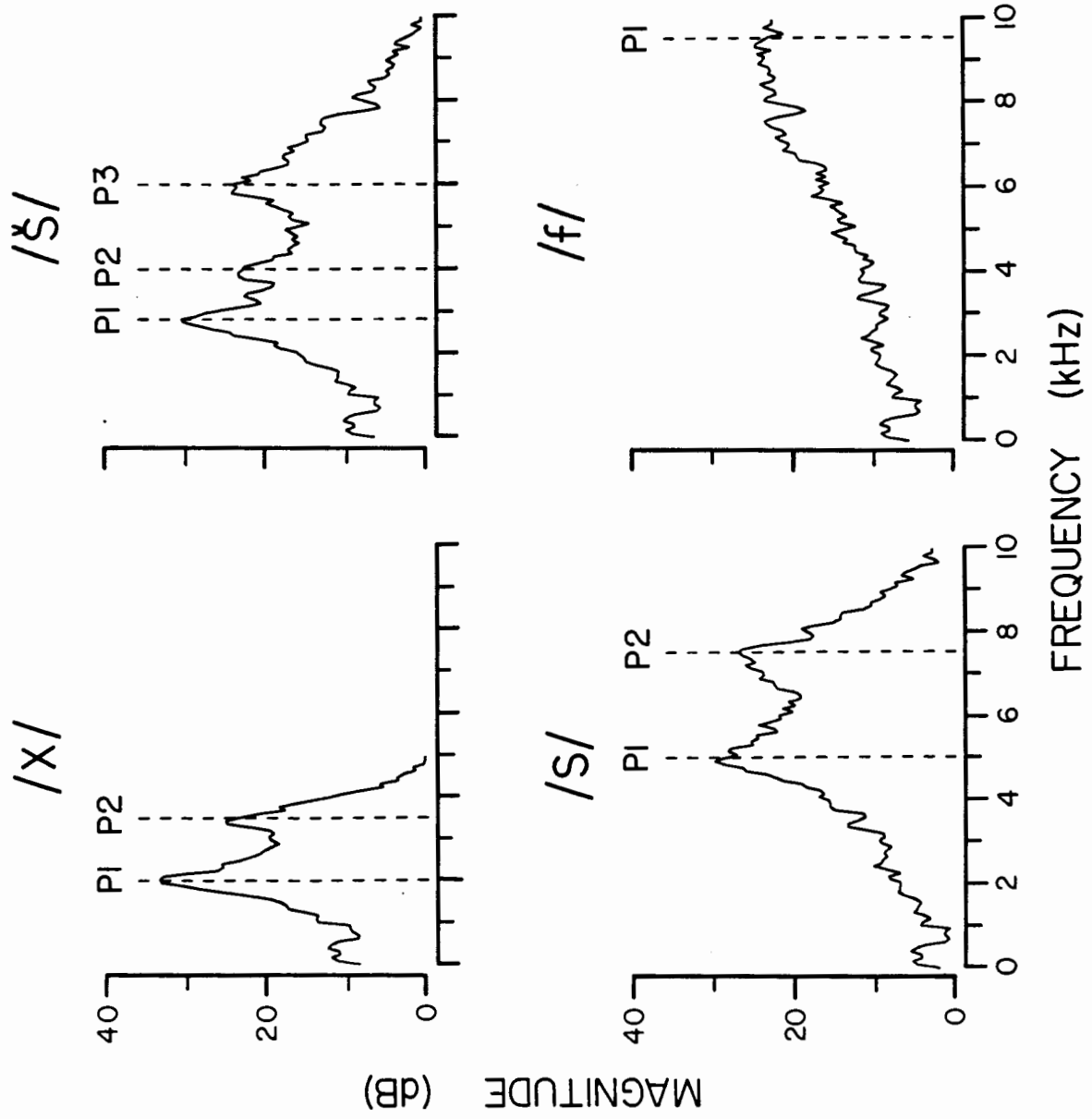


Fig. 2

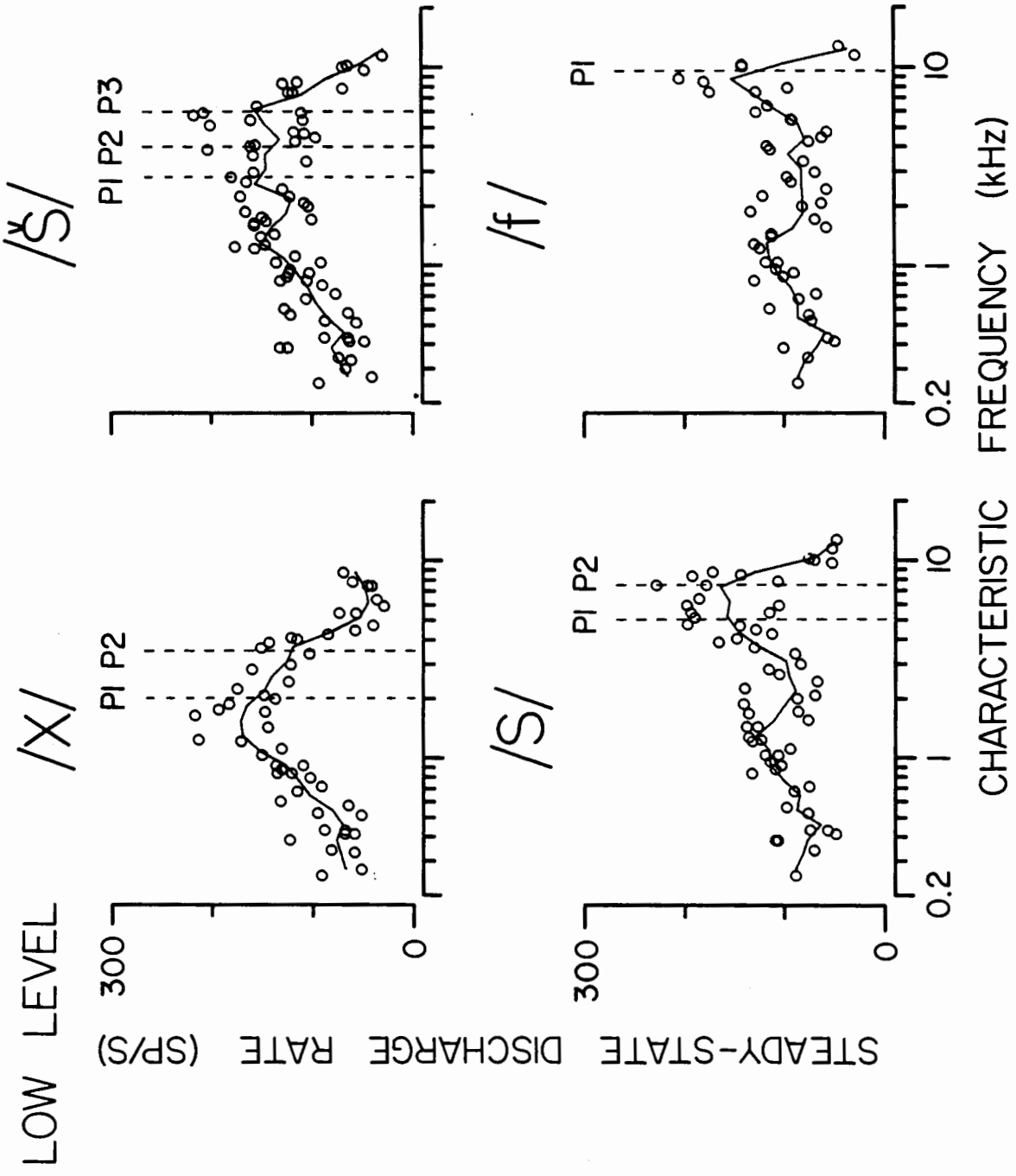

LOW LEVEL
/X/
/S/
/S/
/f/
P1 P2
P1 P2 P3
P1 P2
P1
300
0
0.2
1
10
STEADY-STATE DISCHARGE RATE (SP/S)
CHARACTERISTIC FREQUENCY (kHz)

Fig. 3

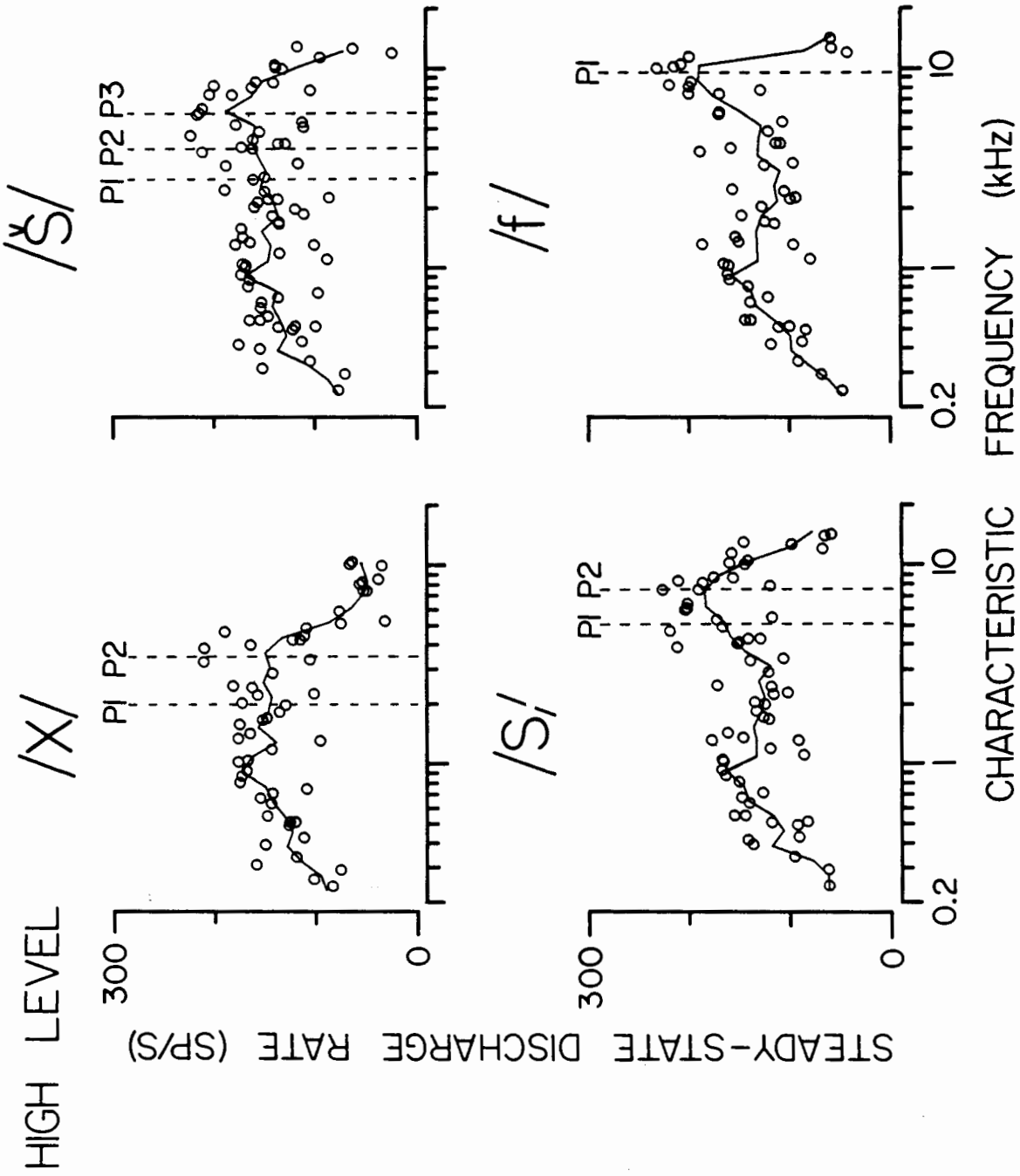

HIGH LEVEL
/X/
/S/
/S'/
/f/
P1 P2
P1 P2 P3
P1
300
0
0.2
1
10
STEADY-STATE DISCHARGE RATE (SP/S)
CHARACTERISTIC FREQUENCY (kHz)

Fig. 4

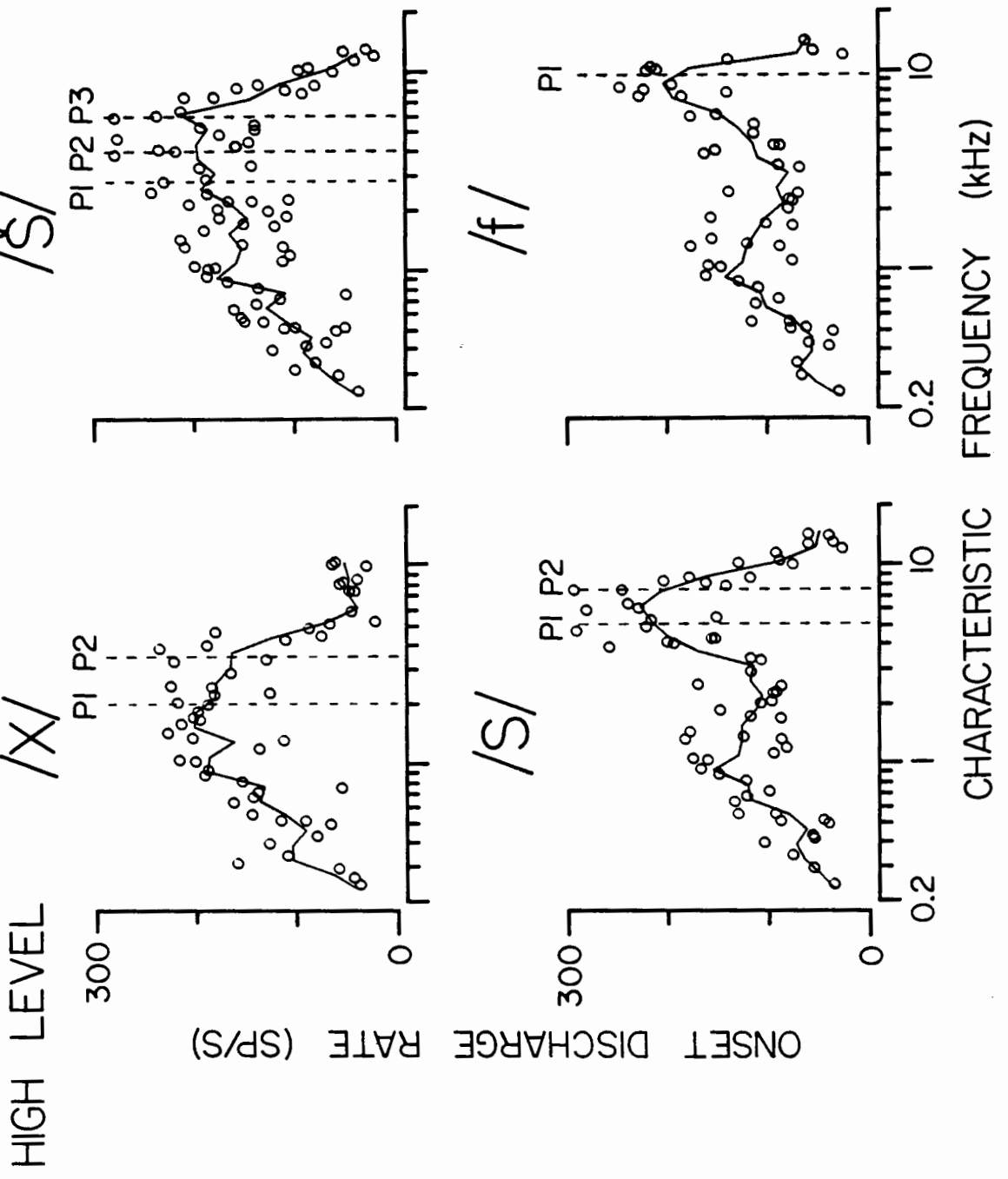

HIGH LEVEL
/ʃ/
/s/
/x/
/f/
P1 P2 P3
P1 P2
P1
P1 P2
300
0
ONSET DISCHARGE RATE (SP/S)
0.2
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 5

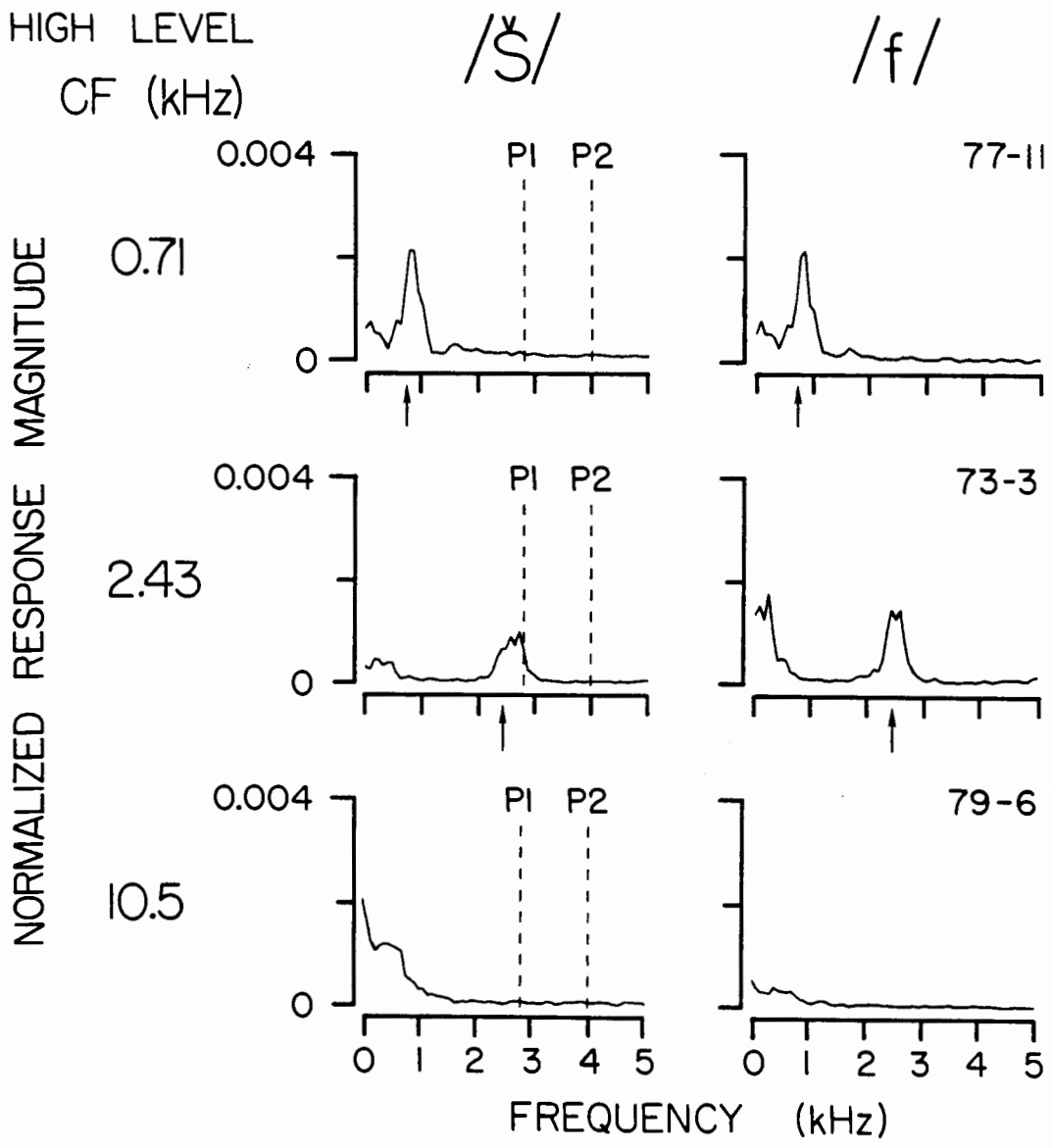

HIGH LEVEL
CF (kHz)
/ʃ/
/f/
0.004
P1
P2
77-11
0.71
0
73-3
2.43
79-6
10.5
NORMALIZED RESPONSE MAGNITUDE
0 1 2 3 4 5
FREQUENCY (kHz)

Fig. 6

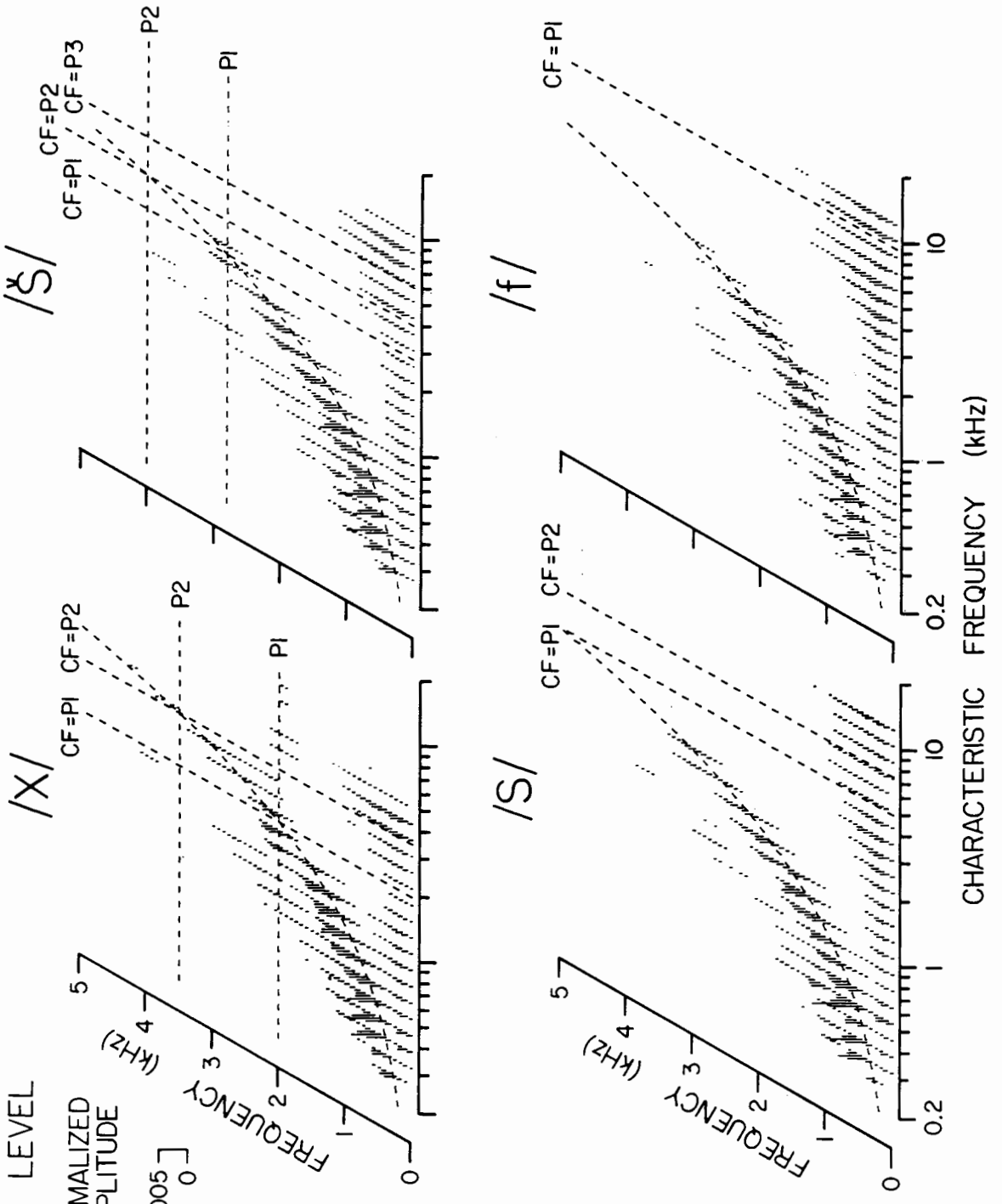

LOW LEVEL
NORMALIZED AMPLITUDE
0.005
0
/X/
/S/
/S/
/f/
CF=P1
CF=P2
CF=P3
P1
P2
FREQUENCY (kHz)
0
1
2
3
4
5
0.2
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 7

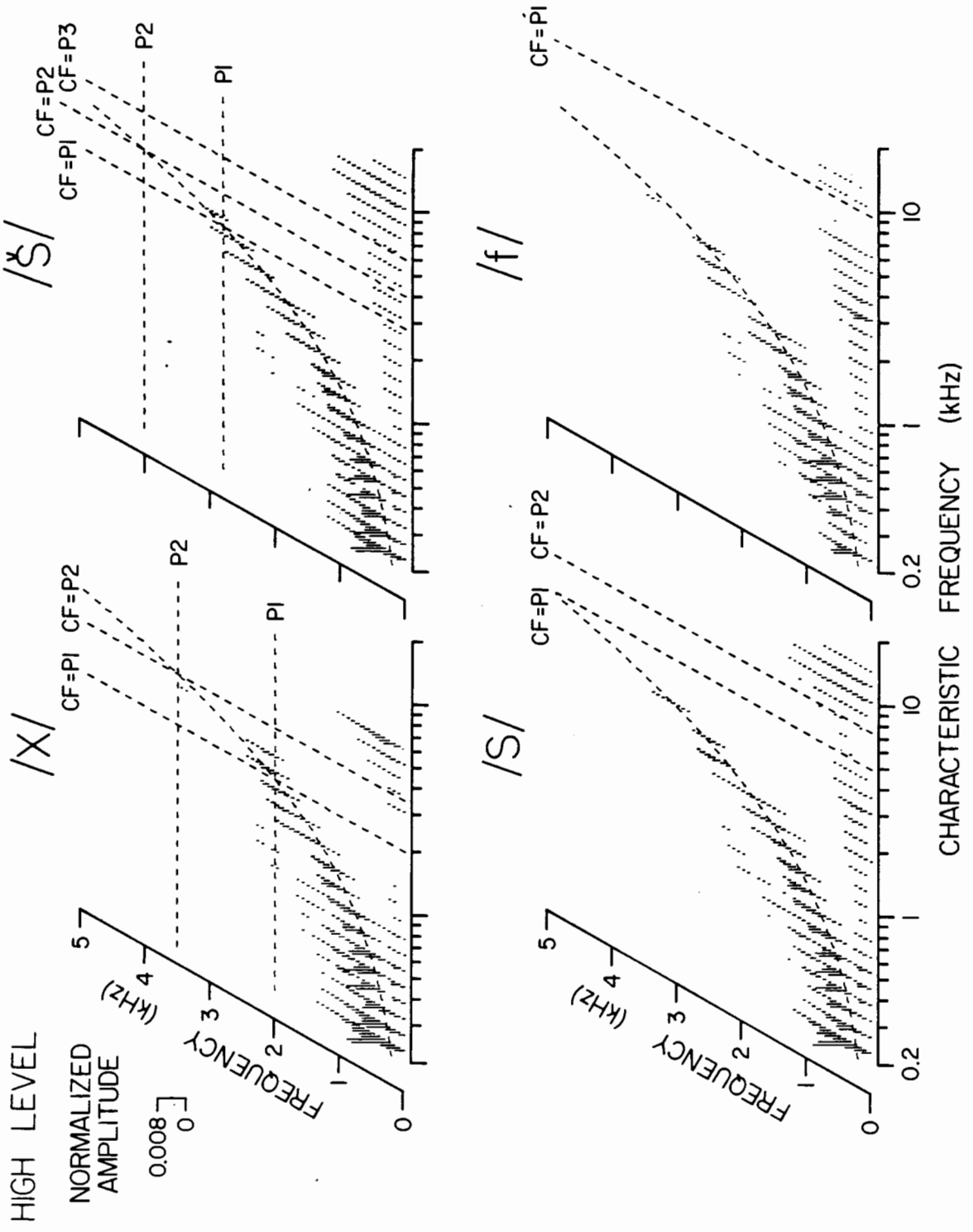

HIGH LEVEL
NORMALIZED AMPLITUDE
0.008
0
/X/
/S/
/S/
/f/
CF=PI
CF=P2
CF=P3
P2
PI
FREQUENCY (kHz)
0
1
2
3
4
5
0.2
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 8

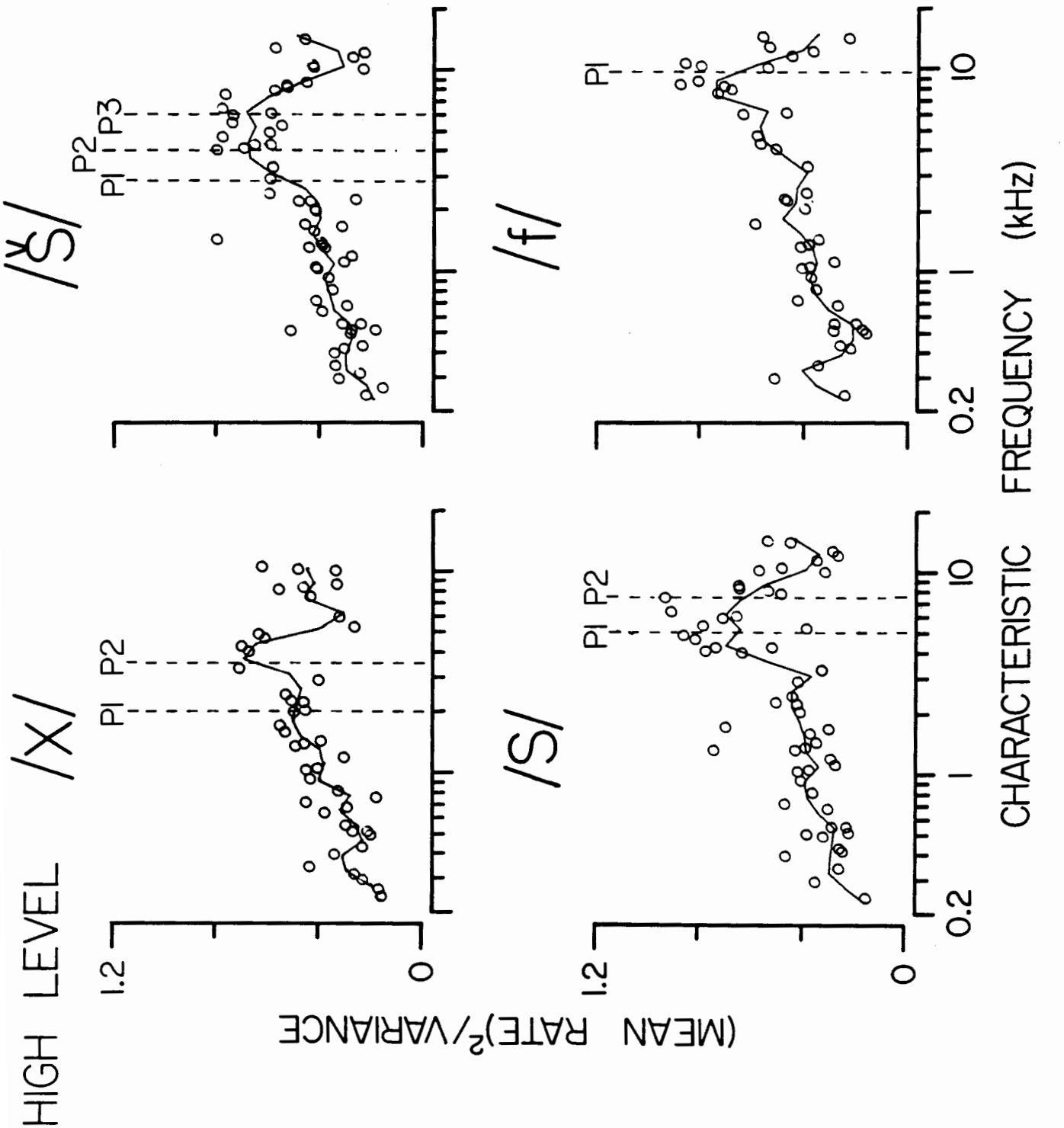

HIGH LEVEL
/x/
/s/
/s/
/f/
P1
P2
P3
1.2
0
0.2
1
10
(MEAN RATE)²/VARIANCE
CHARACTERISTIC FREQUENCY (kHz)

