77-24,893

MARCELLINO, George Raymond, 1949-
TIME AND FREQUENCY CUES IN PHONEME IDENTIFICATION BY NORMAL AND HEARING-IMPAIRED LISTENERS.

City University of New York, Ph.D., 1977
Health Sciences, audiology

Xerox University Microfilms, Ann Arbor, Michigan 48106

TIME AND FREQUENCY CUES IN PHONEME IDENTIFICATION BY NORMAL AND HEARING-IMPAIRED LISTENERS

by

GEORGE RAYMOND MARCELLINO

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1977

This manuscript has been read and accepted by the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

6/14/77 — James R. Lang

date — Chairman of Examining Committee

6/31/77 — Irving Hochberg

date — Executive Officer

Dr. Irving Hochberg, Graduate School

Dr. Juergen Tonndorf, Columbia University

Dr. James Pickett, Gallaudet College

Supervisory Committee

The City University of New York

TABLE OF CONTENTS

A C K N O W L E D G E M E N T S

The author wishes to acknowledge Dr. James K. Lang, chairman of the advisory committee. Dr. Lang is an exceptional teacher, an understanding and sympathetic man-my close friend. Thanks are due Dr. Irving Hochberg and Dr. Juergen Tonndorf, who, as members of my advisory committee, gave their much needed advice and support.

Dr. Robert Port introduced me to, and fascinated me with the area of categorical perception. For his help at Haskins Laboratories and tutorial guidance, I express my deep appreciation.

I am grateful to Drs. Katharine Harris and Alvin Liberman for permitting me to use the facilities at Haskins Laboratories.

To my subjects, especially to Mr. Joseph Wieser, who endured the seemingly endless hours of the pilot research without complaint, indeed with enthusiasm, I am thankful.

I wish to express my heartfelt appreciation to my wife, Ellen, who above anyone else, provided me with the encouragement, warmth, understanding, love and support I needed to complete this project.

And finally, to my children, Andy and Tracy, you're wonderful - I love you.

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

A decrement in the discrimination for speech typically results from cochlear pathology. This fact is evident in both clinical experience and research findings. It is important to realize that this decrement is distinct from, and in addition to, a similar loss of the ability to process pure tones. For example, in cochlear pathology altered processing of acoustic stimuli are found (in addition to the loss of pure tone hearing and speech discrimination):

1. abnormal temporal integration functions (Wright, 1968; Martin and Wofford, 1970; Gengel and Watson, 1971 and Olsen et al., 1974);
2. increases in the width of the critical-band (Scharf and Hellman, 1966; Jerger et al., 1974 and DeBoer and Bouwmeester, 1974);
3. poor discrimination for frequency (Butler and Albrite, 1956; DiCarlo, 1962; Jerger and Jerger, 1967 and Gengel, 1973) and
4. greater-than-normal spread of masking (Jerger, 1960; Rittmanic, 1962 and Martin and Pickett, 1970).

These findings clearly suggest that an altered processing of signals at the cochlear level affects the neural input to the central nervous system.

More to the point of the present discussion is that a relationship may exist between altered speech discrimination and altered processing of non-speech stimuli. For example, the findings of a number of authors (Danaher and Pickett, 1975; DiCarlo, 1962; Gengel, 1973; Nabelek and Hirsh, 1969 and Fulton and Waryas, 1974), suggest a relationship between frequency discrimination and that of speech. That is, if frequency discrimination is generally poor, then an individual so afflicted will be unable to discriminate the frequency components of speech, especially noise bursts and frequency transitions that are essential cues in the recognition of consonants. Since speech discrimination is the major concern of the present research, then speech sounds (as contrasted with pure tones or complex non-speech stimuli) should be employed as experimental stimuli.

Time and Frequency Trade-Off

The Fourier analysis of acoustic steady-state signals rests on the assumption that the duration of signals is infinite. Since all signals are of finite duration, the assumption just stated is unrealistic. Gabor, in 1946 (Littler, 1965 p. 313) related the Heisenberg Uncertainty Principle of quantum mechanics to Acoustics, leading him to the equation $\Delta f \times \Delta t \cong 1$, where Δf = signal bandwidth and Δt = signal duration. Thus, signal bandwidth is inversely proportional to signal duration. With respect to acoustic analyzers, time resolution improves as bandwidth increases. Stated somewhat differently, greater precision in frequency analysis can only be gained at the expense of time resolution. It appears likely that the peripheral auditory system is subject to the same time-frequency trade-offs as electronic or mechanical frequency analyzers. Typical findings in sensorineural hearing loss include larger-than-normal difference limens for frequency, increases in the width of the critical-band, greater-than-normal spread of masking effects and shorter-than-normal integration times. These findings support the conclusion that cochlear pathology which is accompanied by better time resolution may result in a decreased frequency resolution. The present author believes that the application of Gabor's equation to speech perception in general, and in particular to speech perception by persons with cochlear hearing losses, will aid in our understanding of normal hearing as well as hearing disorders.

Use of Real Speech Stimuli

One of the major problems in the interpretation of responses to real speech items lies in the effects of uncontrolled variables such as relevance and redundancy which confound the issue. These confounding variables probably exist at both acoustic and linguistic levels of analysis. Essentially three approaches have been used in attempts to control these variables: (1) the use of a confusion matrix as an analytic tool, (2) the use of specially constructed speech samples, and (3) the use of synthetic speech signals generated by means of computer controlled speech synthesizers.

Confusion matrices as a method of signal presentation and data analysis have been employed by Miller and Nicely (1955), and by Wang and Bilger (1973) and others; but this procedure fails to sort out data in a manner which is useful when one wants to study the nature of the processing difficulties that result from cochlear pathology. Furthermore, with a confusion matrix, neither relevance nor redundancy are controlled at the linguistic level unless nonsense syllables are used.

Control of Linguistic Redundancy

Speaks and Jerger (1965) developed the Synthetic[1] Sentence Identification test (SSI) in an effort to control linguistic redundancy. In this test, real speech items are used to form ten sentences. The sentences are approximations of real English sentences based on the transitional probabilities linking adjacent words. The problem with these test sentences is that the signal parameters (frequency, duration and amplitude) are virtually uncontrolled. Since speech constitutes a time-varying signal marked by rapid changes in the frequency, amplitude and time domains, it is difficult to describe or to control

---

1 The word synthetic in this context should not be confused with synthetic speech as generated by a computer controlled speech synthesizer.

these parameters precisely. Hence, when a subject with sensorineural hearing-impairment incorrectly identifies a particular sentence, no information is gained as to what abnormal perceptual analysis of the acoustic properties of the speech items led to the misidentification. The relationship between (1) auditory frequency, amplitude and time analyses and (2) the detection, discrimination and identification[2] of speech items in persons with sensorineural hearing-impairments is unknown.

The use of computer generated consonant-vowel syllables appears to be an alternative method of testing. By reducing the number of possible phoneme identifications to, for example, three (a closed response set), linguisitic relevance and redundancy are kept at a minimum. Hence, the use of computer generated speech permits precise control and easy manipulation of the acoustic parameters of the test items.

Synthetic Speech Stimuli

Researchers at the Haskins Laboratories, and others, have extensively studied the perception of computer generated

---

[2] Throughout this paper the term speech detection refers to a listeners awareness that a given signal consists of speech (as opposed to non-speech). Discrimination refers to the determination that two or more speech items that differ in some acoustic dimension are judged by the listener to be the same or different. Identification refers to the categorization of speech stimuli into discrete phonemic classes. Speech perception is the generic term collectively referring to detection, discrimination and identification.

(synthetic) speech stimuli by normal hearing listeners. However, synthetic speech sounds have not yet been employed in the evaluation of subjects with hearing impairment with exception of the studies of Pickett and his colleagues at Gallaudet College.

Through the use of synthetic consonant-vowel syllables (CV), most frequently /b/ , /d/ and /g/ followed by a steady-state vowel it has been established that:

1. auditory processing of speech is different from the processing of non-speech stimuli;
2. identification of consonants that are distinctly categorized with respect to the target frequency of the second-formant transition (F2TR) and
3. discrimination of stimuli that vary in target frequency of the F2TR along a continuum, and are also within categories is at a chance level of performance, while there are peaks or maxima in discrimination scores at phonemic boundaries (Liberman et al., 1954; Liberman et al., 1967; Studdert-Kennedy et al., 1970 and Pisoni, 1971).

Aims of the Present Research

Martin, Pickett and Colten (1972), Danaher, Osberger and Pickett (1973) and Danaher and Pickett (1975) have been concerned with the problem of discrimination of formant transitions by listeners with sensorineural hearing-impairment. It was the intent of the present investigation: (1) to relate Gabor's time-frequency trade-off to phoneme identification for synthetic CV syllables that vary in target frequency and duration of the F2TR; and (2) to compare results derived from subjects with normal hearing with those obtained in a population of subjects with cochlear hearing losses.

Hypotheses

The following hypotheses were formulated in order to test the application of Gabor's time-frequency uncertainty principle to the perception of synthetic speech items by listeners with normal hearing and by listeners with cochlear hearing-impairments.

Subjects with Normal Hearing

1. The phoneme boundary (in Hertz), will shift toward higher frequencies as a function of decreasing duration (in msec), of the second-formant transition.
2. As target frequency of the F2TR increases, sensitivity with respect to duration of the F2TR will increase. This may be evidenced by more abrupt phoneme boundary shifts as a function of duration, for high target frequencies than for the low target frequencies.

Subjects with Cochlear Hearing-Impairment

1. Under the assumption that frequency resolution becomes poorer in subjects with cochlear hearing losses than in normal hearing listeners, it is hypothesized that subjects with cochlear hearing

loss will show phoneme boundaries (in Hertz) at higher frequencies than do the subjects with normal hearing.

2. Under the further assumption that poor frequency resolution may be accompanied by an improvement in time resolution, it is hypothesized that: (a) more pronounced durational effects will be observed and (b) durational effects will be evident for a greater range of target frequencies of the F2TR than is found in subjects with normal hearing.

CHAPTER II

REVIEW OF THE LITERATURE

There were two assumptions underlying the present investigation. First, it was assumed that pathology at the cochlear level interferes with the normal processing of acoustic signals in a way that alters the input to the central nervous system. Secondly, it was assumed that cochlear pathology results in an increase in the bandwidth for frequency analysis, i.e., of the critical bandwidth, which would entail an improvement in time resolution. The present author believes that changes in the bandwidth for cochlear frequency analysis, and integration time (a possible measure of time resolution), play a major role in the deterioration of speech perception in listeners with cochlear hearing-impairment. In the sections that follow, evidence from various psychoacoustic disciplines will be presented in support of the two assumptions stated. The review will then proceed to a discussion of the categorical perception of speech-like items.

Time-Frequency Trade-Off

According to the Fourier analysis of complex acoustic signals, any periodic waveform is composed of the sum of its individual sinusoidal components. Thus, any periodic waveform may be written as:

$$s(t) = \sum_{n} A_n \cos \left(2\pi \frac{nt}{To} - \phi_n\right) \qquad (2.1)$$

where, A= amplitude
n= harmonic number
t= an instant in time
T_o= period and,
φ= phase angle (Cherry, 1965 p. 313)

However, an important condition of the Fourier analysis is that the signal must be of infinite duration. This strict, but unrealistic notion led Gabor in 1946 (Littler, 1965, p. 313) to relate the Heisenberg Uncertainty Principle of quantum mechanics to acoustics. In this scheme,

$$\Delta f \cong \frac{1}{\Delta t}, \text{ or} \tag{2.2}$$

$$\Delta f \times \Delta t \cong 1, \tag{2.3}$$

where, Δf = signal bandwidth, Δt = signal duration. Stated in another way, the minimum bandwidth of a signal is inversely proportional to its duration. Thus, in one extreme a sinusoidal signal of infinite duration has a bandwidth of zero Hz, and in ther other, a unit-impulse (Dirac pulse or delta function) of infinitely short duration has energy spread throughout the frequency domain. In analyzing instruments like filters or resonators, the limits of time resolution are set by the bandwidth (Δf). Wider bandwidths result in improved accuracy in the time domain at the expense of accuracy with respect to frequency. Thus, in a practical sense, when signals are neither infinitely long nor infinitely short, frequency and or time cannot be precisely measured (uncertainty).

Gabor's formulation is illustrated by the analysis performed by the sound spectrograph. This device was developed at approximately the same time as Gobor's paper was written, for the purpose of analyzing speech in the time, frequency and amplitude domains. In this instrument two bandwidths of analysis are available; namely, 300 Hz and 60 Hz. Thus, their time resolutions are approximately 3 msec and 25 msec respectively. When one inspects the speech spectrogram for a wide-band analysis, vertical striations are seen in the record that occur at multiples of the period of the glottal waveform. In the narrow-band analysis, these striations disappear, but greater precision is obtained with respect to the harmonics and principal frequencies of the formants. Thus, in the graphic record obtained from the speech spectrograph, time resolution is gained at the expense of frequency resolution.

Georg Ohm applied the Fourier analysis of complex signals to the operation of the ear. He suggested that the ear might be insensitive to changes in phase spectrum and thus might analyze complex signals into their individual sinusoidal components (Ohm's Acoustic Law). Helmholtz (1875), prompted by the work of Ohm, believed that frequency analysis occured primarily, and perhaps solely, at the cochlear level. He viewed the operation of the cochlear in terms of a series of tuned resonators. Thus, for analysis of complex signals, resonators tuned to the frequency component of that signal would resonate. In addition, based on Müller's Doctrine of Specific Nerve Action, Helmholtz believed that each resonator was supplied by independent nerve groups that could transmit the results of the spectral analysis to the brain.

However, the difference limen for frequency is small, i.e., the Δf is narrow: on the order of 1-3 Hz, at 1 kHz, 40 dB SL. This suggests that the resonators have little or no damping; but little damping must result in substantial ringing. We know today that the basilar membrane is less than critically damped (Békésy, 1960). Furthermore, contrary to Ohm's assumptions, the ear is sensitive to phase spectrum.

Just as the relationship $\Delta f \times \Delta t \cong 1$ relates to the performance of electronic or mechanical analyzers on signals, it also relates to sensory systems which perform an analysis of the stimulus. Suppose that a system, such as the ear, is to analyze a complex stimulus. Is it reasonable to assume that it could (as suggested by a narrow interpretation of Helmholtz's theory), perform the analysis of frequency without regard to the duration of the stimulus? In the sections to follow, evidence will be presented that: (1) shows that the auditory system performs a dual time-frequency analysis and (2) cochlear pathology results in an improved time resolution at the expense of frequency resolution.

Difference Limen for Frequency (DLF)

The concept of the difference limen for frequency

(DLF) is relevant to the present research since several authors, including Fulton and Waryas (1974), Nabelek and Hirsh (1969), and Gengel (1973), have suggested a relationship between an individual's ability to detect a change in frequency, and his ability to discriminate among a variety of speech sounds. Fulton and Waryas (1974) and especially Nabelek and Hirsh (1969) pointed to discrimination of formant transitions as being essential to the perception of speech. It seems quite logical to conclude that if the ability of the hearing-impaired listeners to detect frequency changes and especially formant transitions is altered, then the discrimination of speech will deteriorate.

The aim of the section to follow is to present the research findings on the topic of the (DLF), under various listening conditions, for normal listeners and listeners with sensorineural hearing-impairment. Several sources of confusion have developed in this important area of research. As the following review will reveal, the confusions are linked to stimulus differences, methodological differences and intersubject variability.

Shower and Biddulph (1931), of the Bell Telephone Laboratories were the first to measure the DLF rigorously. In terms of the physical characteristics of their stimuli, they attempted to control items such as "transients generated by a change of frequency, harmonics due to non-linearity of the oscillator or receiver, inflexibility of apparatus and the attainment of intensity and frequency ranges sufficient to cover the auditory area." They used a frequency modulation (FM) technique and established the optimum rate of frequency modulation for their experiment. For a signal of 1 kHz, at 40 dB, when $\frac{\Delta f}{f}$ (relative DLF)[3] was plotted against the rate of frequency change, a broad minimum occurred at 2-3 Hz. As the rate of frequency modulation increased, the relative DLF increased from a minimum value of .003 at a modulation rate of 2 Hz, to .0095 at a rate of 5.33 Hz.

Particularly relevant to the present research are the findings of Shower and Biddulph at 500, 1,000 and 2,000 Hz; namely, the frequency range for speech often cited as the range most essential for discrimination accuracy. The relative and absolute DLF's at a sensation level of 30 dB were: .0055 or 2.75 Hz at 500 Hz, .0036 or 3.60 Hz at 1,000 Hz, and .0019 or 3.80 Hz at 2,000 Hz respectively. The overall results may be summarized as follows:

1. Absolute sensitivity with respect to frequency differences (DLF) improves with increased sensation level to 40 dB and remains constant thereafter.

2. Above 500 Hz, the minimum detectable percent change in frequency is approximately constant, but the absolute DLF increases with increasing frequency.

3. Below 500 Hz, the absolute DLF becomes approximately constant except for the very low frequencies, e.g., 62 Hz or less.

---

[3] In this section of the review the term Δf is synonomous with absolute DLF. $\Delta f / f$ (Weber fraction) refers to the relative DLF.

4. With binaural presentation the absolute DLF is reduced.

According to Harris (1969)

> . . . whereas with pitch modulation (FM) the Weber fraction is approximately constant above 1 kHz, and the differential threshold is constant below 1 kHz, the data from pitch-memory are unanimous in showing that the Weber fraction is approximately constant down to at least 3 octaves lower.

Although Harris' statement is undoubtedly correct, the FM stimuli of Shower and Biddulph closely approximate the frequency changes in the formant transitions seen in speech, something pure tone stimuli cannot achieve. Thus, the use of FM signals appears important with respect to the present study.

For frequencies above 1 kHz the FM data of Shower and Biddulph (1931) and the pure tone data of Harris (1952) are in close agreement. General trends were the same in terms of the effects of loudness and frequency, but below 1 kHz, the values of the relative DLF ($\Delta f / f$) were quite different; that is, when pure tones were used they were smaller than those obtained by the FM method. The differences are probably attributable to the fact that Harris employed pairs of pure tones as stimuli (rather than FM) and thus used a psychophysical procedure that was different from that employed by Shower and Biddulph.

In an attempt to reconcile differences in the estimated size of the DLF (both relative and absolute), Rosenblith and Stevens (1953), using two subjects, measured the DLF for the pure tone frequencies of 250 Hz (30 dB SL), 1,000 Hz (30 and 60 dB SL) and 4,000 Hz (30 dB SL) employing the ABX method and the method of constant-stimulus-differences.

Both the absolute and relative measures of the DLF were consistently lower for the both the ABX and constant-stimulus-difference methods than the results of Shower and Biddulph. However, the ABX method yielded absolute DLF's that were on the average 2.6 times greater than those obtained by the other method. Further, there were large individual differences between subjects. In table I a comparison is given between the average data of Shower and Biddulph (1931) and the data from the better subject of Rosenblith and Stevens (1953).

Table I

Absolute and Relative DLF from Shower and Biddulph (1931) and Rosenblith and Stevens (1953)

Frequency in Hz	dB SL	Shower and Biddulph Abs. DLF	Shower and Biddulph Rel. DLF	Rosenblith-Stevens* Abs. DLF	Rosenblith-Stevens* Rel. DLF
250	30	2.73	.0109	0.34	.0014
1,000	30	3.60	.0036	1.20	.0012
1,000	60	3.40	.0034	0.90	.0009
4,000	30	10.80	.0027	8.70	.0022
		Rosenblith and Stevens**			
250	30	.72	.0029		
1,000	30	2.50	.0025		
1,000	60	1.90	.0019		
4,000	30	1.60	.0004		

* constant-stimulus-difference procedure
** ABX procedure

In conclusion, it appears correct to state that, at present, there is no generally agreed upon value for the DLF. In addition to physical variables, the DLF is influenced by individual differences and differences in the psychophysical tasks. It is generally thought that an ABX procedure is more difficult than a method of constant-stimulus-differences. Thus, the more difficult the task, the larger the estimate of the DLF.

The Difference Limen for Frequency as a Function of Duration

Discrimination of frequency changes, as a function of time, is an important measure from the standpoint of the present research, because (a) many components of speech signals are brief in duration and (b) because a major independent variable of the present study is the duration of the second-formant transition.

Turnbull (1944) was among the first to study the problem of duration on frequency discrimination. He found, over a limited range of frequencies, that as duration of the stimulus to be discriminated was decreased below 150 msec, the absolute DLF increased. This result suggested a trading relationship in the form $\Delta f \times \Delta t \cong 1$.

Moore (1973) using a Parameter Estimation by Sequential Testing procedure, investigated the DLF for short duration tones in terms of a spectral vs. time analysis. On the basis of Zwicker's model for detection of a frequency change (Zwicker, 1970), he calculated for a pure auditory spectral analysis that $\Delta f \times \Delta t \geq 0.024$. According to Moore:

> . . . He (Zwicker) suggests that changes in a stimulus will be detected whenever the pattern of excitation changes by 1 dB or more anywhere along the basilar membrane, and that, for a frequency change, detection will take place at the point of steepest slope on the low-frequency edge of the excitation pattern.

Moore predicted that if a time mechanism were responsible for detection of brief frequency changes, then the empirical DL's should yield products of duration and frequency substantially below 0.024. In addition, if there were any overlap between time and spectral analyses, the DL plotted as a function of frequency should show an inflection in the curve. Moore found that the DLF increased with inverse duration across frequencies. For frequencies below 4-5 kHz the product of time and frequency in some instances was an order of magnitude smaller than that predicted by a place model. For normalized DL's ($DL_{norm} = \frac{DL \; 200\ \text{msec}}{DL \times \text{msec}}$) an inflection appeared in the region of 4-5 kHz suggesting the action of a spectral analyzer. It should be noted that for 200 msec stimuli results were in close agreement with Rosenblith and Stevens (1953). Moore suggested that for frequencies less than 4 kHz the auditory analysis was primarily temporal. In the region of 4-5 kHz a combined time-frequency analysis took place; and for frequencies greater than 5 kHz the analysis was purely spectral. These suggestions are in close agreement with Dallos (1976) in terms of the response of single units of the VIIIth cranial nerve.

Frequency Resolution as a Function of Pathology

Several dependent variables are either direct or indirect measures of frequency resolution; namely, (1) the difference limen for frequency, (2) width of the critical-band and (3) spread of masking effects. These psychoacoustic phenomena will be reviewed, with special emphasis on the findings in sensorineural hearing losses.

The DLF in Hearing-Impairment

Butler and Albrite (1956), investigated the pitch discriminative capacity in groups of listeners with conductive and sensorineural hearing losses. They found that, for the conductive loss group, the DLF was consistent with those previously obtained for normal hearing listeners. For the group with sensorineural hearing losses, the DLF was always larger than normal. The divergence between the two groups increased with increasing frequency. Authors did not report any details on the pathology of their subjects, nor did they provide data on the hearing loss. This considerably weakened the possibility of a useful interpretation of their research findings.

Gengel (1973) stated that:

> . . . One could hypothesize that the frequency uncertainty due to sensorineural hearing impairment, which is manifested as a reduced ability to discriminate small differences in

> frequency, also might reduce a person's ability to discriminate among vowels with closely adjacent formant frequencies and among spectrally similar consonants such as those contained with the categories nasals, unvoiced stops, voiced stops, and unvoiced fricatives. Thus discrimination of frequency differences and discrimination of speech might be highly correlated, not because good speech discrimination is dependent on good frequency discrimination, but because both types of discrimination are affected similarly by the degree of frequency uncertainty produced by sensorineural hearing impairment.

Gengel investigated the DLF as a function of duration in normal hearing listeners and in listeners with sensorineural hearing-impairments. He found that an increase in the DLF, at short durations, i.e., less than 150 msec, was

> . . . due in part, to the physical characteristics of short tones; i.e., the bandwidth of the signal increases as its duration decreases. The resulting uncertainty of the frequency of the signal is manifested as an increase in the size of the DLF.

Jerger and Jerger (1967), presented evidence that the DLF increased in cochlear and retrocochlear pathology. Among other psychoacoustic comparisons, the DLF for pure tones and for filtered noise was obtained from (a) a normal hearing listener, (b) a subject with presumed cochlear pathology and (c) a subject with a surgically confirmed eighth nerve lesion. The DLF's plotted in Figure 1 were obtained at 50 dB SPL for the subject with normal hearing. For the subjects with cochlear and retrocochlear hearing losses the levels were 80 dB SPL at

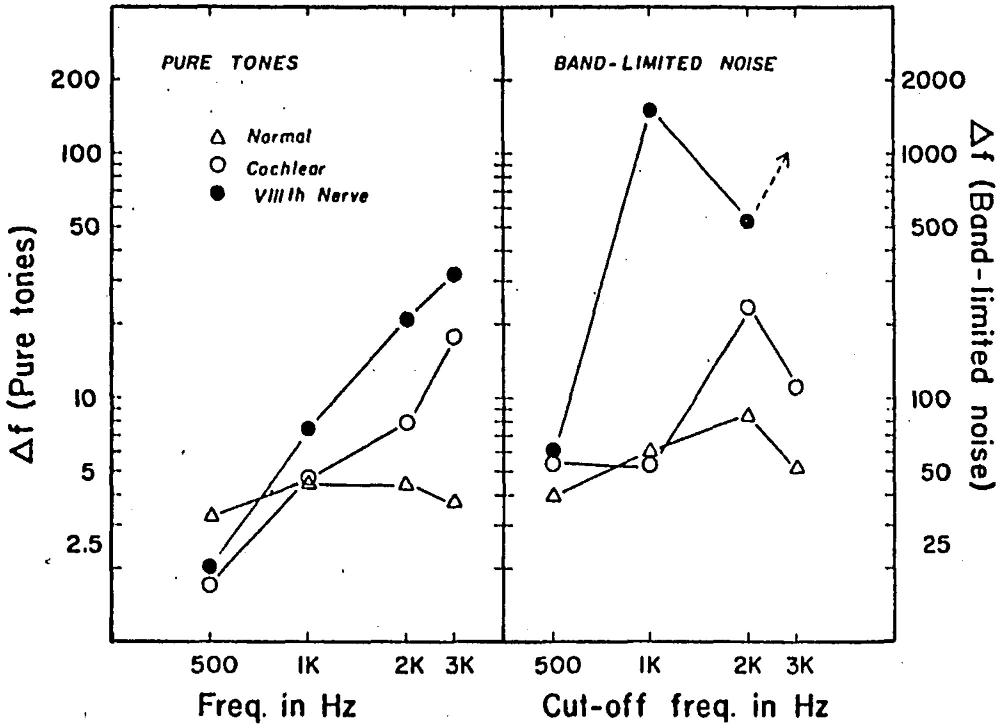

FIGURE 14. Δf as a function of frequency for pure tones and filtered noise.