Fig. 9

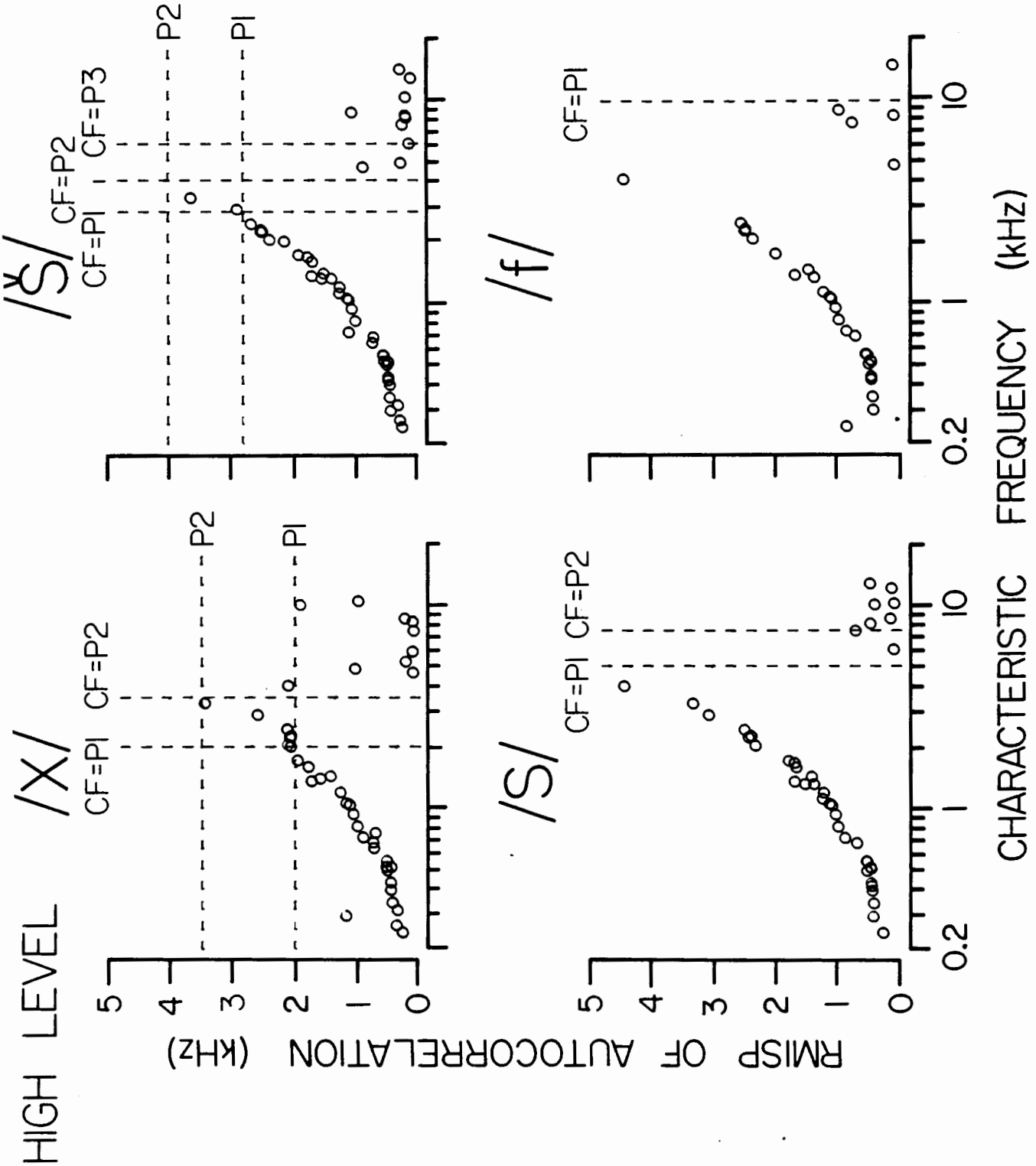

HIGH LEVEL
/X/
/S/
/ʃ/
/f/
/S/
CF=P1
CF=P2
CF=P3
P1
P2
RMISP OF AUTOCORRELATION (kHz)
CHARACTERISTIC FREQUENCY (kHz)
0
1
2
3
4
5
0.2
1
10

Fig. 10

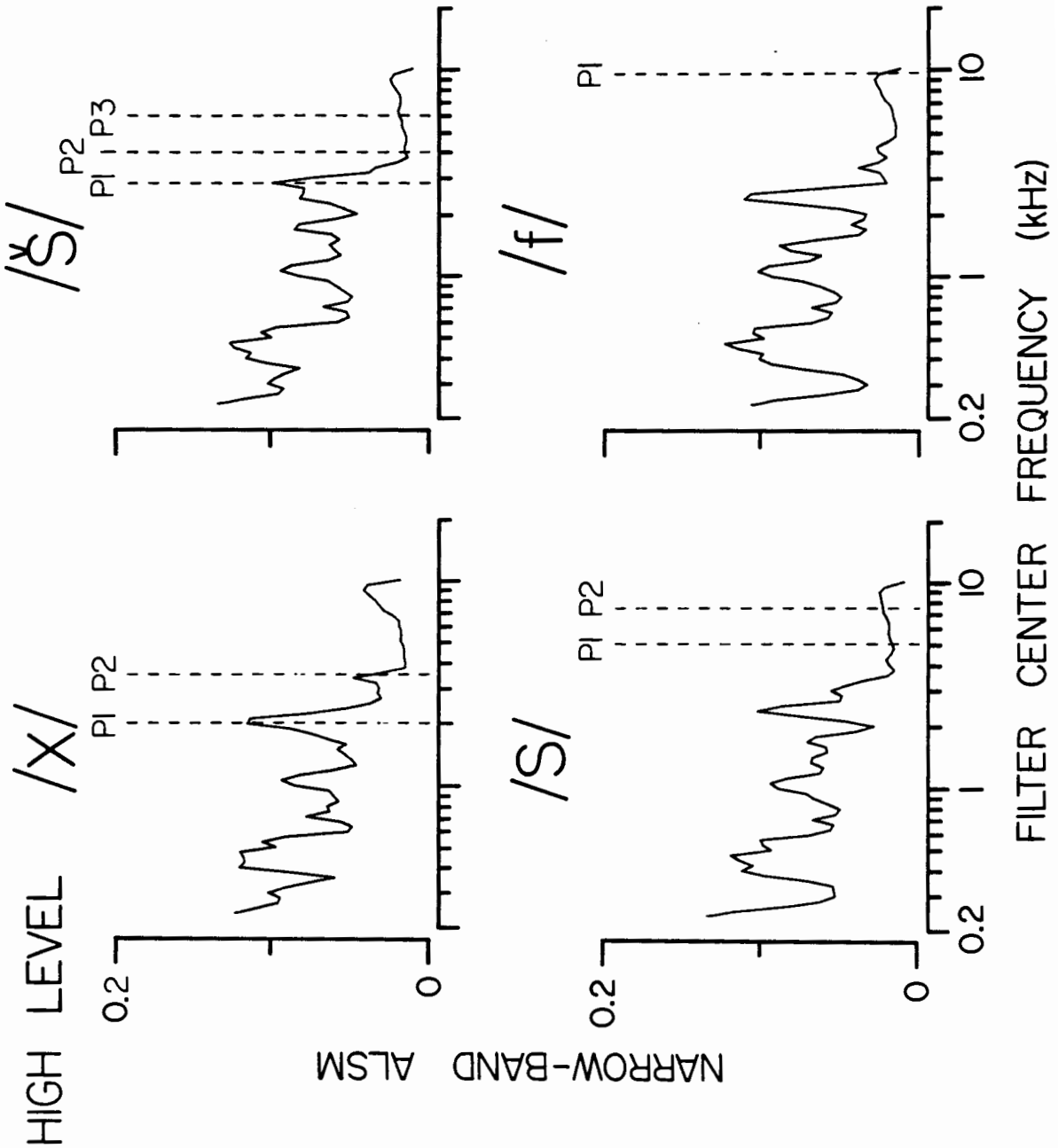
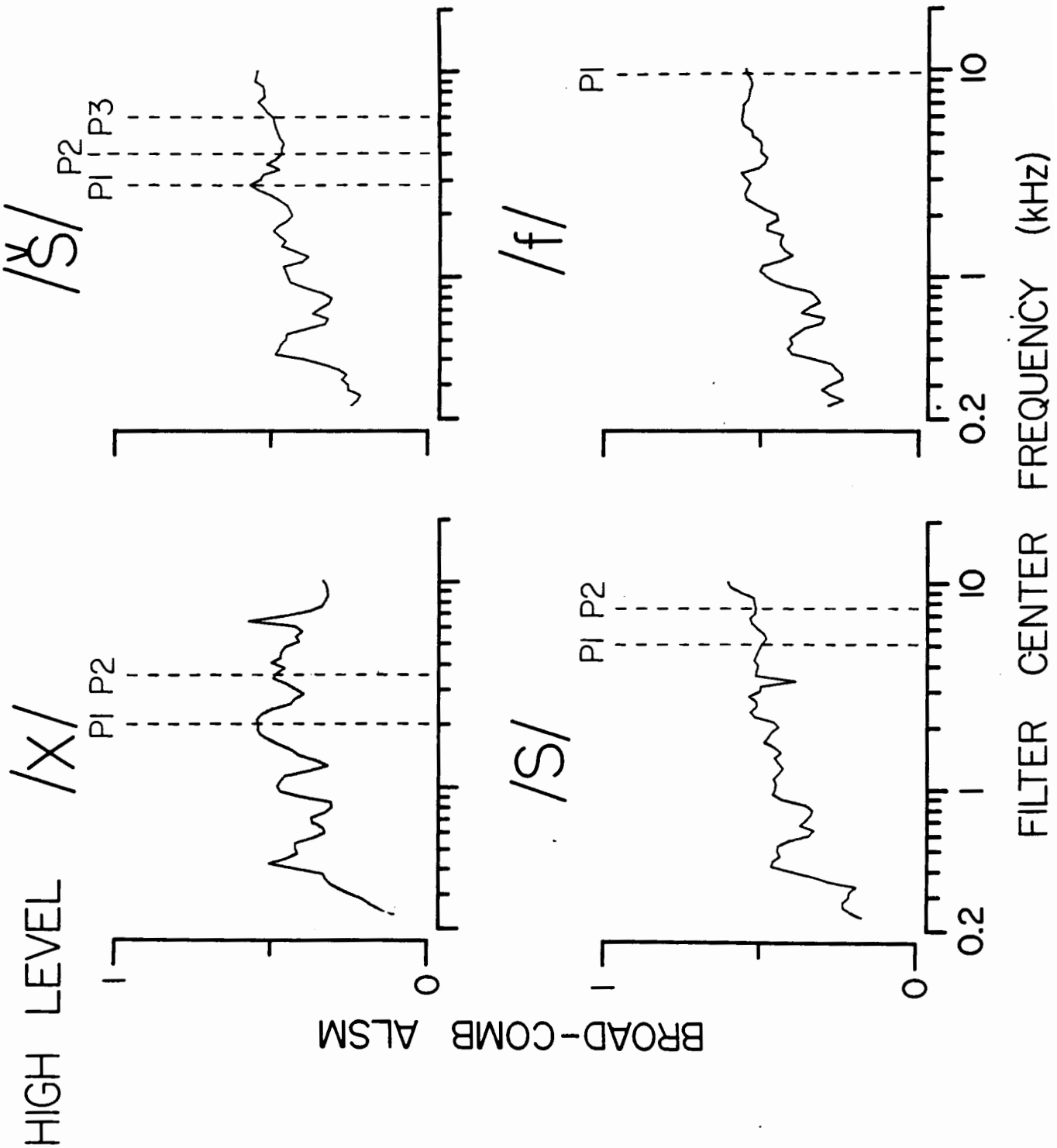

HIGH LEVEL
/X/
/S/
/S/
/f/
P1
P2
P3
0.2
0
1
10
NARROW-BAND ALSM
FILTER CENTER FREQUENCY (kHz)

Fig. 11


HIGH LEVEL
/X/
P1 P2
/S/
P1 P2 P3
/S/
P1 P2
/f/
P1
1
0
0.2
1
10
BROAD-COMB ALSM
FILTER CENTER FREQUENCY (kHz)

CHAPTER III

CODING OF SOUNDS WITH SPEECH-LIKE DYNAMIC CHARACTERISTICS IN THE AUDITORY NERVE

INTRODUCTION

Consonants differ from vowels in that they show rapid changes in amplitude and spectral characteristics. Portions of the speech signal in the vicinity of these rapid changes are rich in information about phonetic distinctions among consonants (Stevens, 1980). For instance stop, nasal and affricate consonants show an abrupt increase in amplitude, whereas fricatives and glides have a more gradual onset (Liberman et al., 1956; Cutting and Rosner, 1974). For stop consonants the abrupt onset is accompanied by a rapid spectral change, whereas the gradual onsets of glides are associated with relatively slow spectral changes. Voicing distinctions for stop and possibly fricative consonants are based on the timing of acoustic events occurring near the rapid spectral changes (Lisker and Abramson, 1964; Stevens, 1980). Nasal consonants differ from stops in that the spectral change is preceded by considerable low-frequency energy (Cooper et al., 1952). Place of articulation distinctions among stop and nasal consonants are based on the spectrum sampled at the onset associated with consonantal release, and on the direction of formant-frequency changes following the onset

(Cooper et al., 1952; Halle et al., 1957; Blumstein and Stevens, 1979). Regions of rapid changes are also important for certain distinctions among fricative consonants (Heinz and Stevens, 1961; Stevens, 1981), and in some cases spectral changes influence vowel identification (Lindblom and Studdert-Kennedy, 1967; Strange et al., 1976).

In spite of their heavy functional load, dynamic characteristics of speech sounds have received very little attention from auditory physiologists. Experiments with simple acoustic stimuli suggest that the responses of auditory-nerve fibers are not the same for rapid changes in amplitude and spectrum as for steady-state stimuli (Kiang et al., 1965; Smith and Zwislocki, 1975; Møller, 1976; Smith, 1979; Delgutte, 1980; Smith and Brachman, 1980a). The present paper is based on two sets of speech-like stimuli whose dynamic characteristics differ along dimensions that are important for phonetic distinctions. The first set of stimuli consists of the fricative /$\overset{\vee}{s}$/ and the affricate /$\overset{\vee}{c}$/, which have similar spectra but differ in the duration of the rise in amplitude at the onset of the stimulus. Stimuli with rise times shorter than about 40 ms are heard as /$\overset{\vee}{c}$/, whereas stimuli with longer rise times are heard as /$\overset{\vee}{s}$/ (Cutting and Rosner, 1974). A goal of these experiments is to find correlates of this phonetic distinction in the discharge patterns of auditory nerve which would be invariant with respect to stimulus level.

The second set of stimuli consists of consonant-vowel or vowel-consonant-vowel sounds which all have changes in formant frequencies appropriate for the syllable /da/. By manipulating the spectral characteristics of the context preceding the /da/ formant transitions, it is possible to elicit the perception of the nasal consonant /n/, the fricative consonants /ʃ/ and /s/, and fricatives followed by the stop /t/ (Cooper et al., 1952). Studies with simple stimuli suggest that the responses of auditory-nerve fibers to a stimulus depends on the acoustic signal during the 200-odd ms preceding that stimulus (Smith, 1977; Harris and Dallos, 1979; Delgutte, 1980). Some of the experiments reported in this paper were designed to study how fiber responses to the /da/-like formant transitions depend on the context and how these context dependencies may provide information for phonetic distinctions among consonants. Because previous studies have shown that both average discharge rate and fine time patterns of discharge are important for phonetic distinctions (Kiang and Moxon, 1974; Sachs and Young, 1979; Young and Sachs, 1979; Chapters I and II), the effect of context is studied for both types of response measures.

I METHODS

A. Stimuli

The bottom panels of Fig. 1 show the waveforms of the /ʃ/ and /ʧ/ stimuli. The two stimuli were generated by

passing the same 200-ms broadband noise waveform through three bandpass filters modelling the resonances of the vocal tract. The bandpass filters are implemented as described in Chapter II. Their center frequencies (formant frequencies) P1, P2 and P3 are 2.8, 4.0 and 6.0 kHz respectively, so that the stimuli have most of their energy between 2 and 7 kHz. The waveform of the filtered noise is then multipled by a 200-ms trapezoidal weighting function. The stimuli differ in that the rise-fall time of the weighting function is 75 ms for /$\overset{\curvearrowright}{s}$/ and 10 ms for /$\overset{\curvearrowright}{c}$/.

Figure 2 shows the spectrograms of 10 stimuli that have da-like changes in spectral characteristics. These stimuli were generated by a speech synthesizer described in Appendix A. The simplest stimulus is /da/, which consists of a 50-ms interval during which the three formant frequencies vary linearly with time, followed by a 150-ms interval during which the formant frequencies remain at values appropriate for the vowel /a/. Specifically, F1 rises from 0.45 to 0.7 kHz, F2 falls from 1.4 to 1.2 kHz, and F3 falls from 2.8 to 2.6 kHz. The fundamental frequency is 125 Hz during the first 120 ms, and then falls gradually to 110 Hz.

All the other stimuli include a 50-ms interval during which the formant frequencies vary as in /da/, and a steady vowel segment /a/. They differ in that the /da/-like formant transitions are preceded by different contexts. The characteristics of the nine contexts are listed in Table A-I.

For /ada/, the context is a 100-ms segment with formant frequencies appropriate for /a/, followed by 40-ms formant transitions and a 100-ms silence. This silent interval is essential for hearing the sound /da/ in /ada/. The /ida/ and /uda/ stimuli are similar to /ada/, except for the values of the formant frequencies during the initial vowel. The last stimulus on the top row of Fig. 2 is also heard as /da/, but it differs from the basic /da/ in that the formant transitions are preceded by a 14-ms burst of noise having most of its energy between 4 and 8 kHz. In English, /d/ sounds preceding a stressed vowel or following a pause have a prominent burst, but there are contexts in which the /d/ burst is weak, so that a /da/ with no burst is also acceptable (Zue, 1976).

The five stimuli on the bottom row of Fig. 2 do not evoke a /da/ percept, even though the formant frequencies during the transition interval are the same as in the top row. For /na/ the context is a 100-ms periodic nasal murmur consisting mostly of frequency components below 1.5 kHz. To make the stimulus sound like a /na/, it was necessary to modify somewhat the formant amplitudes during the transition interval. For /ʃa/, the context is a 150-ms fricative noise with significant energy between 2 and 7 kHz. To provide good perceptual continuity, there is a brief interval during which voice and noise excitations are superimposed, so that the amplitude of F3 during the transitions is somewhat higher than in /da/. The context for the /sa/ stimulus is similar to the

one of /ʏša/, except it has most of its energy between 4 and 8 kHz. The contexts of the /ʏšta/ and /sta/ stimuli begin with the same 150-ms fricative noise as /ʏša/ and /sa/ respectively, but the noise is separated from the formant transitions by a 110-ms silent interval, so that the stop consonant /t/ is heard.

The left panels of Fig. 3 show the waveforms of the /da/, /ada/, /na/ and /ʏša/ stimuli. For all stimuli, there is an increase in amplitude during the interval of formant transitions indicated by dotted lines. The amplitude at the beginning of the transitions is somewhat higher for /na/. The waveforms of these stimuli during the formant transitions are shown in greater detail in the left side of Fig. 4. The magnitude spectra of these waveform segments are shown in the right side. The spectral envelope shows clear peaks near the first three formant frequencies at 0.5, 1.3 and 2.7 kHz. The greater amplitude of the third formant for /ʏša/ is apparent.

B. Experimental procedures and data processing

The experimental results to be reported are based on recordings from single auditory-nerve fibers in anesthetized cats. The preparation of the animals, the stimulus generation system, and the recording procedures are as described in Chapter I. Recordings were restricted to the most sensitive population of auditory-nerve fibers, those with spontaneous discharge rates greater than 18 spikes/s (Liberman, 1978).

Stimulus levels for /ʃ/ and /ʧ/ were set as described in Chapter II. The levels of the stimuli with /da/-like formant transitions were set relative to the average of the transfer characteristics of the acoustic system over the frequency range 0.1-5 kHz. Because the frequency dependence of the transfer characteristics varies somewhat from animal to animal, the relative stimulus levels vary from one experiment to another. However, for the 27 cats that were used in this study, the mean of the magnitude of the transfer ratio was nearly flat, and the standard deviation of the interanimal variations did not exceed 2 dB for frequencies below 8 kHz.

The stimuli were presented repeatedly for the computation of post-stimulus time (PST) histograms (Gerstein and Kiang, 1960). In order to leave the fibers sufficient time for recovery, the duration of the silent interval between each presentation was always at least twice the duration of the stimulus. Stimulus levels for the /da/-like stimuli were such that the central portion of the vowel /a/ would be at 60 or 75 dB SPL. The /ʃ/ and /ʧ/ stimuli were presented at 45 and 60 dB SPL, which are appropriate levels for speech in which the vowels would be at 60 and 75 dB SPL respectively.

Processing of the responses to /ʃ/ and /ʧ/ are based on PST histograms computed with a 0.25-ms bin width from 200 to 400 presentations. These histograms were used to estimate discharge rate in two different intervals. The "steady-state" rate was measured by weighting the contents of the histogram

bins by a trapezoidal window centered 100 ms after the onset of the stimulus, and with a central duration of 50 ms and a total duration of 150 ms (effective duration 100 ms). The "onset" rate was measured using a window of value 1 from 0 to 10 ms, then decreasing linearly to reach 0 at 50 ms (effective width 30 ms). To obtain data at regularly-spaced samples along the log CF dimension, the PST histograms of all fibers whose CF lies in a 0.5-octave band of frequencies were averaged, and such characteristic-frequency bands were sampled every half octave. Band averaging was also applied to onset discharge rate and steady-state rate, using the windows specified in the figure captions.

PST histograms in response to the /da/-like stimuli were computed with a bin width of 0.1 ms from 75 to 300 presentations. Typical PST histograms for a fiber in response to the /da/, /ada/, /na/ and /ša/ stimuli are shown in the right side of Fig. 3. The dotted lines mark a 51.2-ms "test interval" that corresponds to the formant transitions in the stimulus. In order to take into account the CF variations in latencies of auditory-nerve fiber responses, the onset of the test interval for each CF is equal to the onset of the formant transitions delayed by the click latency at that CF minus 0.5 ms, as given by Kiang et al. (1965). PST histograms during the test interval are shown with fine time resolution in the left side of Fig. 4. The bin contents during the test interval of each PST histogram were then weighted by the half

raised-cosine window shown at the top right of Fig. 3. This window has an effective duration of 25.6 msec, and has its center of gravity 15 ms after the beginning of the test interval. The 51.2-ms windowed histograms were used to compute the average discharge rate during the test interval and the correlation index C_i between the windowed histogram for stimulus i and the windowed histogram for /da/:

$$C_i = \sum_{0<t_n<51.2} H_{da}(t_n)\, H_i(t_n) \,/\sqrt{S_{da}\, S_i} \qquad (1)$$

where t_n is time in ms, H_i is the weighted histogram for stimulus i, and S_i is given by:

$$S_i = \sum_{0<t_n<51.2} H_i^2(t_n) \qquad (2)$$

The correlation index, a number varying between 0 and 1, is a rate-independent measure of similarity between the response to /da/ and the response to stimulus i during the test interval.

Discrete Fourier transforms of the windowed histograms were computed with a frequency resolution of about 20 Hz (Oppenheim and Schafer, 1975). The magnitudes of the Fourier transforms of the histograms of Fig. 4 are shown in the right side of Fig. 4. The Fourier transforms were used to compute an Average Localized Synchronized Measure (ALSM), as described in Chapter I. This measure, similar to the ALSR proposed by Young and Sachs (1979), is obtained by filtering auditory-nerve response patterns by a narrow bandpass filter whose center frequency f_c is near the fiber characteristic frequency. Specifically, the result A_j of the filtering operation for fiber j was computed from:

$$A_j = \sum_{0<f_k<5} P_j(f_k)\, H(f_k) \qquad (3)$$

where f_k is frequency in kHz, H(f) is the transfer function of a Gaussian bandpass filter, and $P_j(f_k)$ is the magnitude square of the response spectrum for fiber j normalized by the mean square discharge rate during the test interval. The filter transfer function is given by:

$$H(f) = \exp -\pi[(f-f_c)/b_c]^2 \qquad (4)$$

where the filter bandwidth b_c is set to 0.116 f_c (1/6 octave). The ALSM was computed by averaging the results of the filtering operations for all fibers whose CF is in a narrow band of frequencies centered at f_c, using a trapezoidal weighting window with a central width of 1/6 octave and a total width of 1/2 octave. This computation was made for values of f_c ranging from 0.1 to 5 kHz, in 1/12-octave intervals.

II RESULTS

A. Abrupt and gradual onsets

The top panels in Fig. 1 show the response patterns of an auditory-nerve fiber for the /ʃ/ and /ʧ/ stimuli presented at 45 dB SPL. The characteristic frequency of the fiber is in the frequency region where the stimuli have most of their energy. Because the stimuli are pseudo-random noise, the response patterns have a complex fine time structure. The envelope of the response pattern for the /ʧ/ stimulus shows a sharp peak in discharge rate immediately after the onset of

the stimulus, whereas the envelope of the response pattern for /ʃ/ has a broader peak occurring later after the onset. The middle panels show the response patterns when the stimulus level is raised to 60 dB SPL. Though the fine time patterns of discharge remain somewhat different for the two stimuli, the response patterns for both stimuli show a prominent peak in discharge rate at the onset. Thus, for the fiber of Fig. 1, the distinction between /ʃ/ and /ʧ/ is not as clear at 60 dB as at 45 dB. However, at 60 dB SPL, other fibers have response patterns with clear cues for distinguishing the two stimuli.

Figure 5 shows band-average PST histograms for 0.5-octave CF bands in response to the /ʃ/ and /ʧ/ stimuli presented at 45 and 60 dB SPL. At the 45-dB level, fibers whose CF is in the frequency region where the stimuli have most of their energy show clear cues for the distinction between /ʃ/ and /ʧ/ in their response patterns. In contrast, fibers whose CF is far from that region respond with a low discharge rate to both stimuli, and do not show a prominent peak in their response pattern at the onset of either stimulus. At the 60-dB level, the response patterns of fibers whose CF is in the frequency region where the stimuli have most of their energy, show a sharp peak at the onset of the stimulus for both /ʃ/ and /ʧ/, though some small differences remain. However, for fibers whose CF is on either side of that frequency region, the response pattern to /ʧ/ shows a

more prominent peak in discharge rate than the response pattern to /ʏ̯š/.

Thus, for both stimulus levels, there is a population of auditory-nerve fibers whose response patterns have clear information about the distinction between /ʏ̯š/ and /ʏ̯č/. However the population of fibers with the strongest cues changes with stimulus level, so that it may not be advantageous to distinguish the two stimuli on the basis of a fixed set of fibers within a restricted CF region. Instead, it might be preferable to scan the entire array of auditory-nerve fibers and focus on the CF regions that provide unequivocal cues.

Figure 6 shows the ratio of onset rate to steady-state rate plotted against CF for the /ʏ̯š/ and /ʏ̯č/ stimuli at both stimulus levels. Because the steady-state rate is about the same for both stimuli, this ratio is a measure of the prominence of the peak in discharge rate at the onset of the stimulus. For both stimulus levels, the ratio of discharge rate in response to /ʏ̯č/ is at least equal to 1 throughout the range of CF's, though the range over which it is greater than 1 is wider at 60 dB. In response to /ʏ̯š/, there is always a CF region where the ratio is less than 1, indicating that the peak in discharge rate occurs more than 15-30 ms after the onset of the stimulus or that there is no clear peak. Thus, in spite of considerable variations in the response patterns with stimulus level, the profile formed by

the ratio of onset rate to steady-state rate shows features which are characteristic of /ʃ/ and /ʧ/, at least for the two stimulus levels that have been investigated.

A second possible approach to the search for invariant response properties is illustrated by Fig. 7, which shows the "grand average" response patterns for the /ʃ/ and /ʧ/ stimuli at both stimulus levels. This grand average is obtained by adding the response patterns for all the CF bands in Fig. 5. For both stimulus levels, there is a rapid decrease in discharge rate following the peak at the onset of the /ʧ/ stimulus, whereas for /ʃ/ the maximum in discharge rate is more like a plateau.

B. Effect of context on the representation of spectral changes

1. Short-time average discharge rate

For all the stimuli with /da/-like formant transitions, the acoustic events in the vicinity of the transitions are of great phonetic importance. Variations in short-time average discharge rate near the test interval are likely to contain information about these events. The right side of Fig. 3 shows the response patterns of an auditory-nerve fiber for the /da/, /ada/, /na/ and /ʃa/ stimuli presented at 75 dB SPL. The CF of the fiber is near the second formant frequency. For the /da/ stimulus there is a prominent peak in discharge rate during the test interval.