(after Jerger and Jerger, 1967)

Figure 1

500, 1,000 and 2,000 Hz, and 90 dB SPL at 3,000 Hz. Stimulus durations were 200 msec. For pure tones higher than 1 kHz it was clear that the data for the subject with cochlear hearing loss and the subject with retro-cochlear hearing loss diverged from the data for the subjects with normal hearing. For the band-limited noise condition, where the low pass cutoff of the variable stimulus was manipulated with respect to the low pass cutoff of the standard, the subject with cochlear impariment showed a divergence from normal above 1 kHz.

For the band-limited noise condition, where the low pass cutoff of the variable stimulus was manipulated with respect to the low pass cutoff of the standard, the subject with cochlear impairment showed a divergence from normal above 1 kHz.

The subject with eighth nerve pathology showed exceedingly large DLF's above 500 Hz. For example, all three subjects had a DLF for band-limited noise of 40-60 Hz at 500 Hz. At 1 kHz the subject with eighth nerve impairment had a DLF of approximately 2 kHz. The relative DLF for band-limited noise, as a function of sound pressure level, indicated that, at all sound pressure levels, the subject with cochlear impairment showed poorer performance than did the normal hearing subject.

The Critical-Band in Sensorineural Hearing Loss

The critical-band is a model representing a filtering mechanism for the summation of loudness of complex tones. It has been empirically shown that in normal hearing listeners, the critical-band increases with frequency in bands approximately one-third octave wide. In man, there are 16 critical-bands (barks) that encompass the audible range of frequencies (Zwicker, 1957). For multitone complexes, loudness judgments remain constant until the frequency spacing among the components exceed the critical bandwith. At this point the loudness of the multitone complex increases.

The critical-band model has also been applied to masking. It has been determined that only a narrow band of frequencies, generated by a random noise source and surrounding a test tone, serves to effectively mask the tone (Hawkins and Stevens, 1952; Zwicker, 1957 and Greenwood, 1961). This range of effective masking frequencies is the critical-band.

The critical-band is related to several well documented psychophysical and anatomical facts. In man, the critical-band is approximately 30 times greater than the difference limen for frequency at any frequency (Zwicker et al., 1957). However, the width of the critical-band parallels a plot of the DLF as a function of frequency. In addition each critical-band: (1) approximated a pitch differential of 100 mels (Licklider, 1951), (2) corresponds to an approximate length of 1.2 mm along the basilar membrane (Zwislocki, 1965) and (3) contains approximately 1,300 receptor cells (Zwislocki, 1965 and Greenwood, 1961).

Thus, it appears that the critical-band is functionally related to psychophysical tasks in which frequency analysis is required. It also appears to require an anatomical integrity at a cochlear level. It is reasonable to assume that hair cell pathology impairs the critical-band mechanism and therefore affects frequency resolution. Earlier in this review, evidence was presented suggesting the occurrence of a greater-than-normal DLF in sensorineural hearing losses. In the paragraphs to follow evidence showing wider-than-normal critical-bands in sensorineural hearing losses will be presented.

Scharf and Hellman (1966) measured the loudness of multitone complexes as a function of Δf (frequency separation) and sensation level in the presence of intense masking noise, for normal-hearing subjects, subjects with conductive hearing loss, and subjects with cochlear hearing-impairment. At equivalent sensation levels, the subjects with conductive hearing loss summated loudness in the same way as the normal subjects when not subjected to masking. For the subjects with cochlear pathology, loudness was independent of Δf. In addition, these subjects differed from the normal hearing subjects in which loudness of the multitone complex was determined in the presence of an intense masker. Although the masker produced elevated thresholds and steep loudness functions, loudness increased, at moderate sensation levels, when the critical-band was exceeded. The authors concluded that: (1) Zwicker's model of loudness summation did not predict the invariance of loudness with Δf in cochlear pathology although the model yielded calculations in excellent agreement with the data for the subjects with conductive hearing loss and the normal, masked subjects and (2)

> ...the model did provide the basis

> for the tentative hypothesis that the invariance of loudness in cochlear pathology may result primarily from alterations of the critical-band mechanism and a widening of the critical-band.

According to DeBoer and Bouwmeester (1974), "...the slopes of the critical-band indicate the power of the frequency resolution of the ear." The slopes of the critical-band were measured by band-rejection filtering of a wide-band masker, and obtaining thresholds for pure tones presented in the frequency gap (tone in gap masking). Thresholds were determined as a function of the width of the gap. Results for subjects with presumed cochlear hearing losses indicated that: (1) in some subjects the width of the critical-band was normal, (2) in some other subjects the critical-band mechanism did not seem to be functioning at all and (3) the high frequency slope was generally more affected than the low frequency slope.

In yet another experimental study, Jerger et al., (1974), applied the critical-band model to the summation of loudness for elicitation of the acoustic reflex. The purpose was to develop a clinical test that would predict the degree of hearing handicap for difficult-to-test populations. The authors based their predictions on a study by Flottorp, Djupesland and Winther (1971) which showed the effect of the critical-band mechanism on the acoustic reflex, and another study by Deutsch (1972), which showed that the SPL required to elicit the acoustic reflex using pure tone stimuli is greater than that required for broad-band noise. Jerger et al., (1974), reasoned that broad-band noise has a supercritical bandwidth and hence, is louder than a pure tone at the same SPL. Since the acoustic reflex is functionally related to loudness rather than to SPL, the authors further reasoned that mechanisms affecting loudness should in turn affect the acoustic reflex. One such mechanism is the critical-band. If the critical-band is wider in subjects with sensorineural hearing losses than it is in normal hearing listeners, then the loudness of a broad-band noise will be less for the listeners with sensorineural hearing losses. Thus, the difference in SPL between a pure tone and broad-band noise, required to elicit the acoustic reflex, should decrease with increasing width of the critical-band. Authors predicted the severity of the hearing loss in a series of 918 patients by comparing the SPL required to elicit the acoustic reflex by pure tones and by wide-band noise. They found that the severity increased as the difference in SPL between pure tones and wide-band noise decreased.

Spread of Masking Effects

Some authors have demonstrated abnormally large spreads of masking effects in subjects with sensorineural hearing losses. Spread of masking may be viewed as another means of measuring the frequency resolution of the peripheral auditory system. Abnormal spread of masking effects may be related to the critical-band mechanism.

Rittmanic (1962), obtained thresholds, utilizing a Békésy tracking procedure, in the presence of narrow-bands of filtered thermal noise. The nominal center frequencies were 250 Hz, 500 Hz, 1,000 Hz, 2,000 Hz and 4,000 Hz. The subjects were divided into three groups: (1) normal listeners, (2) normal listeners wearing earplugs (thus, simulating conductive hearing loss) and (3) listeners with sensorineural hearing losses. Analysis of the results indicated that for all noise bands, the sensorineural group of subjects exhibited a greater spread of masking, on the order to 10 to 20 dB, than the normal or normal subjects wearing earplugs. Rittmanic attributed the downward spread of masking to nonlinear mechanical effects in the inner ear, but failed to provide an explanation for the upward spread of masking. The present author believes that both the greater-than-normal upward and downward spread of masking effects may be attributed to wider-than-normal critical-bands. Jerger et al., (1960), similarly, found a greater-than-normal spread of masking in their subjects with sensorineural hearing losses.

Martin and Pickett (1970), suggested that the spread of masking effects is influenced by the choice of the masking procedure. That is, the choice of equivalent masking levels or SPL as independent variables affect the results. In addition, authors felt that a reduced dynamic range in sensorineural hearing losses may tend to obscure trends in the data. In other words, small threshold shifts might really represent a great deal of masking. These authors proposed measurement of the rejection slope for the spread of masking effects, for a low-pass (250 Hz) noise at SPL's of 77, 97 and 107 dB as an alternative. Subjects were classified according to their degree of hearing loss and audiometric contour. Martin and Pickett (1970), summarizing their results stated:

> . . . there are marked differences in masking spread within the sensorineural population. Some types of losses exhibit greater-than-normal masking spread; others exhibit the same or slightly less. In addition, it appears that in some sensorineural losses, there is not the normal, positive relation between level of masking and amount of masking spread.

The consensus of opinion appears to be that the critical-band is wider-than-normal in sensorineural hearing losses. However, this finding is by no means universal. The Gabor time-frequency trade-off would predict for persons who demonstrate abnormally wide critical-bands better than normal time resolutions. Thus, the next section of this review will explore the temporal integration of acoustic energy as a measure of time resolution in the auditory system.

Time Resolution in Hearing-Impairment

The temporal summation of loudness refers to the integration of acoustic power over time by the auditory system (Olsen et al., 1974). The time constant of integration has been estimated to be on the order of 200 msec for listeners with normal hearing (Zwislocki, 1960; Olsen and Carhart, 1966; Dallos and Olsen, 1964 and Plomp and Bouman, 1959). This follows from the fact that, as duration is reduced below 200 msec, stimulus magnitude must be increased in order to maintain a threshold level of response. When threshold is plotted in dB as a function of duration on rectilinear coordinates, a negative exponential function is obtained which becomes assymptotic at 200 msec. Thus, for threshold values, the product of duration and intensity (power) is nearly constant. For perfect energy integration the following equation may be written (Garner, 1947):

$$I \times t = k \qquad (2.4)$$

where I = stimulus intensity, t = duration and k = the threshold energy. Since the energy integration of the ear is not perfect, Dallos and Olsen (1964), modified Garner's equation to fit their data as follows:

$$\log\,(I-I_o) = -\log T + \log k, \qquad (2.5)$$

where I_o and k are constants.

When duration is shortened to one-tenth, e.g., 200 down to 20 msec, stimulus magnitude must be increased by

10 dB. However, Zwislocki (1960), suggested that the slope of the temporal integration function was not constant. His results indicated that for a range of 200 to 100 msec a 1.5 dB increase was necessary to maintain threshold, but that a 3 dB increase was necessary in the range of 100-20 msec, while a 4.5 dB increase in magnitude occurred for durations less than 20 msec.

Gengel and Watson (1971) obtained temporal integration functions for four subjects with normal hearing and three subjects with sensorineural hearing losses at various frequencies in the audiometric range. Graphic analysis indicated overlap between groups of subjects and extreme intersubject variability. There was, however, a significant difference in the slope of the temporal integration function with respect to frequency. That is, slope decreased with increasing frequency. In addition, it appeared that the time constant of integration was less than 200 msec for the subjects with sensorineural hearing losses. Further examination of eleven subjects with hearing-impairment led the authors to the following conclusion: ". . . we interpret our data to suggest reduced efficiency for detecting acoustic energy by some hearing-impaired subjects."

Wright (1968 a,b), on the basis of temporal integration functions obtained in subjects with sensorineural hearing losses, concluded that, "hearing loss that is cochlear in origin will definitely affect the slope of the temporal integration function and the slope is not necessarily frequency dependent. The extent of the effect was not uniquely related to the amount of hearing loss present in the individual."

Martin and Wofford (1970) and Olsen et al., (1974), agreed that the variability and extreme overlap in the temporal integration functions between subjects with normal hearing and subjects with cochlear hearing-impairment, preclude clinical interpretation with respect to differential diagnosis. These authors showed, however, that on the average, the slope of the temporal integration function was depressed and the time constant of integration was shorter in their groups of cochlear hearing-impaired subjects than in their normal hearing subjects.

The present author disagrees with the notion that a depressed slope of the temporal integration function is equivalent to an inefficient integration of acoustic power by listeners with sensorineural hearing losses. An alternative interpretation is that a reduced slope: (1) suggests that the time constant of integration is less than 200 msec and, (2) shows that listeners with sensorineural hearing-impairment are, by virtue of their short time constant of integration, more resistant to threshold shift (in dB SPL), as a function of decreasing duration below 200 msec than subjects with normal hearing.

Simon (1963), linked the width of the critical-band to the time constant of integration. He reasoned, as does the present author, that . . ." the time constants of that system (auditory) are directly and inversely related to its bandwidth." Measures of the width of the critical-band were obtained for subjects with normal hearing and those with sensorineural hearing-impairment. Thresholds were assessed at 1 kHz and at 4 kHz in the presence of a broad-band masker (100-6,000 Hz). Signal duration varied from 12.5 to 800 msec. The spectrum levels of the masker were 52.5 and 62.5 dB. Simon's data show that for his subjects with normal hearing the critical-band was inversely proportional to signal duration. For his subjects with cochlear pathology, the critical-band at 4 kHz was wider, and the integration time shorter (100 msec); Recruitment was present at that frequency.

Categorical Perception

In chapter one the advantages of synthetic speech items over real speech items were discussed. Synthetic speech-like items are most frequently used in research dealing with the categorical perception of speech. The use of synthetic speech-like items and analysis of data in accord with categorical perception is potentially useful in research / clinical audiology because it provides a powerful method for the analysis of speech perception data and their interpretation. For purposes of the present research, synthetic, speech-like items offer the advantage of being easily manipulated and controlled with respect to the time and frequency parameters of the stimuli.

The methodology and the analysis of data in categorical perception experiments focus on the relationship between specific acoustic features of a given synthetic speech item and the response, specifically the identification of the stimulus as a particular phoneme and/or discrimination of the acoustic features. Identification of synthetic speech-like items in speech perception experiments has not yet been investigated with populations of subjects with sensorineural hearing-impairment. In sensory sciences, research on pathological sensory systems often yields insight with respect to normal processes. Furthermore, use of synthetic speech items on persons handicapped by cochlear hearing loss and application of the categorical perception type of analysis, may provide valuable information as to the nature of the speech perception difficulties experienced by this group of persons.

The literature review to follow first discusses the results of the basic categorical perception experiments, then describes the experimental methodology and data analysis. At every opportunity the relationship between the work cited and the present research effort will be elaborated.

The search for cues underlying the perception of speech began with the development of the sound spectrograph (Koenig, Dunn and Lacey, 1946), and of pattern playback (Cooper, 1950) devices. The sound spectrograph permitted investigators to analyze speech signals simultaneously in terms of frequency, amplitude and time and to display the results of such analysis graphically. The pattern playback, through the use of an optical scanning device, provided for conversion of hand-painted spectrograms to synthetic speech signals. It was in this way that acoustic variables of the speech signal became subject to experimental manipulation and control.

The principal frequencies of formants[4] and formant-transitions[5] were quickly recognized as relevant variables with respect to phoneme identification. The frequency relationship between the first-formant (F1) and the second-formant (F2) is the major determinant of vowel identity. For example, if F1 = 489 Hz and F2 = 1620 Hz, listeners will identifty the stimulus as the vowel /ae/ (Pisoni, 1971). The first-formant transition (F1TR) has been identified as the cue by which listeners infer manners of articulation (class of stops, nasals etc.). The second-formant transition (F2TR) is the basis for inferences about the place of articulation (Liberman, 1967). Thus, in the synthesis of consonant-vowel (CV) syllables, the appropriate choice of formant and formant transition frequencies can produce test items that may be identified by listeners as, for example, the stop consonants followed by the steady-

[4] Formants are resonances in the speech spectrum determined by the vocal tract transfer function.

[5] Formant transitions represent changes in the vocal tract resonances as the articulators change position during an utterance.

state vowel /ae/ (/bae/, /dae/ or /gae/ Pisoni, 1971).

Acoustic Invariance

It was found in early speech perception studies that it was not readily possible to segment the acoustic signal we know as speech into discrete units (phonemes), and this was especially difficult with consonants. The formant transition cues both the consonant and vowel to follow simultaneously (parallel transmission of information). When formant transitions are isolated from a speech sample, listeners identify the transitions as chirps, glissandos or clicks, but not as speech sounds (Liberman, 1970). Thus, the consonant cannot be separated from the vowel. Furthermore, the acoustic properties of a particular consonant vary with the following vowel, i.e., there is a marked lack of acoustic invariance. In the now classic example, the F2TR fall in frequency for /d/ followed by /u/, but rises in frequency when /d/ is followed by /i/, (Liberman et al., 1954). The acoustic properties of the consonant are context conditioned.

Categorical Perception-Methods and Data Analysis

The experimental methods that evolved with the use of synthetic speech stimuli employed identification and discrimination of the stimuli by subjects with normal hearing. Analysis of the data has revealed that consonants are identified discretely, i.e. categorically. Subjects discriminate acoustic differences, e.g. various F2TR target frequencies, between stimuli only as well as they can identify a given stimulus as being a member of one phoneme set or another. If, for example, a synthetic speech signal is modified, under computer control, with respect to the target frequency of the F2TR while F1, F1TR and F2 remain constant, it can be shown that peaks in the discrimination function occur at the boundaries[6] of phoneme identification (Liberman et al., 1957; Liberman et al., 1967; Studdert-Kennedy et al., 1970 and Pisoni, 1971). Figure 2 is representative of the relationship between identification and discrimination of synthetic speech stimuli--an ideal case. The categories A and B could represent the phoneme categories /b-d/ or /d-g/ . There is an abrupt change in the identification (measured in percent) from category A to B at the stimulus value of 4.5 along the abscissa The value of 4.5 may be considered the category boundary. The abscissa is representative of an acoustic continuum, target frequency of the F2TR, for example. The discrimination between stimuli (right-hand coordinate) is maximum at 4.5, while the discrimination between stimuli that are within categories is at chance level. The relationship between the maxima and minima of the discrimination functions and the boundaries on the identification functions

---

[6] The phoneme boundary is that point on the identification function where the curves from two successive phonemic categories cross.

THEORETICAL NOTES

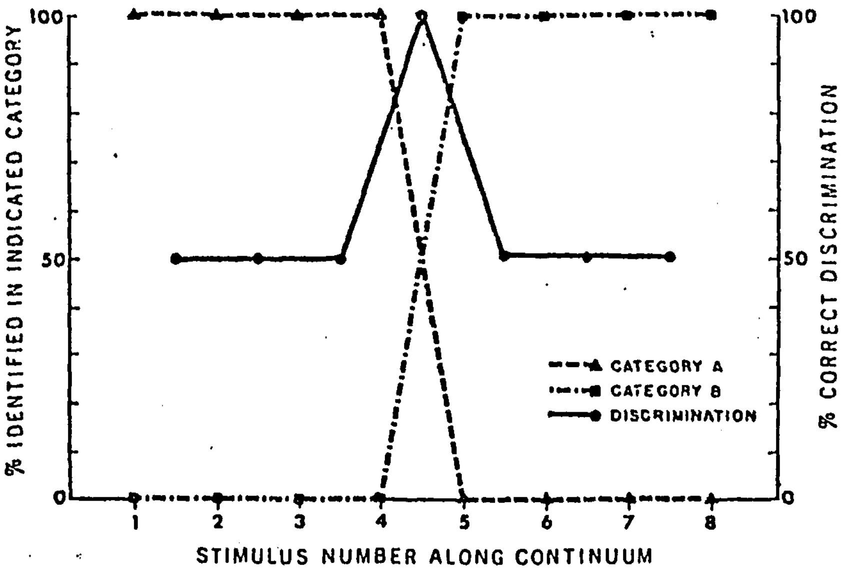

FIG. 1. Idealized identification functions (left ordinate) and discrimination function (right ordinate) to illustrate categorical perception of eight stimuli distributed at equal intervals along a physical continuum.

(after Studdert-Kennedy et al., 1970)

Figure 2

is so reliable that a very close agreement exists between discrimination scores obtained and discrimination scores predicted from the identification function (Liberman et al., 1957; Studdert-Kennedy, 1970 and Pisoni, 1971). Thus, categorical perception represents a perceptual process whereby: (1) there are abrupt shifts (boundaries) in the identification of stimuli along a continuous physical dimension, (2) discrimination between stimuli peaks at the same point along the physical dimension at which the phoneme boundary occurs and (3) discriminatuon within categories is at a chance level of performance.

Of major importance is the relationship between the discrimination of frequency transitions within the speech signal and phoneme identification. The present research is based on the assumption that whatever mechanism is responsible for poor frequency discrimination or for the lack of frequency resolution in listeners with cochlear hearing losses, it may also be responsible for poor speech processing capabilities. That is, poor frequency resolution may be related to changes in phoneme identification and/or changes in the position of the phoneme boundary in the frequency domain.

Discrimination of Formant Frequencies by Listeners with Hearing-Impairment

In the section to follow, previous research related to the discrimination of formant transitions by listeners with hearing-impairment will be presented. In the studies to be reviewed, formant frequency, formant duration and formant transition duration were among the variables investigated. These variables are particularly relevant to the present study. The relationship between discrimination of formant freqeuency and phoneme identification has been established. Available data suggest that a decrement in discrimination may lead to alterations in phoneme identification. If short duration formant transitions are more difficult to discriminate than long ones, it is conceivable that phoneme identifications and phoneme boundaries may likewise be affected by duration.

To the best knowledge of this author, researchers at Gallaudet College were the only ones who have so far utilized synthetic speech items for the evaluation of speech perception in populations with hearing-impairment. The experimental strategy used by these investigators has been to measure the discrimination of formant transitions.

Martin, Pickett and Colten (1970), used an adaptive procedure to determine the discrimination of the vowel formant transition (F2TR alone) in a normal and severely hearing-impaired population. The transitions had final frequencies of 1.2, 2.0 and 3.0 kHz. The transition threshold was estimated as the smallest transition the subject could discriminate in 75% of the trials. All transitions were rising in frequency. Discrimination was also measured at transition durations of 50, 100, 300 and 500 msec. For both groups of subjects, transition thresholds were largest for the shortest duration (50 msec) and became smaller as duration was increased. At 500 msec the normal hearing listeners and the listeners with hearing-impairment yielded similar discrimination thresholds of 50 Hz. However, at the shorter durations the listeners with hearing-impairment yielded transition thresholds that were substantially higher than those obtained for the subjects with normal hearing, and at the same durations. In addition, there appeared to be an interaction between duration and frequency. For example, the F2TR with a final frequency of 3 kHz at

50 msec yielded at Δf of .78 kHz and at 300 msec a Δf of .2 kHz. For a final frequency of 1.2 kHz at 50 msec Δf = .3 kHz and at 300 msec, Δf = .1 kHz, Thus, with increasing frequency the difference between thresholds obtained at 50 msec and 300 msec increased. According to Martin et al., (1970), ". . . for formant transitions in speech, we would expect sensorineural subjects to have somewhat poorer discrimination for high frequency transitions than for low frequency ones."

The Effect of the First Formant on the Discrimination of the Second-Formant Transition

In an additional experiment by Martin et al., (1970), the discrimination of F2TR in the presence of F1 was measured. The vowel duration equalled 300 msec. Transitions were either 300 or 100 msec in duration. The second-formant transition reached a final frequency of 1.5 kHz, with transition either rising or falling to that frequency. The first-formant was held constnat at 489 Hz without any transition. First, a most comfortable level (MCL) was established for F1 and F2 presented simultaneously. Then F1 was removed and the threshold for F2 alone was established. Subjects reported that the stimulus was "soft" in comparison to the presentation of F2 alone in the previous experiment. The results are summarized in Table II below. When the results in Table II are compared with results obtained for

Table II

Discrimination Threshold (Δf) for F2 Transitions for Subjects with Sensorineural Hearing-Impairment (after Martin et al., 1970)

Transition duration	F2 comfortable	F2 soft	F1 and F2
300 msec	50 Hz	50 Hz	200 Hz rising 300 Hz falling
100 msec	70 Hz	70 Hz	300 Hz rising and falling

normal hearing subjects, it can be seen that the effect on Δf, when F1 and F2 are present, is much greater for the hearing-impaired group than for the other group. In addition, the effects of duration were more pronounced. It was suggested by the authors that the change in Δf (F2TR) in the presence of F1 may have been attributable to the upward spread of masking by the more intense F1.

The Effect of a First-Formant Transition on Discrimination of the Second-Formant Transition

Danaher, Osberger and Pickett (1973), reported the results of a three-part series of experiments concerned with the masking effects of F1 on F2TR discrimination. In their first experiment the synthetic vowels /i/ and /a/ were presented to normal hearing listeners and to severely hearing-impaired ones. The hearing-impaired group consisted of subjects with flat audiograms and others with high frequency hearing losses evidenced by sloping audiograms. Discrimination functions were obtained for the conditions listed below.

/i/	/a/
F2TR alone	F2TR alone
F1 and F2	F1 and F2

The /a/ thresholds for F2 alone were approximately the same for all groups. The /i/ threshold for F2 alone was different for the two groups, and it seemed that discrimination was related to the degree of hearing loss. As high frequency hearing loss increased, the level of the discrimination threshold increased (see Figure 3). It should be noted that the /i/ second-formant is higher in frequency than the /a/ second-formant.

In the second experiment by Danaher et al., (1973), a small frequency transition was added to F1 in order to determine the effects of F1 transitions on the discrimination of the F2TR. Two transitions were presented, a 50 Hz frequency excursion and a 100 Hz excursion. It was found that discrimination of F2 transitions which occurred simultaneously with F1 transitions was poorer than F2TR discriminations in the presence of low frequency formants that did not change in frequency.

In the third experiment by Danaher et al., (1973),

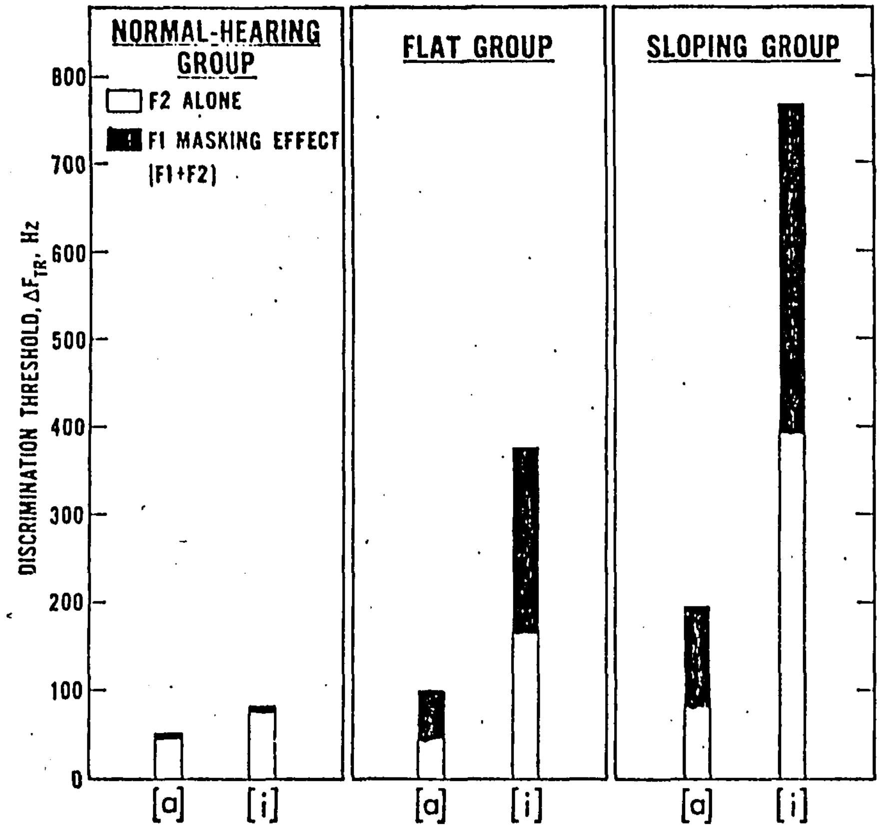

FIGURE 3. Discrimination of F_2 transitions in [ɑ] and [i]. Mean discrimination thresholds, ΔF_{TR}, for one normal-hearing group and two hearing-impaired groups. The upper limits of the unfilled areas denote ΔF_{TR} when F_2 was presented alone. The upper limits of the shaded areas denote ΔF_{TR} when both F_1 and F_2 comprised the signal. Size of the shaded areas describes the amount of masking produced by F_1.

(after Danaher et al., 1973)

Figure 3

the amplitude of F1 was reduced relative to F2. The authors reasoned that if the spread of masking of F1 on the F2TR was similar to the upward spread of masking of pure tones, then a reduction of the F1 amplitude should minimize masking, and discrimination of F2TR should improve. The data indicated that subjects with flat and sloping audiograms showed improved discrimination when the F1 amplitude was reduced. In the group with sloping audiograms it was found that, when F1 had a frequency transition, the discrimination thresholds were consistently worse than when F1 had no transition. It appeared that the magnitude of the F1 transition had little, if any effect on discrimination. This conclusion was supported by an independent study of Dorman (1974). Dorman used a signal detection paradigm, and found that, for a synthetic speech stimulus, the discrimination of F2TR with intensity differences over a 9 dB range in the presence of F1 was near chance ($d' < 1$). However, when the F2TR was presented in isolation, with intensity increments in the range of 9 dB, intensity discrimination was much improved ($d' > 3$).

Danaher et al., (1973) suggested that a transition in F1 gives the stimulus a speech-like quality which may account for the poorer discrimination in the sense of the poor intraphonemic discrimination observed by Liberman et al., (1967). However, authors then proceeded to dismiss this explanation in favor of the upward spread of masking.

<u>Monotic and Dichotic Discrimination of the Second-Formant Transition</u>

Danaher and Pickett (1975), investigated the discrimination of F2TR under various conditions of monotic presentation and dichotic presentation (F1 to one ear F2 to the other), and a F1 time cutback, i.e. F1 had a delayed onset with respect to the onset of F2. They found that, for subjects with flat and sloping audiograms and for both vowels (/a/ and /i/), discrimination was best when F2TR was presented alone. The amount of masking by F1 varied, but discrimination was always worse when F1 was present. Thresholds were worst with monotic presentation. Discrimination improved substantially when F1 and F2 were split and directed independently to either ear i.e. dichotic presentation. Under this condition, masking was still evident but had decreased in magnitude. The authors concluded that, in the monotic condition, poor discrimination of the F2TR in the presence of F1TR and F1 was primarily a function of the upward spread of masking. In the dichotic condition, the residual masking was attributed to central masking.

Backward Masking in Discrimination of the Second-Formant Transition

For subjects with hearing-impairment, whose audiograms were flat or sloping, it was found that the discrimination thresholds monotonically improved as F1 was cutback in time, that is, delayed by 0, 50, 100 or 200 msec. Authors felt that this was related to backward masking as seen in normal hearing subjects for noise bands presented at high intensities.

Danaher and Pickett (1975) made the following point:

> . . . some of the subjects volunteered the information that the stimuli with the F1TR sounded like speech and when

> asked to imitate what they heard, they usually produced words or syllables beginning with /b/, such as /bap/ or /bla/. No subject reported that stimuli without F1TR sounded like speech.

Authours suggested on account of these anecdotal remarks that hearing-impaired subjects may have phoneme boundaries similar to normal hearing listeners. However, according to Liberman et al., (1956), a rising second-formant transition of 100 msec duration for the frequency values used, should have been identified by normal hearing listeners as /w/. This appears to suggest that for the hearing-impaired listeners, phoneme boundaries might be different from those found for normal hearing listeners. Danaher and Pickett (1975) modified their former conclusion (Danaher et al., 1973), by making the following statement:

> . . . when considering the data in these experiments and the conclusions we have drawn from them, it is important to note that our tests measured discrimination rather than recognition of speech-sound information. Whereas the masking effects we have identified might reduce discrimination of speech sounds they have not yet been shown to be signficant factors in speech recognition.

To the present author, this later conclusion has greater appeal, since the measurement of recognition (identification), of synthesized speech sounds by hearing-impaired listeners is the major goal of the present research.

<u>Glide Rate as a Variable in DLF and Speech Perception</u>

Certain questions can be raised relative to the fact that, during formant transition, frequency changes occur during a particular time interval. There may be several cues in the transition that are functionally related to phoneme identification. These cues may include: (1) the target frequency of the F2TR, (2) the duration of the transition or (3) the rate of change of frequency (glide rate) of the transition.

Prompted by recent findings in auditory neurophysiology; specifically, the identification of single unit neurons sensitive only to FM stimuli, Pollack (1968), established thresholds for the detection of the direction and rate of frequency changes for frequency modulated signals. Frequencies were in the range of 125-1,000 Hz. Duration of the frequency changes varied from .5 to 4 sec. Pollack found an inverse relationship between the threshold for rate of frequency change (for 75% detection) and the duration (slope = -1), of the change. The sensitivity for the detection of transition differences was "nearly independent" of the sound level over a range of 60 dB. In addition, there appeared to be two ways of detecting changes with FM signals, " . . . for sharp transitions, threshold temporal gradients are proportional to the frequency difference subtended; for slow transitions, threshold temporal gradients are proportional to the temporal difference subtended." These results relate well to the uncertainty relationship between time and frequency of Gabor.

Nabelek and Hirsh (1969), explored a much wider range of durations of changes on the discrimination of frequency transitions than did Pollack. Their motivation for conducting the study stemmed from a study by Liberman et al., (1956), in which the effect of glide rate of the F2TR was studied. The essential findings of Nabelek and Hirsh were as follows. For fixed frequency excursions, the optimum discriminable glide rate expressed in Hz/msec increased with an increasing extent of the frequency transition and with an increasing frequency of the steady-state portion of the stimuli, i.e. the carrier frequency of the signal. However, the optimum glide rate across frequency appeared to be approximately 30 msec, except at 250 Hz for the small transition (optimum duration = 1400 msec). These results tend to support Pollack (1968), in terms of a dual processor for frequency excursions and suggests ". . . that the best discriminability of large frequency excursions for transition duration around 30 msec is a general property of hearing and it does not appear only in connection with speech sounds,"

Liberman et al., (1956) showed that subjects were unable to utilize tempo, that is, rate of frequency change of the F2TR, as a cue for distingusishing between stops and semivowels. The vowel /ɛ/ in the first experiment has held constant. Only the duration of the F2TR was varied over a range of 10 to 300 msec. The CV syllable was identified as /bɛ/ at durations of less than 40 msec; /bɛ/ became /wɛ/ at durations greater than 40 msec; and /wɛ/ became /uɛ/ at approximately 200 msec. The transiton from /gɛ/ to /jɛ/ occurred at approximately 210 msec. A conclusion, i.e. that duration has a significant effect on phoneme identification for the stop semivowel distinction can be drawn from this experiment. With the vowel formants held constant, however, the question of glide rate, as opposed to duration of the F2TR as a signficant cue, could not be answered. In a second experiment, Liberman et al., (1956), attempted to determine the distinction between /b/ and /w/ for various vowels. When the percent judgements /b/ were plotted against transition duration, with vowels as parameters, it appeared that duration was the significant cue. For comparison, when the data were re-plotted with percent /b/ judgments as a function of the rate of the F2TR in Hz/msec, with vowels as parameters, it was clear that the curves separated, but they had essentially the same slope as before. Thus, Liberman et al., (1956) concluded that "duration rather than rate was the controlling cue for distinguishing between stop and semivowel.

Summary of the Review of the Literature

The Gabor trade-off between time and frequency was discussed. The present author assumes that there is a gain in time resolution at the expense of frequency resolution for complex stimuli as a consequence of cochlear pathology. Support for this notion derived from data on: (a) the difference limen for frequency, (b) the critical-band mechanism, (c) spread of masking effects and (d) temporal integration with normal hearing listeners and listeners with sensorineural hearing-impairment. In general results published by others, indicated that: (1) the DLF becomes larger, (2) the critical-band becomes wider, (3) there is a greater spread of masking and (4) the slope of the temporal integration function becomes depressed and is accompanied by a shorter time constant of integration for subjects with sensorineural hearing-impairment as opposed to subjects with normal hearing. Because bandwidth and duration are inversely proportional and since: (1) frequency resolution and time resolution abilities differ in persons with cochlear hearing-impairment from those with normal hearing and (2) several authors have suggested a relationship between frequency resolution and speech discrimination, it is important to investigate those time and frequency paramters of the speech signal that relate to speech discrimination and/or speech identification. Two such variables are the target frequency and duration of the second-formant transition.

Because real speech items are difficult to manipulate and to control with respect to their time and frequency parameters, the use of computer generated speech-like items was offered as an alternative. The speech-like items /bae/ , /dae/ and /gae/ have been extensively used before in categorical perception experiments. Hence, a great deal of data are available on the perception of these test stimuli by listeners with normal hearing. In particular, the target frequency of the F2TR has been shown to be an important cue related to phoneme identification in terms of the places of articulation. In addition, it was shown that there are abrupt boundaries along a continuous physical dimension, e.g. target freqeucny of the F2TR, with respect to phoneme identification. Discrimination of acoustic changes within a given boundary were shown to be at a chance level, while discrimination showed a maximum at the region of the frequency where the phoneme boundary occurred.

Finally, studies by Pickett and his colleagues were reviewed. These authors measured discrimination of F2TR under various listening conditions in subjects with normal hearing and in those with sensorineural hearing losses. They found that in general, discrimination of the F2TR is poorer in subjects with sensorineural hearing loss than in normal hearing listeners. Furthermore, they found that discrimination of F2TR deteriorates to a much greater degree in subjects with sensorineural hearing losses than it does in normal hearing listeners, when the F2TR was discriminated in the presence of F1 and the F1TR. They attributed differences in the discrimination between their groups of subjects to greater-than-normal upward spread of masking effects of the F1 and F1TR on the F2TR. If discrimination of the F2TR is poorer in persons with sensorineural hearing losses than in subjects with normal hearing, it appears likely that phoneme boundaries will likewise be altered with respect to those phoneme boundaries obtained for subjects with normal hearing. Recall that in the categorical perception of synthetic speech-like items, discrimination between stimuli along a continuous frequency dimension is no better than the subjects' ability to identify these same stimuli phonemically.

Chapter III

EXPERIMENTAL METHOD

Subjects

The subjects were divided in two groups: (1) those with normal hearing and (2) those with sensorineural hearing losses most likely of cochlear origin. The type and degree of hearing-impairment were substantiated through audiometric tests conducted prior to the collection of experimental data.

The age of the subjects with normal hearing varied from 23 to 52 years. Three females and two males volunteered as subjects. All subjects were native speakers of American English. Only the right ear was measured experimentally. Pure tone thresholds (at audiometric octave frequencies) were $\leq$ 15 dB HL (ANSI, 1969). Speech discrimination scores for PB-50 words (W-22 word lists, presented via monitored live voice) were equal to or greater than 92% bilaterally at 30 dB SL, for each ear. The speech reception thresholds (W1 and W2 word lists, presented via monitored live voice) agreed in all instances with the average of the three speech frequencies (500, 1,000 and 2,000 Hz) within $\pm$ 5 dB. In none of the subjects was a history of chronic otologic pathology present. Tympanograms were normal bilaterally.

For the hearing-impaired group, subjects were chosen who demonstrated a bilateral, symmetrical, mild to moderately severe, sensorineural hearing loss of a type suggestive of a cochlear site. Subjects beyond 55 years of age were excluded in order to avoid the complications by presbycusis. Subject SS, although 55 years old was included in the study. He had a twenty year history of sensorineural hearing loss of unknown etiology. Pure tone threshold had remained stable ($\pm$ 10 dB) for the past fifteen years.

Bilateral symmetrical hearing losses were necessary so as to avoid the use of masking in the experiment. Like the subjects with normal hearing, all hearing-impaired subjects were native speakers of American English.

The subjects with hearing-impairment underwent a series of audiometric tests including air and bone conduction thresholds, speech reception thresholds, and a performance intensity function for PB-50 words over a range of 10 to 40 dB SL. In addition a site of lesion test series was performed which included the short increment sensitivity index, Békésy tracking (pulsed and continuous), tone decay (Carhart test), tympanometry and measurement of the acoustic reflex.

The short increment sensitivity index (SISI) was administered to the hearing-impaired subjects at .5, 1.0 and 2.0 kHz. The signal consisted of a 1 dB increment superimposed on a 20 dB carrier signal. A SISI score greater than 70 % is considered a positive indication of cochlear pathology (Jerger et al., 1959). Except for subject JKL, who showed negative SISI scores for all frequencies, the subjects with hearing-impairment showed positive SISI scores for a minimum of two frequencies.

Thresholds were also obtained using a Békésy tracking method for pulsed and continuous tones. These audiograms were classified according to the recommendations of Jerger (1960). Type II and type III Békésy audiograms are considered an indication of cochlear pathology. Subject JKL had a type I audiogram, the remaining subjects showed separation between pulsed and continuous thresholds at approximately 1 kHz (continuous dropped below pulsed), and a narrowing of the continuous tone excursions above 1 kHz, i.e. type II audiogram.

Tympanograms were obtained with an American Electromedics Corp. model 83 impedance bridge for all subjects except JW. JW experienced a vertiginous episode upon insertion of the probe. This subject has a history of Meniere's disease and endolymphatic sac decompression surgery. Thus, there was no further attempt to obtain tympanograms or acoustic reflex measurements for this subject. Tympanometry was felt to be necessary in order to evaluate the status of the middle ear mechanism and to determine the pressure at which maximum compliance occurred for subsequent acoustic reflex measurements. A normal tympanogram is one where maximum compliance occurs at a pressure of 0 mm $\pm$ 50 mm H_2O. Jerger (1972) classified normal tympanograms as type A. For the subjects tested, tympanograms were type A.

The presence of recruitment is commonly accepted as a symptom of cochlear pathology. Comparison of the sound pressure levels required to obtain auditory threshold with the sound pressure level required to elicit the acoustic reflex is an objective method for determining the presence of recruitment. Acoustic reflex threshold measurements (minimum detectable meter deflection) were obtained for all subjects except JW at .25, .50, 1.0, 2.0 and 4.0 kHz. Elicitation of the acoustic reflex at levels less than or equal to 60 dB SL (Jerger et al., 1972) is considered a positive indication of recruitment. For those subjects tested, the acoustic reflex was elicited at sensation levels of less than or equal to 60 dB at a minimum of two test frequencies.

Tone dcay was measured for all subjects with hearing-impairment, following the Carhart procedure (Carhart, 1957). All subjects showed less than 20 dB tone decay for the audiometric octave frequencies in the range of 500-4,000 Hz. Thus, tone decay was considered negative for these subjects. The audiometric test results of the better ear for these subjects are summarized in Tables III to V . The better ear, in terms of speech discrimination scores was measured experimentally.

Dependent Variables

The primary dependent variable was a probability score for identification of the four synthetically produced speech-like items /bae/ , /dae/ , /gae/ or /ae/ .

Table III

Air Conduction, Bone Conduction and Acoustic Reflex Thresholds for the Subjects with Hearing-Impairment (test ear) (ANSI, 1969)

Subject		Frequency in kHz .125	.25	.5	1.0	2.0	4.0	8.0	PTA
JW	AC	-	70	70	70	75	70	55	72
	BC	-	NR	NR	70	70	60	-	
	AR	CNT	CNT	CNT	CNT	CNT	CNT		
JKL	AC	35	45	55	65	65	70	NR	63
	BC	-	NR	55	70	70	NR	--	
	AR	-	110*	115	110*	NR	NR		
SB	AC	-	20	25	35	70	55	60	43
	BC	-	20	25	35	65	55		
	AR	-	90	85*	95*	95*	105*		
SS	AC	25	20	20	15	45	70	65	27
	BC	-	15	15	20	45	65	-	
	AR	-	95	90	90	105*	110*	-	

Table IV

Speech Discrimination Scores for Subjects with Hearing-Impairment (percent responses to W-22 word lists)

Subject	Sensation Level in dB 10	20	30	40
JW	46	68	78	CNT
JKL	40	64	72	68
SB	64	72	86	88
SS	50	68	74	70

CNT- could not test

*- Recruitment

Table V

Results of the Site of Lesion Test Battery

Subject	Békésy Audiogram	Tone Decay (Carhart)	SISI .5k	SISI 1k	SISI 2k	Tympanogram type
JW	Type II	negative	+	+	+	CNT
JKL	Type I	negative	-	-	-	Normal (A)
SB	Type II	negative	-	+	+	Normal (A)
SS	Type II	negative	-	+	+	Normal (A)

The identification function was obtained by plotting these probabilities graphically as a function of target frequency of the F2TR. It has been determined by Liberman et al., 1957; Liberman et al., 1967; Studdert-Kennedy et al., 1970 and Pisoni, 1971, that there is a rather abrupt change in the probability of response to a particular phoneme category at certain target frequencies. The frequency (in Hertz) at which two consecutive categories cross is known as the phoneme boundary. The phoneme boundary, as an alternate way of stating the dependent variable, was chosen as the major dependent variable for the purpose of the present study.

Independent Variables

The independent variables were: (1) target frequency of the second-formant transition (F2TR), measured in Hertz, and (2) duration of the F2TR, measured in msec. The target frequency varied, from 1312 Hz for a rising transition, to 2234 Hz for a falling transition at intervals of approximately 75 Hz. There was a total of thirteen intervals. The duration of the F2TR varied from 30 to 5 msec in 5 msec steps for each target frequency.

Simulus Preparation

The test items were prepared at the Haskins Laboratories, New Haven, Connecticut with the cooperation of Dr. Alvin Liberman and with support from the National Institute of Child Health and Development. The synthetic speech items were modifications of those perpared by Pisoni,(1971). They consisted of a continuum with respect to the target frequency of the F2TR. The samples prepared led to the identification of the stimuli by normal hearing listeners, as the voiced-stop-consonant-vowel syllables /bae/ , /dae/ and /gae/ . There was no third-formant, and thus no third-formant transition, no noise bursts preceeding the formant transitions, or any prevoicing prior to the first-formant transition. These parameters were excluded to maintain acoustic redundancy at a minimum. As mentioned above, the duration of the F2TR varied in 5 msec steps from 30 to 5 msec.

Stimulus preparation was aided thorugh the use of the Glace-Holmes parallel resonance synthesizer under computer control. The executive program, in use at the computer facility, permitted precise control of the parameters of the test items. The parameters and their values are listed in Table VI.

The onset of the F2TR was delayed 15 msec with respect to the onset of the F1TR and the amplitude was programmed so as to decline linearly after 215 msec. This was done in an effort to give the stimuli a more natural quality. Both the F1TR and the F2TR varied linearly toward their steady-state frequencies.

Verification of the stimulus values was obtained: (1) in the form of a computer printout of the coded stimulus values and (2) as a sound spectrogram produced as a hard-copy, computer print-out. For additional verification, wide-band (300 Hz) sound spectrograms were obtained for each stimulus with a Kay model 6310 sound spectrum analyzer. Samples of these spectrograms are provided in Figures 4-6 (a-f). The number on top of each spectrogram is the identification number of each stimulus. Table VII lists the stimulus identification numbers with their respective target frequency and duration values for Figures 4-6. For example, stimulus identification # 100 represents a second-formant target frequency of 1312 Hz at 30 msec duration.

The spectrograms depict only the extreme values of the F2TR and a mid-value transition. Specifically, 1312 Hz is an example of a maximum rising transition, 1920 Hz of a moderate falling transition. The wide-band spectrograms visually correlated with the highly schematized computer

Table VI

Control Parameters of the Stimuli

Parameter	Value
1. Fundamental frequency	114 Hz
2. F1 and F2 amplitude	constant at 0 dB max
3. Steady-state frequency of the first-formant	743 Hz
4. Bandwidth of the first-formant	60 Hz
5. Steady-state frequency of the second-formant	1620 Hz
6. Bandwidth of the second-formant	90 Hz
7. Target frequency of the first-formant transition	154 Hz
8. Duration of the first-formant transition	45 msec
9. Target frequency of the second-formant transition	a continuum from 1312 to 2234 Hz in 13, 75 Hz steps.
10. Duration of the second-formant transition	varied over the range of 30 to 5 msec in 5 msec steps.
11. Duration of the steady-state segment	230 msec
12. Overall amplitude	0 dB max which declines linearly after t = 2.5 msec to -22 dB

Table VII

Stimulus Identification Numbers

Stimulus Identification Number	Frequency in Hz	Duration in msec
100	1312	30
101	1312	25
102	1312	20
103	1312	15
104	1312	10
105	1312	5
148	1920	30
149	1920	25
150	1920	20
151	1920	15
152	1920	10
153	1920	5
172	2234	30
173	2234	25
174	2234	20
175	2234	15
176	2234	10
177	2234	5

generated spectrograms. Variation in target frequency and duration may be found in Figures 4-6 (a-f). On the basis of inspection of the spectrograms, stimulus # 140 (1772 Hz, 10 msec) was rejected from analysis because it was contaminated by a click occurring 10 msec after the onset of the F2TR. Figure 7 is a waveform of stimulus # 172. The upper trace is the electrical waveform across TDH-49 headphones. The lower trace is the pressure waveform obtained from the output of a Bruel and Kjaer model 2203 sound level meter.(linear scale). The headphone was mounted

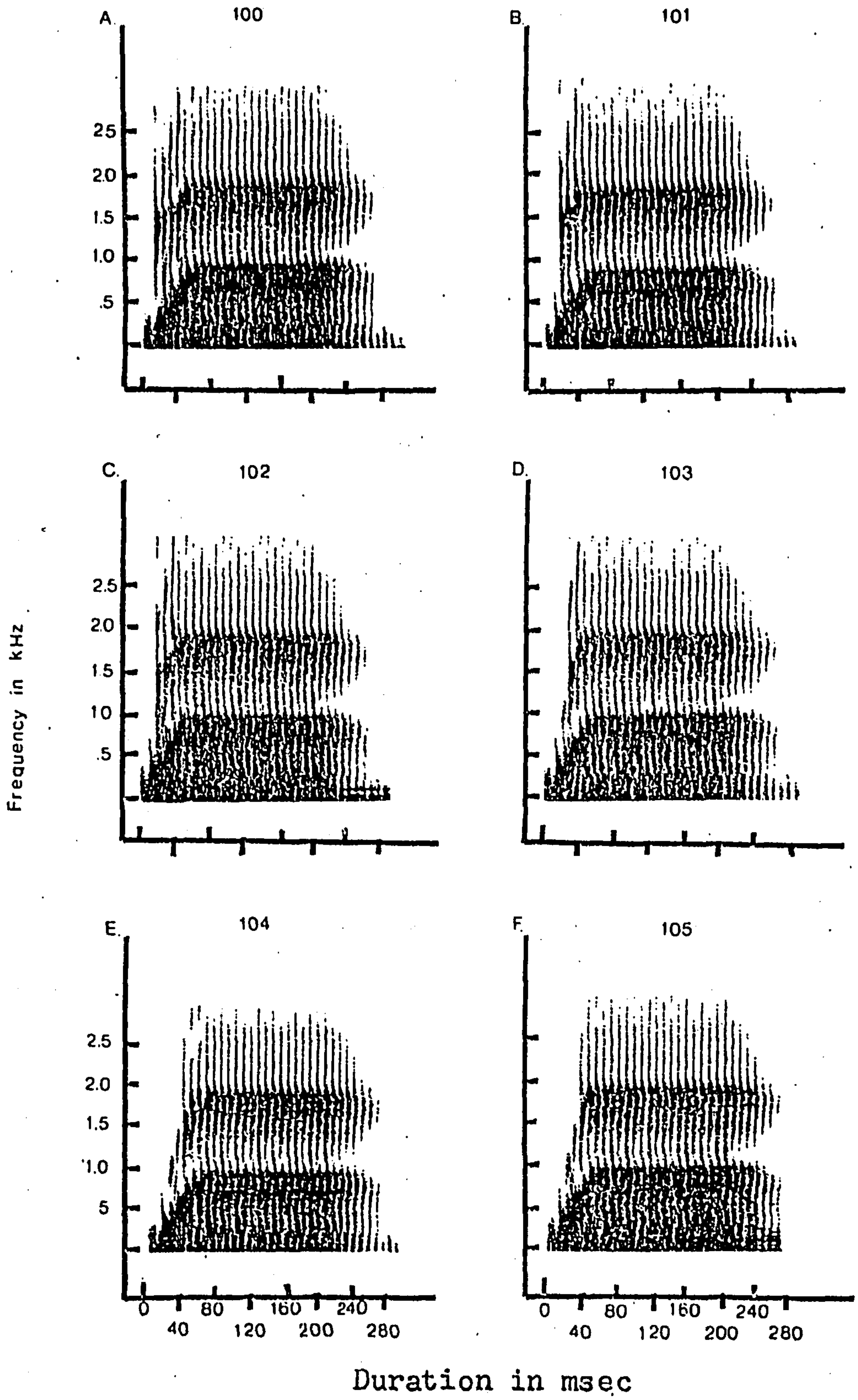

Figure 4. Sound spectrograms of stimuli 100-105. Target frequency of the F2TR = 1312 Hz, duration = 30 msec at A and decreases in 5 msec steps through F.