For the /ada/ stimulus, the fiber responds with a peak in discharge rate at the onset of the stimulus, adapts during the initial /a/ segment, and then recovers partially during the silence so that there is a second peak in discharge rate during the test interval. However, the size of the peak during the test interval is somewhat lower than for the /da/ stimulus. For the /na/ stimulus, the fiber responds strongly to the /n/ segment and shows no peak in discharge rate during the test interval. For the /$\overset{v}{s}$a/ stimulus, the fiber responds moderately to the high-frequency /$\overset{v}{s}$/ segment, and shows a small peak in discharge rate during the test interval. Thus, discharge rate during the test interval depends on the context even though the formant frequencies during the test interval are the same for all stimuli. In spite of this context-dependence, the formant transitions are marked by a peak in the response patterns of this fiber for all stimuli except /na/.

Figure 8 shows profiles of discharge rate during the test interval plotted against characteristic frequency for the /da/, /ada/, /na/ and /$\overset{v}{s}$a/ stimuli presented at 75 dB SPL. For /da/, the profile is essentially lowpass, as is typical for voiced speech sounds at high stimulus levels (Sachs and Young, 1979), though small peaks near the first two formant frequencies can be seen. The profiles for the other stimuli differ considerably from the profile for /da/. The data are plotted in Fig. 9 in a way that makes these differences apparent.

The top left panel of Fig. 9 shows the ratio of the discharge rate during the test interval for /ada/ to the rate during the test interval of /da/ plotted against characteristic frequency for the 75-dB condition. The other panels show comparable ratios for the other stimuli. For the /ada/, /ida/ and /uda/ stimuli, the ratio of discharge rates is slightly below 1 throughout the range of CF's, indicating that the profile of discharge rate against CF during the transition interval has about the same shape for /da/ and these stimuli. In contrast, for the other stimuli, the ratio of discharge rates varies strongly with CF, indicating that the shape of the profile of discharge rate against CF is not the same as for /da/. Thus, the profile of discharge rate against CF for the test interval retains its shape for the stimuli in which a /da/ with no burst is heard, whereas it is strongly different for the other stimuli, even though the formant frequencies are the same in all cases. These differences in the profile of discharge rate during the test interval provide information about the spectrum of the context.

For the /na/ stimulus, in which the context has mostly low-frequency components, the ratio of discharge rates is small for low-CF fibers and approaches one for high-CF fibers. The ratio is smallest near 0.3 and 1.4 kHz, which are the frequencies of the first and second formant of /n/, respectively. In contrast, for the /ša/ stimulus in which the

context has mostly high-frequency components, the ratio of discharge rates is small only for high-CF fibers. For the /sa/ stimulus, the decrease in discharge rate due to the context is even more restricted to high-CF fibers. For */da/, in which the burst has about the same spectrum as /s/, the decrease in discharge rate affects almost the same fibers as for /sa/, but the ratio is larger, probably because the /d/ burst is shorter and less intense than the /s/ noise. For /ʃta/ and /sta/, the greatest decrease in discharge rate also occurs for high-CF fibers, but it is not as great as for /ʃa/ and /sa/, probably because the units recover during the 110-ms silent interval. In all cases, the effect of the context is strongest for fibers whose CF is in the frequency region where the context has considerable energy.

Figure 10 shows that these general trends apply to the 60 dB stimuli, though the CF regions of maximum effect are narrower. Another difference is that the ratio of discharge rates is not as flat as at 75 dB for the /ada/, /ida/ and /uda/ stimuli. For these three stimuli, the profile shows small dips for CF's near the first and second formant frequencies of the initial vowel. At this stimulus level, the profile of average discharge against CF during the /a/, /i/ and /u/ segments has peaks near the formant frequencies, whereas such peaks are not apparent at 75 dB.

The ratio plotted in Fig. 9 and 10 demonstrates that discharge rate during the test interval depends systematically

on the spectrum of the context. From the point of view of central processing, this ratio is not a good response measure because it depends on responses to two different stimuli, while speech syllables can usually be identified without external reference to other speech sounds. Figure 11 plots the ratio of discharge rate during the test interval to the rate during the steady /a/ segment following the transitions for the 9 stimuli presented at 75 dB SPL. Because discharge rate during /a/ is about the same for all stimuli, this ratio is a measure of the prominence of the peak in discharge rate during the test interval. The ratio for /da/ (shown by dashed lines) is greater than 1 over a wide range of CF's and shows broad peaks near the average values of the first and second formant frequencies during the transitions. The ratio of discharge rates for the other stimuli roughly follows the ratio for /da/ in frequency regions where the context has little energy, but is considerably decreased in frequency regions where the context has most of its energy. With the exception of /na/, the ratio is considerably greater than 1 in at least some CF region, indicating that a peak in discharge rate is present in the response patterns of many fibers. For /na/, the ratio is never much larger than 1, but there is often an increase in discharge rate during the test interval (not shown), so that the formant transitions are also marked in the short-time average variations in discharge rate. At 60 dB SPL, context-dependent changes in the ratio of rate during the test interval to rate during /a/ follow the same pattern

as at 75 dB, but the decrease in rate with context is generally lower (not shown). Thus, for all stimuli and both stimulus levels, there is a peak in discharge rate or an increase in discharge rate in the discharge patterns of auditory-nerve fibers whose CF's are in frequency regions where the interval of formant transitions has considerably more energy than the context.

In this discussion, the burst of the */da/ stimulus was considered as a context affecting the response to the formant transitions. However, the auditory system may not process separately the burst and transition information about the identity of stop consonants. Instead, it has been proposed that stop-consonant identity depends primarily on the gross shape of the spectrum during the interval following the consonantal release (Blumstein and Stevens, 1979). This interval would include the burst and part of the formant transitions. To obtain a response measure reflecting this hypothesis, the 51.2-ms window used to estimate discharge rate during the test interval was shifted by 14 ms so that its beginning would coincide with the onset of the /d/ burst in */da/. Figure 12 shows discharge rate at the consonantal release plotted against CF for the */da/ and /da/ stimuli presented at 60 and 75 dB SPL. Discharge rate for */da/ is about 2/3 of the rate for the test interval of /da/ for low-CF fibers, but, at least at 75 dB, it is larger than the rate during the test interval of /da/ for high-CF fibers. Thus the

shape of the profile of discharge rate against CF at the onset of a /da/ with a burst differs considerably from the profile at the onset of a /da/ stimulus that has formant transitions but no burst, so that the rate profile at consonantal release is not a good candidate for an invariant cue to the identity of stop consonants.

2. Fine time patterns of discharge

Because fine time patterns of discharge are rich in information about the spectra of low-frequency stimuli (Young and Sachs, 1979), the effects of the context on fine time patterns during the test interval were investigated. The upper left panel of Fig. 13 shows the correlation index between the response to the test interval of /ada/ and the response to the test interval of /da/ plotted against CF for the 75 dB condition. The other panels show the correlation indices between /da/ and the other stimuli. For all stimuli, the correlation index remains close to 1 except in the CF regions where discharge rate during the test interval is considerably decreased by the context. Correlation indices below 0.5 are found only in the CF region near 5 kHz for /ʃa/, and above 5 kHz for /sa/ and */da/. For the 60-dB condition, low values of the correlation index are also unusual (not shown).

To elucidate the nature of the differences in fine time patterns of discharge associated with low correlation

indices, the spectra of the response during the test intervals were compared for /da/ and the other stimuli. The right side of Fig. 4 shows discrete Fourier transforms of the PST histogram during the test interval for one auditory-nerve fiber in response to the /da/, /ada/, /na/ and /ša/ stimuli presented at 75 dB SPL. The characteristic frequency of the fiber is near the second formant frequency. For the four stimuli, the spectra shows clear peaks near the second formant frequency and its harmonics, though the relative amplitudes of the various components differ somewhat. For all stimuli except /na/, there are also small peaks at the 125-Hz fundamental frequency and its low-frequency harmonics. In general, differences between response spectra for /da/ and the other stimuli were limited to variations in the amplitude of different components, without changes in the identity of the largest components.

Because the speech processing schemes that were proposed in chapter I for the estimation of formant frequencies depend primarily on the most intense components of the response spectra, one would expect that these schemes would not be strongly affected by the context preceding the test interval. For instance, Figure 14 shows an Average Localized Synchronized Measure (ALSM) plotted against the center frequency of a 1/6-octave bandpass filter for the test intervals of /da/ and the other 9 stimuli presented at 75 dB SPL. For /da/, the profile has major peaks near the first and

second formant frequencies, and smaller peaks at the frequencies of the first three harmonics of the 125-Hz fundamental frequency. Whenever there are enough data to cover the relevant frequency regions, the same peaks are apparent in the profiles for the other nine stimuli, though the relative amplitudes of the peaks associated with the formant frequencies vary somewhat. For /$\underset{\vee}{s}$a/, there is an additional peak near the third formant frequency, probably because of the higher amplitude of this formant in the stimulus. The ALSM's plotted in Fig. 13 are normalized by the mean square discharge rate. The shapes of the profiles for unnormalized ALSM's are more dependent on the context because of the large variations in discharge rate shown in Fig. 9, but peaks at the first two formant frequencies remain apparent in all cases. The identity of the major peaks in the profiles for ALSM's based on comb filtering schemes are also roughly invariant to changes in the context.

III DISCUSSION

A. Short-term adaptation and responses to speech stimuli

Many of the effects observed in this study, both for the /$\underset{\vee}{s}$/ and /$\underset{\vee}{c}$/ stimuli, and for the stimuli with /da/-like formant transitions, can be related to short-term adaptation of auditory-nerve fibers (Kiang et al., 1965; Smith, 1979). Smith and Brachman (1980a) studied the response of auditory-nerve fibers to tone bursts with varying rise times.

The shape of the response patterns, and the finding that the peak in discharge is more prominent and occurs earlier at high stimulus levels are similar to our observations with the /ʃ/ and /ʧ/ stimuli. Smith and Brachman interpreted these results in terms of a saturating nonlinearity followed by an adaptation mechanism. According to this interpretation, the saturation effectively decreases the stimulus rise time at high intensities, so that the response patterns for long rise times become more similar to response patterns for short rise times.

The lower discharge rate during the test interval for fibers in CF regions where the context has considerable energy resembles the decreased response to a test tone occurring after an adapting tone burst (Smith, 1977; Harris and Dallos, 1979). The decrease in discharge rate for the test tone is greater when the adapting stimulus has a longer duration or a greater intensity, when the adapting frequency is close to the fiber CF, and when the silent interval between adapting stimulus and test stimulus is shorter. All these trends are apparent in the data of Fig. 9 and 10. The relatively small differences in fine time patterns of discharge during the test interval for stimuli with different contexts (Fig. 13) is consistent with the fact that synchronization index for low-frequency tone bursts remains approximately constant during short-term adaptation (Delgutte, 1980).

These similarities suggest that the decrease in discharge rate during the test interval for the stimuli with /da/-like formant transitions are primarily due to short-term adaptation by the context. For the /na/ and /ʃa/ stimuli, this conclusion needs to be qualified because the differences in stimulus waveform during the formant transitions between /da/ and these stimuli make it impossible to attribute with absolute certainty the differences in response to the effect of the context. However, the amplitude of the first formant during the transitions is greater for the /na/ stimulus than for the /da/ stimulus, so that, if one were to attribute differences in response to differences in the stimuli during the transitions, one would expect discharge rate during the test interval to be greater for /na/ than for /da/ for fibers whose CF is near the first formant. A similar argument can be constructed for fibers with CF's near the third formant for /ʃa/. In both cases, the observed decrease in discharge rate goes counter to the increase that would be expected from the differences in waveforms during the transition intervals, so that it seems likely that adaptation by the context is the dominant effect.

B. Context-dependencies and speech processing

A general consequence of short-term adaptation is that in continuous speech, the activity of the auditory nerve at any point in time depends on previous stimulation. The results of Fig. 8 and 9 demonstrate that fibers whose CF is in

the frequency region where the context has most of its energy have a decreased responsiveness during the test interval. Thus, short-term adaptation tends to increase contrast between successive speech segments separated by an abrupt change in spectral characteristics. The time course of this effect is about the duration of a speech sound. Rupert et al. (1977) have demonstrated more complex contextual effects for vowel stimuli in the cochlear nucleus. These context-dependencies at peripheral levels of the auditory system are at variance with "spectrogram" or "filter bank" models of auditory processing according to which successive segments of the speech signal are processed independently except for the brief memory associated with the filtering elements (Pols, 1971; Searle et al., 1980; Bladon and Lindblom, 1981; Klatt, 1980a).

Perceptual experiments have shown that the phonetic value of an acoustic event often depends on the neighboring speech segments (Liberman and Pisoni, 1977; Dorman et al., 1979; Mann and Repp, 1980). For instance, the distinction between /ʃ/ and /tʃ/ is usually based on the duration of the rise time of the frication noise. However, when preceded by a vowel, the same noise can be heard as /ʃ/ or /tʃ/ depending on the duration of the silent interval between the end of the vowel and the onset of frication (Dorman et al., 1979). One would expect that, by increasing the duration of the silent interval, auditory-nerve fibers would recover more from adaptation by the preceding vowel, so that the peak in

discharge rate at the onset of the noise would be more prominent. This effect would resemble the effect of decreasing the rise time of the frication noise. Thus, because of short-term adaptation, two apparently unrelated cues to /ʧ/ would have similar effects at the level of the auditory nerve.

Another example concerns the spectra at the release of stop and nasal consonants. Acoustic measurements show that the spectra at the release of nasals have more intense low-frequency components than the spectra of stops with the same place of articulation (Blumstein and Stevens, 1979). Our results show that, at moderate speech levels, discharge rate of low-CF fibers is lower at the release of a /na/ than for a /da/ having similar formant transitions. It is possible that this reduced prominence of low frequencies at the level of the auditory nerve for nasals would compensate for the greater intensity of these components in the stimulus, and would make the auditory representations of the spectra at the release of nasal consonants more similar to those for stops with the same place of articulation. Though short-term adaptation in auditory-nerve fibers clearly cannot account for all the complex context dependencies found in speech perception, these examples suggest that the possibility of context-dependent auditory effects has to be considered carefully before ad hoc explanations involving specialized speech decoding mechanisms are proposed.

The decrease in discharge rate following intense adapting stimuli resembles forward masking. The direction of variation of the adaptation effect as a function of stimulus parameters, and its frequency selectivity are qualitatively consistent with the psychophysical phenomena (Harris and Dallos, 1979). However, as pointed out by Smith (1979), it is difficult to make detailed comparisons between forward masking and short-term adaptation because one would have to know what features of the response are most important for the central processor. In the case of speech stimuli, the problem is further complicated by the possibility that different response measures would be used to make different types of phonetic distinctions. For instance, the normalized ALSM's, which are useful measures for distinctions between the short-term spectra of low-frequency stimuli, do not seem to be strongly affected by short-term adaptation.

C. Short-time average discharge rate and phonetic distinctions

The present results show that variations in short-time average discharge rate provide information about the rapid changes in amplitude and spectral characteristics that occur in speech. Because the relative amplitudes of different response components can remain nearly constant during short-term adaptation (Fig. 14), response measures based on fine time patterns of discharge would have information about these changes only if they also depended on discharge rate. Previous studies had shown that fine time

The decrease in discharge rate following intense adapting stimuli resembles forward masking. The direction of variation of the adaptation effect as a function of stimulus parameters, and its frequency selectivity are qualitatively consistent with the psychophysical phenomena (Harris and Dallos, 1979). However, as pointed out by Smith (1979), it is difficult to make detailed comparisons between forward masking and short-term adaptation because one would have to know what features of the response are most important for the central processor. In the case of speech stimuli, the problem is further complicated by the possibility that different response measures would be used to make different types of phonetic distinctions. For instance, the normalized ALSM's, which are useful measures for distinctions between the short-term spectra of low-frequency stimuli, do not seem to be strongly affected by short-term adaptation.

C. Short-time average discharge rate and phonetic distinctions

The present results show that variations in short-time average discharge rate provide information about the rapid changes in amplitude and spectral characteristics that occur in speech. Because the relative amplitudes of different response components can remain nearly constant during short-term adaptation (Fig. 14), response measures based on fine time patterns of discharge would have information about these changes only if they also depended on discharge rate. Previous studies had shown that fine time

patterns of discharge provide a better representation of the formant frequencies of vowel-like stimuli than the profile of average discharge rate against CF (Sachs and Young, 1979; Young and Sachs, 1979). In chapter II, we showed that the response measures based on fine time patterns of discharge that are useful for vowel stimuli do not provide a representation of phonetically-important features of the spectra of fricative consonants. Thus the present results further suggest that no response measure that has been proposed would suffice to represent all the acoustic information that is important for phonetic distinctions.

Comparison of rate and synchrony response measures is of particular interest for speech segments at the onset of stop consonants because these segments are important for place-of-articulation distinctions (e.g. distinctions between /b/, /d/ and /g/). According to one view (Halle et al., 1957; Blumstein and Stevens, 1979), these distinctions are based primarily on the gross shape of the short-time spectrum sampled at the release of the consonant with a 20-ms window, while according to another view (Cooper et al., 1952; Dorman et al., 1977), they depend more on the directions of changes in the formant frequencies following the release. From the first point of view, the profile of short-time average discharge rate against CF sampled at the release seems to be a useful measure because it represents both the low-frequency information associated with the voiced formant transitions and

the high-frequency information associated with the burst. Even at 75 dB SPL, the profile of discharge rate against CF at the onset of /da/ shows peaks at the places of the first two formant frequencies (Fig. 8). However, this profile can be grossly different for two versions of a /da/ syllable (Fig. 12), so that it is not clear that place-of-articulation for stop consonants could be identified from the rate profile alone. The low-frequency formants are represented clearly in response measures based on fine time patterns of discharge such as the ALSM (Fig. 14), so that such measures may be preferable to detect directions of formant transitions (Voigt et al., 1981; Sinex and Geisler, 1981). For both discharge rate (Fig. 8) and synchrony (Fig. 14) measures, the representation of the third formant at the onset of the /da/ stimulus appears weak. In the human, the resonance of the ear canal in the third-formant region (Mehrgardt and Mellert, 1977) might improve the representation of the third formant in the auditory nerve.

The present experiments have shown that prominent peaks in discharge rate occur in the response patterns of auditory-nerve fibers in specific CF regions when the speech signal shows a rapid increase in amplitude or a rapid spectral change. If such peaks were detected by the central processor, they could be used as markers to regions of the spatio-temporal pattern of auditory-nerve activity that are rich in information about phonetic distinctions. The time

windows over which discharge rate should be averaged for optimal detection of these peaks would have a duration of 10-20 ms, consistent with the rapid phase of short-term adaptation (Smith and Brachman, 1980a).

The detailed characteristics of these peaks in discharge rate provide information about many phonetic distinctions. For instance, at moderate speech levels, the distinction between /$\overset{v}{s}$/ and /$\overset{v}{c}$/ could be made on the basis of the ratio of discharge rate during the peak ("onset rate") to the rate 50-100 ms after the peak ("steady-state rate") in certain CF regions (Fig. 6) or for the "grand average" response pattern of the entire population of fibers (Fig. 7). More generally, these cues could be used to distinguish between sounds with an abrupt onset (stops, nasals and affricates) and sounds with a gradual onset (fricatives, glides and vowels). It is possible however that these strategies would not hold over a broader range of stimulus levels. The presence or absence of a peak in discharge rate at consonantal release in the low-CF region provides information about the distinction between /da/ and /na/ (Fig. 11) and, possibly, between stop and nasal consonants in general. Similarly, for fricative-vowel syllables, the peak in discharge rate at consonantal release would be more restricted to the low-CF region than for a stop-vowel syllable (Fig. 11). Obviously, for these last two examples, information about the identity of the consonants would also be

present in the interval preceding the release. Voicing distinctions between stop consonants (Lisker and Abramson, 1964), and possibly certain fricative consonants (Stevens, 1980) involve the timing of the onset of low-frequency periodicity associated with voicing relative to the onset of the high-frequency burst of noise at consonantal release. One would expect that each of these two events would result in a peak in discharge rate in the responses of auditory-nerve fibers in certain CF regions, but, for voiced consonants, the second peak in rate would occur before the decrease in rate following the first peak is terminated, so that these two peaks would merged into one for windows of sufficiently long duration.

Thus, in contrast with results for steady-state vowel stimuli (Sachs and Young, 1979), variations in the short-time average discharge rate of auditory nerve fibers are rich in information about phonetic distinctions among consonants. Further insights into the usefulness of various response measures for central processing can be obtained by comparing their behavior with psychophysical data under conditions that degrade the speech signal. In the next chapter, the effect of moderate background noise on short-time average discharge rate and fine time patterns of discharge will be described for vowel stimuli.

APPENDIX A: GENERATION OF THE STIMULI WITH CHANGING SPECTRA

The speech synthesizer used to generate the stimuli with /da/-like formant transitions is largely modelled on the one of Klatt (1980b). It consists of a voicing source, a frication source, and 5 bandpass filters connected in parallel. The voicing source generates pulses with frequency F0 and amplitude AV. The pulses are lowpass filtered at 0.3 kHz by a first-order filter intended to model glottal source spectrum and radiation characteristics. The frication source generates Gaussian-like random noise with amplitude AF. Each bandpass filter is implemented by a second-order difference equation whose coefficients are defined by the formant frequency F_i, the formant bandwidth BW_i, and the relative formant amplitude A_i (the first formant amplitude is always 1). The output of the voicing source is sent directly to the first bandpass filter, but the first difference is taken before sending it to the second and third bandpass filters in order to make low-frequency components more realistic (Klatt,1980b). The output of the frication source is sent to the second to fifth bandpass filters. The filter outputs are added with alternate phases. In order to generate flat noise spectra, the frication source can also bypass the filters and be added to the output of the synthesizer with an amplitude ABP. The 18 control parameters (F0, AV, AF, ABP, F_i, BW_i and A_i) are reset every 2 ms. Table A-I lists the values of some of the control parameters for the segments preceding the /da/-like formant transitions in the nine stimuli.

The computations are made in floating-point arithmetic on a PDP 11/34 computer, and then the output is quantized to a 12-bit integer sequence with a 20-kHz sampling rate. The stimuli are converted to analog signals, lowpass filtered at 10 kHz, attenuated and presented to the earphone so that positive voltages represent condensation.

Table A-I: Synthesis parameters for the contexts of the /da/-like stimuli

STIMULI	F1 (kHz)	F2 (kHz)	F3 (kHz)	F4 (kHz)	F5 (kHz)	durat. (ms)	silence durat. (ms)	rms level* (dB)
/ada/	0.7	1.2	2.6			136	104	-3
/ida/	0.3	2.2	3.0			136	104	-6
/uda/	0.3	0.8	2.2			136	104	-6
*/da/				5.0	7.5	14		-27
/na/	0.3	1.4	2.8			96		-12
/$\overset{v}{s}$a/			2.8	4.0	6.0	150		-18
/sa/				5.0	7.5	150		-18
/$\overset{v}{s}$ta/			2.8	4.0	6.0	150	110	-18
/sta/				5.0	7.5	150	110	-18

*Levels are given relative to the level of the /a/ segment in /da/.

FIGURE CAPTIONS

Fig. 1

Response patterns of an auditory-nerve fiber for the /ʃ/ and /ʧ/ stimuli presented at 45 and 60 dB SPL. The stimulus waveforms are shown in the bottom panels. The PST histograms are computed with a bin width of 0.25 ms from 400 stimulus presentations and are smoothed by convolution with a 2-ms Kaiser window.

Fig. 2

Spectrograms of the 10 stimuli with /da/-like formant changes. The spectrograms are measured by a Voiceprint model 4691A machine, using 300-Hz filters and 12 dB/octave emphasis of high-frequency components between 0.3 and 3 kHz.

Fig. 3

Stimulus waveforms and response patterns of an auditory-nerve fiber for /da/, /ada/, /na/ and /ʃa/ presented at 75 dB SPL. The left side shows the waveforms of the electrical signals to the earphone with a normalized amplitude scale. The interval of formant transitions is marked by dotted lines. The PST histograms shown at the right are computed with a bin width of 0.5 ms from 100 to 200 stimulus presentations. The histogram for /ada/ is truncated before the end of the stimulus. The interval of the histograms corresponding to the formant

transitions in the stimulus is marked by vertical dotted lines. The curve over this test interval is a 51.2-ms half raised-cosine window used to weight the PST histograms prior to the computation of average discharge rate, correlation index and discrete Fourier transform.

Fig. 4

Fine time patterns of discharge and normalized response spectra for an auditory nerve fiber in response to the /da/, /ada/, /na/ and /ʃa/ stimuli presented at 75 dB SPL. The left side shows PST histograms during the test interval for the same fiber as in Fig. 3. The histograms are computed with a 0.1-ms bin width. The stimulus waveform for the interval of formant transitions is shown above the histogram for each stimulus, using the same amplitude scale as in Fig. 3. The right side shows the magnitudes of the discrete Fourier transforms of the PST histograms and stimulus waveforms shown at the left. In both cases, the weighting function shown in Fig. 3 is applied prior to the computation of the Fourier transform. The spectra of the histograms are plotted with a linear magnitude scale and are normalized by the average discharge rate, while the spectra of the stimulus waveforms are plotted with a dB scale and are normalized by the amplitude of the first formant.

Fig. 5

Pseudo-perspective displays of band-average PST histograms for 0.5-octave CF bands for the /ʃ/ and /ð/ stimuli presented at

45 and 60 dB SPL. The band-average PST histograms are plotted with time along the oblique axis and discharge rate along the vertical axis. The spontaneous discharge rate is subtracted from each histograms. The histograms have a bin width of 0.25 ms and are smoothed by convolution with a 4-ms Kaiser window. The CF bands have a width of 1/2 octave and are sampled every half octave. The positions of the formant frequencies P1, P2 and P3 along the CF dimension are marked by oblique dashed lines.

Fig. 6

Ratio of onset rate to steady-state rate plotted against CF for the /ʃ/ and /ʧ/ stimuli presented at 45 and 60 dB SPL. Discharge rates are measured as described in Sec. IB. Each circle represents the ratio for one auditory nerve fiber. The continuous lines represent the ratios of the band-average discharge rates, sampled every quarter octave. The band-average rates are computed using a trapezoidal window with a central width of 0.25 octave and a total width of 0.55 octave. The positions of the formant frequencies along the CF dimension are marked by vertical dashed lines.

Fig. 7

Grand-average response patterns for the /ʃ/ and /ʧ/ stimuli presented at 45 and 60 dB SPL. Each grand average is obtained by averaging the response patterns for all the 0.5-octave CF bands shown in Fig. 5.

Fig. 8

Discharge rate during the test interval plotted against CF for the /da/, /ada/, /na/ and /ʃa/ stimuli presented at 75 dB SPL. Each circle represents discharge rate for one auditory-nerve fiber. The continuous lines are band-averages of the data points, obtained using a trapezoidal weighting function with a central width of 1/6 octave and a total width of 2/3 octave. The band averages are sampled every 1/6 octave.

Fig. 9

Ratio of discharge rate during the test interval to the rate during the test interval for /da/ plotted against CF for the nine stimuli presented at 75 dB SPL. Each point represents the ratio of discharge rates for one auditory-nerve fiber. The continuous line is the ratio of rates for the band-average data obtained as described in Fig. 8.

Fig. 10

Same as Fig. 9 for the 60 dB stimuli. The data are obtained for a narrower range of CF's than in Fig. 9 because little response was obtained during the test interval for CF's above about 4 kHz.

Fig. 11

Ratio of discharge rate during the test interval to rate during the steady segment /a/ plotted against CF for the nine stimuli presented at 75 dB SPL. Discharge rate during /a/ is averaged over a 64-ms interval starting 96 ms after the onset

of the test interval. Each point represents the ratio for one fiber, and the continuous lines represent the ratio of band-average rates obtained as described in Fig. 8. The dashed lines represent the ratio of band-average rates for the /da/ stimulus.

Fig. 12

Discharge rate at the onset of the stimulus plotted against CF for the /da/ and */da/ stimuli presented at 60 and 75 dB SPL. For both stimuli, discharge rate is averaged over a 51.2-ms half raised-cosine window beginning at the onset of the stimulus. Each circle represents discharge rate in response to */da/ for one fiber, and the continuous lines represent band-average discharge rates for */da/, obtained as in Fig. 8. The dashed lines represent the band-average discharge rates for /da/.

Fig. 13

Correlation index between the PST histogram during the test interval and the histogram for the /da/ stimulus plotted against CF for the nine stimuli presented at 75 dB SPL. Each circle is the correlation index for one fiber, and the continuous line represents the band-average correlation index computed with the same window as in Fig. 8.

Fig. 14

Average Localized Synchronized Measure (ALSM) plotted against the center frequency of a 1/6-octave Gaussian bandpass filter

for the nine stimuli presented at 75 dB SPL. The ALSM for /da/ is shown above the ALSM for /ada/.

Fig. 1

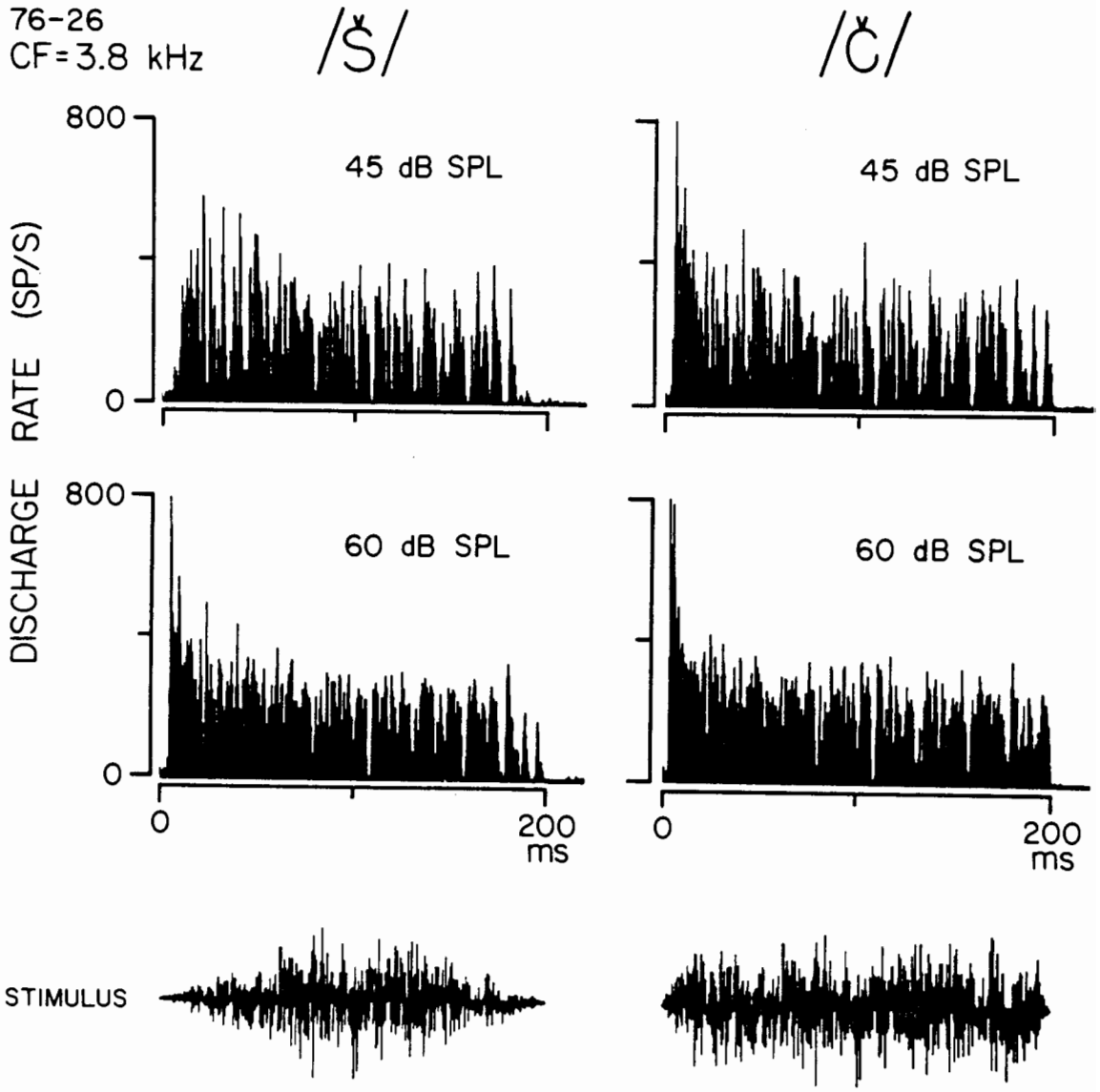

76-26
CF=3.8 kHz
/š/
/č/
800
0
45 dB SPL
45 dB SPL
DISCHARGE RATE (SP/S)
800
0
60 dB SPL
60 dB SPL
0
200
ms
0
200
ms
STIMULUS

Fig. 2

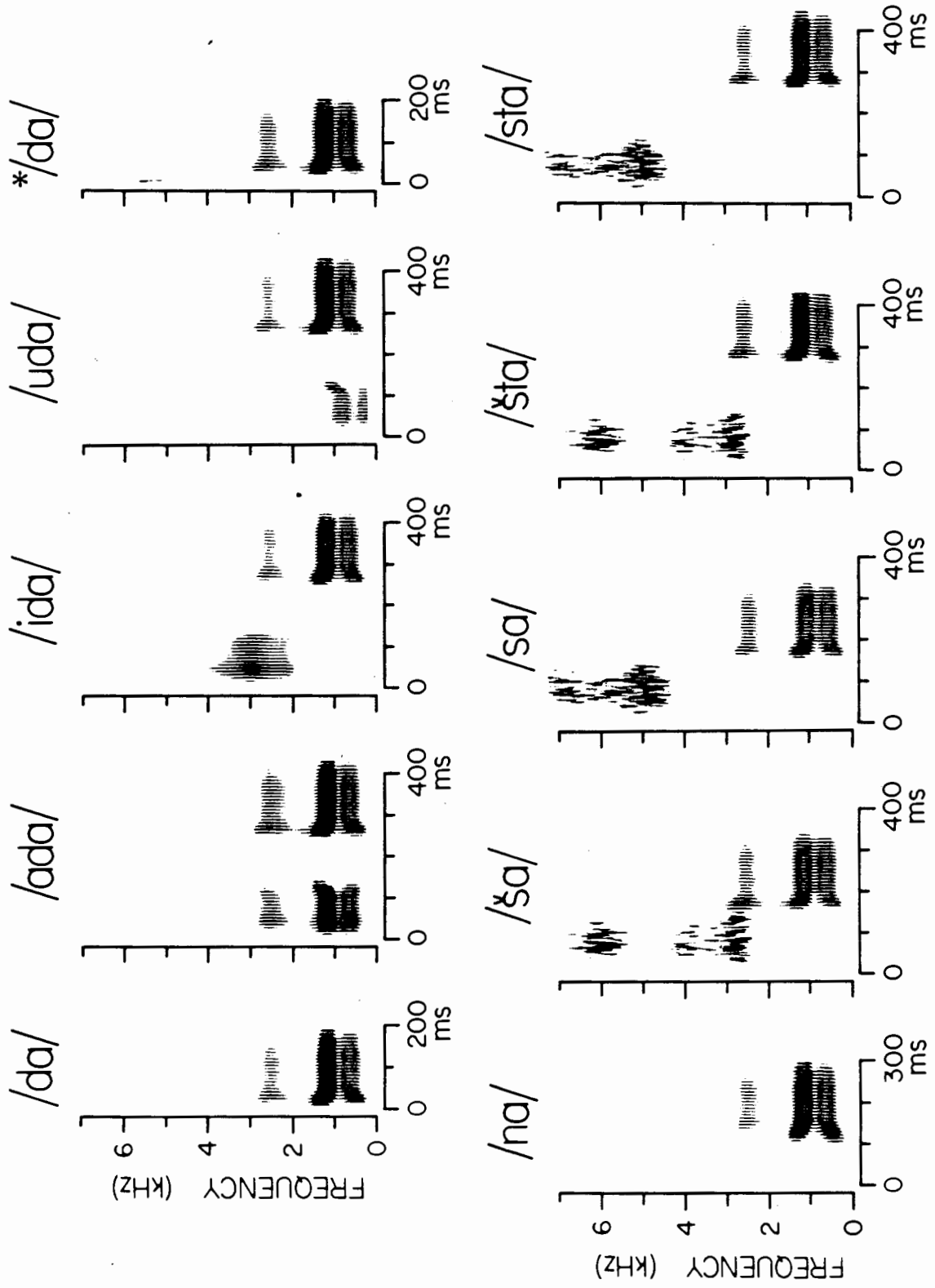

/da/
/ada/
/ida/
/uda/
*/da/
/na/
/ša/
/sa/
/šta/
/sta/
FREQUENCY (kHz)
0
2
4
6
0
200 ms
400 ms
300 ms

Fig. 3

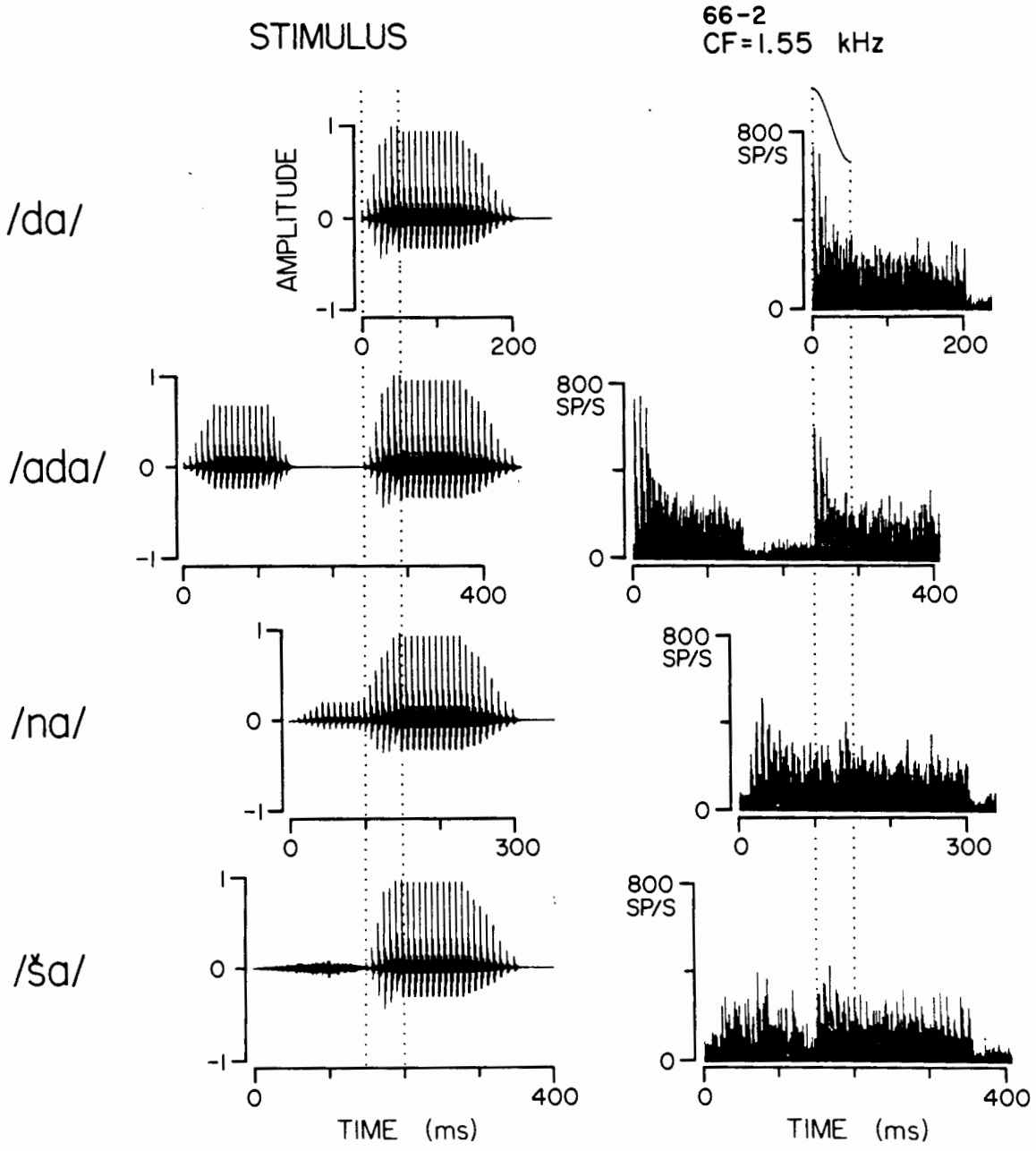
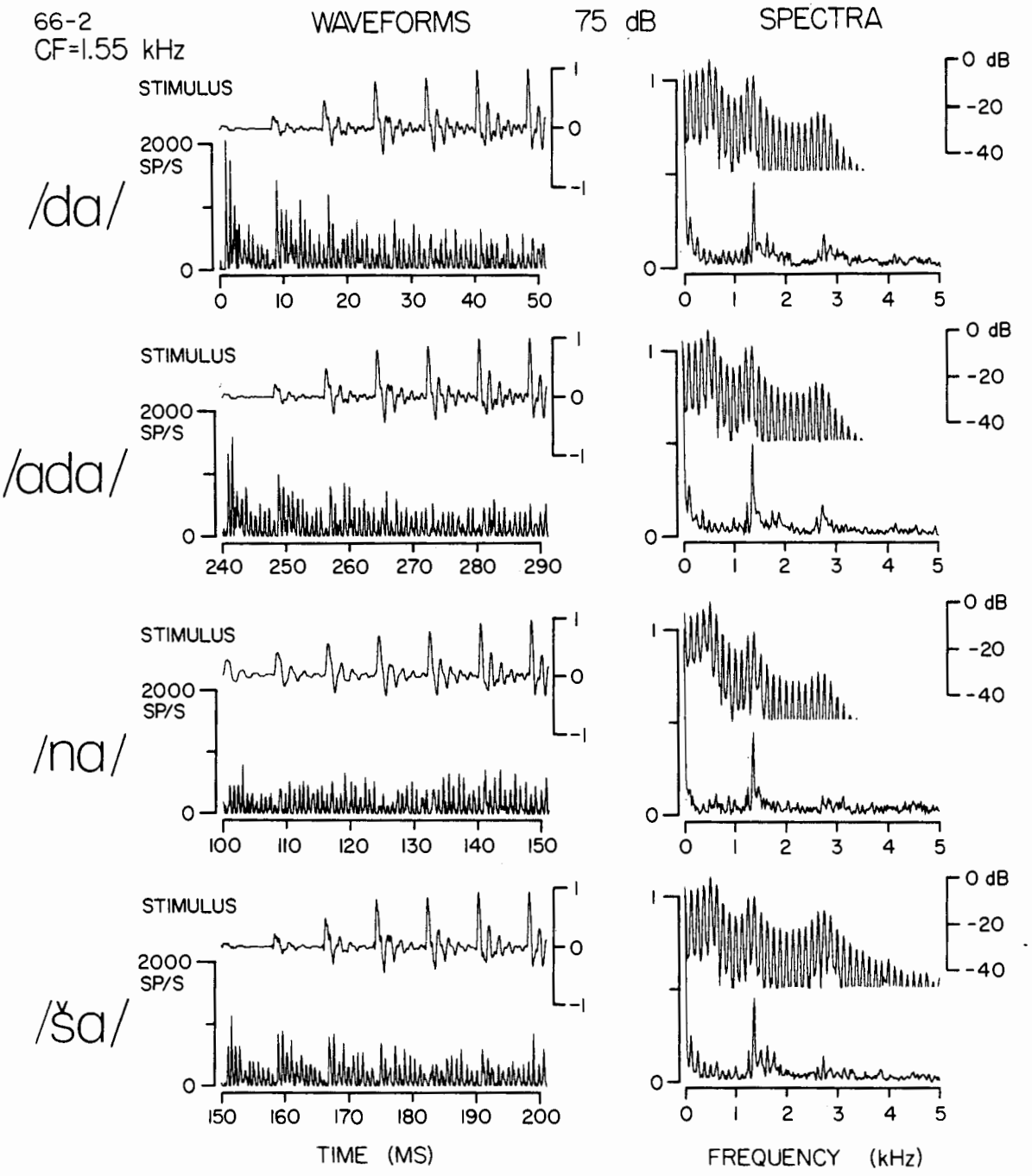

STIMULUS
66-2
CF=1.55 kHz
/da/
/ada/
/na/
/ša/
AMPLITUDE
1
0
-1
800
SP/S
0
200
400
300
TIME (ms)

Fig. 4


66-2
CF=1.55 kHz
WAVEFORMS
75 dB
SPECTRA
/da/
/ada/
/na/
/ša/
STIMULUS
2000
SP/S
0
0 dB
-20
-40
TIME (MS)
FREQUENCY (kHz)

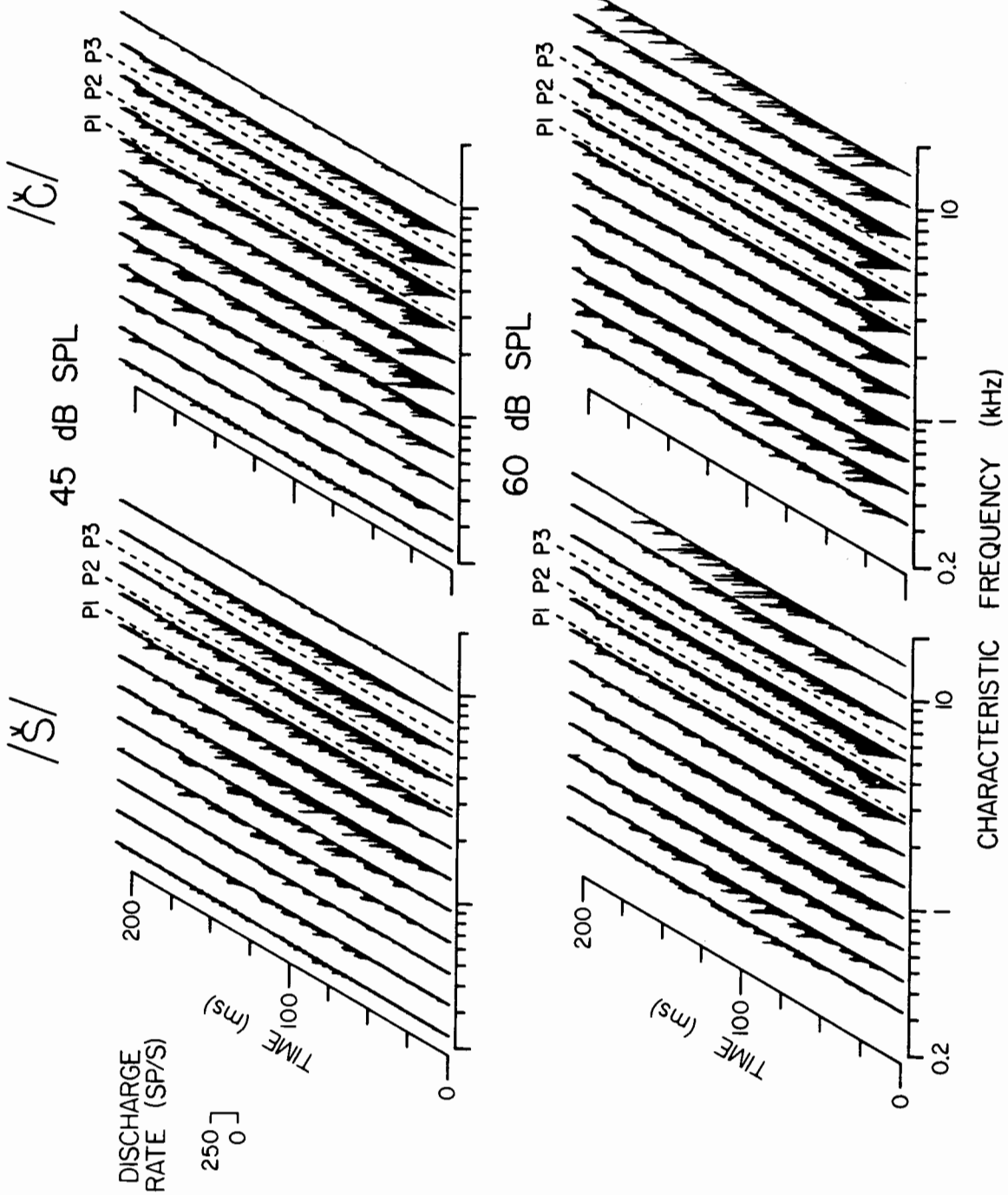

/š/
/č/
45 dB SPL
60 dB SPL
P1 P2 P3
DISCHARGE RATE (SP/S)
250
0
200
100
0
TIME (ms)
0.2
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 5

Fig. 6

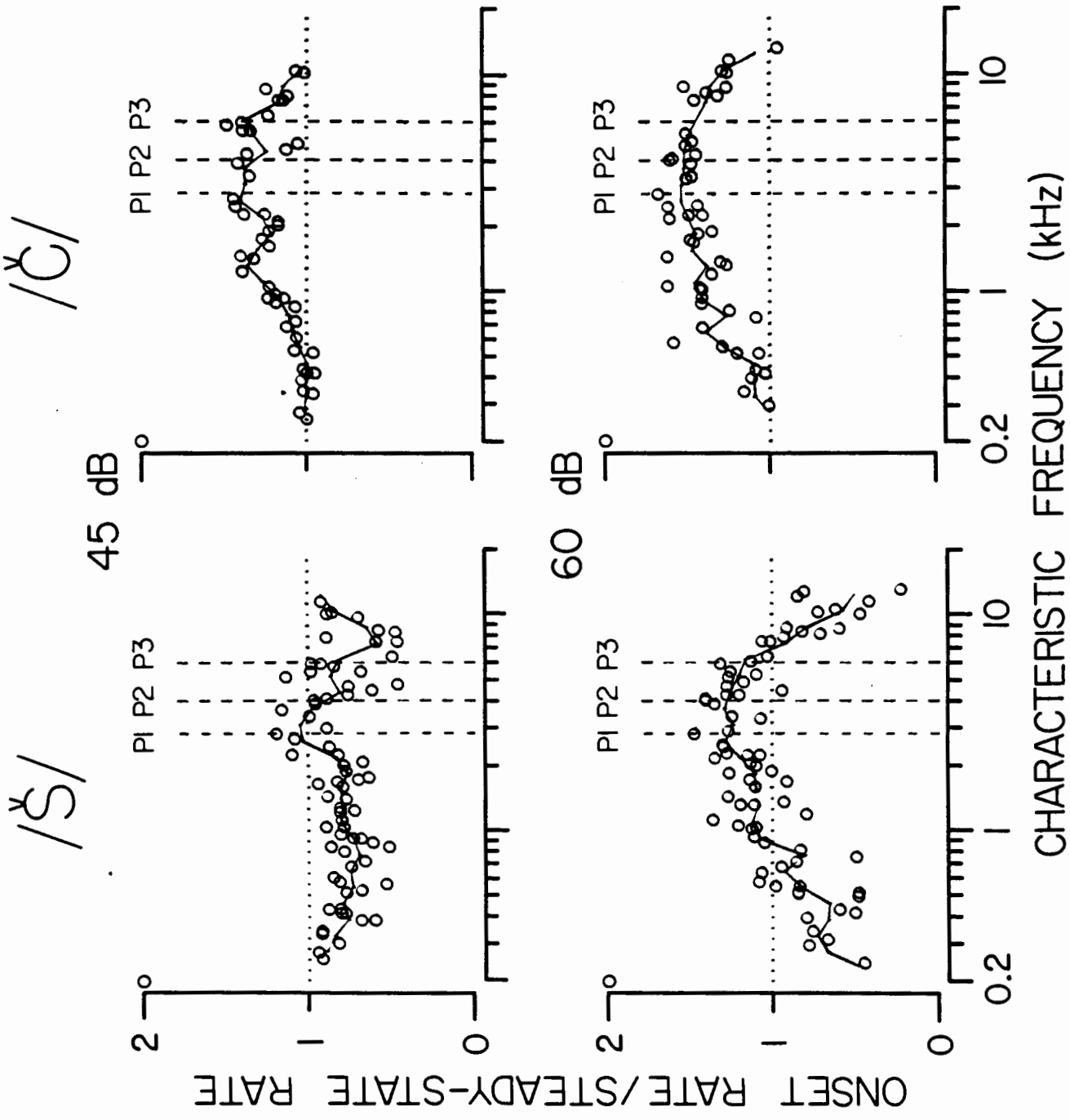

/S/
/C/
45 dB
60 dB
P1 P2 P3
ONSET RATE/STEADY-STATE RATE
CHARACTERISTIC FREQUENCY (kHz)
0.2
1
10
0
1
2

Fig. 7

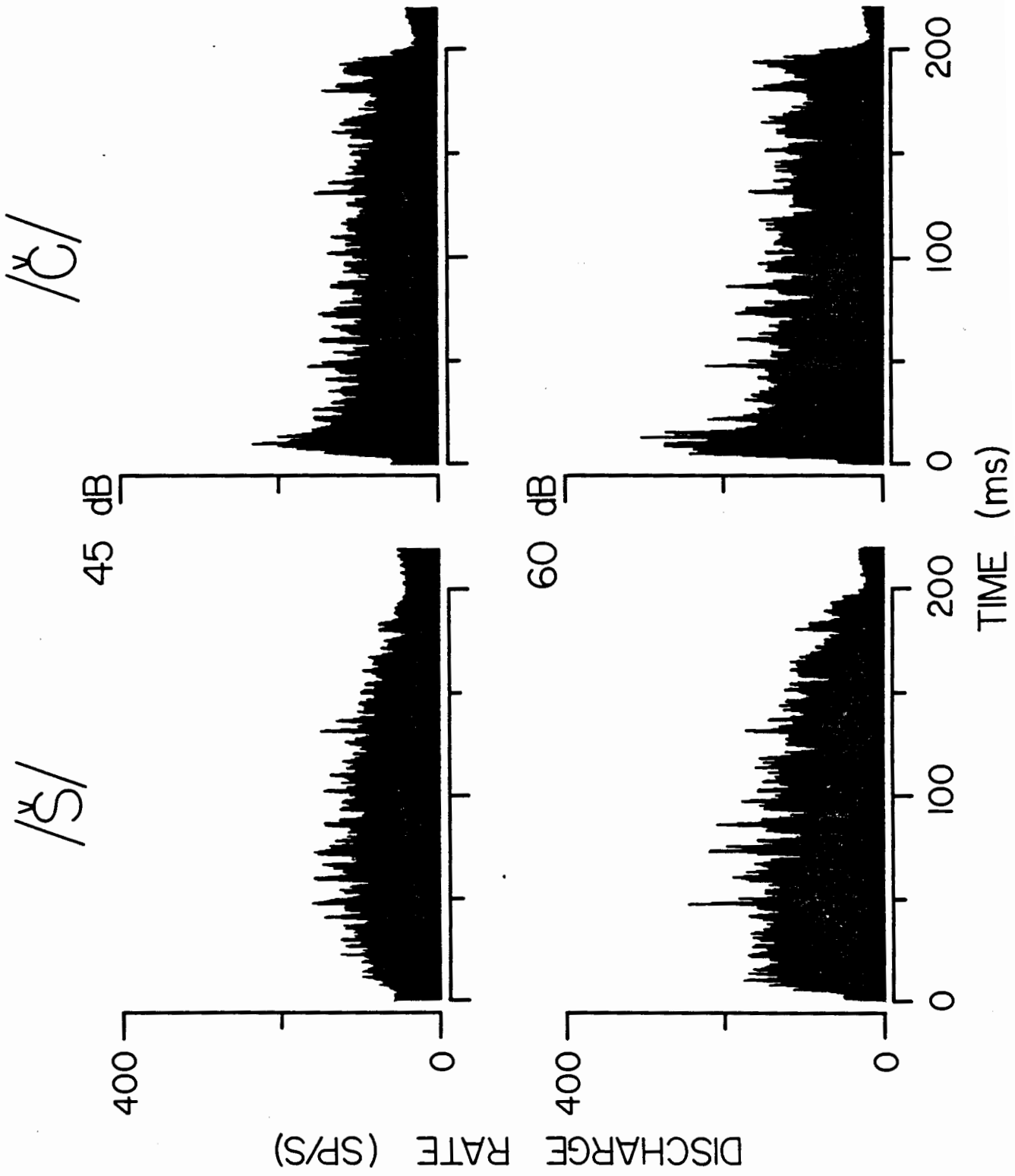


Fig. 8

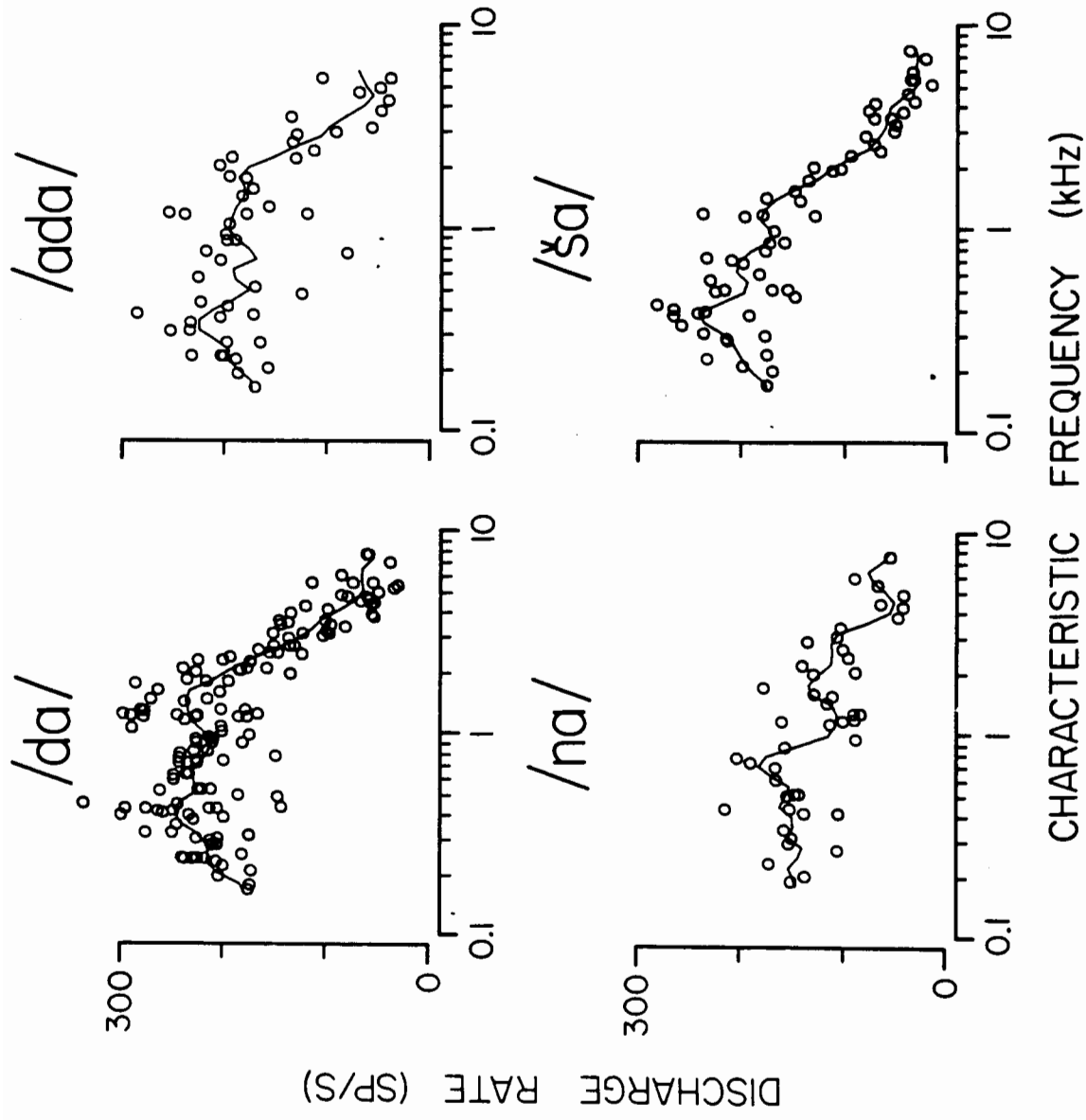

/da/
/ada/
/na/
/ʃa/
300
0
0.1
1
10
DISCHARGE RATE (SP/S)
CHARACTERISTIC FREQUENCY (kHz)