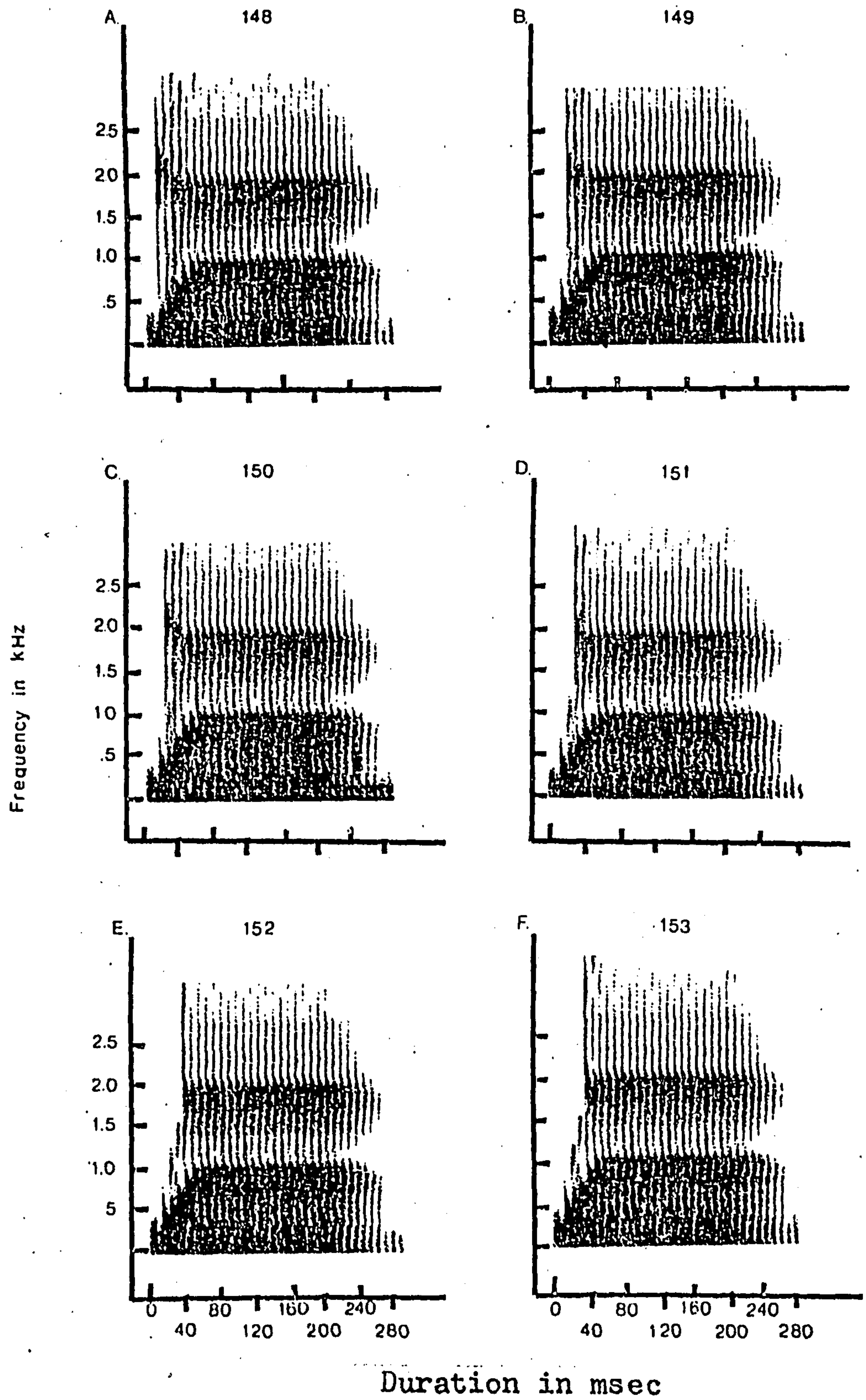

Figure 5. Sound spectrograms of stimuli 148-153. Target frequency of the F2TR = 1920 Hz, duration = 30 msec at A and decreases in 5 msec steps through F.

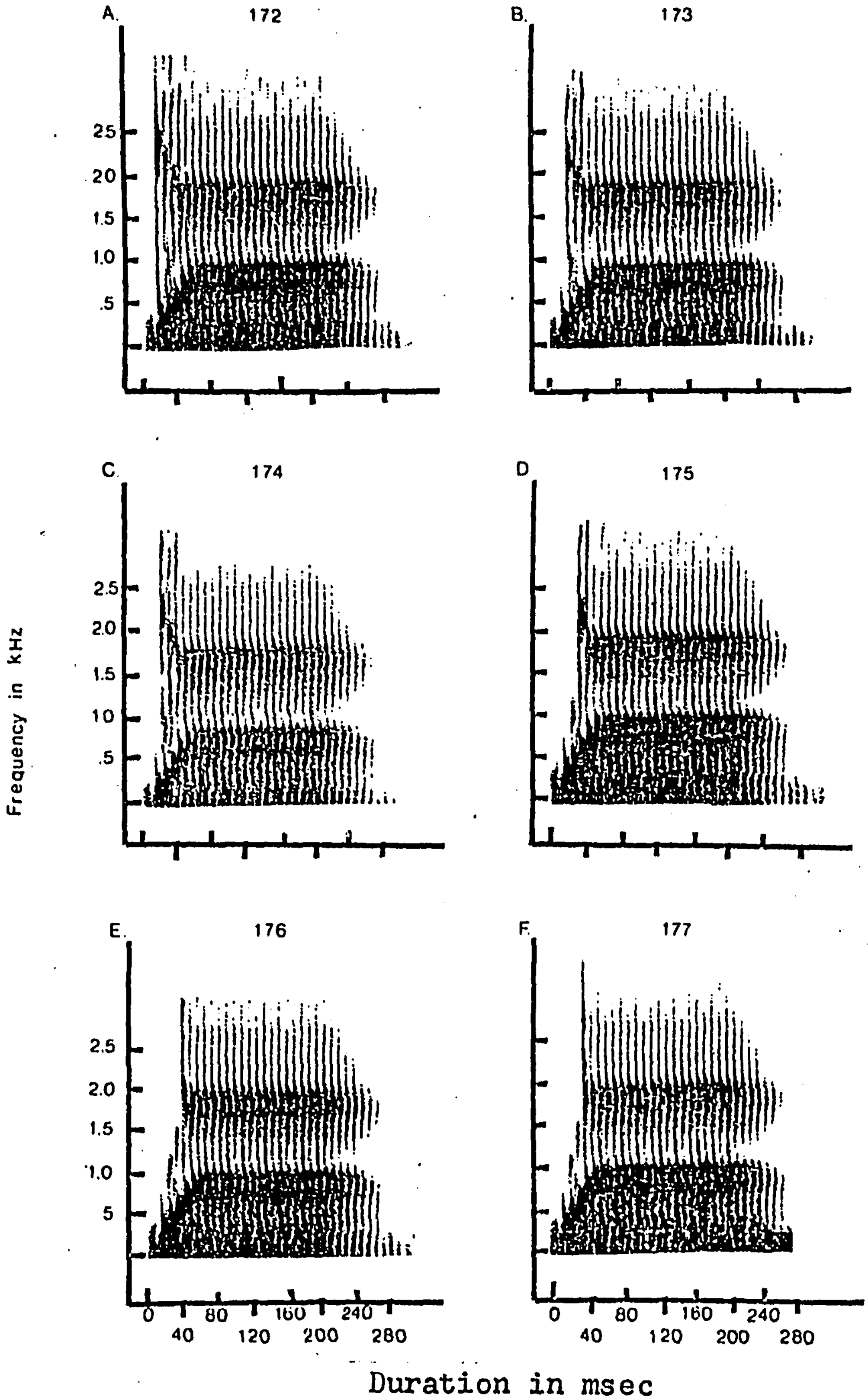

Figure 6. Sound spectrograms of stimuli 172-177. Target frequency of the F2TR = 2234 Hz, duration = 30 msec at A and decreases in 5 msec steps through F.

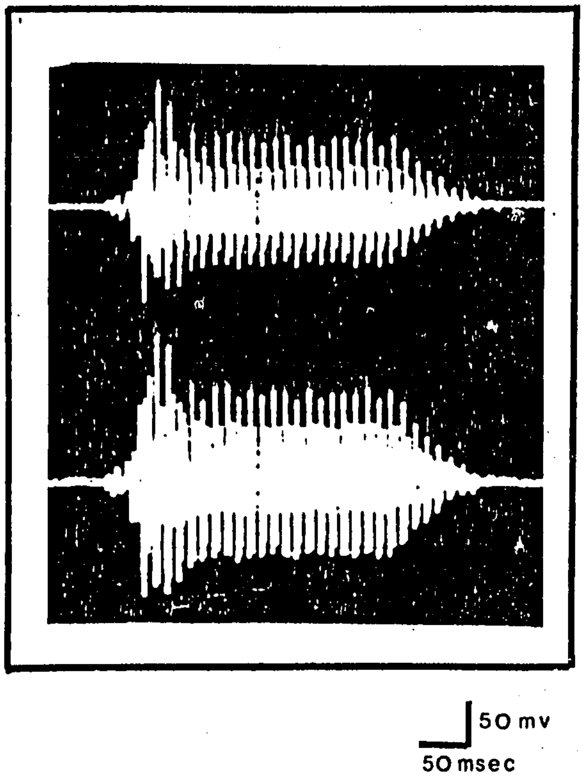

Figure 7. Waveforms of stimulus number 172 (f = 2234 Hz, t = 30 msec). The upper trace is the electrical waveform across TDH-49 headphones. The lower trace is the pressure waveform measured in an NBS-9A coupler.

in an NBS-9A coupler. It may be seen that: (a) there was no substantial ringing or distortion in the lower waveform and (b) there was a linear decline of amplitude after t = 215 msec.

There were a total of 13 target frequencies in the range of 1312 to 2234 Hz. Each target frequency had six durations. A steady-state vowel /ae/ was also synthesized and included as a control and calibration signal.

The stimuli (a total of 79) were sequenced in the following way. All synthesized items were stored in successive memory locations of the computer disc file. The stimuli were then pulse code modulated (PCM), i.e. converted to digital waveforms. The sampling rate was 8 kHz. After digitization, the signals were stored in successive PCM locations (I.D. #'s 100-178). The PCM computer program permits sequencing of the items by first listing all the locations and the number of repetitions of each item (in this study 5 x per tape) in each list. The stimuli were shuffled electronically in a pseudo-random order. At this point, interstimulus and interblock intervals were inserted under computer control. The stimuli were arranged in blocks of 11, with an interstimulus interval of 2 sec and an interblock interval of 4 sec. Two sequences were prepared in this way. The digital waveforms were then converted, via a digital-to-analog converter, to analog form.

Calibration

After conversion back to analog form, the test items were recorded on magnetic audio tape at 0 VU, with a tape speed of 7½ ips. Five one-second steady-state vowels (/ae/) were recorded at 0 VU prior to the recording of the test stimuli. These calibration signals were of one-second duration in order to overcome the VU meter and true rms voltmeter ballistics. On playback, the signals from the tape deck were passed through an amplifier to a 600 ohm attenuator (providing 5 dB steps), then to another 600 ohm attenuator (providing 1 dB steps). Each attenuator had an output impedance of 10 kohm. This was an appropriate input impedance for the tape input of a second amplifier. Levels were adjusted to produce 100 mv (rms) across a 10 ohm TDH-49 headphone. Voltages were measured with a Bruel and Kjaer model 2409 true rms voltmeter. This 100 mv level produced a sound pressure level of 106 dB in a Bruel and Kjaer model 4152 artificial ear. The headphone was mounted in an Mx-41 AR cushion. Sound pressure level were monitored with a Bruel and Kjaer model 2203 sound level meter (linear scale).

Procedure

Subjects were seated in a room that conformed to ANSI, 1969 standards for audiometric testing. The training consisted of a two hour session during which time the subjects were exposed to both tapes twice. Each tape had a different randomization of test items. Signals were presented to all subjects at 30 dB above their speech reception thresholds. There were two reasons for this choice of level. All subjects with normal hearing had speech discrimination scores of at least 92% at 30 dB SL, and the hearing impaired subjects showed no evidence of roll-over on the performance intensity function at this level. A sensation level of 30 dB was established by converting the speech reception threshold to sound pressure levels and adding 30 dB.

The subjects were provided with an answer sheet whereon they were instructed to check an appropriate box under the column heading /bae/ , /dae/ , /gae/ or /ae/ each time the stimulus was presented. This, therefore, constituted a closed response set.

CHAPTER IV

RESULTS

Analysis of the data was primarily graphic. The data from the subjects with normal hearing, as well as from those with cochlear hearing-impairments were averaged separately. Although there were individual differences in the performance among the members of both groups, it was felt that an analysis of the differences would not aid in the interpretation of the data in this initial study. However, future experiments should be designed to evaluate individual differences. In no instance did a subject provide responses that could be interpreted as qualitatively different from those of other, similar subjects (see Appendices A-D). In addition, statistical analysis was not applied to the data because, in the opinion of the author, the conclusions drawn were obvious from inspection of the graphs and the experimental design precluded statistical inference.

Frequency Boundaries for the Subjects with Normal Hearing

The raw data consisted of a tally of the number of phoneme identification responses made by each subject for the categories /bae/ , /dae/ , /gae/ and /ae/ , for each stimulus. The raw data were converted to a probability score by dividing the number of responses for each category by the number of times the stimulus was presented. Then, the individual probability scores were averaged across subjects (N= 5).

Table VIII depicts the average probability of indentification responses for the phoneme categories /bae/ , /dae/ and /gae/ , made by the subjects with normal hearing. For example, for a target frequency (F2TR) of 1312 Hz, the probability of a /b/ response at a duration of 30 msec equalled 1.00. By reading down the /b/ column at 30 msec, it may be observed that the probability of the /b/ response decreased with increasing target frequency of the F2TR. Inspection of the /d/ column at 30 msec reveals a minimum average probability score of 0.00 at 1312 Hz; a maximum average probability score of .90 at 1695 Hz and 1772 Hz; and a minimum probability score of 0.00 at 2156 and 2234 Hz. Reading across the table at a particular target frequency, one may observe the distribution of the phoneme identification responses, in terms of the average probability score, as a function of duration of the F2TR. The /ae/ responses were not included in the table since: (1) the subjects identified the stimulus with no transitions as /ae/ with p = 1.00, and, (2) in no instance was a stimulus that had a F2TR, identified as /ae/ .

Identification functions (Figures 8-10) were plotted, using the probability scores from Table VIII, each of them as a function of target frequency of the F2TR, with duration of the F2TR as the parameter.

Table VIII

Average Probability of Response for Five Subjects with Normal Hearing

Target Frequency F2TR	Probability of Response											
	Duration of the Second-formant transition											
	30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	1.00	.00	.00	.98	.02	.00	1.0	.00	.00	.98	.02	.00
1387	.96	.04	.00	1.0	.00	.00	.98	.02	.00	1.0	.00	.00
1465	.92	.06	.00	.96	.04	.00	.98	.02	.00	.98	.02	.00
1541	.48	.52	.00	.70	.30	.00	.68	.32	.00	.80	.20	.00
1620	.30	.70	.00	.22	.78	.00	.36	.64	.00	.36	.62	.02
1695	.10	.90	.00	.08	.92	.00	.12	.86	.02	.14	.84	.02
1772	.04	.90	.06	.04	.94	.02	.10	.88	.02	.20	.80	.00
1845	.04	.61	.34	.00	.88	.12	.10	.76	.14	.22	.72	.06
1920	.02	.52	.46	.02	.74	.24	.04	.70	.26	.08	.80	.12
1996	.00	.20	.78	.18	.52	.30	.08	.44	.48	.20	.62	.18
2078	.00	.10	.90	.00	.44	.56	.02	.36	.62	.26	.34	.40
2156	.00	.00	1.0	.00	.06	.94	0.0	.10	.90	.16	.58	.26
2234	.00	.00	1.0	.00	.08	.92	.02	.20	.78	.20	.48	.32

continued on next page

Table VIII continued

Average Probability of Response for Five Subjects with Normal Hearing

Target Frequency F2TR	Probability of Response					
	Duration of the Second-formant transition					
	10 msec			5 msec		
	b	d	g	b	d	g
1312	.96	.04	.00	.94	.06	.00
1387	.96	.04	.00	1.0	.00	.00
1465	.98	.02	.00	1.0	.00	.00
1541	.84	.16	.00	.88	.06	.06
1620	.62	.38	.00	.72	.28	.00
1695	.52	.48	.00	.58	.40	.02
1772	-	-	-	.52	.48	.00
1845	.28	.72	.00	.34	.60	.06
1920	.48	.52	.00	.46	.54	.00
1996	.44	.56	.00	.78	.20	.02
2078	.50	.48	.02	.46	.52	.02
2156	.56	.38	.06	.62	.38	.00
2234	.62	.38	.00	.72	.24	.04

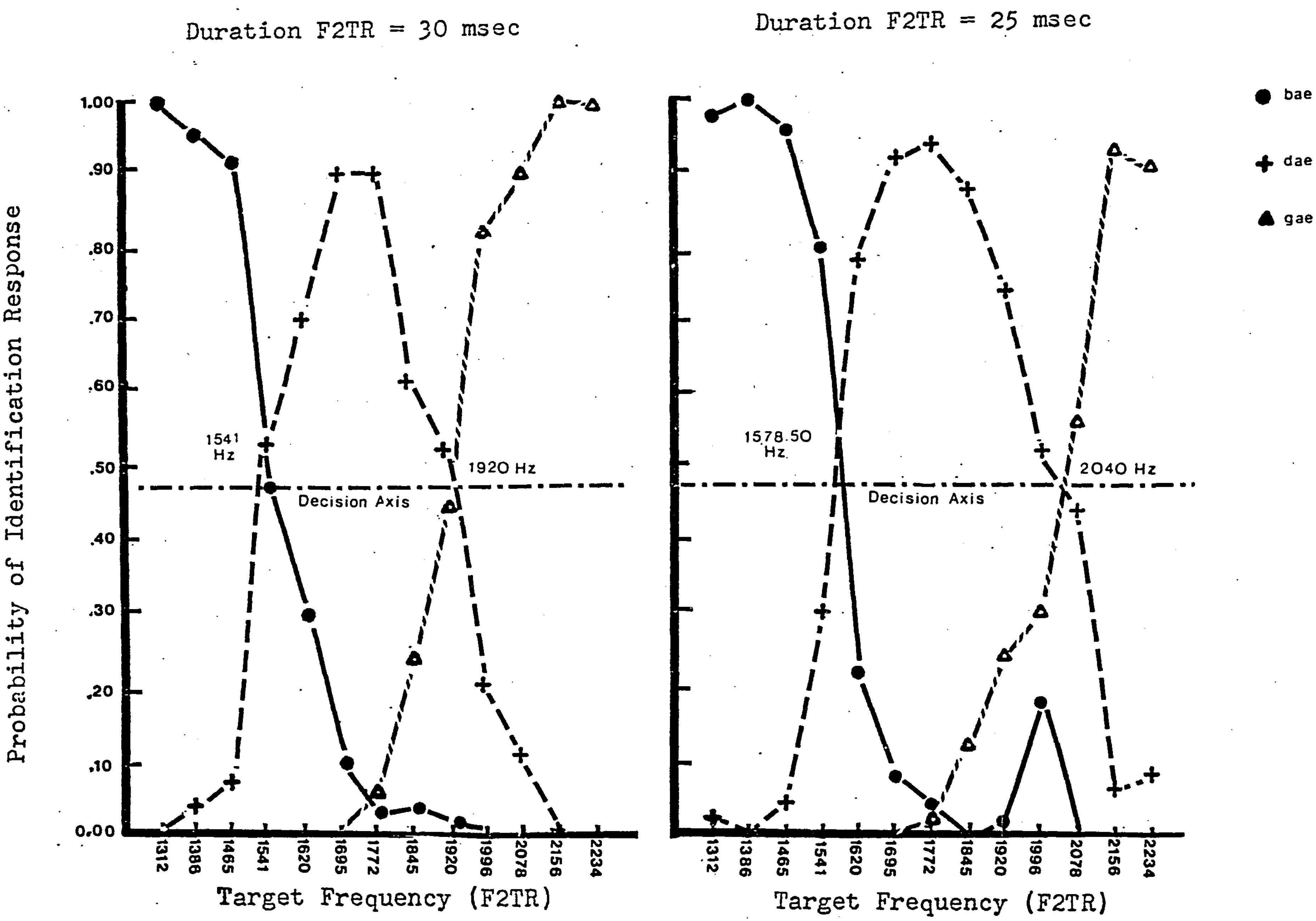

Figure 8. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with normal hearing.

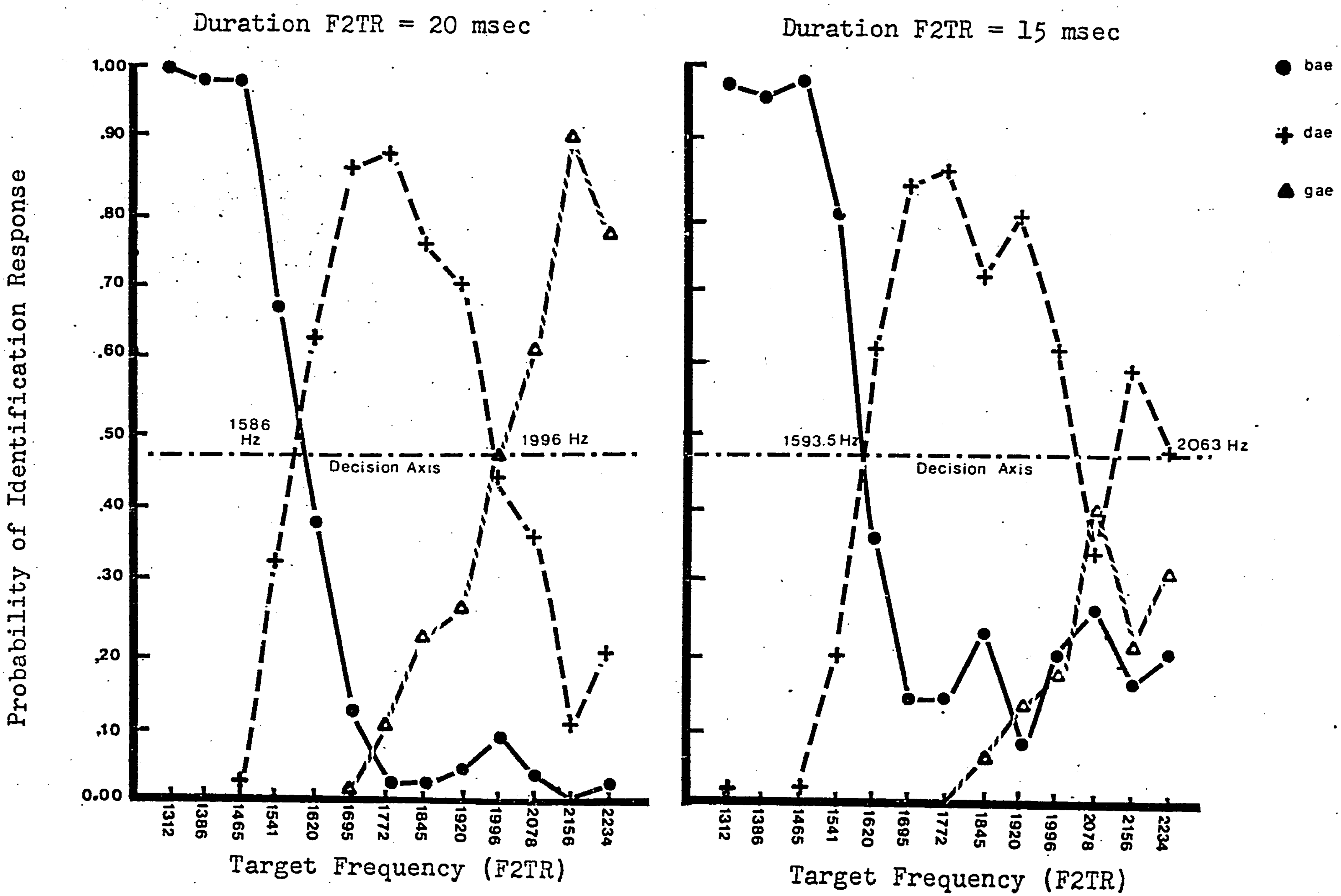

Figure 9. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with normal hearing.

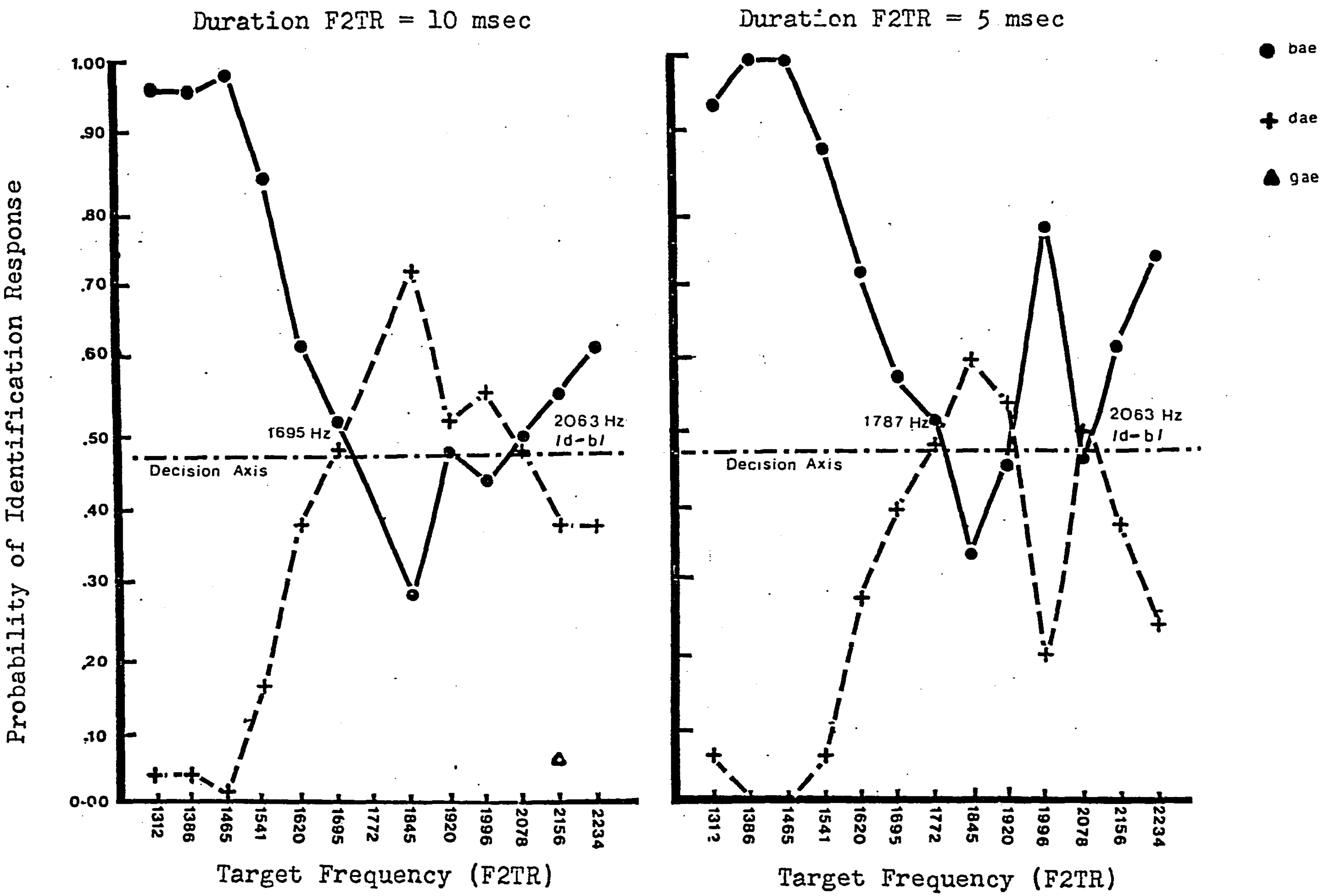

Figure 10. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with normal hearing.

The phoneme boundary is that point on the identification function where the curves from two successive categories cross. In Figures 8-10, the values for the /b-d/ boundary and the /d-g/ boundary are shown, in Hz, immediately adjacent to the boundary. These frequency boundaries were determined by dropping a perpendicular to the abcissa and converting to frequency. Each data point on the identification function represents a probability score based on 50 responses. According to Natrella, 1963 (p. T-45), for an N of 50 and a three-choice decision task, a one-sided probability of .47 is a significant departure from chance at a 99 percent level of confidence. This probability is shown as a decision axis (broken horizontal line) across each graph.

In chapter I (p. 6) it was hypothesized for the subjects with normal hearing, that the phoneme boundary, in Hz, would shift toward higher frequencies as a function of decreasing duration of the F2TR. Support for this hypothesis was found by inspection of the identification functions (Figures 8-10) and of Figure 11.

Figures 8-10 indicated that the subjects distinctly categorized the stimuli phonemically with respect to the target frequency of the F2TR. For the /b-d/ phoneme boundary there was a monotonic increase in the value of the frequency boundary with decreasing duration of the F2TR. The frequency domain /b-d/ boundary values varied from 1541 Hz obtained at 30 msec, to 1787 Hz obtained at 5 msec.

Hence, the /b-d/ frequency boundary and the duration of the F2TR were inversely related. The /d-g/ frequency boundary also increased in frequency with decreasing duration of the F2TR. The boundary values varied from 1920 Hz, obtained at 30 msec, to 2063 Hz, obtained at 15 msec. However, at 20 msec there was a decrease in the value of the /d-g/ frequency boundary. At 15 msec the probability of the /gae/ response dropped well below $p = .47$. As a consequence, the /dae/ response predominated, that is, $p \geq .47$, through the target frequency range formerly occupied by /gae/. At durations of 10 and 5 msec, the probability of the /bae/ response increased at the higher target frequencies of the F2TR, thereby significantly reducing the probability of the /dae/ response in the mid and high frequency range. Thus, a /d-b/ phoneme boundary emerged.

In Figure 11, the /b-d/ and /d-g/ boundaries, in Hz, are plotted as a function of duration of the F2TR. The curves are a visual best fit to the data. Each data point is the average of five boundary determinations at each duration. It was evident that for both the /b-d/ and the /d-g/ phoneme boundaries, the frequency boundary increased with decreasing duration of the F2TR.

These results suggest that the subjects with normal hearing were using duration as well as target frequency of the F2TR, as cues for phoneme identification.

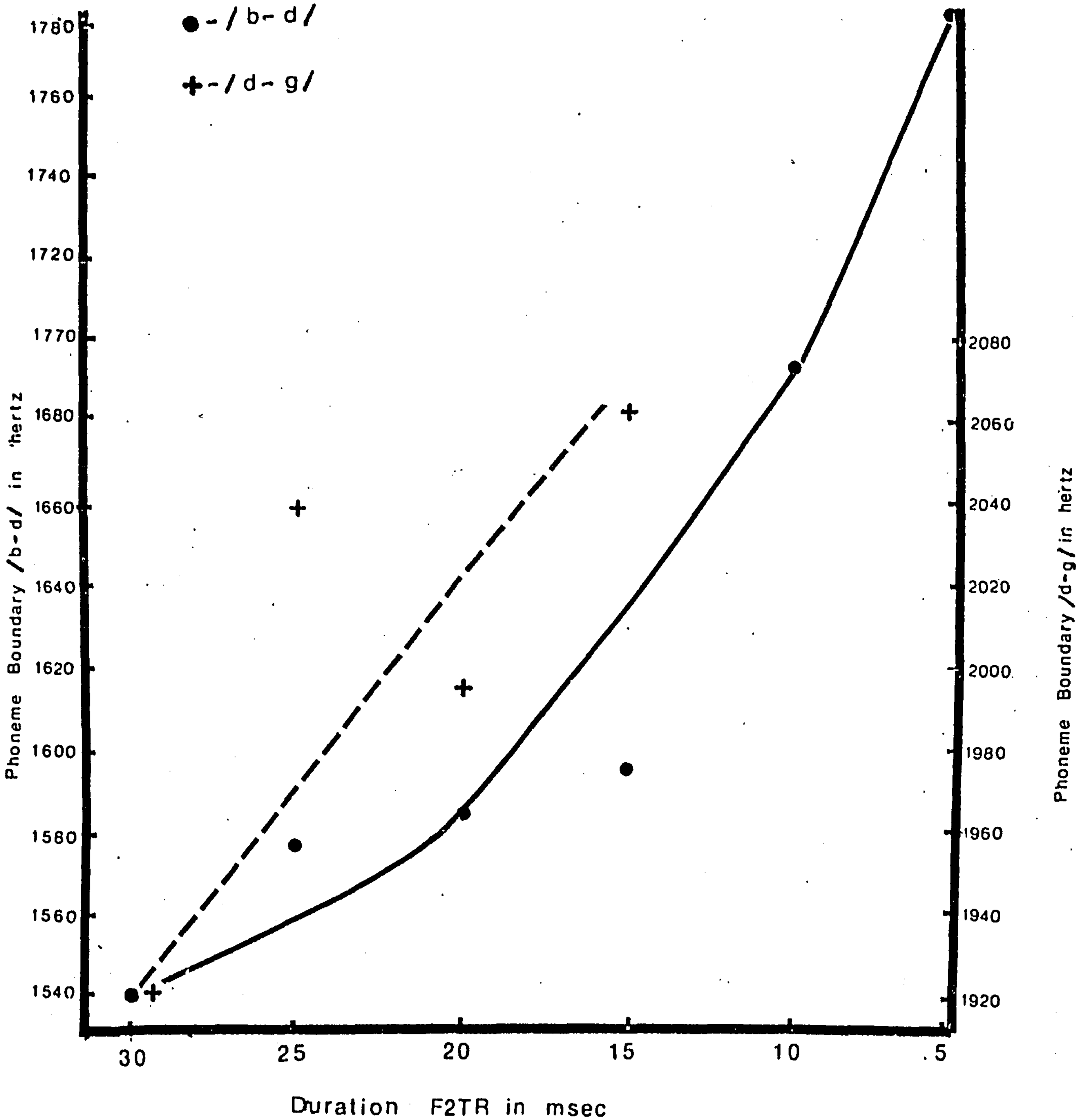

Figure 11. Phoneme Boundary in Hz, as a Function of the Duration F2TR in msec for subjects with normal hearing.

Frequency Boundaries for Subjects with Cochlear Hearing-Impairment

Table IX depicts the average probability scores for the phoneme categories /bae/ , /dae/ and /gae/ for all stimuli as perceived by the subjects with cochlear hearing-impairments. The probability scores for individual subjects are presented in tabular form and may be found in appendices A through D. The /ae/ responses were eliminated from the table, and thus from further consideration, for the same reasons as those given with respect to the subjects with normal hearing (see p. 70). The probability scores were derived in the same way as for the subjects with normal hearing.

The following hypothesis was presented in chapter I (p. 6): Under the assumption that frequency resolution is poorer in cochlear hearing losses than in normal hearing, it was hypothesized that subjects with cochlear hearing losses would show phoneme boundaries, in Hz, at higher target frequencies of the F2TR than subjects with normal hearing.

It is clear that the identification functions (Figures 12-14), were quite different from those assessed in the subjects with normal hearing (Figures 8-10 pp. 73 to 75). It was apparent, by inspection of Figure 12 (duration F2TR = 30 msec), that the subjects with cochlear hearing-impairment identified all target frequency variables as /gae/

Table IX

Average Probability of Response for Four Subjects with Hearing-Impairment

Target Frequency F2TR in Hz	Probability of Response — Duration of the second-formant transition: 30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	.23	.05	.73	.55	.43	.02	.75	.25	.00	.55	.45	.00
1386	.23	.02	.75	.33	.40	.25	.33	.58	.10	.53	.48	.00
1465	.20	.10	.70	.33	.43	.24	.38	.55	.08	.53	.43	.04
1541	.13	.10	.78	.30	.68	.02	.48	.50	.02	.50	.43	.07
1620	.10	.20	.70	.13	.30	.55	.38	.58	.04	.53	.43	.04
1695	.08	.30	.63	.28	.63	.10	.30	.58	.13	.48	.53	.00
1772	.02	.33	.65	.10	.78	.13	.18	.70	.15	.33	.68	.00
1845	.00	.35	.65	.23	.68	.15	.23	.55	.23	.33	.65	.02
1920	.02	.48	.50	.35	.43	.23	.30	.58	.13	.50	.43	.07
1996	.05	.25	.70	.13	.55	.33	.40	.45	.15	.35	.58	.07
2078	.00	.13	.88	.10	.58	.33	.28	.50	.23	.40	.58	.02
2156	.00	.08	.90	.00	.43	.58	.25	.50	.25	.45	.53	.02
2234	.00	.15	.85	.05	.45	.50	.15	.65	.20	.43	.53	.04

Table IX continued

Average Probability of Response for Four Subjects with Hearing-Impairment

Target Frequency F2TR in Hz	Probability of Response — Duration of the second-formant transition: 10 msec			5 msec		
	b	d	g	b	d	g
1312	.50	.50	.00	.63	.35	.03
1386	.60	.40	.00	.65	.33	.03
1465	.65	.35	.00	.78	.22	.00
1541	.83	.17	.00	.58	.40	.02
1620	.80	.20	.00	.90	.10	.00
1695	.48	.52	.00	.53	.45	.02
1772	-	-	-	.78	.23	.00
1845	.48	.50	.02	.65	.35	.05
1920	.75	.23	.02	.70	.28	.02
1996	.60	.38	.02	.73	.25	.02
2078	.70	.28	.02	.65	.35	.00
2156	.78	.18	.04	.88	.10	.02
2234	.73	.25	.02	.78	.18	.04

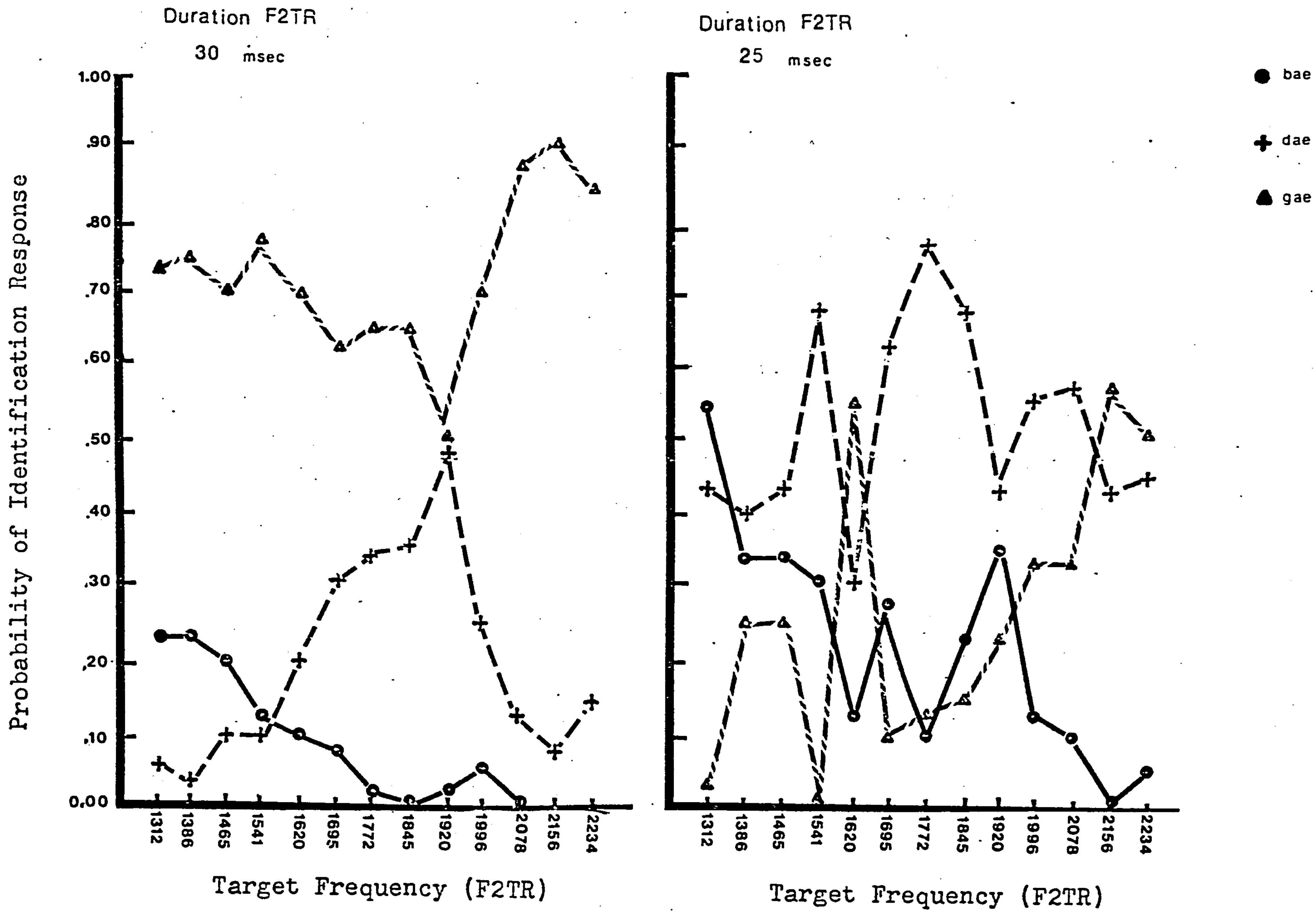

Figure 12. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with cochlear

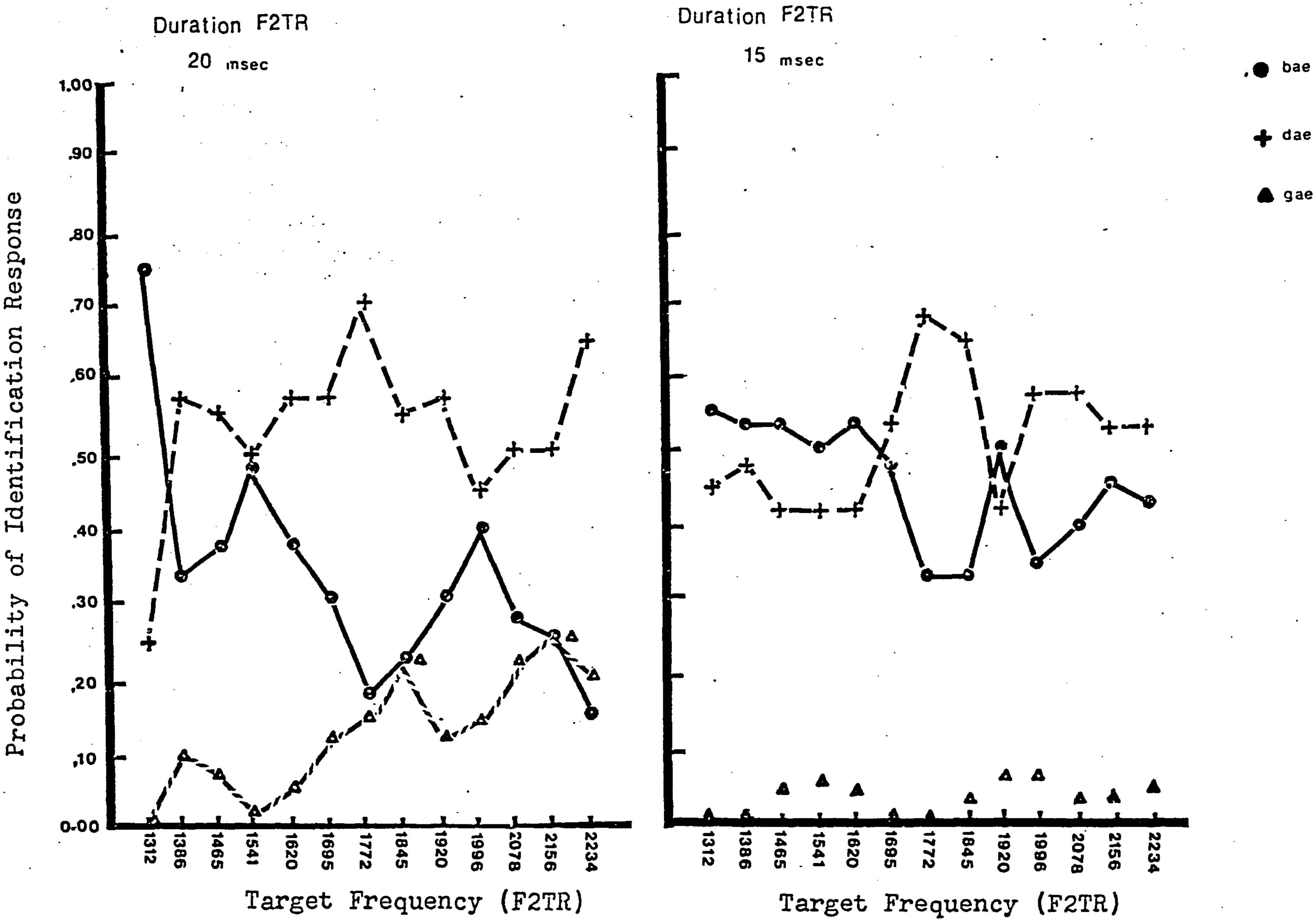

Figure 13. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with cochlear hearing-impairment.

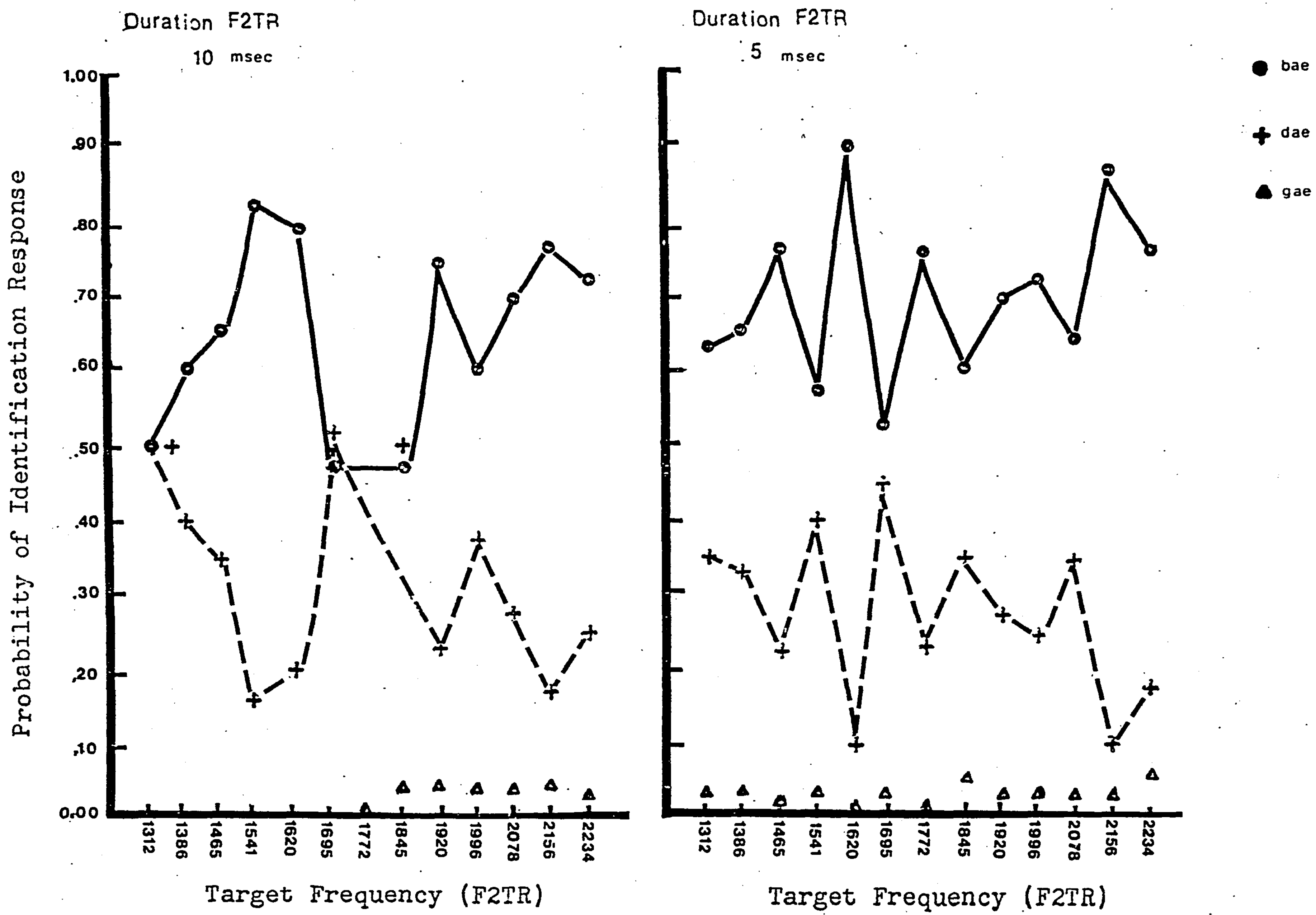

Figure 14. Probability of Identification Responses as a Function of Target Frequency (F2TR) for subjects with cochlear hearing-impairment

with an average probability score greater than or equal to .50 and less than or equal to .90 ($.50 \leq p \leq .90$); whereas, the subjects with normal hearing distinctly identified the same stimuli as /bae/ , /dae/ and /gae/ as the case might be. At the F2TR duration of 25 msec (Figure 12), the stimulus identification abruptly changed to /dae/ over a target frequency range of 1312 Hz through 2078 Hz ($.30 \leq p \leq .80$). It should be noted that a 5 msec decrease in the duration of the F2TR, caused a significant change in stimulus identification. For the subjects with normal hearing, the same 5 msec decrease in duration resulted only in an increase in the position of the /b-d/ and /d-g/ boundaries in the frequency domain (Figure 8). Similarly, in Figure 13 (duration F2TR = 20 msec), the /dae/ response was most likely ($.25 \leq p \leq .70$), while at 15 msec /bae/ and /dae/ were equiprobable. At F2TR durations of 10 and 5 msec (Figure 14), the /bae/ response was maintained throughout the entire target frequency range of the F2TR with, $.50 \leq p \leq .90$, and $.50 \leq p \leq .83$ respectively.

The data for the subjects with cochlear hearing-impairment was interpreted as supporting the hypothesis that phoneme boundaries along a continuous frequency dimension for the subjects with cochlear hearing-impairments would differ from those boundaries obtained for subjects with normal hearing. Indeed, frequency domain phoneme boundaries could not be established for the group of subjects with cochlear pathology.

Phoneme Boundaries in the Time Domain

It was apparent that both groups of subjects utilized duration of the F2TR as a cue for phoneme identification. In order to examine the relationship between duration and phoneme identification more carefully, the probability scores listed in tables VII and IX, were replotted as a function of the logarithm of duration of the F2TR, with target frequency of the F2TR as the parameter.

Subjects with Normal Hearing

Recall that in chapter I (p.6) it was hypothesized for the subjects with normal hearing, that as target frequency of the F2TR increases, sensitivity with respect to duration of the F2TR would also increase. It was suggested that greater time sensitivity would be demonstrated by more abrupt phoneme boundary shifts (in the time domain), for larger target frequencies than for smaller ones.