Fig. 9

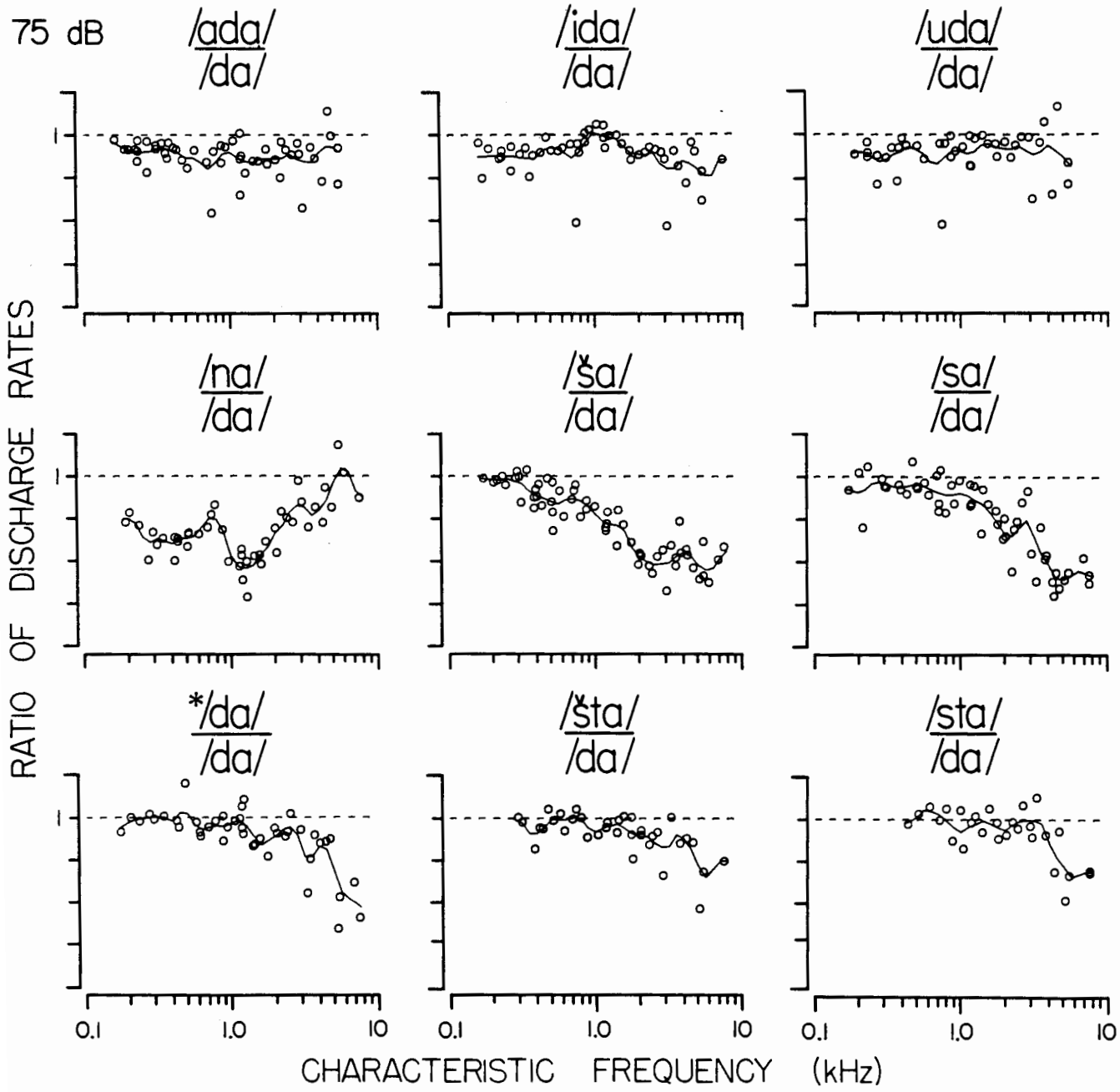

75 dB
/ada/
/da/
/ida/
/da/
/uda/
/da/
/na/
/da/
/ša/
/da/
/sa/
/da/
*/da/
/da/
/šta/
/da/
/sta/
/da/
RATIO OF DISCHARGE RATES
0.1
1.0
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 10

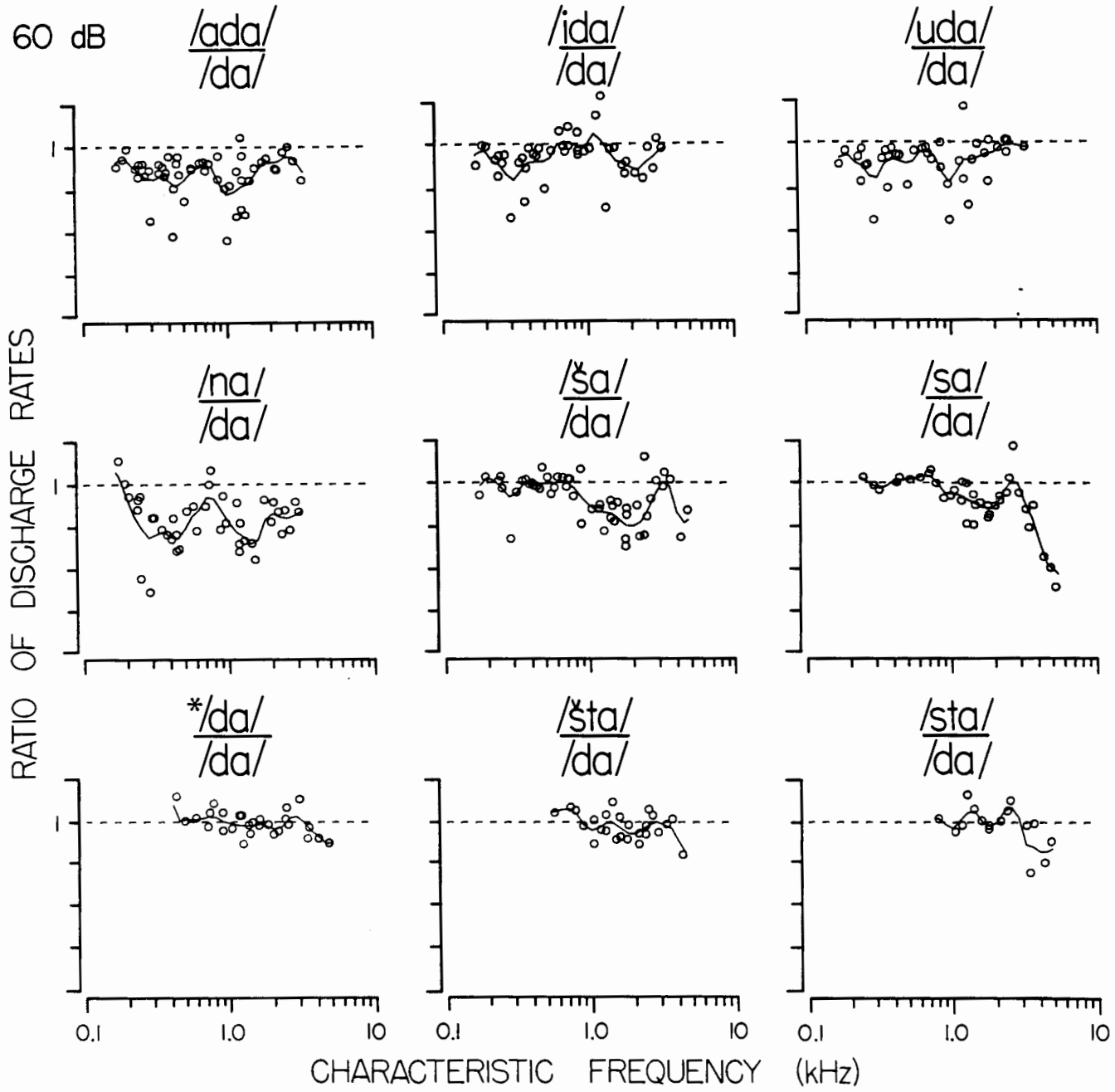
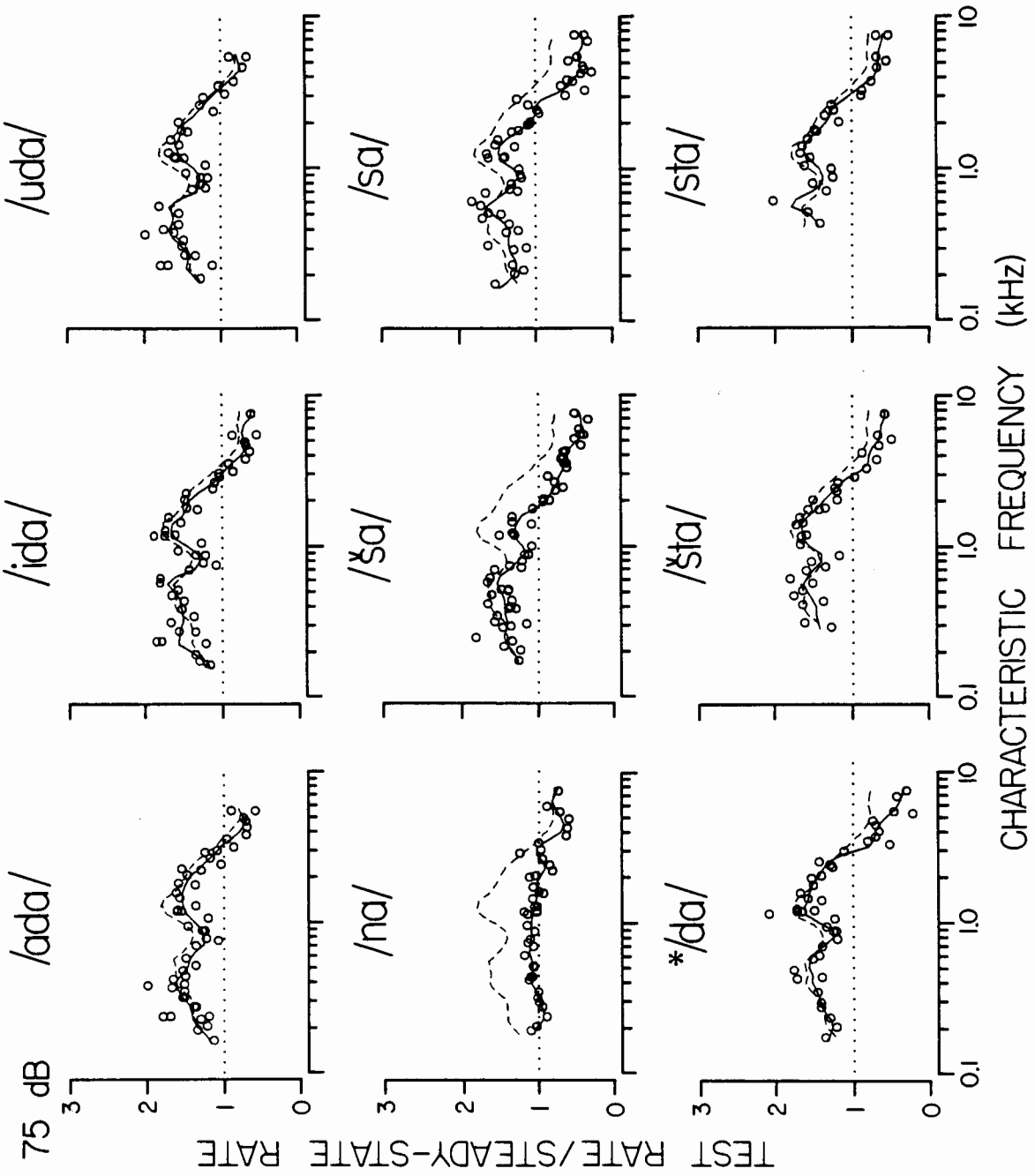

60 dB
/ada/
/da/
/ida/
/da/
/uda/
/da/
/na/
/da/
/ša/
/da/
/sa/
/da/
*/da/
/da/
/šta/
/da/
/sta/
/da/
RATIO OF DISCHARGE RATES
0.1
1.0
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 11


75 dB
/ada/
/ida/
/uda/
/na/
/ša/
/sa/
*/da/
/šta/
/sta/
TEST RATE/STEADY-STATE RATE
CHARACTERISTIC FREQUENCY (kHz)
0
1
2
3
0.1
1.0
10

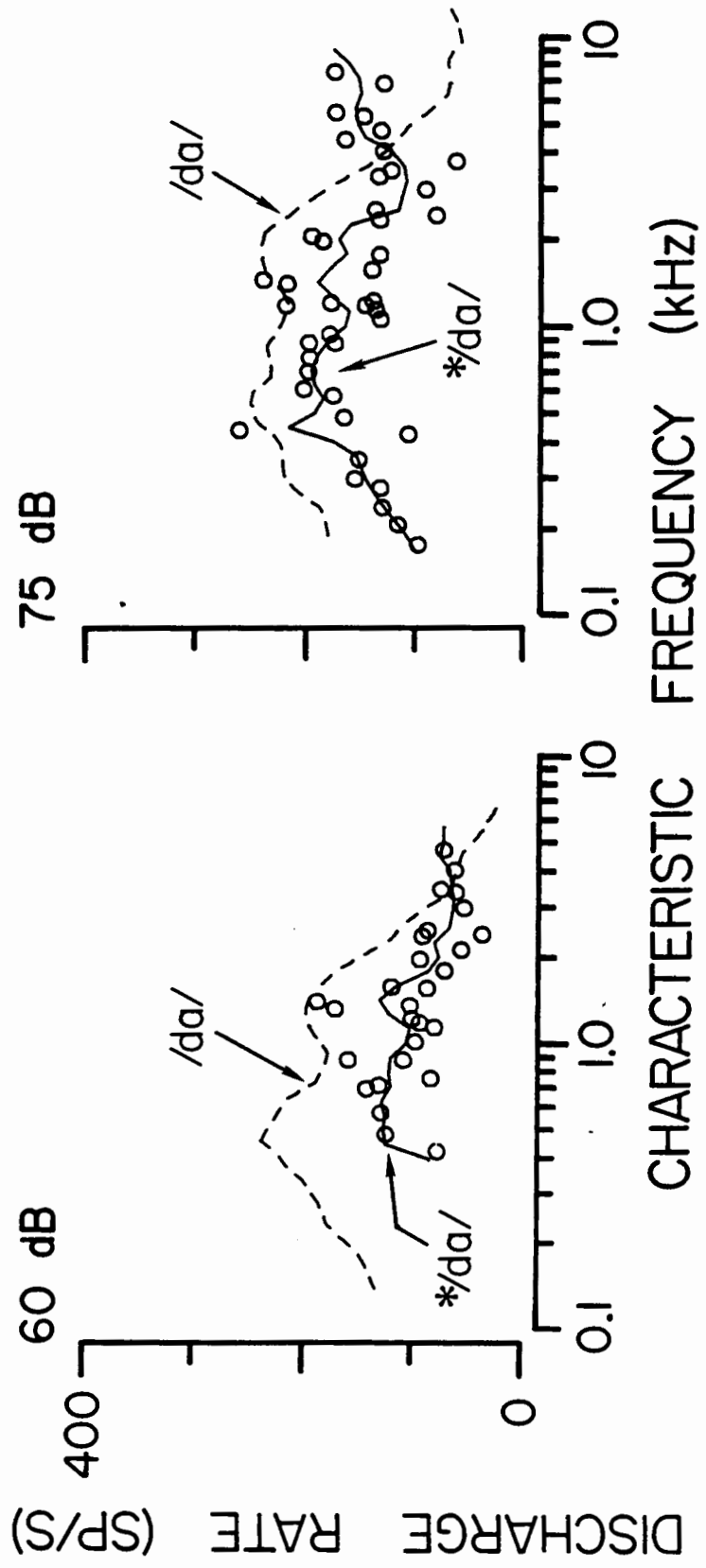

DISCHARGE RATE (SP/S)
400
0
60 dB
75 dB
/da/
*/da/
0.1
1.0
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 12

Fig. 13

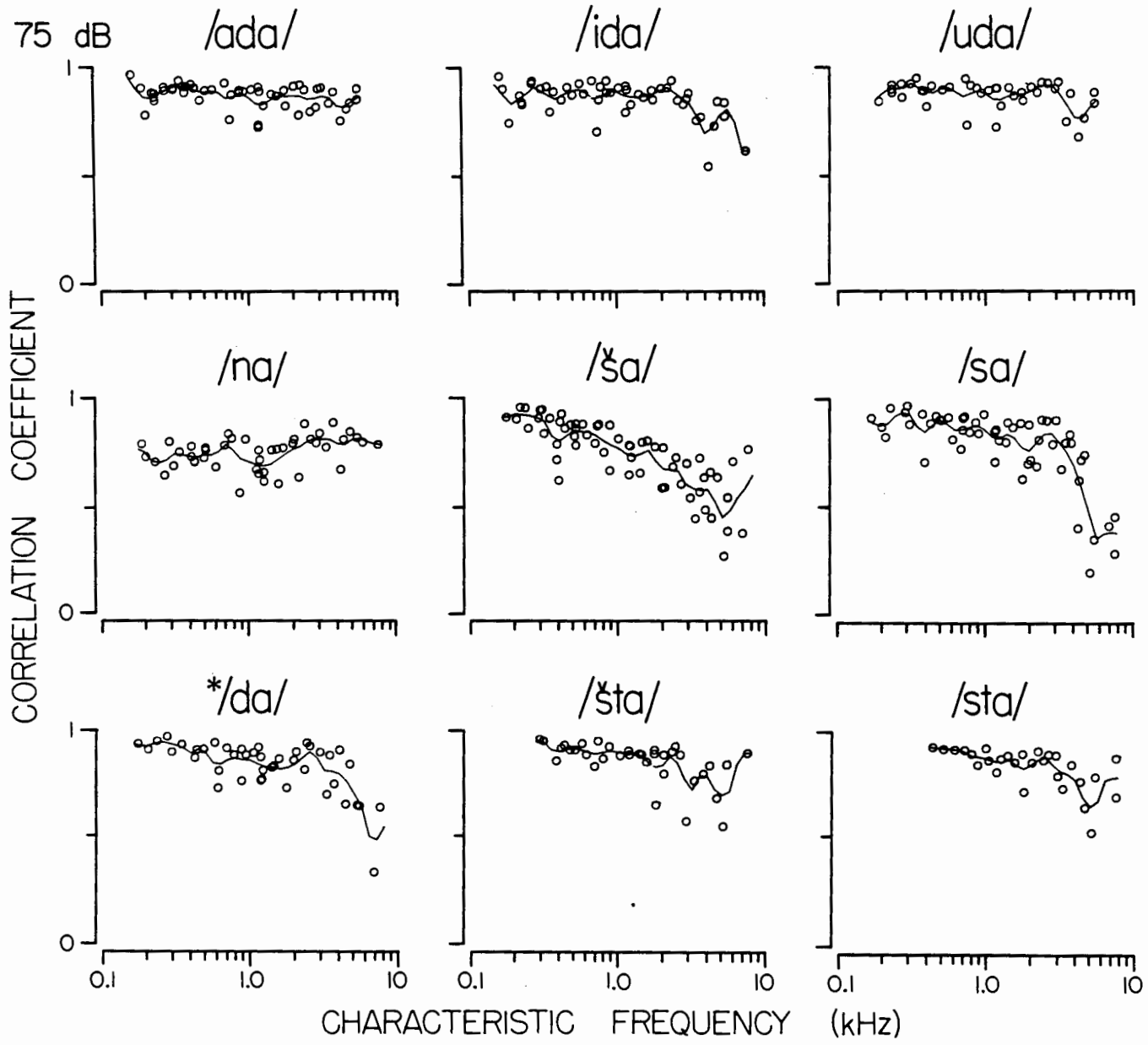
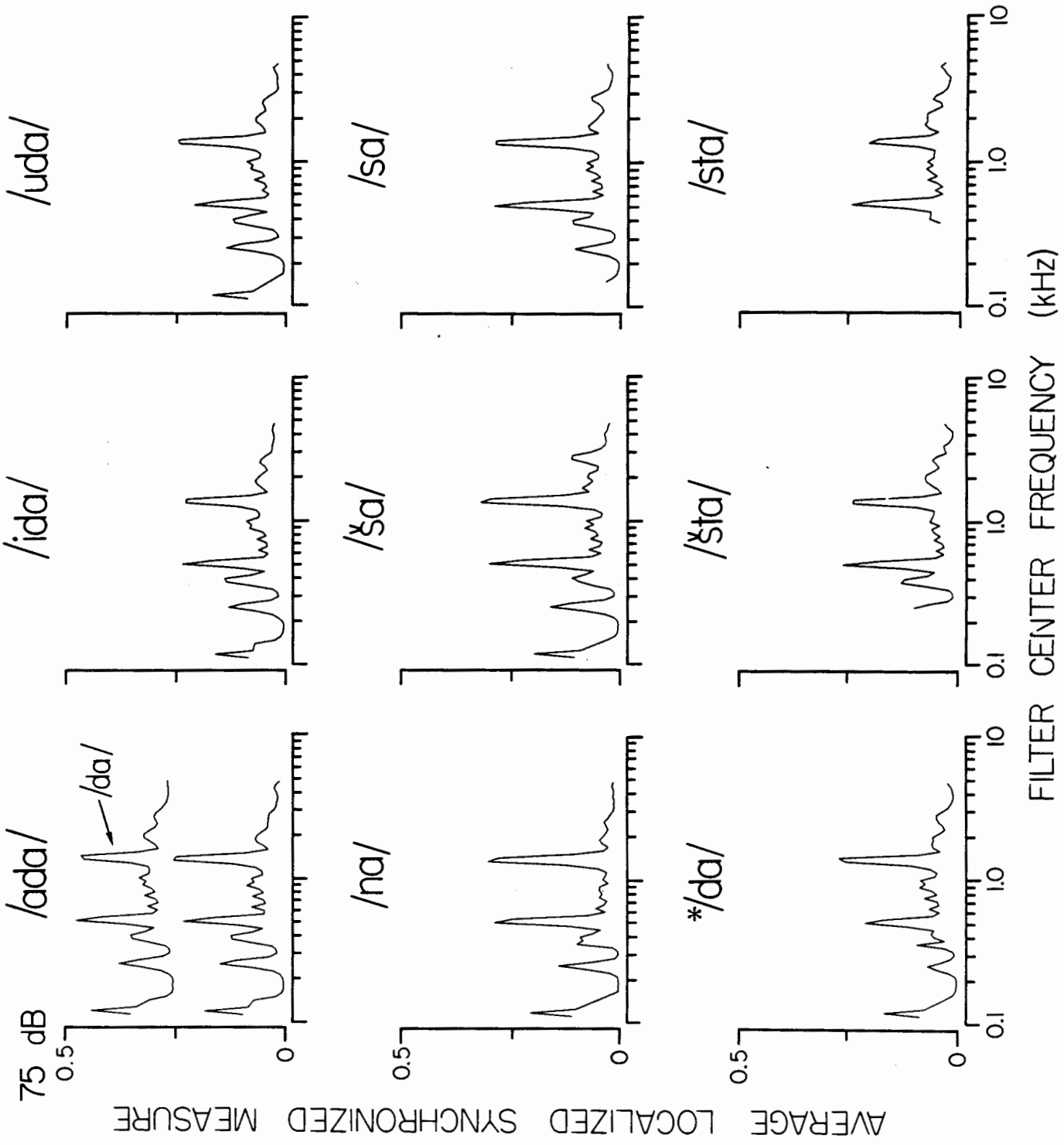

75 dB
/ada/
/ida/
/uda/
/na/
/ša/
/sa/
*/da/
/šta/
/sta/
CORRELATION COEFFICIENT
1
0
0.1
1.0
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 14


75 dB
/ada/
/da/
/ida/
/uda/
/na/
/ša/
/sa/
*/da/
/šta/
/sta/
0.5
0
0.1
1.0
10
AVERAGE LOCALIZED SYNCHRONIZED MEASURE
FILTER CENTER FREQUENCY (kHz)

CHAPTER IV

RESPONSES OF AUDITORY-NERVE FIBERS TO VOWELS IN BACKGROUND NOISE

INTRODUCTION

Because speech communication usually takes place in a noisy background, it is essential for a description of stimulus coding in the auditory nerve to study fiber responses to speech-like stimuli presented in noise. Data on perceptual confusions between speech sounds presented in noise show a definite pattern of errors, speech sounds that share certain acoustic properties being more easily confused than sounds sharing other properties (Miller and Nicely, 1955; Picket, 1957; Nooteboom, 1968). Studies of the effect of noise on responses of auditory-nerve fibers to speech-like stimuli might contribute to an understanding of the physiological basis of this pattern. Another motivation for studying responses in background noise is to evaluate the speech processing schemes that have been proposed in previous chapters for extracting features of the speech signal that are important for phonetic distinctions. Speech-processing schemes that would be strongly affected by noise at signal-to-noise ratios that do not influence perception could not be good models of central processing.

Studies of responses of auditory-nerve fibers to acoustic stimuli presented in background noise have been reported (Kiang et al., 1965; Kiang and Moxon, 1974; Smith, 1979; Rhode et al., 1978). These studies suggest that different aspects of fiber discharge patterns are affected differently by the background noise. Experiments with a speech utterance have shown that fine time patterns of discharge of low-CF fibers can retain considerable information about the stimulus at signal-to-noise ratios for which variations in short-time average discharge rate are minimal (Kiang and Moxon, 1974). Recently, Voigt et al. (1980) showed that, for steady-state vowel stimuli, a response measure based on fine time patterns of discharge had clear maxima for fibers with CF's near the formant frequencies, even at signal-to-noise ratios for which the profile of average discharge rate against CF were basically flat.

The present paper is a more detailed study of the effects of background noise on responses of auditory-nerve fibers to vowel-like stimuli. Vowels were chosen because they typify a broad class of speech sounds, the "sonorants", which are characterized by a well defined formant structure and a low-frequency periodicity at the fundamental frequency. In addition, pilot experiments suggested that different components in the spectra of the response patterns for vowel stimuli can be affected differently by background noise (Delgutte, 1980).

To simulate conversational or environmental noise, a background noise with a low-pass spectrum was used. The signal-to-noise ratio was set at 10 dB, consistent with typical measurements of this parameter in every-day situations (Pearsons et al., 1976). In these conditions, the vowel stimuli remain readily identifiable and have a clear voice pitch, so that the discharge patterns must preserve considerable information about formant frequencies and fundamental frequency.

I METHODS

A. Stimuli

The vowel stimuli are the same set of nine two-formant, steady-state stimuli that were used in Chapter I. The vowels are 184-ms bursts of sound with a fundamental frequency of 125 Hz, and a rise-fall time of 2 ms. The background noise is generated by passing a sequence of approximately white, Gaussian noise samples through a first-order digital lowpass filter with a cutoff frequency of 0.5 kHz. Thus, the noise spectrum roughly matches the long-term spectrum of speech (Dunn and White, 1940). The noise waveform, which is the same for all stimuli, is added to the waveform of each vowel so that the signal-to-noise ratio is 10 dB for all stimuli. The noise waveform is continued after the offset of the vowel burst to produce a stimulus with a total duration of 600 ms. During experiments, this stimulus

is presented without interruption to the animal so that the background noise is in effect continuous.

Figure 1 shows the harmonic spectra of the /i/, /ae/ and /u/ stimuli superimposed on the power spectrum of the noise. The noise level exceeds the vowel harmonics only in the high-frequencies, and, for /i/, between 0.5 and 1.2 kHz. These three stimuli are "extreme" vowels, in the sense that the formant frequencies of the other six vowels are between the formant frequencies of these stimuli. Specifically, /i/ has the lowest F1 and the highest F2, /ae/ has the highest F1, and /u/ has the lowest F2. For convenience, only data obtained in response to these three stimuli will be presented in the figures. Special mention will be made when results for the other stimuli cannot be "interpolated" from those for /i/, /ae/ and /u/.

B. Experimental procedures and data processing

The experimental results to be reported are based on recordings from single auditory-nerve fibers in anesthetized cats. The preparation of the animals, the stimulus generation system, and the recording procedures are as described in Chapter I. Recordings were restricted to the most sensitive population of auditory-nerve fibers, those with spontaneous discharge rates greater than 18 spikes/s (Liberman, 1978). The vowel and vowel-in-noise stimuli were presented at levels of 60 and 75 dB SPL. Only fibers for which data were obtained

both in the quiet and noise conditions were used in this paper.

The vowel stimuli were delivered every 600 ms for the computation of post-stimulus time (PST) histograms (Gerstein and Kiang, 1960). The histograms were computed with a bin width of 0.05 ms on the basis of 200 to 500 stimulus presentations. This bin width is appropriate for estimation of the frequency components of the instantaneous discharge rate up to about 4 kHz (Johnson, 1978). Because the waveforms of both vowel and noise are identical for each stimulus presentation, PST histograms estimate response components synchronized to the noise as well as those synchronized to the vowel. In contrast, because the noise waveform differs for each period of the vowel, period histograms, such as those that were used in Chapter I, would attenuate the response components that are not harmonics of the fundamental frequency of the vowel stimuli (Johnson, 1978), so that they would not provide a valid representation of the effect of noise on the response of auditory-nerve fibers. Thus, all further processing of the PST histograms is made asynchronously with the period of the vowels.

For each PST histogram, a 172-ms analysis interval beginning 12 ms after the onset of the vowel stimulus was defined. The mean discharge rate during that interval will be called "steady-state" rate. The "onset" rate is the mean rate over a segment of the histogram defined by a weighting window

of value 1 from 0 to 10 ms, then decreasing linearly to reach zero at 30 ms. For each fiber and each vowel, the correlation index C_{qn} between the PST histogram in quiet R_q and the histogram in noise R_n was computed from:

$$C_{qn}=\sum_{0<t_i<172} R_q(t_i)\ R_n(t_i)\ /\sqrt{S_q\ S_n} \qquad (1)$$

where t_i is time (in ms) of each histogram bin during the analysis interval, and:

$$S_q=\sum_{0<t_i<172} R_q(t_i)^2 \qquad (2)$$

The correlation index, a number varying between 0 and 1, is a rate-independent measure of similarity between the response in quiet and the response in noise.