Support for this hypothesis may be found by inspection of Figures 15 through 19. The target frequency parameter begins at 1541 Hz because, for the group of subjects with normal hearing, there were no time domain phoneme boundaries below that frequency, i.e. at 1312 Hz through 1465 Hz. It may be observed (Figures 15-19) that abrupt phoneme boundaries occurred in the time domain, and that the shapes of the curves resemble those of Figure 8-10. The phoneme

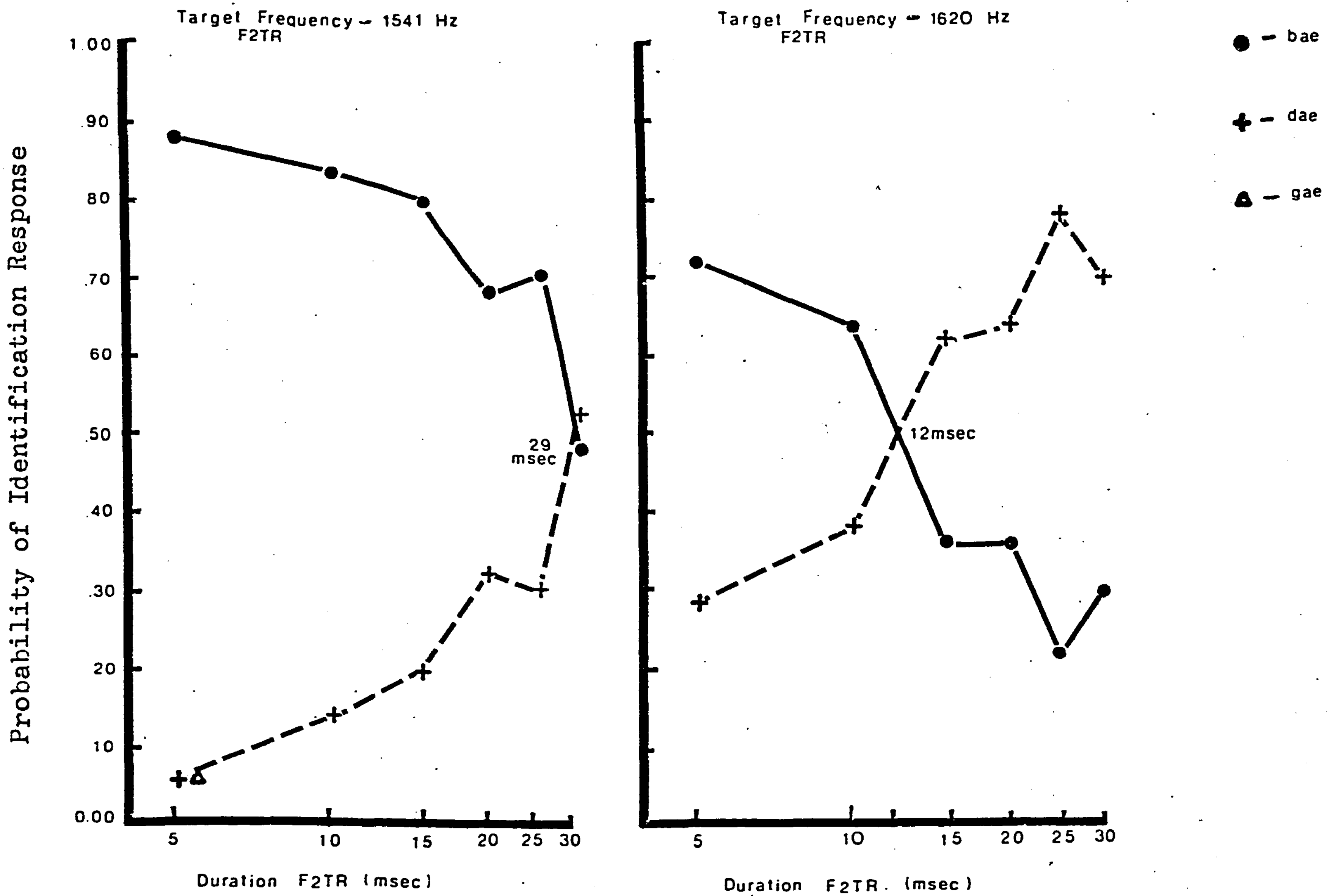

Figure 15. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with normal hearing.

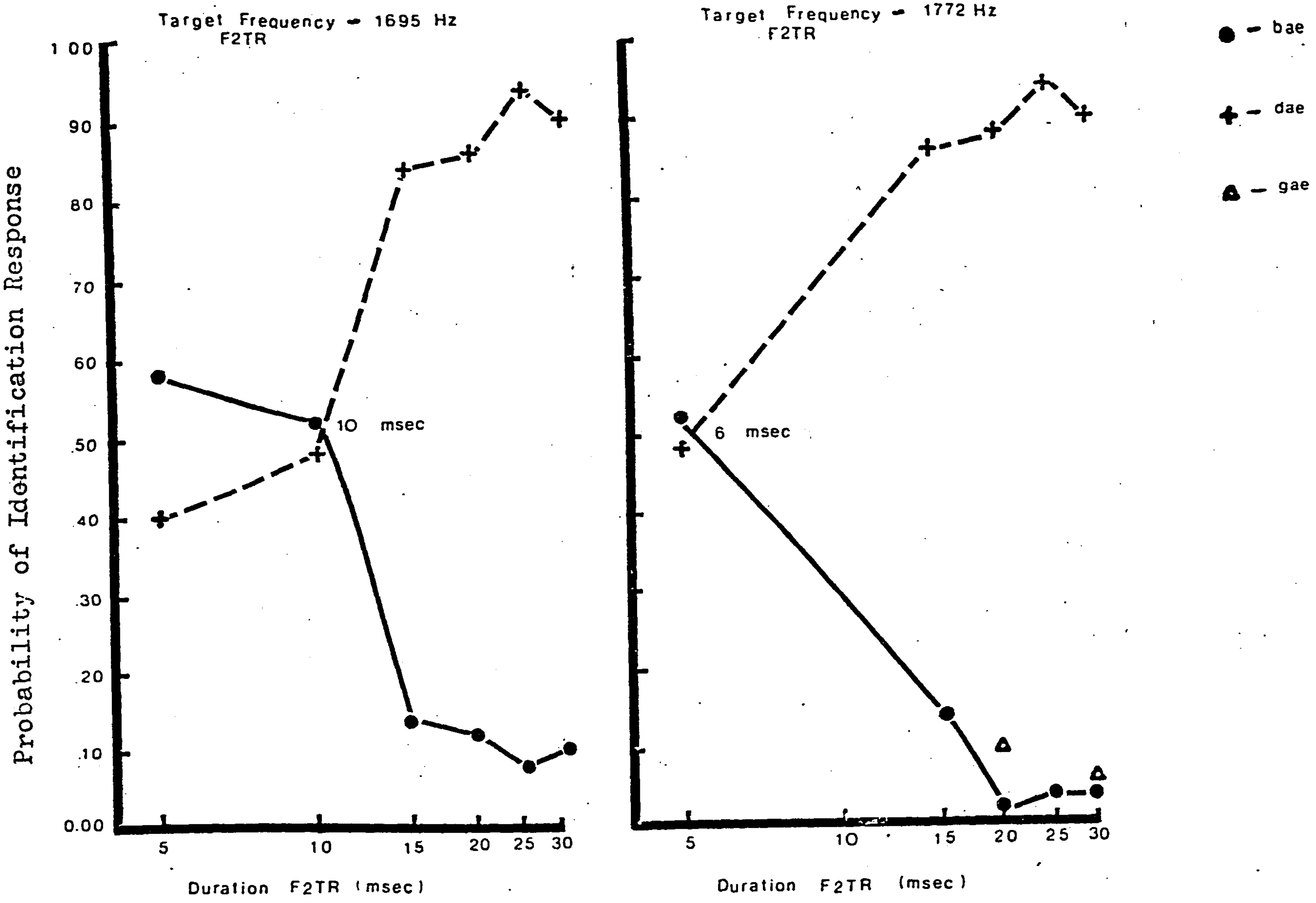

Figure 16. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with normal hearing.

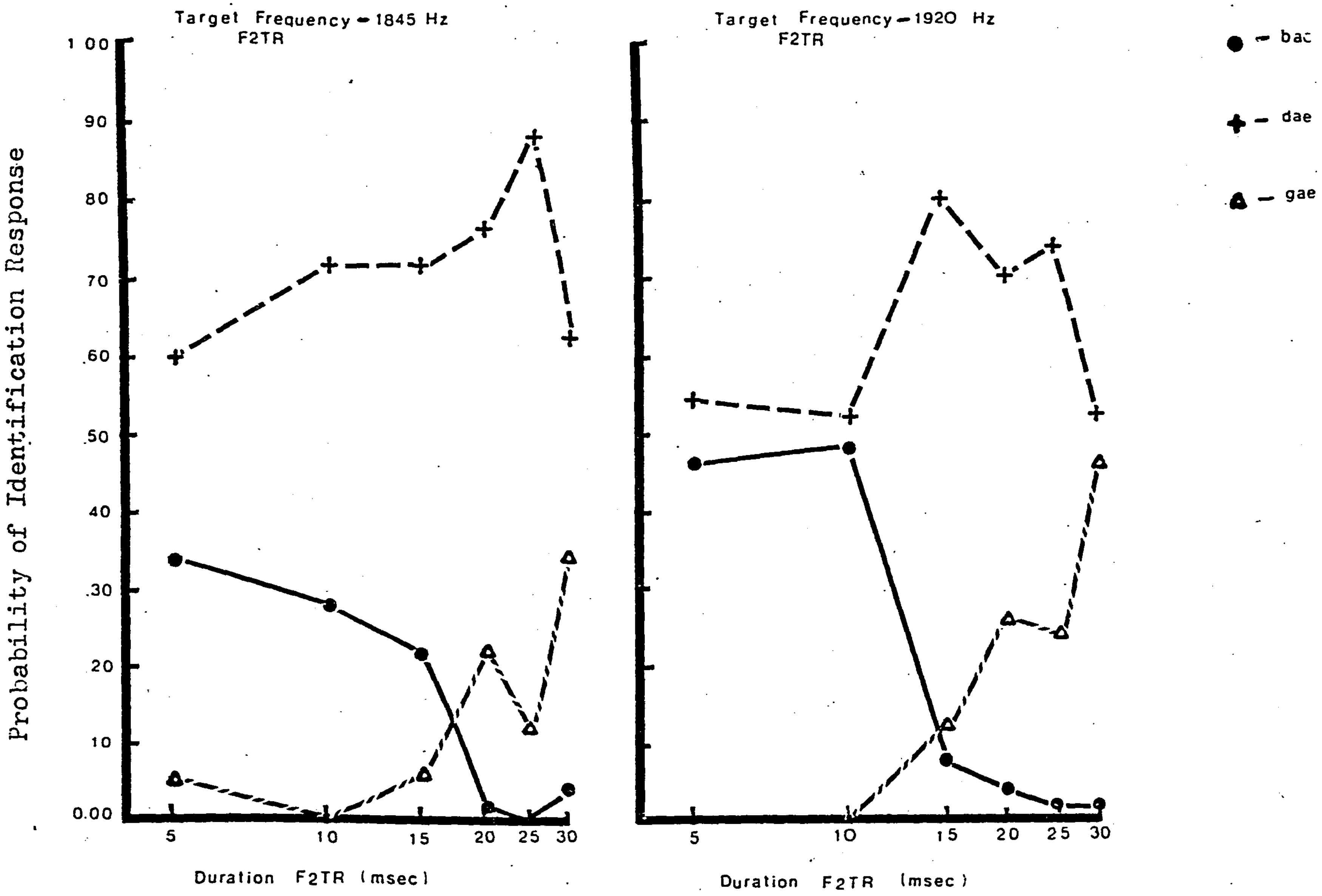

Figure 17. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with normal hearing.

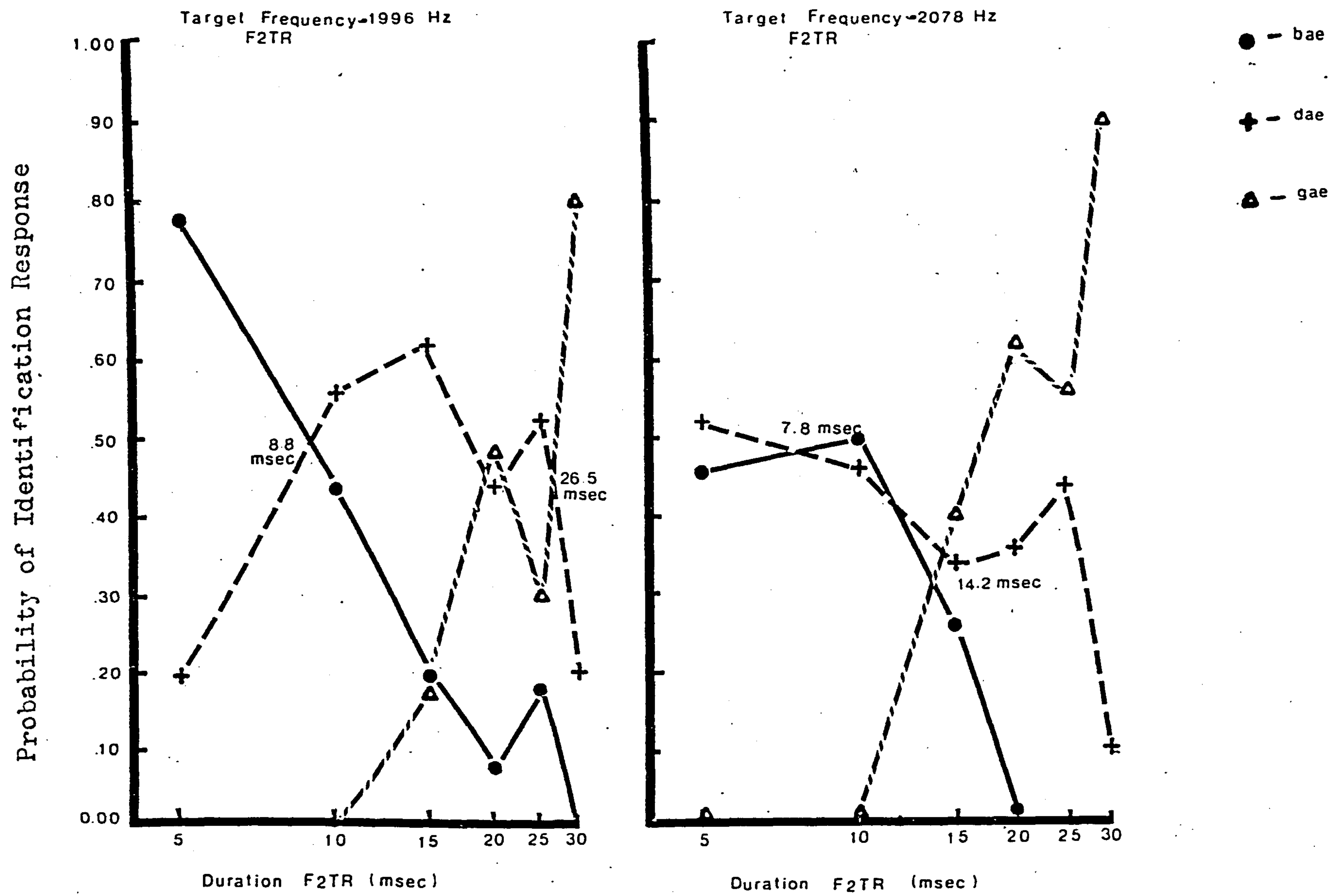

Figure 18. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with normal hearing.

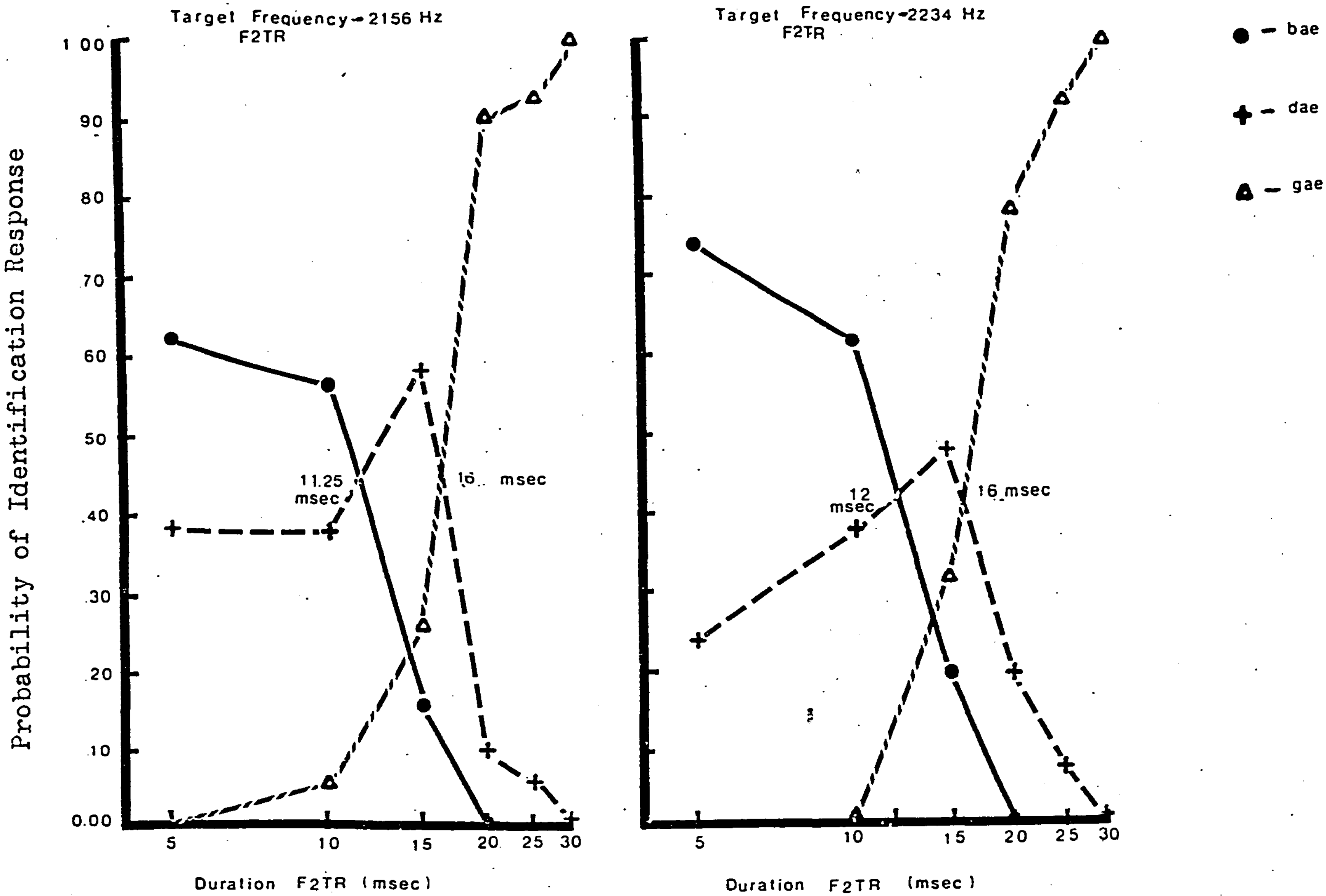

Figure 19. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with normal hearing.

boundary, in msec, is listed adjacent to the intersection of the response categories on each time domain identification function. The boundaries were obtained by dropping a perpendicular, at the intersection, to the x-axis.

At 1541 Hz (Figure 15), the /bae/ response predominated at all durations of the F2TR. However, the probability of the /dae/ response increased from a minimum of .06 at 5 msec, to a maximum of .52 at 30 msec. Thus, a /b-d/ boundary occurred at 29 msec. At 1620 Hz (Figure 15), there were two distinct phoneme categories with a boundary at 12 msec. In Figure 16 (f = 1695 Hz and 1772 Hz) the /dae/ response became increasingly likely, with boundaries at 10 msec and 6 msec respectively. When target frequency of the F2TR was increased to 1845 Hz, and to 1920 Hz (Figure 17), no phoneme boundary occurred. Hence, the boundary may be considered to be at negative infinity. The /dae/ response predominated throughout the range of durations. Inspection of the time domain boundaries of Figure 15-17 reveals a monotonic decrease in the boundary value, with increasing target frequency of the F2TR. Inspection of Figures 18 and 19 reveals that for target frequencies greater than or equal to 1996 Hz, three response categories were utilized; namely, /gae/ at long durations, /dae/ at intermediate durations and /bae/ at short durations. This finding suggests greater sensitivity with respect to durational changes of the F2TR at higher target frequencies of the F2TR.

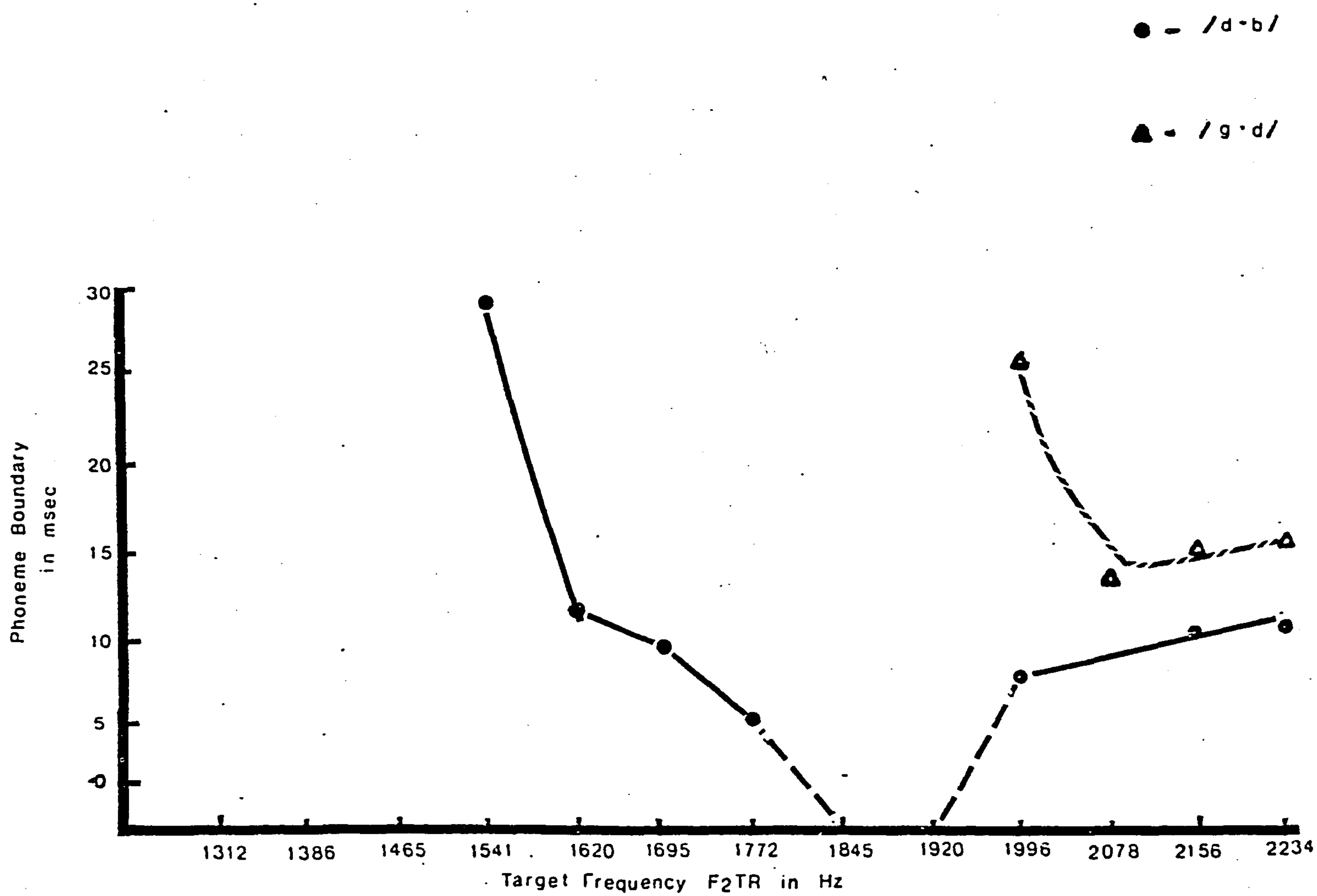

Figure 20. Phoneme Boundaries in msec as a Function of Target Frequency (F2TR) for subjects with normal hearing.

To facilitate further interpretation of the data, the time domain phoneme boundaries, in msec, were plotted against target frequency of the F2TR for the /d-b/ and /g-d/ boundaries (Figure 20). Each data point represents the average boundary value for the five normal hearing subjects. For both the /d-b/ and /g-d/ boundaries, the curves were U-shaped. For the /d-b/ boundary there was a monotonic decrease in the time values from 1541 through 1920 Hz. Above 1920 Hz there was a monotonic increase in the boundary values. However, at 1996 Hz and above, /gae/ responses occurred. It is clear that the /d-b/ and /g-d/ curves converge. Convergence of the curves implies that increasingly smaller changes in the duration of the F2TR resulted in more abrupt phoneme boundaries in time, as target frequency increased.

From these results, it may be concluded that: (1) the monotonic decrease in the time domain boundary from 1541 Hz to 1920 Hz, (2) the existence of three response categories (when target frequency of the F2TR is greater than or equal to 1996 Hz) and (3) the convergence of the /d-b/ and /g-d/ time domain boundaries (as a function of increasing target frequency) support the hypothesis that time sensitivity improves with increasing target frequency of the F2TR for the subjects with normal hearing.

Subjects with Cochlear Hearing-Impairment

In chapter I (p.7) the following hypothesis was presented: Under the assumptions that frequency resolution may be poor and that this poor frequency resolution is likely to be accompanied by improved time resolution, it was hypothesized that: (a) more pronounced durational effects would be observed and (b) durational effects would be evident for a greater range of target frequencies of the F2TR than they are for normal hearing subjects.