The PST histograms were also used to estimate power spectra and autocorrelation functions for the analysis interval. The DC component of the histogram was removed prior to these computations. Power spectra were estimated by (1) dividing the analysis interval of the histogram into 14 overlapping 25.6-ms segments, each of which was weighted by a Kaiser window, and (2) averaging the magnitude square of the discrete Fourier transforms (DFT) of these segments (Oppenheim and Schafer, 1975). The resulting spectra have a frequency resolution of about 40 Hz. Figure 2 shows power spectra normalized by the square of the mean discharge rate. The 25.6-ms Kaiser window has a sufficient duration to resolve the harmonics of the vowel stimuli, as indicated by the peaks at intervals of about 125-Hz in the spectra. If a longer window were used, response components at the harmonics of the vowel

would stand out more relative to the inharmonic components, which, to a large extent, originate from the noise stimulus. The autocorrelation functions of PST histograms were estimated for delays up to 20 ms by the DFT method (Oppenheim and Schafer, 1975).

To obtain data at regularly spaced samples along the log CF dimension, "band-average" power spectra were computed by averaging the power spectra of all units whose CF lies in a narrow frequency band, using a trapezoidal weighting window with a central width of 0.25 octave and a total width of 0.55 octave (effective width 0.4 octave). Whenever data were available, center frequencies of the CF bands were sampled every quarter octave from 0.1 to 10 kHz. The legitimacy of this averaging procedure has been discussed in Chapter I.

The power spectra and autocorrelation functions of PST histograms were used to compute the response measures that were found useful in Chapter I for extracting spectral parameters of vowel stimuli. One of these measures is the Reciprocal of the Mode of the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function. To compute this measure, an histogram of the frequency distribution of the intervals between successive peaks of the autocorrelation function was constructed, weighting each interval by the height of the second peak, and the RMISP was set to the reciprocal of the center of gravity of the intervals within $\pm$12 % of the histogram mode. Other

useful response measures are the Average Localized Synchronized Measures (ALSM), which are similar to the ALSR proposed by Young and Sachs (1979). These measures are obtained by filtering auditory-nerve response patterns by a filter whose center frequency f_c is near the fiber characteristic frequency. The ALSM's differ by the choice of the filter, narrow bandpass, broad comb, or narrow comb. For the ALSM based on a the narrow-band filtering scheme, the result A_j of the filtering operation for fiber j was computed from:

$$A_j = \sum_{0<f_k<5} P_j(f_k)\ H(f_k) \qquad (3)$$

where f_k is frequency in kHz, H(f) is the transfer function of a Gaussian filter, and $P_j(f_k)$ is the power spectrum of the PST histogram for fiber j. The filter transfer function is given by:

$$H(f) = \exp -\pi[(f-f_c)/b_c]^2 \qquad (4)$$

where the bandwidth b_c is 0.116 f_c (1/6 octave) or 50 Hz, whichever is larger. For the ALSM based on a broad comb filtering scheme, the result of the filtering operation for each fiber was obtained by evaluating the autocorrelation function of the PST histogram at time $1/f_c$. The ALSM based on a narrow-comb filtering scheme was realized by a delayed feedback system whose output $y_j(t)$ for fiber j is given by the difference equation:

$$y_j(t_i) = (1-a_{f_c})\ R_j(t_i) + a_{f_c} y_j(t_i - 1/f_c) \qquad (5)$$

where $R_j(t)$ is the PST histogram for fiber j, and the feedback gain a_{f_c} is set to 0.7 for a 1/6-octave filter. The overall

result of the narrow-comb filtering operation is computed from:

$$A_j = \sum_{0<t_i<172} R_j(t_i) y_j(t_i) \qquad (6)$$

For all three filtering schemes, the result of the filtering operation A_j is normalized by the mean square discharge rate for each fiber, and the ALSM is computed by averaging the normalized A_j for all fibers whose CF is in a narrow frequency band centered at f_c, using a trapezoidal weighting window with a central width of 1/6 octave and a total width of 1/2 octave. The ALSM was computed for values of f_c ranging from 0.1 to 5 kHz, in steps of 1/12 octave.

II RESULTS

A. Short-time average discharge rate

Experiments with tone-burst stimuli have shown that the presence of continuous background noise at a moderate level decreases the discharge rate of auditory-nerve fibers during the stimulus (Kiang et al., 1965). The decrease in rate is largest during the short-term adaptation following the onset of the tone burst (Kiang et al., 1965; Smith, 1979). Because the peak in discharge rate following an onset provides considerable information for phonetic distinctions (Chapter III), the effect of background noise on these peaks was studied for vowel stimuli.

In response to the vowel stimuli at 60 and 75 dB SPL, the steady-state discharge rate is decreased by at most 30-40 % by the addition of the background noise (not shown). For the /u/ stimulus, there is a CF region near 4-5 kHz in which steady-state rate is greater in noise than in quiet, probably because the noise has considerably more energy than the vowel in that frequency region. In that CF region, the presence of the vowel can suppress discharge rate in response to the noise.

Figure 3 shows the ratio of onset discharge rate to steady-state rate plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL, both in quiet and in background noise. The ratio in the quiet condition is always greater than 1, and tends to be largest for fibers whose CF's are near the formant frequencies. In contrast, the ratio for the noise condition is close to 1 throughout the range of CF's, and can even be lower than 1 for high-CF fibers in response to /ae/ and /u/. The decrease in ratio due to the noise seems to be slightly greater at 75 dB than at 60 dB, even though the signal-to-noise ratio is the same. Thus, whereas steady-state discharge rate in response to vowel stimuli is only moderately affected by background noise, the peak in discharge rate at the onset of the stimuli almost totally disappears throughout the range of CF's.

B. Fine time patterns of discharge

For tone-burst stimuli, there are situations in which fine time patterns of discharge are less affected by background noise than short-time average discharge rate (Kiang and Moxon, 1974; Delgutte, 1980). For instance, the synchronization index can remain unchanged in the presence of a background noise that considerably reduces discharge rate during short-term adaptation (Delgutte, 1980). However, for sufficiently low signal-to-noise ratios, fine time patterns of discharge are also affected (Kiang et al., 1965). Because of the frequency selectivity of auditory-nerve fibers, the signal-to-noise ratio for vowel stimuli should effectively be higher for fibers with CF's close to the formant frequencies.

Figure 4 shows the correlation index between the PST histogram in quiet and the histogram in noise for the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL. Correlation indices that are much smaller than 1 indicate that the background noise changes considerably the fine time patterns of response to the vowels. The correlation index remains close to 1 for fibers with CF's near the formant frequencies, and becomes lower than 0.5 only in the high-CF region for all vowels, and near 1 kHz for /i/. These frequency regions of maximum effect are those in which the noise level exceeds the signal level (Fig. 1). In certain CF regions (e.g. near the F2 place for /i/ and /ae/, or near 1 kHz for /i/) the correlation indices seem to be somewhat lower

at 75 dB than at 60 dB, even though the signal-to-noise ratio remains the same. Thus, fine time patterns of discharge are significantly affected by background noise but, unlike for short-time average discharge rate, the strong effects are restricted to CF regions that are far from the formant frequencies.

To elucidate the nature of the effect of background noise on the fine time patterns of discharge, power spectra of PST histograms were compared for the quiet and noise conditions. Figure 2 shows normalized power spectra for fibers with various CF's in response to the /i/ and /ae/ stimuli presented at 75 dB SPL. As expected from the high value of the correlation index at the places of the formant frequencies, the power spectra for fibers whose CF is close to the formant frequency of a vowel stimulus are not strongly affected by the presence of the background noise. This applies to the 0.25 and 3.2-kHz units in response to /i/, and for the 0.8 and 1.8-kHz units in response to /ae/.

For fibers whose CF is sufficiently far from the formant frequencies, the response spectra show qualitative changes in the presence of background noise. For instance, the spectrum of the 0.25-kHz unit for /ae/ presented in quiet shows many peaks separated by intervals of 125 Hz, reflecting the harmonic structure of the vowel. The harmonic structure is not as prominent in the noise condition, indicating an increase in nonharmonic response components possibly

originating from components in the noise stimulus. This effect also occurs for the 0.8 and 1.8-kHz fibers in response to /i/, and for the 5.6-kHz fiber in response to both stimuli. Another common change in response is a decrease in the amplitude of low-frequency components, particularly the fundamental of the vowel stimuli. This effect is seen for the 1.8 and 5.6-kHz unit in response to /i/.

The most dramatic type of noise-induced changes is a shift in the identity of the largest response components. For instance, the largest component in the spectrum of the 0.8-kHz unit in response to /i/ presented in quiet is the 2 F1 (0.5-kHz) component. In noise, components near the CF become largest, while the 2 F1 component becomes small. Because low-CF fibers show prominent response components at the CF for broad-band noise stimuli (De Boer and Kuyper, 1968; Møller, 1977), this type of change can be interpreted as a suppression of the response to the vowel by the noise. A major change in the identity of the largest response components is also seen for the 5.6-kHz unit in response to /ae/. The power spectrum in quiet has a peak near the first formant frequency, and shows little activity in the low frequencies. In contrast, the spectrum in noise is dominated by low-frequency components. Because responses of high-CF fibers to broad-band stimuli have their largest components in the low frequencies (Chapter II), this effect can also be interpreted as a suppression by the noise.

Thus the effects of background noise on the power spectra of PST histograms for vowel stimuli vary greatly with fiber CF. Whereas near the place of a formant frequency the changes are small, the spectra of fibers with CF's far from the formants often show a decreased prominence of the harmonic structure and a reduction in the relative amplitude of the fundamental component. For fibers with CF's in regions where the signal-to-noise ratio is low, there can be major changes in the identity of the largest response components.

These effects are summarized in Fig. 5, which shows power spectra for many CF bands in response to the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL, both in quiet and in noise. In describing responses to vowels in quiet in Chapter I, the array of auditory-nerve fibers had been divided into five CF regions characterized by the identity of the most prominent response components. For all vowels there was a CF region near the F1 place in which harmonics closest to F1 were the largest response components, and a region near the F2 place in which harmonics near F2 were the largest components with the possible exception of the fundamental. For vowels with a high F1, there was a low-CF region in which harmonics closest to the CF were the largest components, and, for vowels with a large F2/F1 ratio, there was an intermediate CF region between F1 and F2 in which the fundamental and components near CF were most prominent. For all vowels, there was a high-CF region above the F2 place in which response spectra had many

prominent components with broad peaks at F1, F2 or the fundamental frequency. This basic structure remains apparent in Fig. 5 for the quiet condition, even though the power spectra were estimated asynchronously with the fundamental period of the vowel stimuli, whereas in Chapter I they were estimated from period histograms.

For all three stimuli, the power spectra of fibers with CF's near the formant frequencies are not strongly affected by the background noise, so that the formant-dominated regions centered at the places of the formant frequencies remain clearly apparent. However, the extent of the F2-dominated region is reduced by about 1/2 octave on the high-CF side for /u/. This reduction in the extent of the F2 region also occurs for /e/, and less strongly, for /i/ and /oh/.

For the /ae/ stimulus in quiet, there is a low-CF region below F1 in which the largest response components are the harmonics closest to CF. This region remains apparent in the noise condition, but there is a slight decrease in the prominence of the harmonic structure, as response activity at the frequencies between harmonics is greater. This finding applies to all "open" vowels /ae/, /a/, /eh/ and /oh/, which are characterized by a high first formant.

The intermediate CF region between F1 and F2 for /i/ is strongly affected by the noise. There is an almost

complete loss of harmonic structure, a strong reduction of the fundamental component, and a narrowing of the activity around the f=CF line. For fibers with CF's near 0.6-1 kHz, there is a decrease in the amplitude of the 2F1 component. These findings are typical for the "spread" vowels /i/, /e/ and /ɨ/, which are characterized by a wide separation between F1 and F2. Clearly, these effects occur because the signal-to-noise ratio is low in the frequency region between F1 and F2 when the formants are widely separated.

The effects of the background noise are also great in the high-CF region above the F2 place. For all stimuli, there is a considerable loss of harmonic structure, and, for sufficiently high CF's, a disappearance of peaks near the formant frequencies in favor of low-frequency response components. For /i/ and /u/, the local maxima at the fundamental frequency are also suppressed. In contrast, for /ae/ and all vowels whose F1 is above 0.4 kHz, a peak at the fundamental frequency remains visible. For /u/, strong response components near the fiber CF appear in the CF region between 1.2 and 3 kHz. This feature is also found for /oh/, which has the second lowest F2 after that of /u/. Thus, for high-CF fibers, the power spectra in the noise condition resemble the spectra in response to a broadband noise, so that the high-CF region of responses to vowels is no longer well defined, except, for /ae/, in a narrow CF band near 3 kHz where a spectral peak near F1 remains apparent.

In summary, the basic organization of responses to vowel stimuli into five CF regions generally remains clear in background noise. The main departures from this organization are a reduction in the extent of the F2-dominated region for certain stimuli, and a near take over of the response by the noise for high-CF fibers, and, for spread vowels, for fibers with CF's between F1 and F2. Effects of the background noise on response spectra for the vowels presented at 60 dB SPL follow a similar pattern.

The response component at the fundamental frequency is of particular interest because of its potential role as a cue to voice pitch. Figure 6 shows the normalized amplitude of the fundamental component in the PST histogram plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL in quiet and in background noise. For /i/, the background noise causes a large decrease in the amplitude of the fundamental component throughout the range of CF's except near the formant frequencies. The fundamental remains considerable near the F2 place in the noise condition. These results apply to the other spread vowels /e/ and /ɨ/, though the decrease in the CF region between F1 and F2 is not as strong as for /i/. For the /ae/ stimulus, the noise-induced decrease in the amplitude of the fundamental is not as great as for /i/, and is limited to the CF region near 0.4 kHz and, at 60 dB to the 2-3 kHz region. The fundamental remains large near 3-5 kHz for the 75 dB /ae/ vowel in the noise condition.

The minimal drop in the normalized amplitude of the fundamental in the high-CF region is also found for the more front and open vowels /eh/ and /ax/, which have both relatively large F1 and F2, while the drop in the low-CF region is also found for the other open vowel /a/ at 75 dB SPL. The fundamental of the /u/ stimulus is considerably decreased by the background noise in the high-CF region, and is small throughout the range of CF's in the noise condition. The strong decrease in fundamental is also found for the other back vowels /oh/ and /a/, though the fundamental remains considerable in the high-CF region for the noise condition. Thus many of the temporal cues to fundamental frequency are strongly degraded by noise at a moderate signal-to-noise ratio, particularly for back and spread vowels.

C. Speech processing schemes

Because responses of many fibers show large components at the formant frequencies in response to vowels in noise, one would expect that the processing schemes that were proposed in Chapter I for the estimation of formant frequencies will remain adequate in the noise condition. Figure 7 shows a plot of the Reciprocal of the Mode in the distribution of Intervals between Successive Peaks (RMISP) of the autocorrelation function of the PST histogram plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL, both in quiet and in noise. For all vowels and both stimulus conditions, the RMISP is close to F1 in a wide CF

region centered at the F1 place. For /ae/ and /u/, there is a similar CF region around F2 where the RMISP is close to F2, but the extent of the region is smaller in the noise condition. For /i/, the F2 region is minimal both in quiet and in noise. The effect of noise on the RMISP is maximal in the high-CF region as the peaks of the autocorrelation function become poorly defined and irregularly spaced. For /u/, the RMISP in the noise condition is approximately equal to CF in the frequency range 1-3 kHz. With the possible exception of the second formant of /i/, the formant frequencies could in principle be estimated by picking the two largest modes in the distribution of the RMISP across the entire range of CF's, though the estimation of the second formant might be somewhat more error-prone in the noise condition.

Figure 8 shows a plot of Average Localized Synchronized Measure (ALSM) against filter center frequency for the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL, both in quiet and in noise. The ALSM is plotted for three filtering schemes, narrow-band, broad-comb or narrow-comb. Both in quiet and in noise, the narrow-band ALSM shows peaks at the positions of the high-frequency formants and at the harmonics of the 125-Hz fundamental closest to the low-frequency formants. Non-formant peaks at harmonics 2 and 3 for /ae/, and at 2 F1 for /i/ are also present in both conditions. The presence of noise introduces no additional

non-formant peaks, and even decreases the amplitude of the peak near 2 kHz for /i/. However, the ALSM in noise in considerably larger for /u/ in the 1-3 kHz region where prominent components near CF appear. Overall, the representation of the formants is not degraded by background noise for the narrow-band ALSM.

The broad-comb ALSM also shows peaks near the formant frequencies, except the second formant of /i/, in both conditions. Differences between the noise and quiet conditions are greater than for the narrow-band ALSM, probably because the comb filter extracts the noise-sensitive low-frequency components as well as the more stable components near CF. As a result of these changes, the peak at the second formant of /u/ is less prominent in the noise condition. In the quiet condition, the narrow-comb ALSM shows peaks at basically the same frequencies as the narrow-band ALSM. The background noise reduces the prominence of the F2 peak for /u/, and almost totally suppresses the F2 peak for /i/. Thus, the representation of the formant frequencies in the profile of ALSM against center frequency is more degraded by noise for ALSM's based on comb-filtering schemes than for the narrow-band ALSM.

The narrow-band and narrow-comb ALSM's are also useful for the estimation of the fundamental frequency because they show peaks for many of the low-frequency (<1 kHz) harmonics of the 125-Hz fundamental. Specifically, the ALSM's

have peaks at the frequencies of harmonics 2, 4 and 5 for /i/, at harmonics 2, 3, 4, 6 and 7 for /ae/, and at harmonics 2, 3, 5 and 6 for /u/. Most of these harmonic peaks remain clear in the noise condition, though their amplitude can drop significantly.

In summary, the processing schemes that have been proposed in Chapter I for the estimation of formant frequencies and fundamental frequency seem, in general, to remain useful in background noise, at least for the moderate signal-to-noise ratio that was used. One of the effects of the noise is a decreased prominence of the representation of the second formant of certain vowels for the RMISP of the autocorrelation and the ALSM's based on comb-filtering schemes.

III DISCUSSION

A. Relation to previous studies of stimulus coding

The main effects of moderate-level background noise on responses of auditory-nerve to two-formant vowel stimuli are to (1) suppress the peak in discharge rate at the onset of the stimuli throughout the range of CF's, (2) reduce the amplitude of the response components at the fundamental frequency in CF regions where this component is large in quiet, and (3) to replace components synchronized to formant frequencies by low-frequency components or components near CF

in CF regions where the signal-to-noise ratio is low, particularly in the high-CF region. Many of these effects resemble effects of background noise on responses to simple stimuli (Kiang et al., 1965; Kiang and Moxon, 1974; Rhode et al., 1978; Smith, 1979). The suppression of the peak in discharge rate at the onset of the stimulus has been observed for tone-burst and noise-burst stimuli (Kiang et al., 1965; Smith, 1979). It has been interpreted as being due primarily to short-term adaptation by the continuous background reducing the discharge rates of fibers, and to the background noise driving the fibers at discharge rates close to saturation, so that increases in stimulus amplitude do not cause increases in discharge rate.

It has been suggested previously that response components at the fundamental frequency of vowels originate primarily by rectification of the envelope of the stimulus waveform in the cochlea (Delgutte, 1980; Chapter I). For CF regions in which the signal-to-noise ratio is moderate, the reduction in the amplitude of the fundamental component may be due to the background noise driving fibers close to saturation, so that fluctuations at the fundamental frequency in the stimulus envelope no longer cause fluctuations in discharge rate. For CF regions in which the signal-to-noise ratio is low, the noise suppresses all components synchronized to the vowel, including the fundamental.

Because the signal-to-noise ratio is lowest in the high-frequency region, the noise almost totally suppresses the response components at the formant frequencies in the high-CF region. For the /u/ and /oh/ vowels, which have a very low F2, and for /e/ and /ɨ/, which have a low F1 and a high F2, this suppression is accompanied by a reduction in the extent of the F2-dominated region around the F2 place. For natural vowels, which have more than two formants, the signal-to-noise ratio would be larger in the high-frequency region, so that one would expect the effects of noise for high CF fibers to be lower. Consistent with this notion, Kiang and Moxon (1974) found that high-CF fibers respond clearly to natural vowel stimuli presented in a lowpass background noise falling off rapidly with frequency.

The present results show that the profile of ALSM against center frequency generally shows peaks near the formant frequencies of vowel stimuli presented in background noise for three different filtering schemes. This result extends those of Voigt et al. (1980) for the ALSR, which is a response measure similar to the narrow-band ALSM. In addition, Voigt et al. showed that the ALSR preserves essential information for vowel identification at signal-to-noise ratios as low as -4 dB. Because of their good resistance to noise, the ALSM's remain possible models of central processing, as suggested in Chapter I. However, the ALSM's based on comb-filtering schemes are somewhat more

degraded than the narrow-band ALSM. If these trends continued at lower signal-to-noise ratios, one might have to reject ALSM's based on comb-filtering schemes as possible models of central processing.

In chapter I, it was shown that the narrow-band and narrow-comb ALSM's have clear peaks at many of the low-frequency harmonics of the 125-Hz fundamental frequency of the vowel stimuli. However, the fundamental-frequency parameter entered into the computation of these measures because all processing was based on period histograms. The ALSM's computed in the present study do not suffer from this defect because all processing of PST histograms was done asynchronously with the fundamental frequency. It is reassuring that, both in quiet and in background noise, the narrow-band and narrow-comb ALSM's also show peaks at many of the low-frequency harmonics, though some of them are not as distinct as for the measures computed from period histograms.

In previous studies of responses to speech-like stimuli in background noise, the waveform of the noise was different for each stimulus presentation (Kiang and Moxon, 1974; Voigt et al., 1980). In the present study, computer-generated noise whose waveform is identical for each presentation was used. From the point of view of central processing, pseudo-random noise may be more realistic because the speech sounds have to be recognized from a single presentation of the stimulus for which all fibers are

stimulated by the same acoustic waveform. Thus, spatial summation of responses of different fibers to a vowel in noise is analogous to temporal summation of responses to different presentations of a vowel in pseudo-random noise for the computation of a PST histogram. However, it is in principle possible that results for different samples of pseudo-random noise would be very different. Because the power spectrum of the noise that was used in this study shows no major irregularities, this possibility is unlikely.

B. Relation to psychophysical data

For a 10-dB signal-to-noise ratio, speech intelligibility is nearly the same as in quiet (Miller and Nicely, 1955; Wang and Bilger, 1973; Nooteboom, 1968). Identification errors are limited to a few place-of-articulation errors among stop consonants, and errors involving the weak fricatives /f/, /θ/, and /v/. Identification errors for vowels are obtained only for signal-to-noise ratios lower than -10 dB (Nooteboom,1968). The finding that considerable information about the formant frequencies of vowels remains in the auditory nerve for the noise condition is consistent with the absence of perceptual confusions for vowels. At lower signal-to-noise ratios, confusions between vowels in lowpass noise falling off at 6 dB per octave occur more often among sounds differing by their second formant frequency than among sounds differing by the first formant, particularly for vowels with a low F1 (Pickett,

1957; Nooteboom, 1968). Even for the moderate signal-to-noise ratio that was used, speech processing schemes such as the ALSM's and the RMISP of the autocorrelation show a greater degradation of second-formant information than of first-formant information.

Speech sounds that belong to the same phonetic category are in general more easily confused in background noise than sounds that are less similar in articulation (Miller and Nicely, 1955; Pickett, 1957). In Chapter I, it had been shown that phonetic dimensions for vowels, such as open-close, front-back and spread are correlated with variations in the extent and position of five CF regions over which responses of auditory-nerve fibers to vowels are relatively homogeneous. The present results suggest that responses to vowels that are similar along these phonetic dimensions are affected in a similar manner by the background noise. Such similarities in the response patterns in background noise, if they can be extended to other types of speech sounds and lower signal-to-noise ratios, might constitute the basis for the frequent confusions between sounds belonging to the same phonetic categories.

For the signal-to-noise ratio that was used, listeners still hear a clear pitch at the fundamental frequency of the vowel stimuli. In contrast, the amplitude of the fundamental response component is considerably decreased through wide ranges of CF's, so that many of the periodicity

cues to fundamental frequency are degraded by background noise. However, the profile of ALSM against center frequency for the narrow-band and narrow-comb filtering schemes contains information about the fundamental frequency that is more resistant to noise. These results suggest that perception of the low pitch of a complex periodic tone is likely to be accomplished by a "place" mechanism, in which the fundamental is estimated from peaks at positions of the low-frequency harmonics in a representation of the stimulus spectrum such as the ALSM, rather than by a "periodicity" mechanism according to which periodicities at the fundamental frequency would be detected in the discharge patterns of auditory fibers. This conclusion fits well with current views of pitch perception (Plomp, 1976; Goldstein, 1978).

In chapter III, it was shown that the response patterns of auditory-nerve fibers in specific CF regions show a peak in discharge rate when an abrupt increase in amplitude or change in spectral characteristics occurs in the speech stimulus. It was suggested that these peaks in rate could serve as markers to phonetically-important events in the spatio-temporal pattern of discharges, and that the detailed characteristics of these peaks provide information about many phonetic distinctions among consonants. For instance, it was shown that the distinction between /$\overset{v}{s}$/ and /$\overset{v}{c}$/ could be made on the basis of the prominence of the onset peak in discharge rate in certain CF regions. The present results show that the

peak in discharge rate at the onset of vowel stimuli is almost totally suppressed by background noise throughout the entire range of CF's. The data of Kiang and Moxon (1974) show that many of the short-term variations in discharge rate in response to a speech utterance are flattened out in the presence of background noise at a comparable speech-to-noise ratio. Thus, one would expect that the distinction between /ʃ/ and /ʧ/ would be difficult to make in moderate-level background noise. Indeed, the results of Wang and Bilger (1973) show that /ʃ/ in background noise is more often confused with /ʧ/ than with any other consonant.

More generally, one would expect that even moderate-level background noise would largely eliminate the markers to rapid changes in the short-time spectrum of speech from the spatio-temporal pattern of auditory-nerve discharges. In contrast, our results and those of Voigt et al. (1980) suggest that the fine time patterns of discharge would still contain information about the important features of the stimulus spectrum. It is possible that the confusions among consonants at moderate signal-to-noise ratios are due to the inability of listeners to detect where in the spatio-temporal pattern of discharge the phonetically-important information should be sampled, rather than to the absence of such information.

FIGURE CAPTIONS

Fig. 1

Harmonic spectra of the /i/, /ae/ and /u/ stimuli superimposed on the power spectrum of the background noise. The magnitude scale is computed in pressure units for the 60 dB SPL stimuli, with the assumption that the transfer characteristics of the acoustic system are flat. The power spectrum of the noise is computed by averaging periodograms of overlapping 12.8-ms segments (Oppenheim and Schafer, 1975). The noise level is given for bands of frequencies whose width is equal to the 125-Hz fundamental of the vowel stimuli. The positions of the formant frequencies F1 and F2 are marked by dotted lines. The first formant frequencies of the /i/, /ae/ and /u/ stimuli are 0.25, 0.8 and 0.3 kHz, respectively. Their second formants are 3.2, 1.8 and 0.7 kHz, respectively.

Fig. 2

Normalized power spectra of PST histograms for auditory-nerve fibers with 5 different CF's in response to the /i/ and /ae/ stimuli presented at 75 dB SPL, both in quiet and in noise. The power spectra are normalized by the square of the mean discharge rate, so that the vertical scale has units of Hz^{-1}. The approximate CF of each fiber is listed at the left. Dashed lines show the positions of the formant frequencies, and an arrow marks the position of the fiber CF along the frequency axis.

Fig. 3

Ratio of onset rate to steady-state rate plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL, both in quiet and in background noise. The discharge rates are measured as described in Sec. IB. Each symbol represents the ratio of discharge rates for one auditory-nerve fiber. The places of the formant frequencies along the CF dimension are indicated by dashed lines.

Fig. 4

Correlation index between the PST histogram in quiet and the histogram in noise plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL. Each circle represents the correlation index for one fiber. The places of the formant frequencies along the CF dimension are indicated by dashed lines.

Fig. 5

Pseudo-perspective representation of normalized band-average power spectra for 0.55-octave bands of CF's in response to the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL, both in quiet and in background noise. Each normalized band-average power spectrum is plotted with frequency along the oblique axis and response magnitude along the vertical axis. Spectra are shown for CF bands whose center frequencies are sampled every 1/4 octave. Spectrum points with an amplitude lower than $0.8\ 10^{-3}\ Hz^{-1}$ are omitted for clarity. The places of the formant frequencies along the CF dimension are indicated by

oblique dashed lines, and the positions of the formants along the frequency axis are marked by horizontal dashed lines. The curved dashed lines is the locus of points for which frequency is equal to CF.

Fig. 6

Synchronization index at the fundamental frequency plotted against center frequency of 0.55-octave CF bands in response to the /i/, /ae/ and /u/ stimuli presented at 60 and 75 dB SPL, both in quiet and background noise. For each fiber, the synchronization index is the amplitude of the 125-Hz Fourier component of the PST histogram normalized by the mean discharge rate (Johnson, 1980). The synchronization indices are then band-averaged in the same manner as the power spectra. The places of the formant frequencies along the CF dimension are indicated by dashed lines.

Fig. 7

Reciprocal of the Mode of the Intervals between Successive Peaks (RMISP) of the autocorrelation function of the PST histogram plotted against CF for the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL in quiet and in background noise. Each circle represents the RMISP for one fiber. The places of the formant frequencies along the CF dimension are indicated by vertical dashed lines, and horizontal dashed lines are drawn at the ordinates corresponding to intervals of 1/F1 and 1/F2.

Fig. 8

Average Localized Synchronized Measure (ALSM) plotted against filter center frequency for three different filtering schemes in response to the /i/, /ae/ and /u/ stimuli presented at 75 dB SPL in quiet and in background noise. For all three filtering schemes, the continuous lines in the lower part of each panel represent the ALSM in quiet, and the dotted lines represent the ALSM obtained in background noise. The continuous line on the upper part of each panel is the ALSM in quiet minus the ALSM in noise. The positions of the formant frequencies are indicated by vertical dashed lines.

A. ALSM for a 1/6-octave Gaussian bandpass filtering scheme.

B. ALSM for a cosinusoidal comb filtering scheme.

C. ALSM for a 1/6-octave comb filtering scheme.

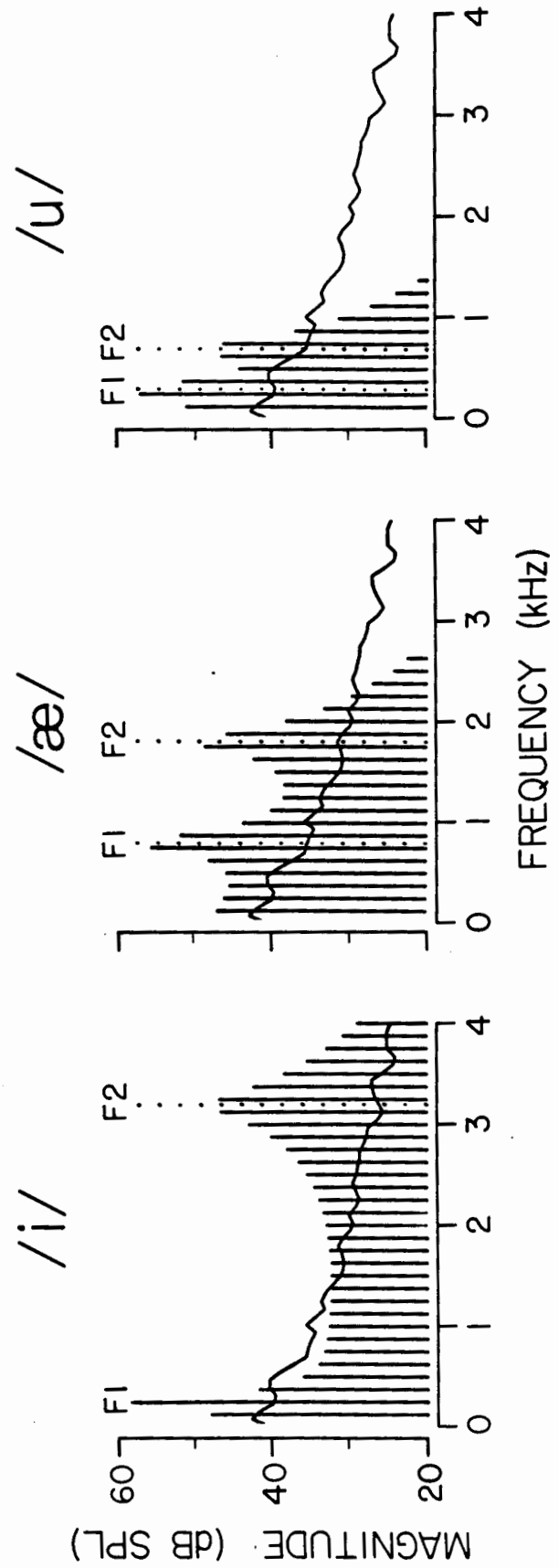


Fig. 1

Fig. 2

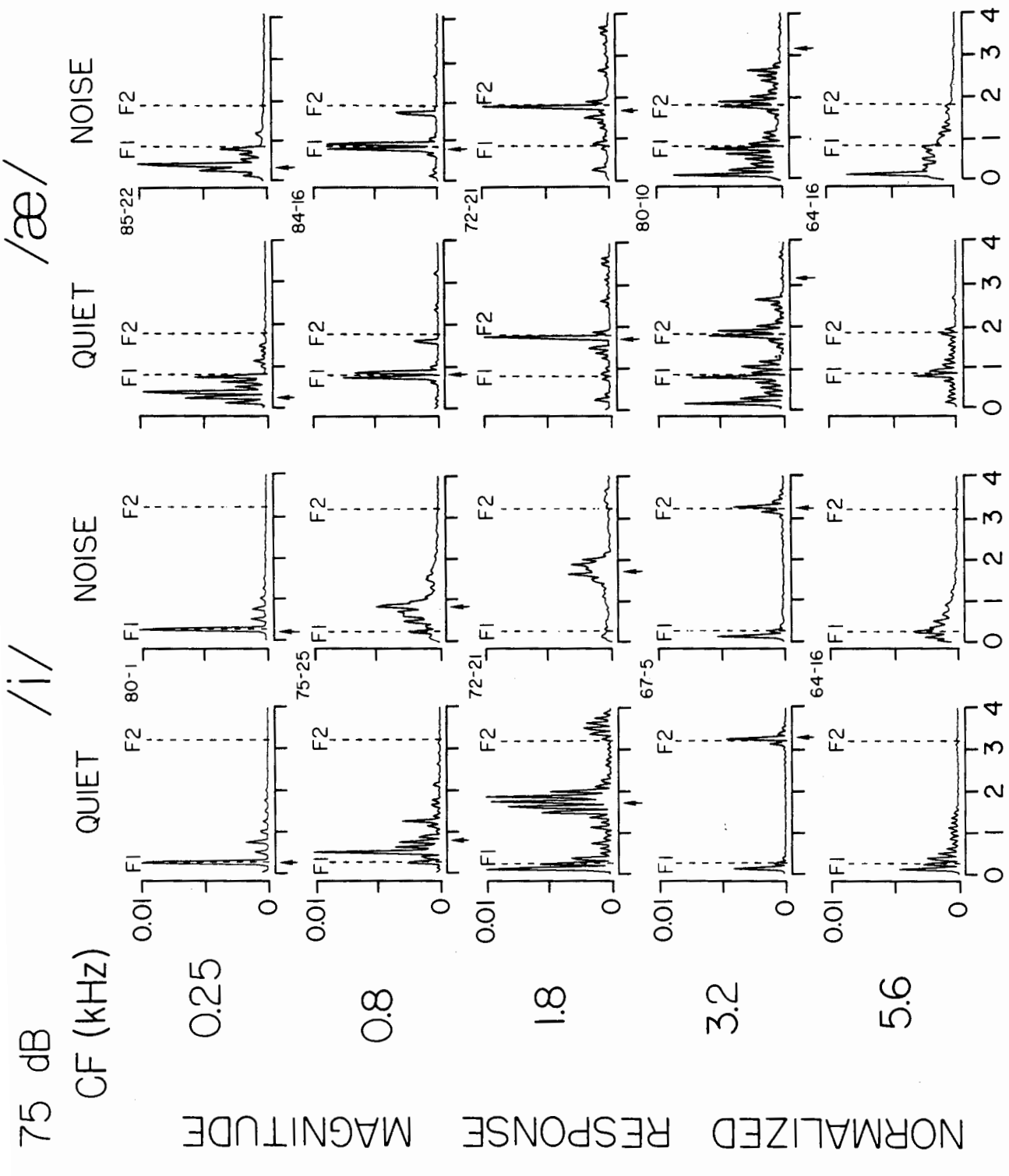
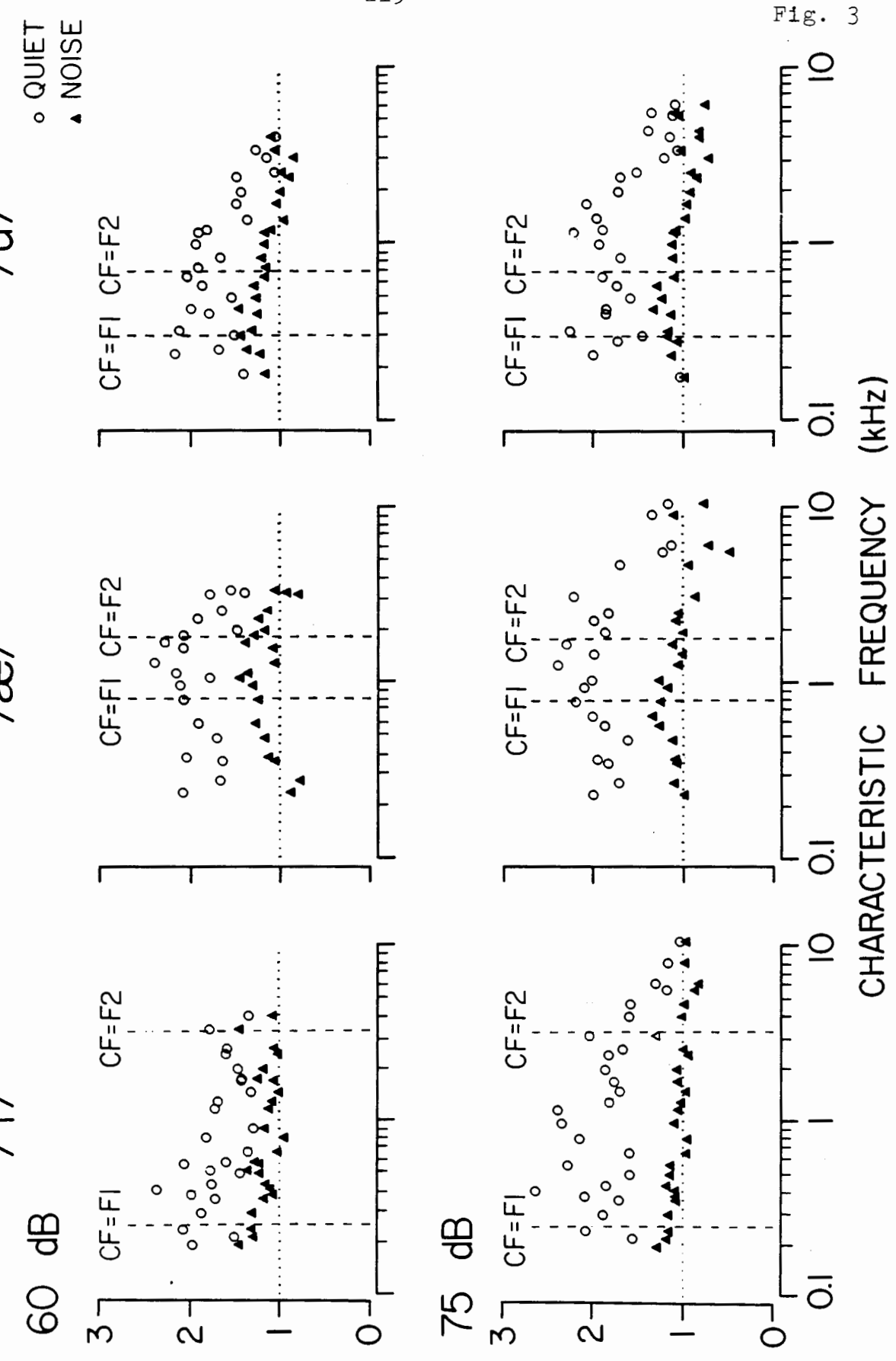

/i/
/æ/
75 dB
CF (kHz)
QUIET
NOISE
0.25
0.8
1.8
3.2
5.6
0.01
0
F1
F2
80-1
75-25
72-21
67-5
64-16
85-22
84-16
72-21
80-10
64-16
0 1 2 3 4
FREQUENCY (kHz)
NORMALIZED RESPONSE MAGNITUDE

Fig. 3


/i/
/æ/
/u/
60 dB
75 dB
CF=F1
CF=F2
CF=F1 CF=F2
QUIET
NOISE
ONSET RATE/STEADY-STATE RATE
CHARACTERISTIC FREQUENCY (kHz)
0
1
2
3
0.1
1
10

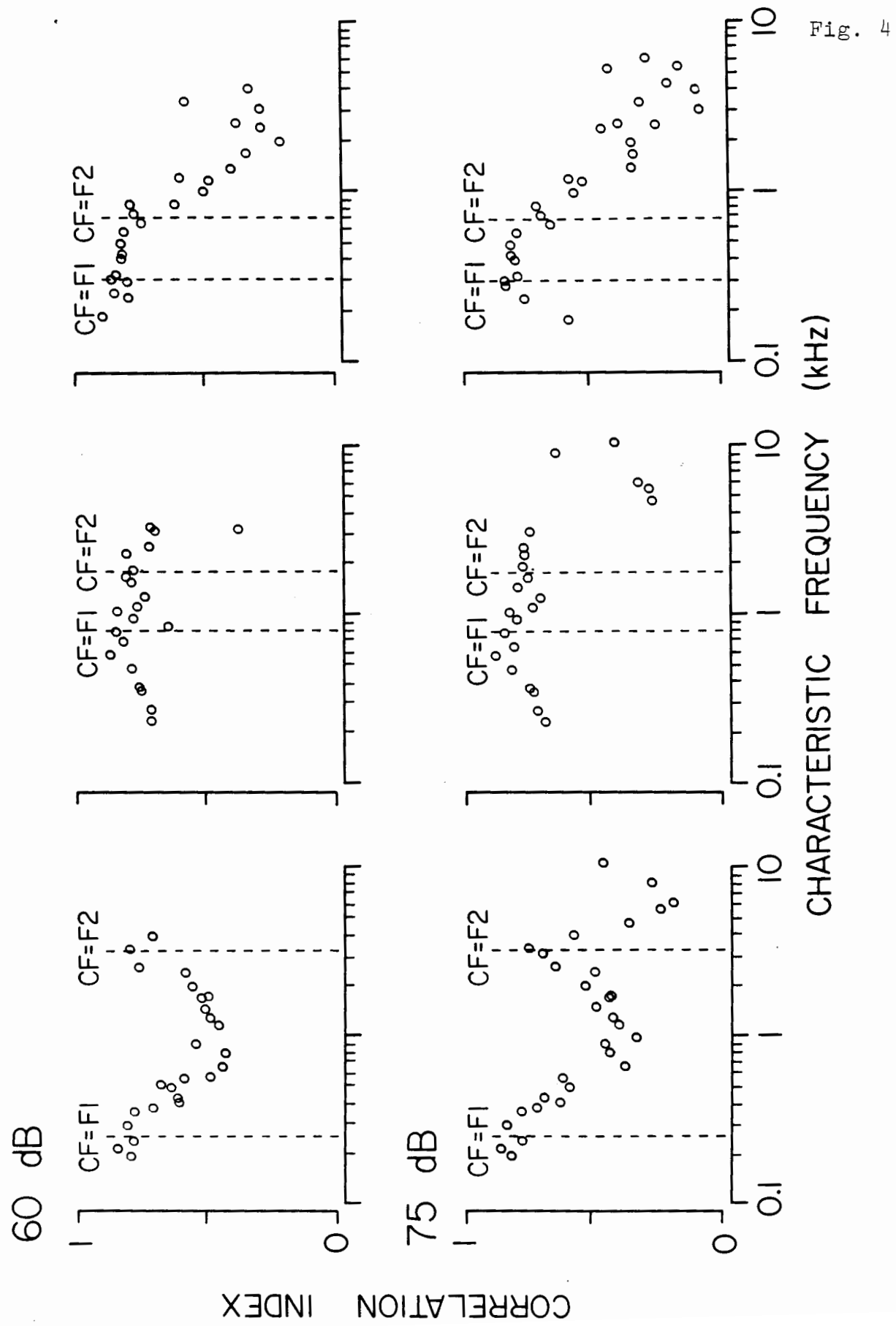

/i/
/æ/
/u/
60 dB
75 dB
CF=F1
CF=F2
1
0
0.1
1
10
CORRELATION INDEX
CHARACTERISTIC FREQUENCY (kHz)

Fig. 4

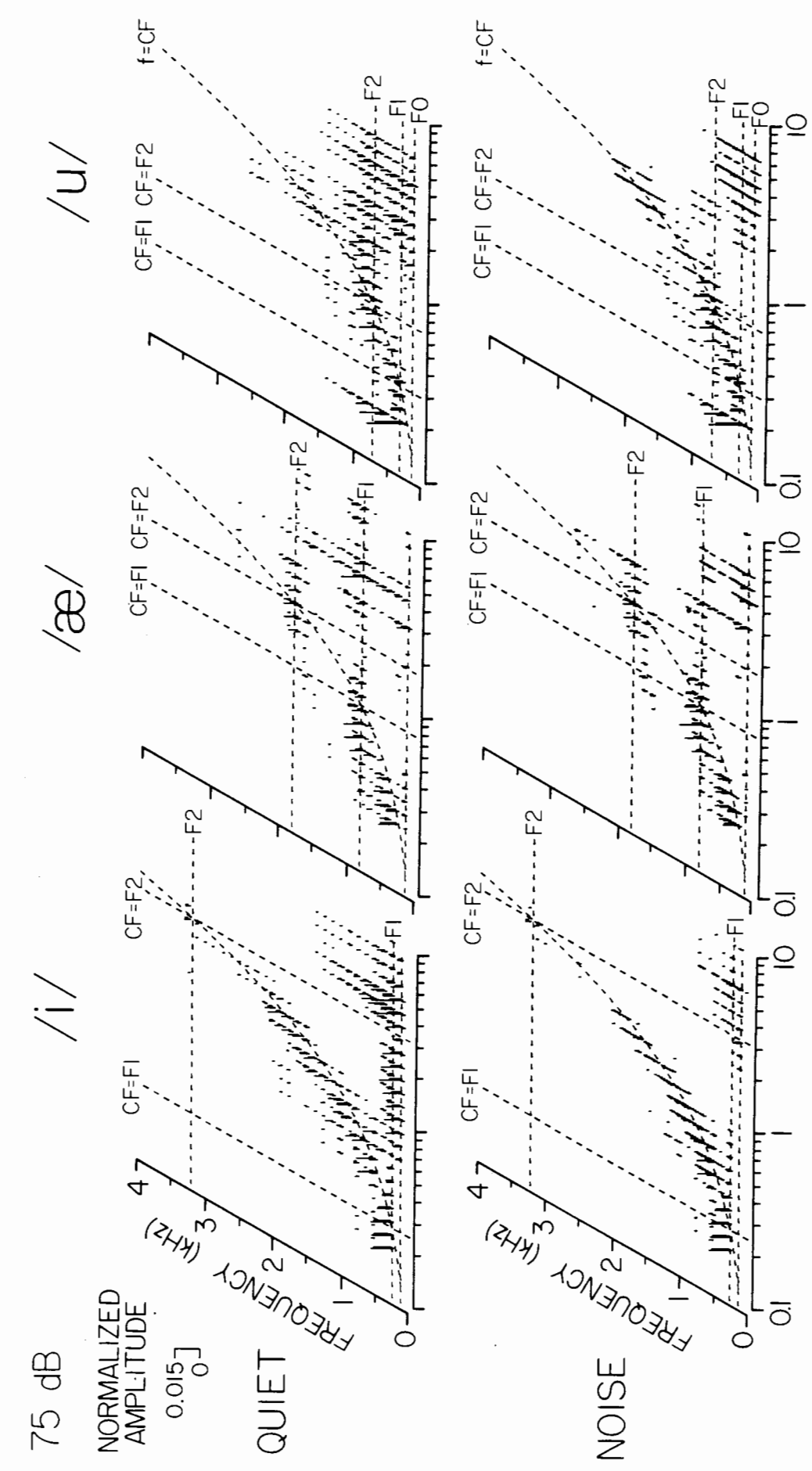

75 dB
/i/
/æ/
/u/
NORMALIZED AMPLITUDE
0.015
0
QUIET
NOISE
FREQUENCY (kHz)
0
1
2
3
4
CF=F1
CF=F2
CF=F2,
F1
F2
F0
f=CF
0.1
1
10
CHARACTERISTIC FREQUENCY (kHz)

Fig. 5

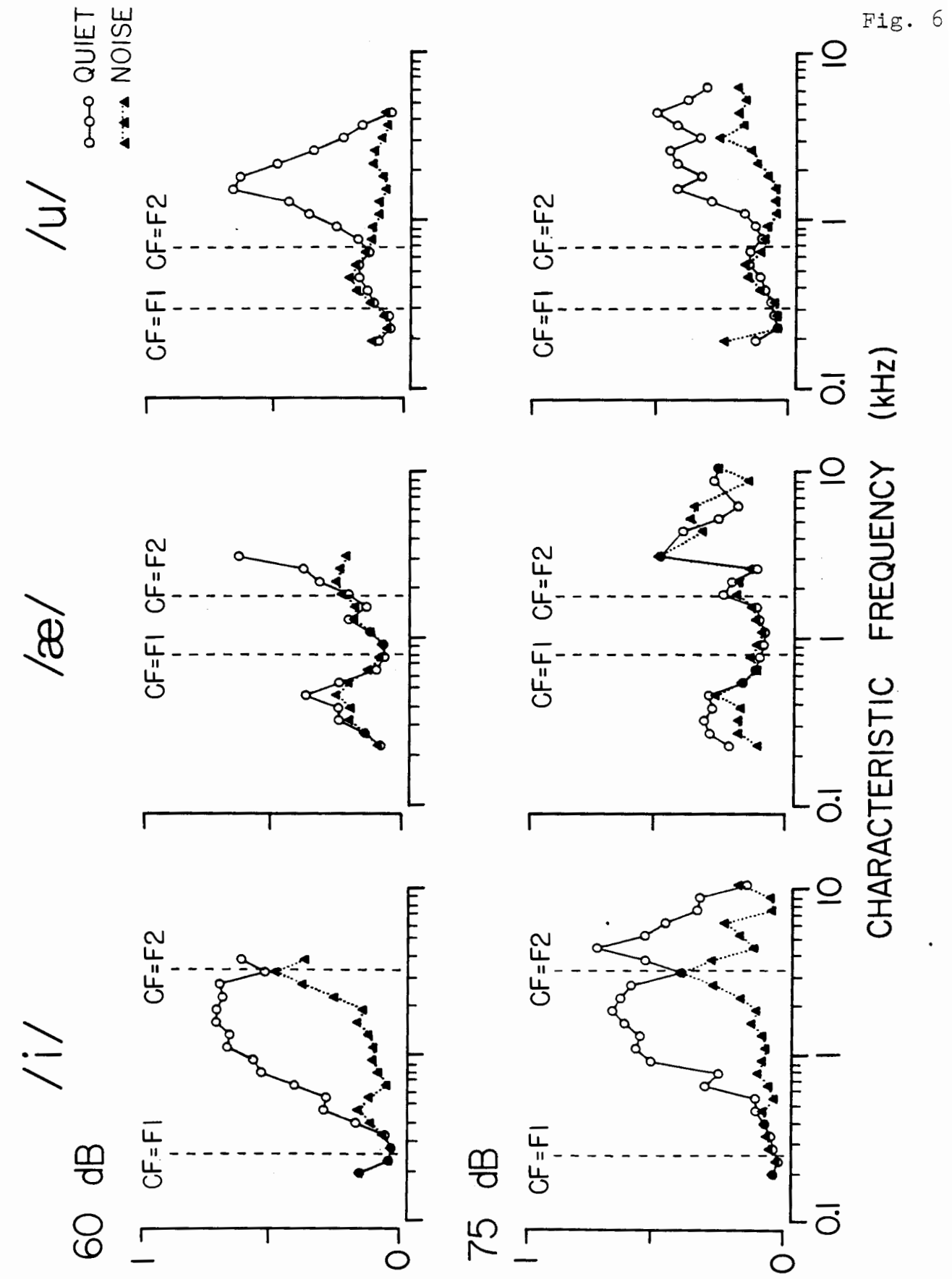

/i/
/æ/
/u/
o—o—o QUIET
▲▲▲ NOISE
60 dB
75 dB
CF=F1
CF=F2
SYNCHRONIZATION INDEX AT FO
CHARACTERISTIC FREQUENCY (kHz)
0.1
1
10
0
1

Fig. 6

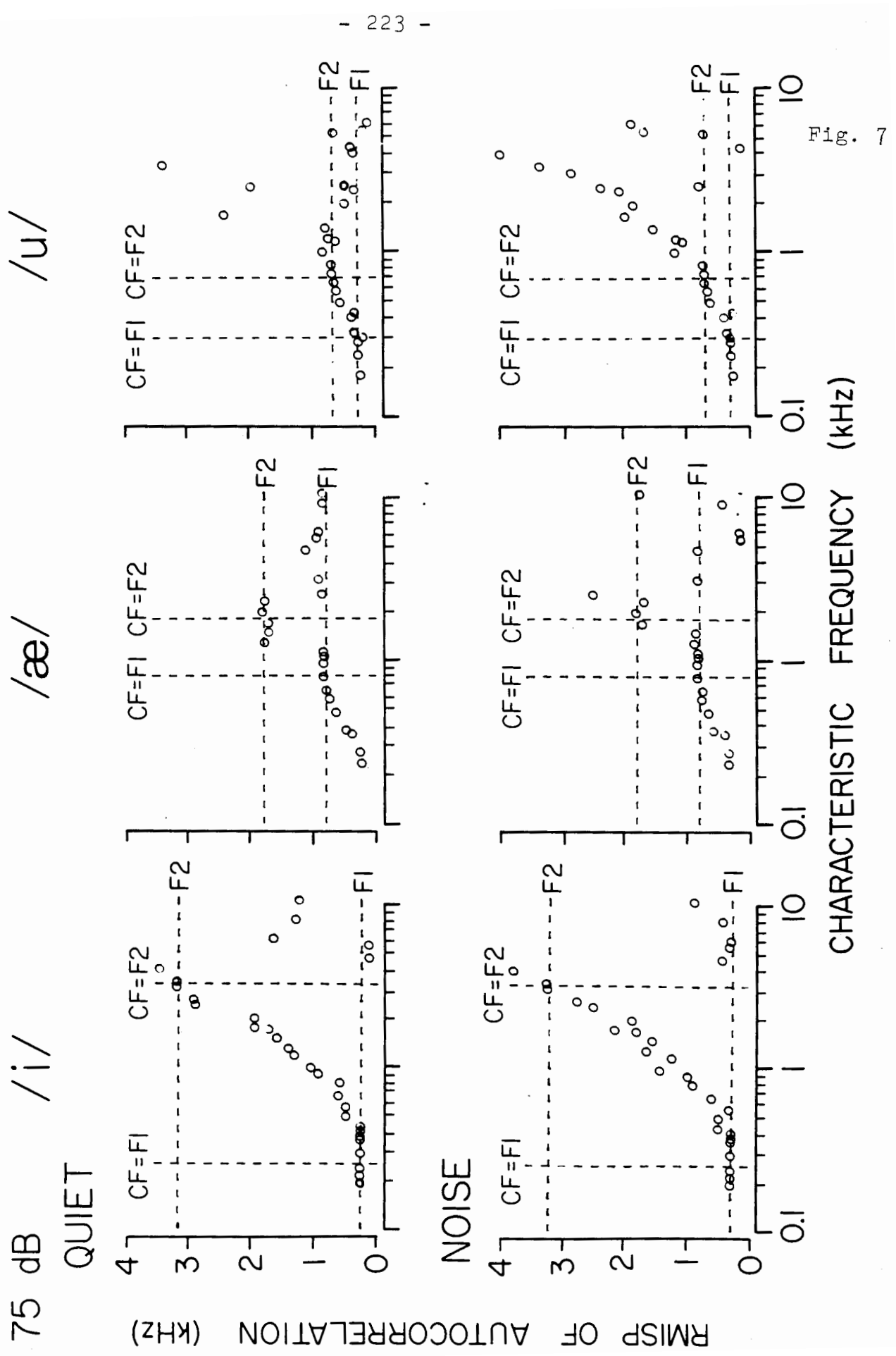

75 dB
/i/
/æ/
/u/
QUIET
NOISE
CF=F1
CF=F2
F1
F2
RMSP OF AUTOCORRELATION (kHz)
CHARACTERISTIC FREQUENCY (kHz)
0
1
2
3
4
0.1
1
10

Fig. 7

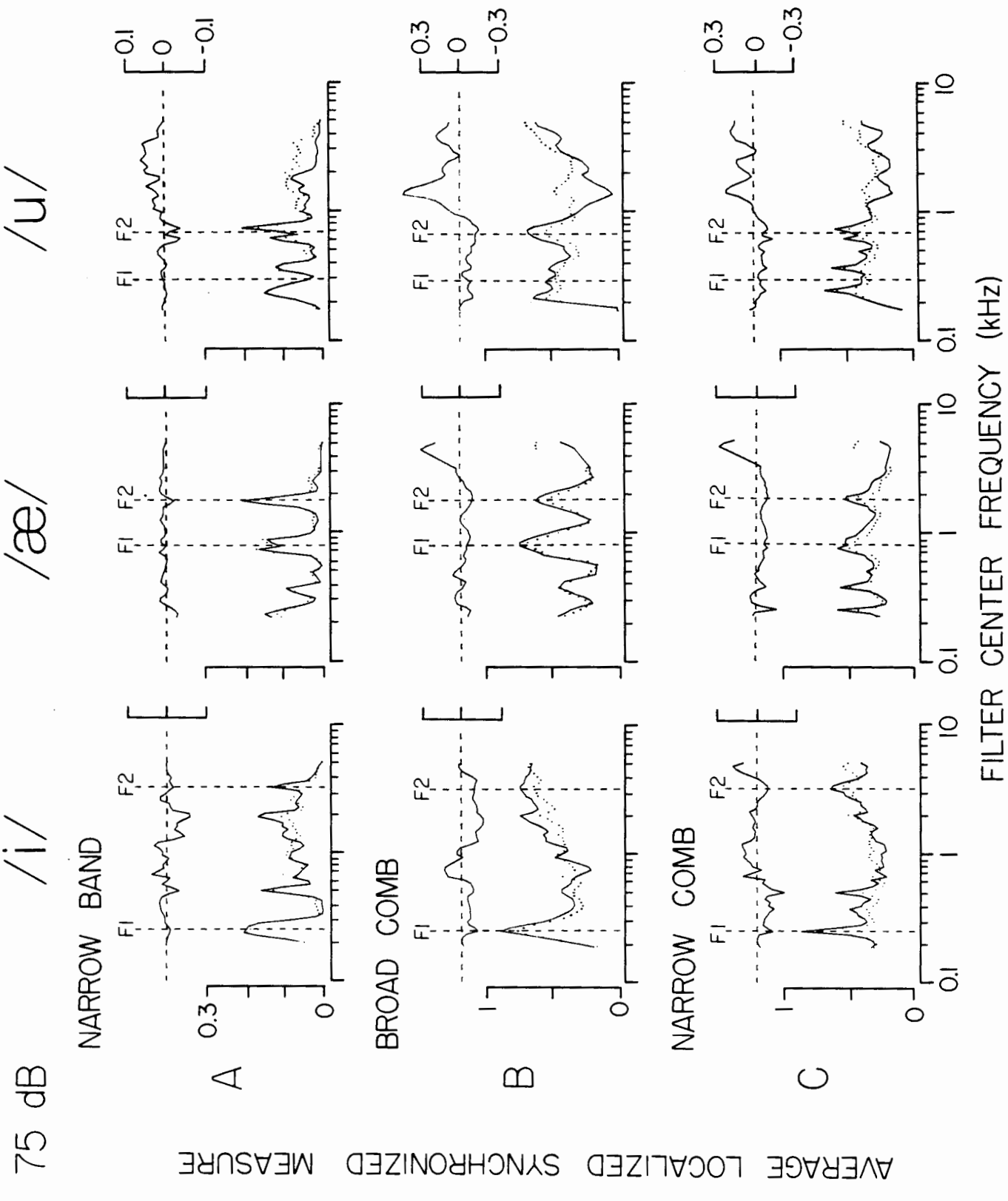

75 dB
/i/
/æ/
/u/
NARROW BAND
BROAD COMB
NARROW COMB
A
B
C
F1
F2
0.3
0
0.1
-0.1
1
10
AVERAGE LOCALIZED SYNCHRONIZED MEASURE
FILTER CENTER FREQUENCY (kHz)

Fig. 8

CONCLUDING REMARKS

The main purpose of this thesis was to describe how acoustic characteristics that are important for phonetic distinctions are represented in the discharge patterns of auditory-nerve fibers. Specifically, we have investigated distinctions among vowels, place-of-articulation distinctions among fricatives, distinctions between sounds that have an abrupt onset and sounds with a more gradual onset, and certain distinctions between stop, nasal, and fricative consonants.