Phoneme identifications as a function of duration of the F2TR were not responses made at random by the subjects with cochlear hearing impairments. Indeed, duration of the F2TR was a major cue in phoneme identification. Evidence in support of the above statement may be found by inspection of Figures 21 through 27. In these figures the probabilities of identification responses were plotted as a function of the logarithm of duration of the F2TR, with target frequency of the F2TR as the parameter. Abrupt phoneme boundaries may be seen in each of these figures. These results are especially important. Firstly, time domain phoneme boundaries, notably /d-b/ and /g-d/ , were established at each target frequency of the F2TR, i.e. over the range of 1312 Hz - 2234 Hz. Secondly, the curves for each graph are nearly identical. Finally there was no systematic variation in the time domain boundary value as a function of target frequency. These findings contrasted with those made on normal hearing subjects, who identified the stimuli as /bae/ , /dae/ and /gae/ in the time domain

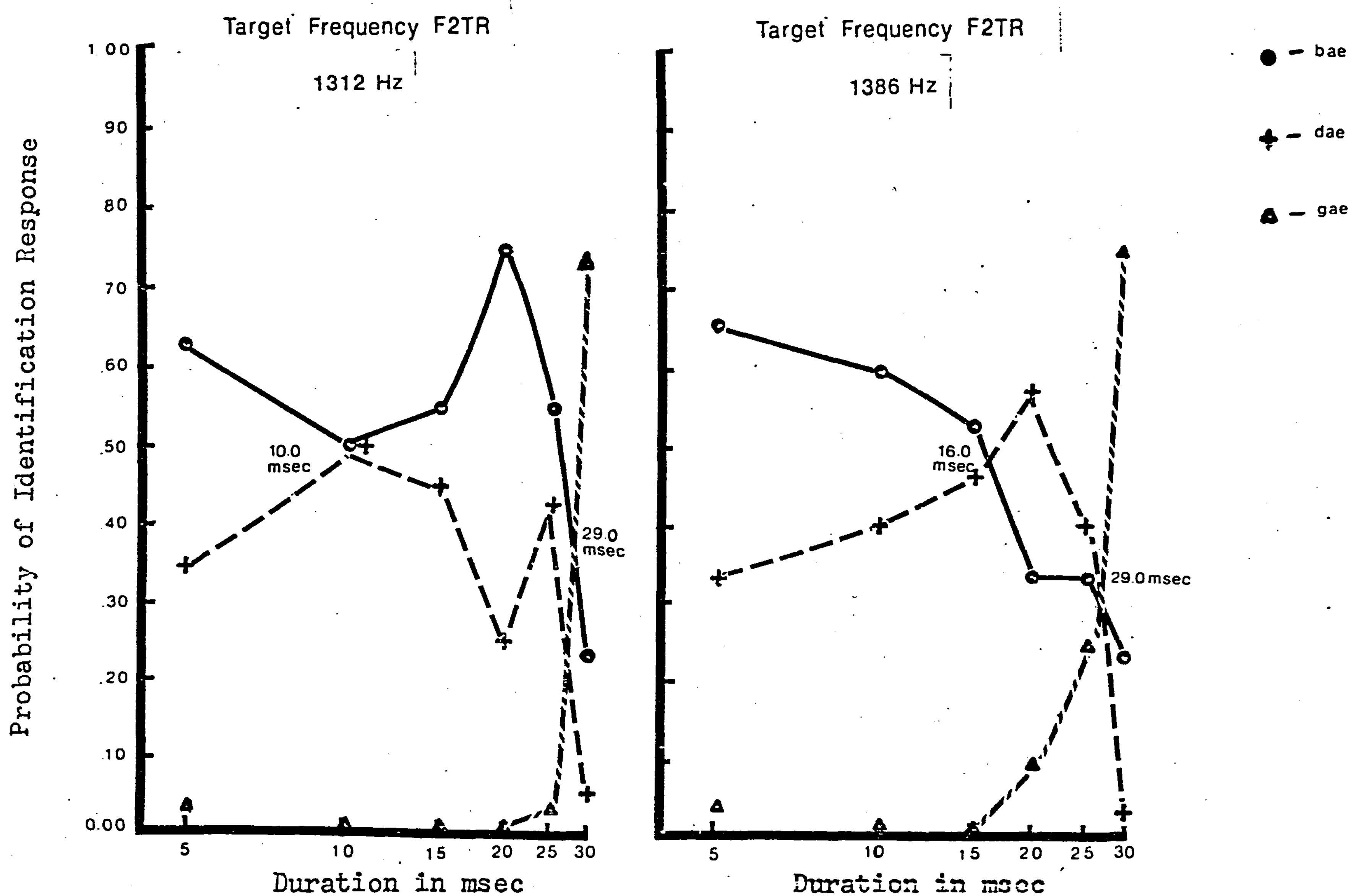

Figure 21. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

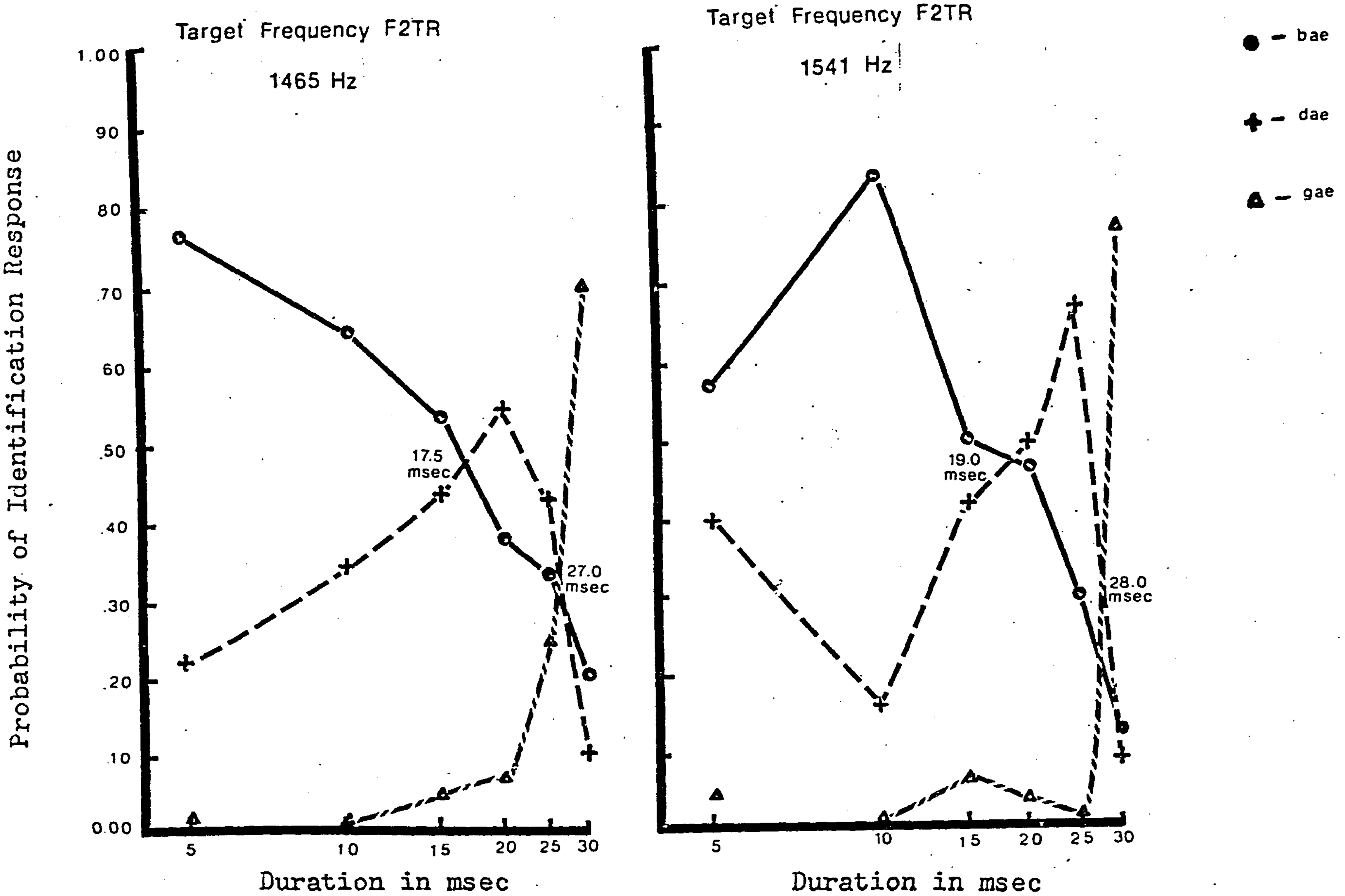

Figure 22. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

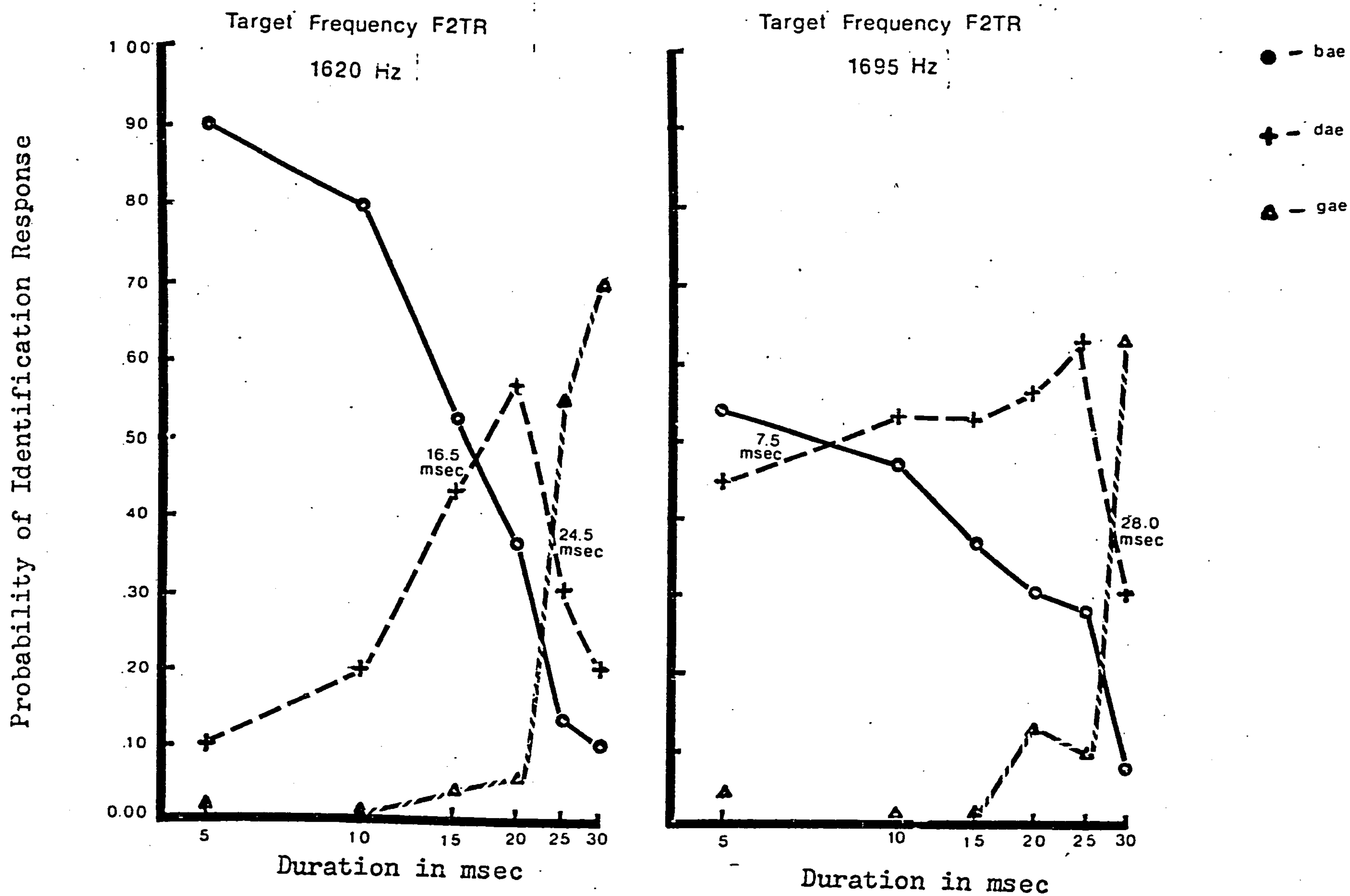

Figure 23. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

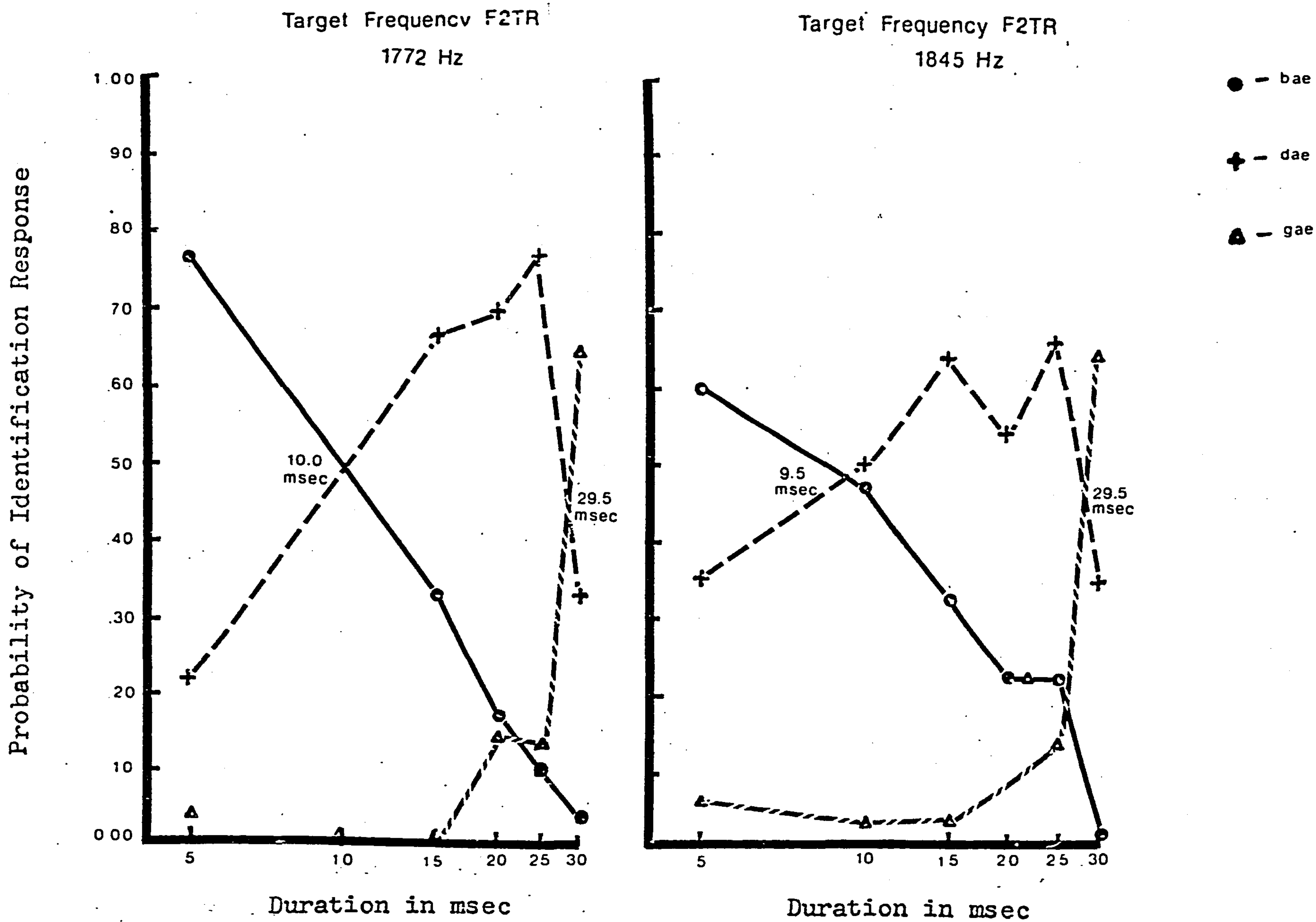

Figure 24. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

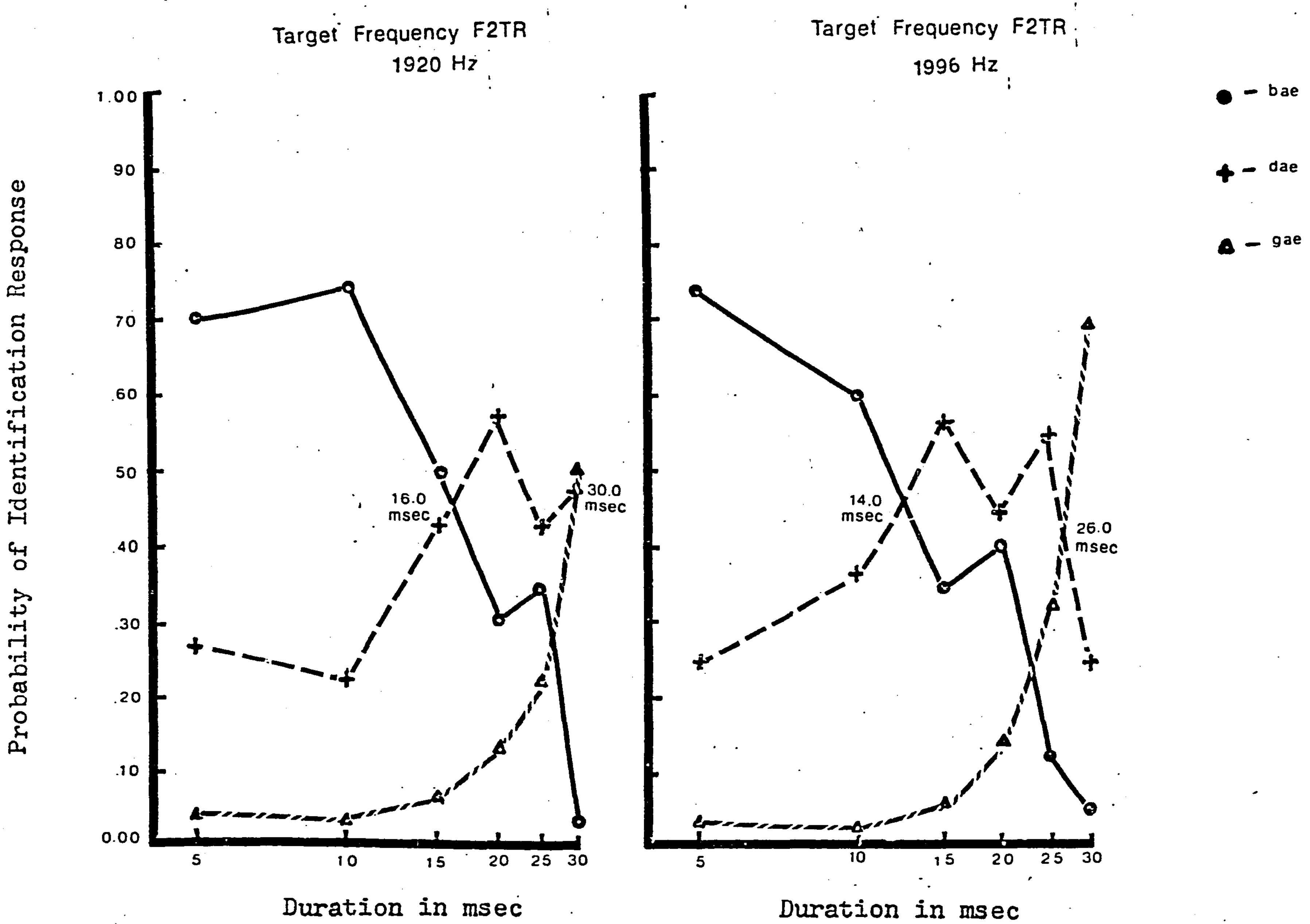

Figure 25. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

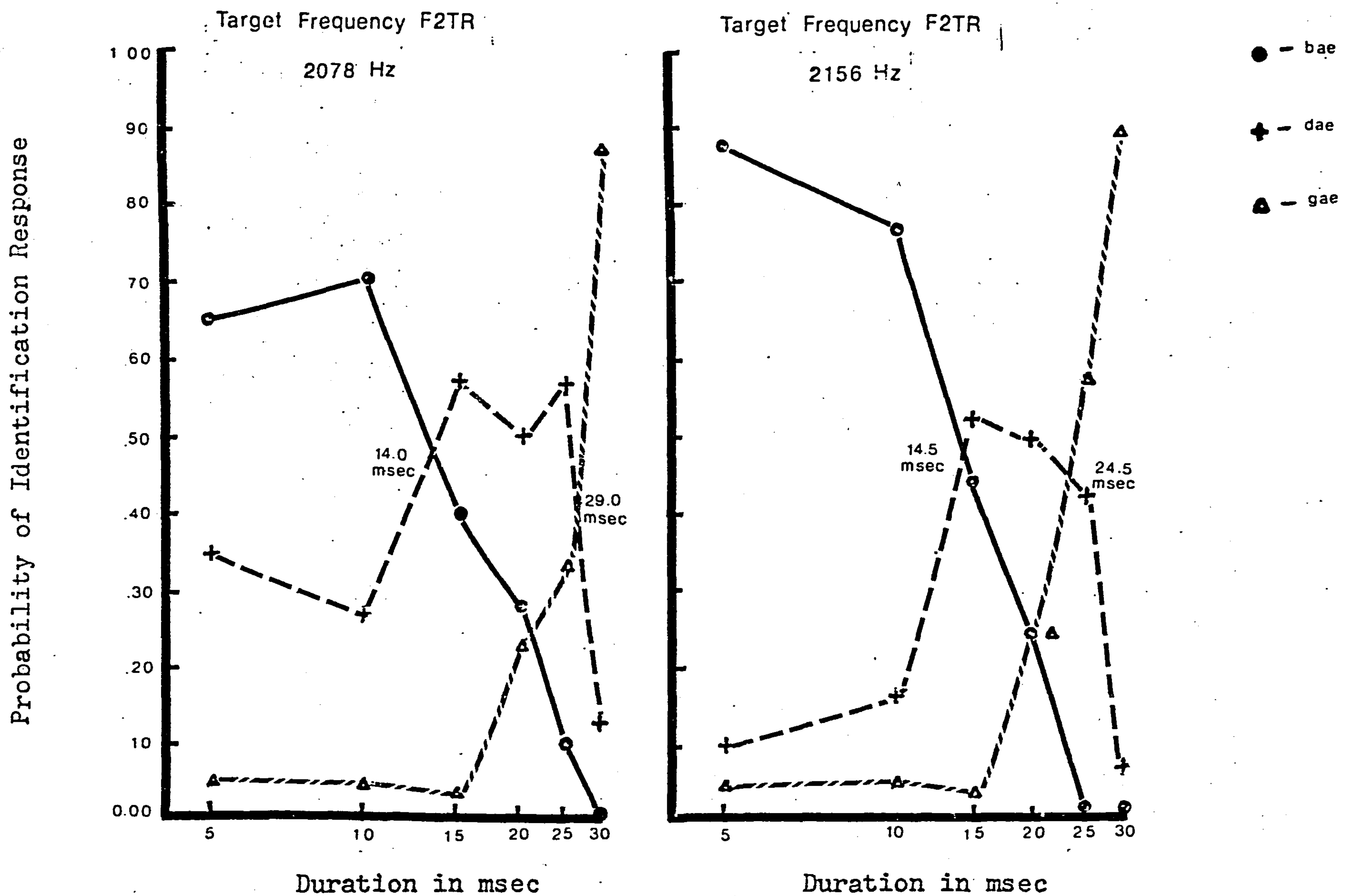

Figure 26. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

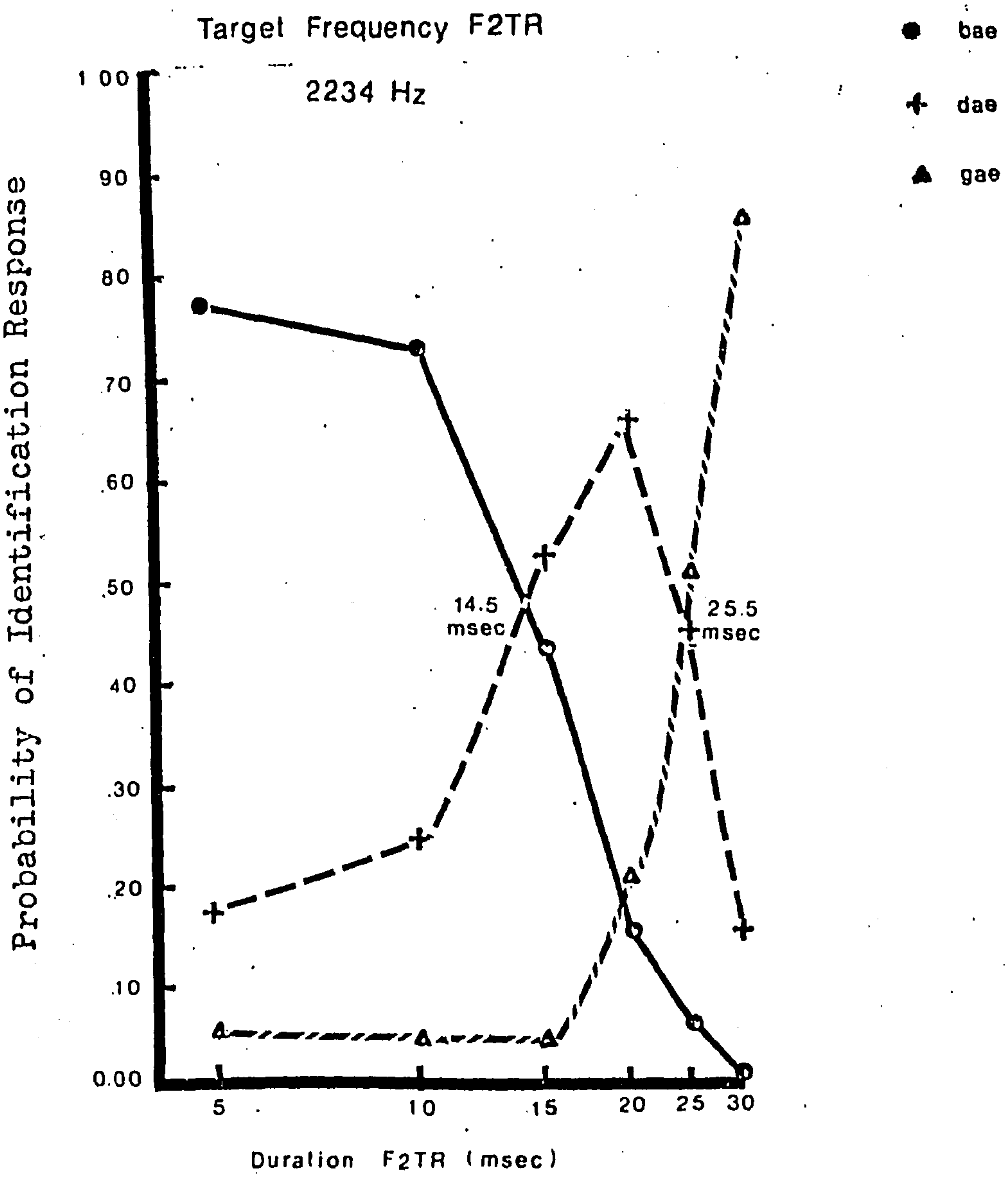

Figure 27. Probability of Identification Responses as a Function of Duration (F2TR) for subjects with cochlear hearing-impairment.

only when the F2TR was greaer than or equal to 1996 Hz, and /dae/ and /bae/ for for target frequencies less than 1996 Hz (see Figures 15-19).

In Table X the average time domain phoneme boundaries, in msec, are presented as a function of target frequency of the F2TR separately for the subjects with normal hearing and the subjects with cochlear hearing-impairments. Although the data were not rigorously analyzed statistically, inspection of the average time boundaries indicated that there was a large difference in the sensitivity to durational changes of the F2TR between the two groups of subjects. For the subjects with normal hearing, the average /b-d/ boundary equalled 9.71 msec and /d-g/ equalled 18.18 msec. For the subjects with cochlear hearing-impairments, the average time boundaries equalled 13.77 and 27.65 msec for /d-b/ and /g-d/ respectively. For both the /d-b/ and /g-d/ boundaries the values were larger for the hearing-impaired subjects. This finding showed that smaller changes in the duration of the F2TR resulted in more abrupt phoneme boundaries for subjects with cochlear hearing-impairments. These observations are viewed as support for the hypothesis stated on pp. 7 and 95.