Vowel identity depends primarily on the formant frequencies. For vowel stimuli at conversational speech levels, the largest components in the response patterns of the majority of auditory-nerve fibers are in the vicinity of the formant frequencies. The formant frequencies appear as local maxima in a profile obtained by processing the response patterns by a filter bank that extracts the response components near the CF of each fiber. The output of this filter bank also shows peaks for center frequencies corresponding to the low-frequency harmonics of the fundamental frequency. These peaks could be used for the identification of the fundamental frequency and for the distinction between voiced and voiceless sounds. These results may apply to sonorant sounds in general provided the formant changes are slow enough.

The average discharge rates in response to sonorants are largest in the CF region below 1 kHz, whereas for obstruents, the largest rates are in the CF region above 1 kHz. The profile of average discharge rate against CF contains the essential information for place-of-articulation distinctions among fricatives. The CF regions in which the largest discharge rates are found correspond roughly with the frequency regions in which the stimuli have most of their energy. In addition, for "compact" obstruents such as /x/ and presumably /k/ and /g/, the output of a filter bank extracting components near CF shows a local maximum at the frequency of the main spectral prominence near 2 kHz.

Many other phonetic distinctions among consonants seem to involve the peaks in discharge rate that occur in fiber response patterns when the speech signal has a rapid change in amplitude or spectrum. For sounds with a gradual onset, like /$\overset{\vee}{\text{s}}$/, there are CF regions in which discharge rate at the onset of the stimulus is lower than the rate 30-odd ms later, whereas for abrupt sounds like /$\overset{\vee}{\text{c}}$/, there is a peak in discharge rate at the onset throughout a broad range of CF's. Fibers in the low-CF region show a prominent peak in rate at the release of a voiced stop consonant, whereas there is no such peak at the release of a nasal. At the release of the fricatives /$\overset{\vee}{\text{s}}$/ or /s/ followed by a vowel, the peaks in discharge rates are more restricted to the low-CF region than at the release of the stop consonant /d/. Though no

experiments have been made, one would expect that for a stop consonant containing a burst, there would be two peaks in discharge rate, one coincident with the consonantal release, and another one, more prominent in the low-CF region, coincident with the onset of voicing. For aspirated stops, such as the English /p/, /t/ and /k/, the first peak would be well separated in time form the second peak, whereas for unaspirated stops, such as the English /b/, /d/ and /g/, the second peak would occur before the rapid decay in discharge rate following the first peak is terminated.

Clearly, considerable work remains to be done to determine auditory-nerve correlates of phonetic distinctions more precisely. In particular, the validity of the proposed correlates in situations more closely resembling free-running speech and in a greater variety of phonetic environments has to be verified.

Data on responses of auditory-nerve fibers to speech-like stimuli can be used to guide the development of functional models of peripheral auditory processing. Many of the concepts that are commonly used to describe responses to simple stimuli, such as frequency selectivity, rate saturation, rectification, synchrony of discharges, suppression, and short-term adaptation have proved useful in predicting trends in response to the speech-like stimuli. More quantitative predictions would require a model of peripheral auditory processing in which all of these

functional properties would interact. A number of efforts have been made to develop models (Weiss, 1966; Siebert, 1968; Colburn, 1973; Hall, 1974,1977; Johnson, 1974; Schroeder and Hall, 1974; Smith and Zwislocki, 1975; Dolmazon, 1980), but many of these models apply only to a fraction of the response properties of auditory-nerve fibers or have not been examined from the point of view of speech processing. A model of peripheral auditory processing that would reproduce the main features of the responses of auditory-nerve fibers to speech-like stimuli would be valuable in testing different speech-processing schemes to extract the phonetically-important information.

The results on the coding of speech in the auditory nerve have been used as a basis to propose models of central processing that would be consistent with physiological mechanisms and psychophysical data. Because the profile of average discharge against CF for intense vowel stimuli does not show peaks at the formant frequencies (Sachs and Young, 1979), several possible processing schemes based on fine time patterns of discharge have been proposed for the estimation of formant frequencies and fundamental frequency. The most parsimonious of these schemes is a bank of central filters whose center frequencies coincide with fiber CF's. This processing scheme is consistent with the psychophysical concept of critical bands, and can be realized using only signal-processing operations that are known to exist at

synapses in the central nervous system. If such central filters exist, one would expect to find somewhere in the central auditory system a homogeneous population of cells that would show sharp frequency selectivity in response to tones, and for which profiles of average discharge rate against CF for vowel stimuli would show clear peaks at the formant frequencies and smaller peaks at the low-frequency harmonics of the fundamental frequency. However, this scheme fails to extract the information that is important for place-of-articulation distinctions among fricative consonants. In contrast, it has been shown that the profile of average discharge rate against CF would be sufficient to make phonetic distinctions between fricatives. Thus, two separate processing schemes seem to be necessary to provide the essential spectral information for distinctions between speech sounds: a central filter bank for the low-frequency information of sonorants and average discharge rate for the high-frequency information of obstruents. This different treatment of low-frequency and high-frequency information is a familiar concept in studies of sound localization (Stevens and Newman, 1936). Single-unit recordings in the brainstem show that, at least in the cat, there are separate auditory nuclei in which units have predominantly low or high CF's (Guinan et al., 1972). The function of these nuclei may include the extraction of properties used in speech communication as well as sound localization.

Several examples of correlation between psychophysical data and responses of auditory-nerve fibers to speech-like stimuli have been given. For instance, for vowel stimuli, the harmonics of the fundamental frequency at which little synchrony of discharges was observed correspond well with harmonics whose deletion from the stimuli fails to be detected by human subjects (Carlson et al., 1979). Our results also suggest that the two most intense harmonics within a spectral peak are the most important for the perception of a low-frequency formant because the response components at these harmonics are considerably more prominent than those at neighboring harmonics. This finding is in good agreement with psychophysical data (Chistovich, 1971; Carlson et al., 1975), but further experiments in which the fundamental frequency is manipulated systematically need to be done. In another example, it was suggested that two apparently unrelated acoustic cues to the distinction between /$\overset{v}{s}$/ and /$\overset{v}{c}$/ would produce similar response patterns at the level of the auditory nerve. Other psychophysical phenomena for which auditory-nerve data would be worthwhile include the minimum height of a formant peak that is necessary for its detection, the effect of increasing formant bandwidths, the effects of relative formant amplitudes and overall spectral tilt, the importance of phase manipulations, and masking by background noise. While some of the psychophysical studies that examine these effects for vowels have been carried out (Carlson et al., 1979; Chistovich and Lublinskaya, 1979),

these experiments need to be extended to include dynamic speech-like stimuli.

Some of our results may contribute to a specification of the properties of the auditory system that constrain the selection of classes of sounds that can be used for speech communication. It has been suggested that the low-frequency information of sonorants and the high-frequency information of obstruents are processed by different mechanisms. Given that all languages have a contrast between obstruents and sonorants, it is possible that this dichotomy in auditory processing was a major factor in the evolution of phonetic systems. Another universal property of languages is that many sounds have an abrupt increase in amplitude or a rapid change in spectral characteristics. The results of the present study suggest that auditory-nerve fibers respond to such transients with a prominent peak in discharge rate. It has been suggested that such peaks in rate could be used by the central processor as pointers to regions of the spatio-temporal pattern of discharges that are rich in phonetic information. It seems plausible that the presence of rapid changes in speech and the concentration of phonetic information near these acoustic events evolved as a result of their distinctiveness to the auditory system. Another result of this study is that, the distribution of spectral response components for vowel stimuli is concentrated around the formant frequencies, whereas the distribution for broad-band

noise stimuli extends over a wide frequency range. It is possible that the formant structure of speech evolved to facilitate distinction between sounds used for speech communication and environmental noise.

REFERENCES

Arnesen, A.R., and Osen, K.K. (1978). "The cochlear nerve in the cat: topography, cochleotopy, and fiber spectrum," J. Comp. Neurol. 178, 661-678.

Arthur, R.M. (1976). "Harmonic analysis of two-tone discharge patterns in cochlear nerve fibers," Biol. Cybernetics 22, 21-31.

Baker, J.M. (1975). "A new time-domain analysis of human speech and other complex waveforms," Ph. D. thesis, Carnegie Mellon U., Pittsburgh.

Bladon, R.A.W., and Fant, G. (1978). "A two-formant model and the cardinal vowels," Speech Transmission Laboratory QPSR 1 (Royal Inst. of Technol., Stockholm), pp 1-8.

Bladon, R.A.W. and Lindblom, B. (1981). "Modeling the judgement of vowel quality differences," J. Acoust. Soc. Am. 69, 1414-1422.

Blumstein, S.E., and Stevens, K.N. (1979). "Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants," J. Acoust. Soc. Am. 66, 1001-1017.

Brugge, J.F., Anderson, D.J., Hind, J.E. and Rose, J.E. (1969). "Time stucture of discharges in single auditory-nerve fibers of the squirrel monkey in response to complex periodic sounds," J. Neurophysiol. 32, 386-401.

Carlson, R., Fant, C.G.M., and Granstrom, B. (1975). "Two-formant models, pitch and vowel perception," in Auditory Analysis and Perception of Speech, edited by C.G.M. Fant and M.A.A. Tatham (Academic, London), pp 55-82.

Carlson, R., and Granstrom, B. (1980). "Model predictions of vowel dissimilarity," Speech Transmission Laboratories QPSR 3-4 (Royal Institute of Technology, Stockholm), pp 84-104.

Carlson, R., Granstrom, B., and Klatt, D.H. (1979). "Vowel perception: The relative perceptual salience of selected spectral and waveform manipulations," Speech Transmission Laboratories QPSR 3-4 (Royal Institute of Technology, Stockholm), pp 73-83..

Chistovich, L.A. (1971). "Problems of speech perception," in Form and Substance, edited by L. L. Hammerich, R. Jakobson and E. Zwirner (Akademisk Forlag, Copenhagen), pp 83-93.

Chistovich, L.A. and Lublinskaya, V.V. (1979). "The 'center ofgravity' effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel-like stimuli," Hearing Res. 1, 185-195.

Chomsky, N., and Halle, M. (1968). The Sound Patterns of English (Harper and Row, New York).

Clopton, B.M., Winfield, J.A., and Flammino, F.J. (1974). "Tonotopic organization: review and analysis," Brain Res. 76, 1-20.

Colburn, H.S. (1973). "Theory of binaural interaction based on auditory-nerve data: I. General strategy and preliminary results on interaural discrimination," J. Acoust. Soc. Am. 54, 1458-1470.

Colburn, H.S., and Durlach, N.I. (1978). "Models of binaural interaction," in Handbook of Perception, Vol IV, edited by (Academic, New York), pp 467-518.

Cooper, F.S., Delattre, P.C., Liberman, A.M., Borst, J.M., and Gerstman, L.J. (1952). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. 37, 318-325.

Cutting, J.E., and Rosner, B.S. (1974). "Catergories and boundaries in speech and music," Percept. Psychophys. 16, 564-570.

De Boer, E. and Kuyper, P. (1968). "Triggerred correlation," IEEE Trans. on Biomed. Engin. BME-15, 169-179.

De Boer, E., and De Jongh, H.R. (1978). "On cochlear encoding: potentialities and limitations of the reverse-correlation technique," J. Acoust. Soc. Am. 63, 115-135.

Delgutte, B. (1980). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am. 68, 843-857.

Dewson, J.H. III (1968). "Efferent olivocochlear bundle: some relationship to stimulus discrimination in noise," J. Neurophys. 31, 122-130.

Dolmazon, J.-M. (1980). "Contribution aux recherches sur l'appareil auditif: Elaboration et exploitation d' un modele de fonctionnement du systeme peripherique," Doctoral disssertation, Grenoble University.

Dorman, M.F., Raphael, L.J. and Liberman, A.M. (1979). "Some experiments on the sound of silence in phonetic perception," J. Acoust. Soc. Am. 65, 1518-1532.

Dorman, M.F., Studdert-Kennedy, M. and Raphael, J. (1977). "Stop-consonant recognition: Release bursts and formant transitions as functionally-equivalent, context-dependent cues," Percept. Psychopphys. 22, 109-122.

Dunn, H.K, and White, S.D. (1940). "Statistical measurements on conversational speech," J. Acoust. Soc. Am. 11, 278-288.

Eggermont, J.J. (1976). "Analysis of compound action potential responses to tone bursts in the human and guinea pig cochlea," J. Acoust. Soc. Am. 60, 1132-1139.

Evans, E.F. (1975). "Cochlear nerve and cochlear nucleus," in Handbook of Sensory Physiology, Vol. V/2: Auditory System, edited by W.D. Keidel and W.D. Neff (Springer Verlag, Berlin), pp 1-108.

Evans, E.F. (1977). Frequency selectivity at high signal levels of single units in the cochlear nerve and nucleus," in Psychophysics and Physiology of Hearing, edited by E.F. Evans and J.P. Wilson (Academic, New York), pp 185-192.

Fant, C.G.M. (1960). Acoustic Theory of Speech Production (Mouton, The Hague).

Fant, C.G.M. (1973). Speech Sounds and Features (MIT Press, Cambridge).

Flanagan, J.L. (1972). Speech Analysis, Synthesis and Perception (Springer-Verlag, New York).

Gerstein, G.L., and Kiang, N.Y.S. (1960). "An approach to the quantitative analysis of electrophysiological data from single neurons," Byophys. J. 1, 15-28.

Goldstein, J.L. (1978). "Mechanisms of signal analysis and pattern perception in periodicity pitch," Audiol. 17, 421-445.

Goldstein, J.L., and Srulovicz, P. (1977). "Auditory-nerve spike intervals as an adequate basis for aural frequency measurements," in Psychophysics and Physiology of Hearing, edited by E.F. Evans and J.P. Wilson (Academic, New York), pp 337-346.

Guinan, J.J.Jr., Norris, B.E., and Guinan, S.S. (1972). "Single auditory units in the superior olivary complex II: Locations of unit categories and tonotopic organization," Intern. J. Neuroscience 4, 147-166.

Gurlekian, J.A. (1981). "On the recognition of the Spanish fricatives /s/ and /f/," J. Acoust. Soc. Am., in press.

Hall, J.L. (1974). "Two-tone distortion products in a nonlinear model of the basilar membrane," J. Acoust. Soc. Am. 56, 1818-1828.

Hall, J.L. (1977). "Two-tone suppression in a nonlinear model of the basilar membrane," J. Acoust. Soc. Am. 61, 802-810.

Halle, M., Hughes, G.W. and Radley, J.-P. A. (1957). "Acoustic properties of stop consonants," J. Acoust. Soc. Am. 29, 107-116.

Harris, D.M., and Dallos, P. (1979). "Forward masking of auditory-nerve fiber responses," J. Neurophysiol. 42, 1083-1107.

Harris, K.S. (1958). "Cues for the discrimination od American English fricatives in spoken syllables," Langu. Speech 1, 1-7.

Hashimoto, T., Katayama, Y., Murata, K., and Taniguchi, L. (1975). "Pitch synchronous response of cat cochlear nerve fibers to speech sounds," Jap. J. Physiol. 25, 633-644.

Heinz, J.M., and Stevens, K.N. (1961). "On the properties of voiceless fricative consonants," J. Acoust. Soc. Am. 34, 179-188.

Jakobson, R., Fant, C.G.M., and Halle, M. (1952). Preliminaries to Speech Analysis (MIT Press, Cambridge).

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

Johnson, D H. (1974). "The response of single auditory-nerve fibers in the cat to single tones: Synchrony and average discharge rate," Ph.D. Thesis, MIT, Cambridge.

Johnson, D.H. (1978). "The relationship of post-stimulus time and interval histograms to the timing characteristics of spike trains," Byophys. J. 22, 412-430.

Johnson, D.H. (1980). "The relationship between spike rate and synchrony in the response of auditory-nerve fibers to single tones," J. Acoust. Soc. Am. 68, 1115-1122.

Jones, D. (1956). Outline of English Phonetics (Heffer, Cambridge).

Kallert, S., David, P., Finkenzeller, P., and Keidel, W.D. (1970). "Two different neuronal discharge periodicities in the acoustical channel," in Frequency Analysis and Periodicity Detection in Hearing, edited by Plomp, R. and Smoorenburg, G.F. (Sijthoff, Leiden), pp. 153-160.

Kiang, N.Y.S. (in press). "Peripheral neural processing of auditory information," in Handbook of Physiology, edited by I. Darian-Smith (Waverly, Baltimore).

Kiang, N.Y.S., Eddington, D.K., and Delgutte, B. (1979). "Physiological considerations in designing auditory implants," Acta Otolaryngol. 87, 204-218.

Kiang, N.Y.S. and Moxon, E.C. (1972). "Physiological considerations in artificial stimulation of the inner ear," Ann. Otol. Rhinol. Laryngol. 81, 714-730.

Kiang, N.Y.S, and Moxon, E.C. (1974). "Tails of tuning curves of auditory-nerve fibers," J. Acoust. Soc. Am. 55, 620-630.

Kiang, N.Y.S., Moxon, E.C., and Levine, R.A. (1970). "Auditory-nerve activity from cats with normal and abnormal cochleas," in Sensorineural Hearing loss, edited by G.E.W. Wolstenholme and J. Knight (Churchill, London), pp 241-273.

Kiang, N.Y.S., and Peake, W.T. (1960). "Components of electric responses recorded from the cochlea," Ann. Otol. Rhinol. Laryngol. 69, 448-458.

Kiang, N.Y.S., Watanabe, T., Thomas, E.C., and Clark, L.F. (1965). Discharge Patterns of Single Fibers in the Cat's Auditory Nerve, MIT Research Monograph No 35 (MIT Press, Cambridge).

Klatt, D.H. (1980a). "Speech perception: a model of acoustic-phonetic analysis and lexical access," in Perception and Production of Fluent Speech, edited by Cole, R.A. (Erlbaum, Hillsdale, N.J.), pp 243-288.

Klatt, D.H. (1980b). "Software for a cascade/parallel formant synthesizer," J. Acoust. Soc. Am. 67, 971-995.

Ladefoged, P. (1971). Preliminaries to Linguistic Phonetics (U. of Chicago, Chicago).

Liberman, A.M. (1957). "Some results of research on speech perception," J. Acoust. Soc. Am. 29, 117-123.

Liberman, A.M., Delattre, P.C., Gerstman, L.J., and Cooper, F.S. (1956). "Tempo of frequency change as a cue for distinguishing classes of speech sounds," J. Exper. Psychol. 52, 127-137.

Liberman, A.M., and Pisoni, D.B. (1977). "Evidence for a special speech-perceiving subsystem in the human," in Recognition of Complex Acoustic Signals, edited by T.H. Bullock (Abakon Verlagsgesellschaft, Berlin), pp 59-76.

Liberman, M.C. (1978). "Auditory-nerve response from cats raised in a low-noise chamber," J. Acoust. Soc. Am. 63, 442-455.

Licklider, J.C.R. (1951). "A duplex theory of pitch perception," Experimentia 7/4, 128-134.

Lindblom, B.E.F., and Studdert-Kennedy, M. (1967). "On the role of formant transitions in vowel recognition," J. Acoust. Soc. Am. 42, 830-843.

Lisker, L. and Abramson, A.S. (1964). "A cross-language study of voicing in initial stops: Acoustical measurements," Word 20, 384-422.

Littlefield, W.M. (1973). "Investigation of the linear range of the peripheral auditory system," Sc.D., Washington Univ., St. Louis.

Malme, C.I. (1959). "Detectability of small irregularities in a broadband noise spectrum," MIT Res. Lab. Electron. Quart. Prog. Rep. 52, 139-142.

Mann, V.A. and Repp, B.H. (1980). "Influence of vocalic context on perception of the [$\overset{v}{s}$]-[s] distinction," Percept. Psychophys. 28, 213-228.

McCasland, G.P. (1979) "Noise intensity and spectrum cues of spoken fricatives," J. Acoust. Soc. Am. 65, S78.

Mehrgardt, S., and Mellert, J. (1977). "Transformation characteristics of the external human ear," J. Acoust. Soc. Am. 61, 1567-1576.

Miller, G.A., and Nicely, P.E. (1955). "Analysis of perceptual confusions among some English consonants," J. Acoust. Soc. Am. 27, 338-353.

Miller, J.D., Watson, C.S., and Covell, W.P. (1963). "Deafening effect of noise on the cat," Acta Otolaryngol. Suppl. 176, 1-91.

Møller, A.R. (1965). "An experimental study of the acoustic impedance of the middle ear and its transmission properties," Acta Otolaryngo. 60, 129-149.

Møller, A.R. (1976). "Dynamic properties of primary auditory fibers compared with cells in the cochlear nucleus," Acta Physiol. Scand. 98, 157-167.

Møller, A.R. (1977). "Frequency selectivity of single auditory-nerve fibers in response to broadband noise stimuli," J. Acoust. Soc. Am. 62, 135-142.

Montandon, P.B., Megill, S.B., Kahn, A.R., Peake, W.T., and Kiang, N.Y.S. (1975). "Recording auditory-nerve potentials as an office procedure," Ann. Otol., Rhinol. and Laryngol. 84, 2-9.

Nooteboom, S.G. (1968). "Perceptual confusions among Dutch vowels presented in noise," IPO Progr. Rep. 3, 68-71.

Oppenheim, A.V., and Schafer, R.W. (1975). Digital Signal Processing (Prentice-Hall, Englewood Cliffs, NJ).

Peake, W.T., Goldstein, M.H., and Kiang, N.Y.S. (1962). "Response of the auditory nerve to repetitive acoustic stimuli," J. Acoust. Soc. Am. 34, 562-570.

Pearsons, K.S., Bennett, R.L. and Fidell, S. (1976). "Conversational speech levels in various environments," Bolt Beranek and Newman Report 3281.

Peterson, G.E., and Barney, H.L. (1952). "Control methods used in a study of the vowels," J. Acoust. Soc. Am. 24, 175-184.

Pickett, J.M. (1957). "Perception of vowels heard in noises of various spectra," J. Acoust. Soc. Am. 29, 613-620.

Plomp, R. (1976). Aspects of tone sensations (Academic, London).

Pols, L.C.W. (1971). "Real time recognition of spoken words," IEEE Trans. C-20, 972-978.

Pols, L.C.W., van der Kamp, L. J. Th., and Plomp, R. (1969). "Perceptual and physical space of vowel sounds," J. Acoust. Soc. Am. 46, 458-467.

Remez, R.E. (1979). "Adaptation of the category boundary between speech and nonspeech: A case against feature detectors," Cognitive Psychol. 11, 38-57.

Rhode, W.S, Geisler, C.D., and Kennedy, D.T. (1978). "Auditory-nerve fiber responses to wide-band noise and tone combination," J. Neurophysiol. 41, 692-704.

Rose, J.E., Brugge, J.F., Anderson, D.J., and Hind, J.E. (1967). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. 30, 769-793.

Ruggero, M.A. (1973). "Response to noise of auditory nerve fibers in the squirrel monkey," J. Neurophysiol. 36, 569-587.

Rupert, A.L., Caspary, D.M. and Moushegian, G. (1977). "Response characteristics of cochlear nucleus neurons to vowel sounds," Ann. Otol. 86, 37-48.

Sachs, M.B. and Young, E.D. (1979). "Encoding of steady-state vowels in the auditory nerve: Representation in terms of discharge rate," J. Acoust. Soc. Am. 66, 470-479.

Sachs, M.B., and Young, E.D. (1980). "Effects of nonlinearities on speech encoding in the auditory nerve," J. Acoust. Soc. Am. 68, 858-875.

Schalk, T.B., and Sachs, M.B. (1980). "Nonlinearities in auditory-nerve fiber responses to bandlimited noise," J. Acoust. Soc. Am. 67, 903-913.

Schroeder, M.R. (1962). "Correlation techniques for bandwidth compression," J. Audio Eng. Soc. 10, 163-166.

Schroeder, M.R., and Hall, J.L. (1974). "Model for mechanical to neural transduction in the auditory receptor," J. Acoust. Soc. Am. 55, 1055-1060.

Schuknecht, H.F. (1974). Pathology of the Ear (Harvard U. P., Cambridge, MA).

Searle, C.L., Jacobson, J.Z., and Kimberley, B.P. (1980). "Speech as patterns in the 3-space of time and frequency," in Perception and Production of Fluent Speech, edited by Cole, R.A. (Erlbaum, Hillsdale, N.J.), pp 73-102.

Sharf, B. (1970). "Critical bands," in Foundations of Modern Auditory Theory, edited by J.V. Tobias (Academic, New York), pp 157-202.

Siebert, W.M. (1968). "Stimulus transformations in the peripheral auditory system," in Recognizing Patterns, edited by P.A. Kolers and M. Eden (M.I.T. Press, Cambridge), pp 104-133.

Sinex, D.G., and Geisler, C.D. (1981). "Response of the auditory nerve to consonant-vowel syllables," J. Acoust. Soc. Am. 69, S54(A).

Smith, R.L. (1977). "Short-term adaptation in single auditory-nerve fibers: Some post-stimulatory effects," J. Neurophysiol. 40, 1098-1112.

Smith, R.L. (1979). "Adaptation, saturation and physiological masking in single auditory-nerve fibers," J. Acoust. Soc. Am. 65, 166-178.

Smith, R.L. and Brachman, M.L. (1980a). "Operating range and maximum response of single auditory-nerve fibers," Brain Res. 184, 499-505.

Smith, R.L. and Brachman, M.L. (1980b). "Response modulation of auditory-nerve fibers by AM stimuli: Effects of average intensity," Hearing Res. 2, 123-133.

Smith, R.L., and Zwislocki, J.J. (1975). "Short-term adaptation and incremental responses of single auditory-nerve fibers," Biol. Cybernetics 17, 169-182.

Stephens, S.D.G., Charlet de Sauvage, R. and Aran, J.M. (1975). "Gross responses from the cochlear nerve in man and in the guinea pig," Symp. Zool. Soc. Lond. 37, 167-186.

Stevens, K.N. (1980). "Acoustic correlates of some phonetic categories," J. Acoust. Soc. Am. 68, 836-842.

Stevens, K.N. (1981). "Evidence for the role of acoustic boundaries in the perception of speech sounds," J. Acoust. Soc. Am. 69, S116.

Stevens, K.N., and House, A.S. (1972). "Speech perception," in Foundations of Modern Auditory Theory, Vol.2, edited by J. Tobias (Academic, New York), pp 1-62.

Stevens, S.S. and Newman, E.B. (1936). "The localization of actual sources of sound," Am. J. Psychol. 48, 297-306.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., and Edman, T.R. (1976). "Consonant environment specifies vowel identity," J. Acoust. Soc. Am. 60, 213-224.

Strevens, P. (1960). "Spectra of fricative noise in human speech," Language and Speech 3, 32-49.

Teas, D.C, Eldredge, D.H, and Davis, H. (1962). "Cochlear responses to acaoustic transients: an interpretation of whole-nerve action potentials," J. Acoust. Soc. Am. 34, 1438-1451.

Trahiotis, C. and Elliot, D.N. (1970). "Behavioral investigation of some possible effects of sectioning the olivocochlear bundle," J. Acoust. Soc. Am. 47, 592-596.

Ver, I.L., Brown, R.M., and Kiang, N.Y.S. (1974). "Low-noise chambers for auditory research," J. Acoust. Soc. Am. 58, 392-398.

Voigt, H.F., Sachs, M.B., and Young, E.D. (1980). "Effects of masking noise on the representation of vowel spectra in the auditory nerve," in Neuronal Mechanisms of Hearing, edited by J. Syko and L. Aitkins (Plenum, 1981).

Voigt, H.F., Sachs, M.B., and Young, E.D. (1981). "Representation of whispered vowels in the temporal patterns of auditory-nerve fiber discharges," J. Acoust. Soc. Am. 69, S53(A).

Wang, M.D., and Bilger, R.C. (1973). "Consonant confusions in noise: a study of perceptual features," J. Acoust. Soc. Am. 54, 1248-1266.

Weiss, T.F. (1966). "A model of the peripheral auditory system," Kybernetic 3, 153-175.

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-nerve fibers," J. Acoust. Soc. Am. 66, 1381-1403.

Young, E.D., Sachs, M.B., and Voigt, H.F. (1981). "Wider dynamic range durind stimulus onset transients: effect on the representation of steady-state vowels in cat auditory-nerve discharge rate profiles," Fourth Midwinter Res. Meet. of Assoc. Res. Otolaryngol. (St Petersburg, Florida).

Zue, V.W, (1976). "Acoustic characteristics of stop consonants: a controlled study," Ph.D. Thesis, MIT, Cambridge.

BIOGRAPHICAL NOTE

Bertrand Delgutte

Born November 10, 1952 in Lille, France.

EDUCATION

1969-1971	Lycee Faidherbe, Lille (France)
1971-1974	Ecole Polytechnique, Paris. "Ingenieur Polytechnicien" degree (1974) (Mathematical and Physical Sciences)
1974-present	Massachusetts Institute of Technology S.M. Electrical Engineering (1976)

EMPLOYMENT

1973-1974	Research Assistant, Centre National d'Etudes des Telecommunications, Issy, France.
1975-1981	Research and Teaching Assistant, Department of Electrical Engineering and Computer Science, and Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts.
1976-1981	Research Assistant in Otolaryngology, Eaton Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts.

PUBLICATIONS

DELGUTTE, B. (1978). "Technique for the perceptual investigation of Fo contours with application to French," J. Acoust. Soc. Am., 64, 1319-1332.

KIANG, N.Y.S., EDDINGTON, D.K., and DELGUTTE, B. (1979). "Fundamental considerations in designing auditory implants, "Acta Otolaryngol., 87, 204-218.

DELGUTTE, B. (1980). "Representation of speech-like sounds in the discharge patterns of auditory-nerve fibers," J. Acoust. Soc. Am., 68, 843-857.