Rate of Change of Frequency (Glide Rate) for Subjects with Normal Hearing

If target frequency of the F2TR alone would have been the significant cue in phoneme identification for normal hearing subjects, then durational changes of the F2TR should

Table X

Time Domain Phoneme Boundaries for Five Subjects with Normal Hearing and Four Subjects with Hearing-Impairment (values represent pooled data relative to each subject group)

Target Frequency F2TR in Hertz	/b-d/ in msec normal	/b-d/ in msec hearing-impaired	/d-g/ in msec normal	/d-g/ in msec hearing-impaired
1312	-	10.0	-	29.0
1386	-	16.0	-	29.0
1465	-	17.5	-	27.0
1541	29.0	19.0	-	28.0
1620	12.0	16.5	-	24.5
1695	10.1	7.5	-	28.0
1772	6.1	10.0	-	29.5
1845	-	9.5	-	29.5
1920	-	16.0	-	30.0
1996	8.8	14.0	26.5	26.0
2078	7.8	14.0	14.2	29.0
2156	11.25	14.5	16.0	24.5
2234	12.0	14.5	16.0	25.5
Average ($\bar{X}$)	9.71	13.77	18.18	27.65

have had no effect on phoneme identification. In contrast if duration alone would have been the significant cue, there should have been no significant effect on phoneme identification as a function of the target frequency of the F2TR. Since phoneme identification was found to vary as a function of both target frequency and duration of the F2TR, an interaction in the form of the rate of change of frequency of the F2TR (glide rate) appeared likely. Therefore data were replotted in terms of the probability of phoneme identification as a function of glide rate of the F2TR for falling transitions, measured in Hz/ msec. Glide rates were obtained by subtracting the second-formant frequency (1620 Hz) from the target frequency of the F2TR and dividing the difference by duration of the F2TR. Because durational effects were not observed for target frequencies less than 1451 Hz, those target frequencies were eliminated from the analysis. Each data point is the average probability of identification for the five subjects and thus, represents 50 observations of the stimulus. The curves are a visual best fit to the data. Probability scores less than .47 were considered to have arisen on the basis of chance. In Figures 28-30 the response categories /bae/ , /dae/ and /gae/ are presented consecutively. Figure 28 depicts the glide rate values for the /bae/ response. For $p \geq .47$, the glide rate ranged from 7.8 to 120 Hz /msec. For more convincing /bae/ responses, that is $p \geq .75$, the glide rate varied from 71 to 120 Hz / msec. In Figure 29 (/dae/)

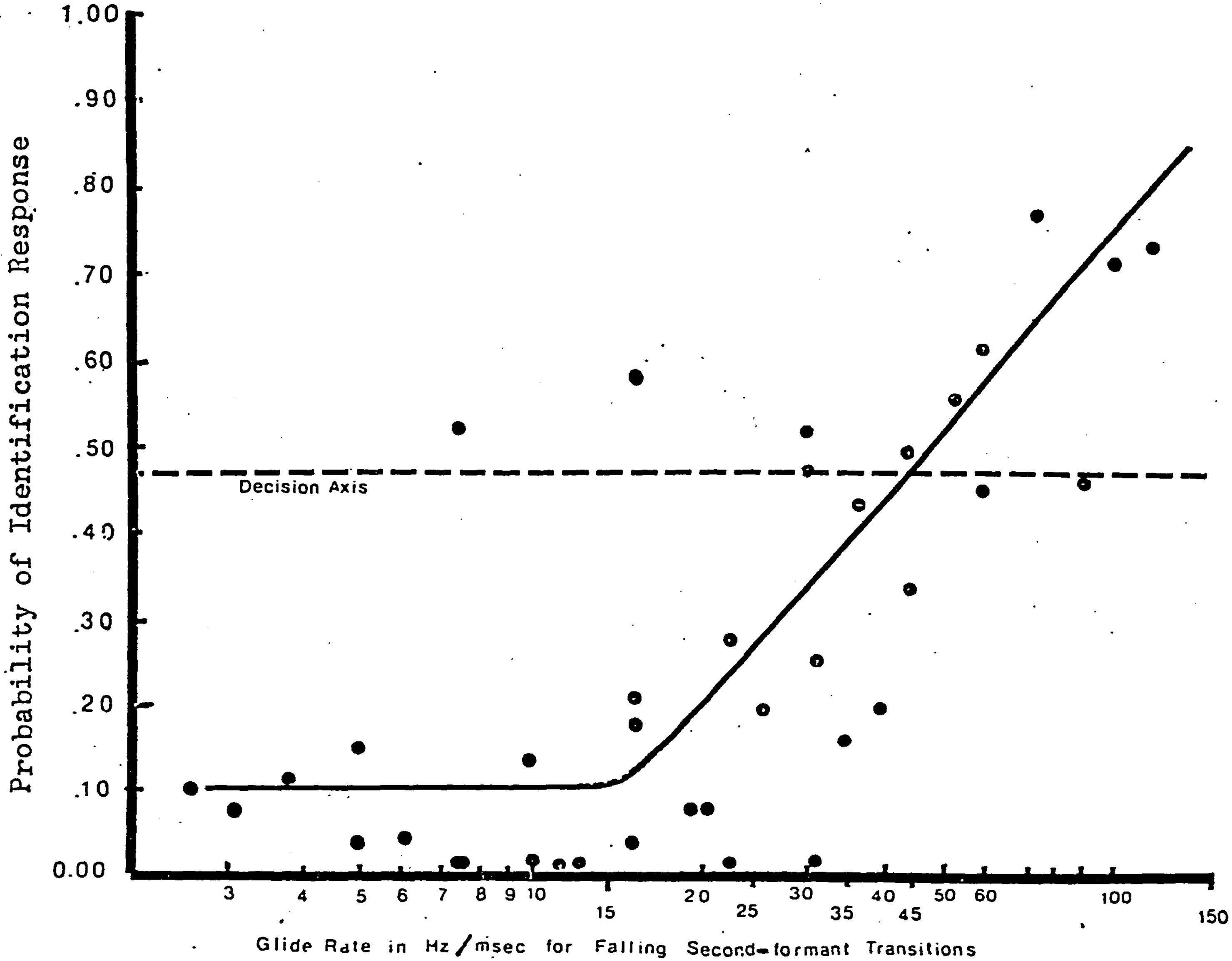

Figure 28 Probability of a /bae/ Response as a Function of Glide Rate

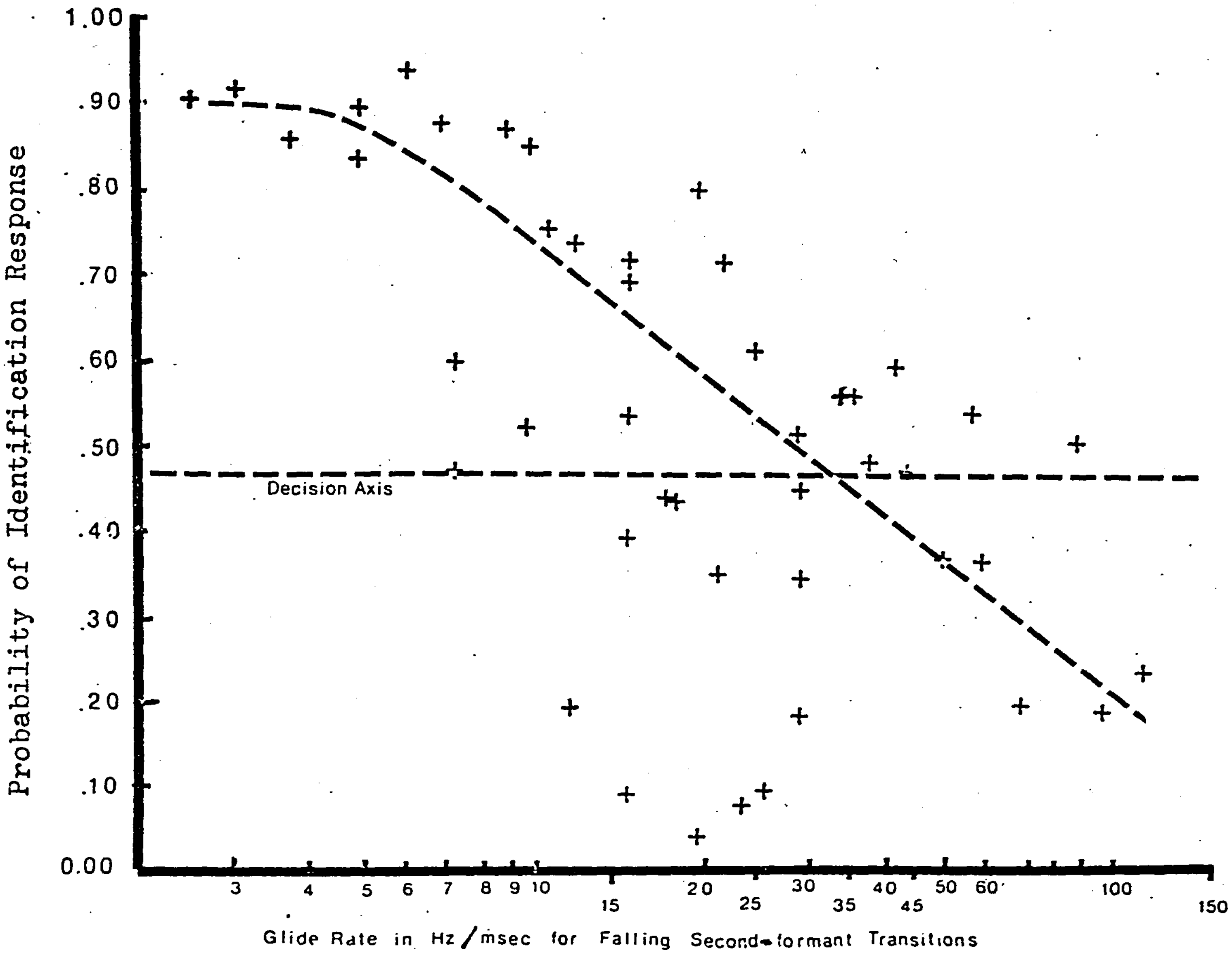

Figure 29 Probability of a /dae/ Response as a Function of Glide Rate

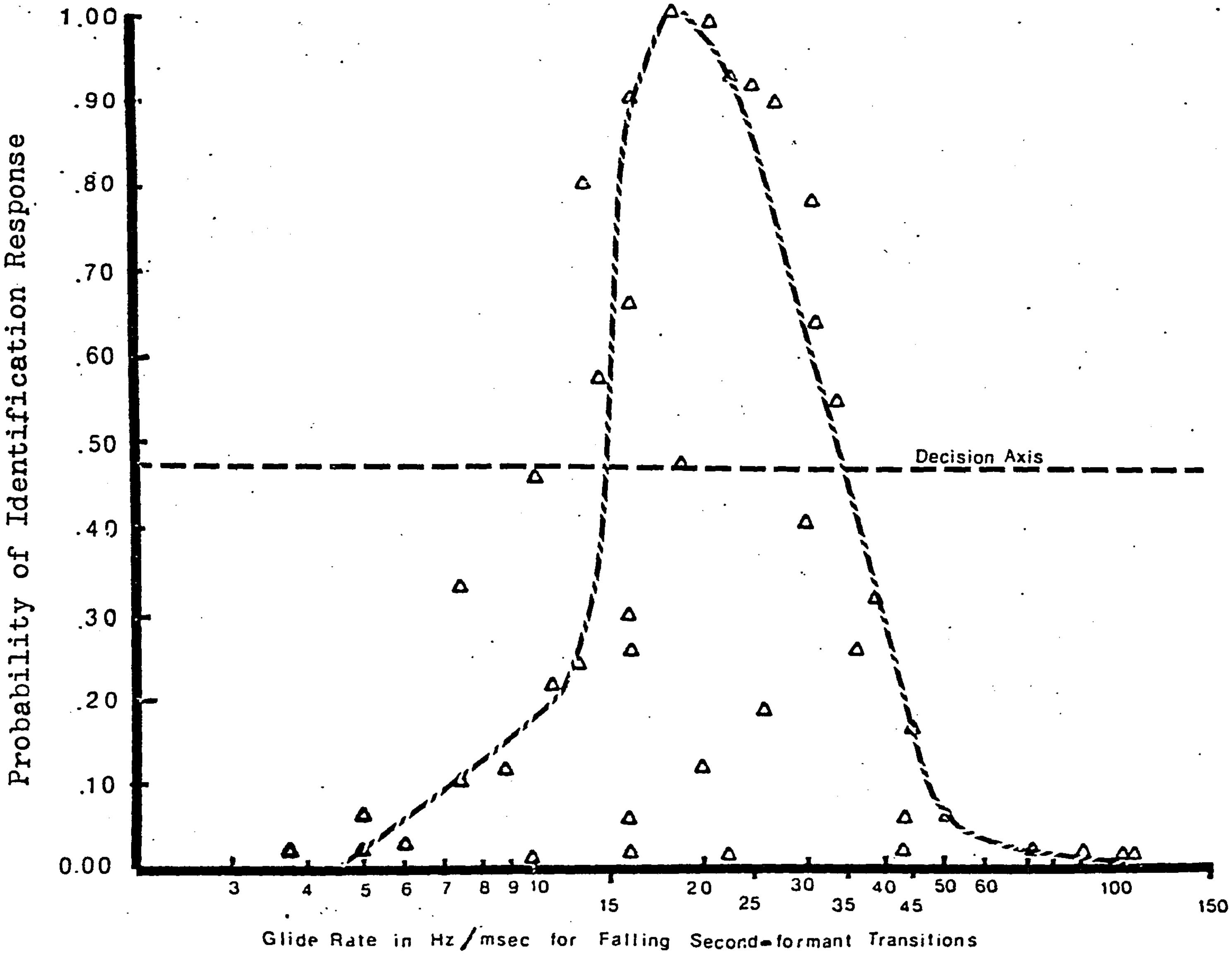

Figure 30 Probability of a /gae/ Response as a Function of Glide Rate

for $p \geq .47$, the glide rate ranged from 2 to 92 Hz / msec, but for $p \geq .75$ the range was reduced to 2 to 22 Hz / msec. For /gae/ (Figure 30), where $p \geq .47$, glide rates were obtained in the range of 13 to 37 Hz / msec; for $p \geq .75$, 15 to 32.5 Hz / msec. It was concluded that for $p \geq .75$ the phoneme categories are fairly well defined in terms of glide rate. Thus, a short duration, but large target frequency excursion, is identified as being similar to a long duration, but small target frequency excursion transition. Therefore, with increasing glide rate, response category varied systematically from /bae/ to /dae/ and, finally to /gae/ .

Glide rates were not computed for the subjects with cochlear hearing losses. Since, over the range of target frequencies studied, these subjects failed to utilize target frequency as a cue for phoneme identification, computation of glide rate would have been meaningless.

Summary of the Results

The hyptheses of the present study were designed to test the application of Gabor's time-frequency trade-off to the perception of synthetic speech-like items by listeners with normal hearing as well as by listeners with cochlear hearing-impairments. The use of synthetic speech-like items permitted easy control of the time and frequency parameters of the stimuli and their manipulation.

Subjects with Normal Hearing

The summary of the results to follow lend support for the hypotheses: (1) that frequency domain phoneme boundaries will shift to regions of higher frequency as duration of the F2TR is decreased and (2) that sensitivity with respect to duration of the F2TR will increase with increasing target frequency of the F2TR. The results were as follows.

1. Frequency domain phoneme boundaries were established as a function of target frequency of the F2TR. However, /b-d/ and /d-g/ boundary values, in Hz, increased with decreasing duration of the F2TR. As a consequence, at durations of less than 20 msec, the /gae/ response disappeared.

2. Analysis of the time domain identifications revealed the existence of phoneme boundaries when the target frequency of the F2TR exceeded 1541 Hz. For target frequencies in the range of 1541 Hz to 1920 Hz the /bae/ and /dae/ response categories occurred. The time domain phoneme boundary values monotonically decreased with increasing target frequency. At 1996 Hz and above, three response categories emerged. For long durations the /gae/ response occurred, followed by /dae/ at intermediate durations, and /bae/ at the shorter durations.

3. The data were also examined with respect to the rate of frequency change of the F2TR expressed in Hz / msec. For probability scores greater than or equal to .75, it was apparent that glide rate was a cue related to phoneme identification.

<u>Subjects with Cochlear Hearing-Impairment</u>

The data supported the hypotheses: (1) that frequency domain phoneme boundaries would occur at locations that differ (presumably being higher in frequency), from those for subjects with normal hearing and, (2) that greater sensitivity to durational changes of the F2TR would be shown than was found in the subjects with normal hearing; this greater sensitivity would occur for a greater range of target frequencies than for the subjects with normal hearing. The results for the subjects with cochlear hearing-impairment may be summarized as follows.

1. Frequency domain phoneme boundaries were not obtained. For the range of target frequencies studied, target frequency of the F2TR was not functionally related to phoneme identification.

2. Clear cut time domain phoneme boundaries occurred at all target frequencies of the F2TR, that is target frequencies in the range of 1312 Hz through 2234 Hz.

3. Comparison of the average time boundaries for the subjects with cochlear hearing-impairment (target frequency 1312-2234 Hz) with the boundaries for the subjects with normal hearing (target frequencies greater than or equal to 1996 Hz), revealed that the boundaries were more abrupt for the hearing-impaired group of subjects. This suggested that the subjects with cochlear hearing-impairment were more sensitive to durational changes of the F2TR than the subjects with normal hearing.

It was thus concluded, that with respect to the identification of the synthetic speech-like stimuli /bae/ , /dae/ and /gae/ , subjects with cochlear hearing-impairment showed better time resolution abilities, at the expense of frequency resolving power, than the subjects with normal hearing. Furthermore, the data supports the assumption that the Gabor Time-Frequency Uncertainty Principle: (1) is useful in the interpretation of speech perception data and (2) is useful with respect to an understanding of the nature of the processing difficulties experienced by persons with cochlear hearing-impairment for the perception of speech-like items.

CHAPTER V

DISCUSSION

In some respects responses by the subjects with cochlear hearing-impairment were similar to the responses by the subjects with normal hearing. Both groups of subjects identified short duration falling F2TRs as /b/ . This was an unusual finding since it is rising F2TRs that that are typically identified as /b/ . The present author can offer no explanation for this observation. When the target frequency of the F2TR was high, i.e. $\geq$ 1996 Hz, both the subjects with normal hearing and those with cochlear hearing-impairment, yielded time domain phoneme boundaries that were similar to each other. However, in contrast to the normal hearing subjects, the subjects with cochlear hearing-impairments showed time domain boundaries for all target frequencies of the F2TR. Furthermore, at 30 msec, the subjects with normal hearing identified target frequencies of the F2TR that were $\geq$1996 Hz as /gae/ . At that duration, the subjects with cochlear hearing-impairments identified all target frequencies of the F2TR as /gae/. The following explanation of this finding is offered: on account of the poor peripheral frequency analysis in cochlear hearing-impairment, the central nervous system interprets the cochlear transformation of the F2TR as a broad-bandwidth neural signal, regardless of the acoustic characteristics of the signal. A broad-bandwidth analysis suggests better time resolution. Thus, when the target frequency of the F2TR $\geq$1996 Hz for the subjects with normal hearing, and for the subjects with cochlear hearing-impairments for all target frequencies of the F2TR, the sensitivity for changes in the duration of the F2TR was at a maximum. Hence, two distinct time domain boundaries (/d-b/ and /g-d/) occurred for these broad-bandwidth signals.

Danaher, Osberger and Pickett, (1973) and Danaher and Pickett, (1975), suggested that greater-than-normal upward spread of masking by the first-formant and the first-formant transition on the second-formant transition may be responsible for the poor discrimination of the F2TR in their populations of subjects with sensorineural hearing-impairment. In terms of the present study, it is possible that an abnormal spread of masking may have been responsible for differences in phoneme identification between the groups of subjects with normal hearing and those with cochlear hearing-impairments. However, the present author feels that greater-than-normal spreads of masking are an indication of poor peripheral frequency resolution. Furthermore, it appears unlikely than an upward spread of masking could account for the fact that subjects with cochlear hearing-impairments made systematic use of duration of the F2TR as a cue in phoneme identification. Nor could upward spread of masking account for the fact that time resolution became increasingly finer with increasing frequency as was observed in subjects with normal hearing. With increasing frequency of the F2TR one would expect less masking, certainly not more. Since the time domain phoneme boundaries for the subjects with cochlear hearing-impairments were similar in form to those of the subjects with normal hearing when $f \geq 1996$ Hz, it appears reasonable to exclude an upward spread of masking from consideration as an explanation for differences among the subjects in the present study.

A more general explanation may be that frequency resolution may be poor in ears with cochlear hearing losses because the critical-band mechanism is either altered or, in some instances, non-functional. An alteration of the critical-band mechanism may result in: (1) poorer-than-normal difference limens for frequency, (2) greater-than-normal spread of masking and (3) an inability to discriminate among target frequencies of the F2TR. Application of the Gabor time-frequency trade-off to the problem of poor frequency resolution in sensorineural hearing-impairment suggests that with increasing bandwidth, time resolution should improve. With populations of subjects with sensorineural hearing-impairments, improved time resolution has been usually found in temporal integration studies (a shorter time constant of integration and a depressed slope of the temporal integration function as compared to the findings in normal listeners), and, as in the present study, an increased sensitivity to changes in the duration of the F2TR.

Theories of Speech Perception

There are several current theories of speech perception. Most notably, these include: (1) the Acoustic Theory (Fant, 1967), (2) the modified Motor Theory (Liberman et al., 1967) and (3) the Analysis-by-Synthesis Theory (Stevens and Halle, 1967 and Stevens, 1972). These theories differ mainly with respect to the manner in which lingusitic information is retrieved by the central nervous system. However, they all share features in common. These include: (1) peripheral auditory analysis of the speech signal, (2) memory and (3) reference to the articulatory mechanism as a mediator in the neural domain, through which the auditory system and the neuromuscular articulatory system are linked. The linkage between the auditory and articulatory systems is the central theme of these three theories; it serves to explain central processing of speech sound information.

In Fant's Acoustic Theory (Fant, 1967), a series of binary decisions is presumed to be available to the listener (distinctive feature analysis); the speech signal is compared to stored neurophysiological maps of distinctive features. Fant maintains that these maps arise from the articulatory stage of speech production. Liberman et al., 1967, suggested that the extracted lingusitic features of the acoustic signal, is compared in the neural domain, with the neuromotor commands to the articulators. Similarly, Stvens and Halle, 1967 p. 88 state that:

> . . . the perception of speech involves the internal sythesis of patterns according to certain rules, and a matching of these internally generated patterns against the pattern under analysis. We suggested, moreover, that, the generative rules utilized in the perception of speech were in large measure identical to those utilized in speech production, and that fundamental to both processes was an abstract representation of the speech event.

Thus, in this view, it is not the acoustic event that is directly analyzed, but rather an abstract speech representation is matched with the phonological rules governing its production. The phonological rules are presumed to exist at the level where patterns of commands are issued to the articulatory mechanism.

In this paper no attempt will be made to model speech perception. However, interpretation of the data obtained in the present study, especially the data for the subjects with cochlear hearing-impairments, raises questions with respect to the role of the peripheral auditory analysis and its neural reference to the articulatory mechanism in the perception of speech.

If the theories of speech perception just cited rely on either neuromotor commands to the articulatory mechanism or on the phonological rules governing the pattern of impulses, then the identification of phonemes must be tied to acoustic signals which accurately reflect articulatory movements. When one views speech spectrograms for real speech, it is readily observed that the duration of the F1TR and F2TR range from 30 to 50 msec. However, in the present study, both the subjects with normal hearing and those with cochlear hearing-impairment, were found to be able to identify the test items phonemically, even when the duration of the F2TR was substantially less than 30 msec, e.g. 5-10 msec. In the production of the stop consonants the articulatory event simply is much longer than 5 -10 msec. In addition, in the present study, the subjects with cochlear hearing-impairment were not responsive to changes in the target frequency of the F2TR as a cue in phoneme identification. Instead, phoneme identification was related to the duration of the F2TR. Thus, both groups of subjects were capable of identifying the test items phonemically even when, for many of the stimuli, no physiological reference to articulation could be made. The theories of speech perception which rely on articulation fail to predict: (1) phoneme identification for short duration F2TRs by subjects with normal hearing, (2) poor frequency resolution as evidenced by the inability of the subjects with cochlear hearing-impairments to use target frequency of the F2TR as a cue for phoneme identification, (3) the systematic time domain phoneme boundaries common to both groups of subjects and (4) the rate of change of frequency (glide rate) as a significant cue in phoneme identification by the subjects with normal hearing. Therefore, future theories of speech perception should: (1) account for time domain phoneme boundaries, (2) predict the speech performance of listeners with hearing-impairment and (3) include analysis of the spatio-temporal characteristics of the speech event.

The Feature Detection Theory of speech perception, first proposed by Abbs and Sussman (1971), suggest the existence of groups of specialized neurons selectively tuned to specific features of the speech signal in terms of spatio-temporal patterns of activity. However, feature detection theory fails (1) to account for how these specialized neurons become tuned to specific features of the speech signal (2) to suggest where in the central nervous system these groups of cells may be found, (3) to adequately explain how the speech signal is retrieved, (4) to account for phoneme identification in certain cases where patterns of spatio-temporal activity were not previously experienced, e.g. short duration F2TRs and, (5) to predict phoneme identification in the time domain.

Areas for Future Investigation

Speech Perception Experiments

Since subjects with cochlear hearing losses may be more sensitive to durational cues than subjects with normal hearing, a boundary determination along a voice-onset-time (VOT) continuum is suggested. The VOT is a durational cue related to the voiced-voiceless distinction among the stops, e.g. /p-b/ , /t-d/ and /k-g/ . The VOT is defined as the time interval, in msec, between the release of the articulators and the onset of voicing (Lisker and Abramson, 1964). The VOT has been investigated with synthetic speech items. It appears that the continuous VOT variable is categorically perceived. Although there are differences in VOT values with place of articulation, a VOT of less than 20 msec is, in general, identified as voiced, while a VOT of greater than 20 msec is identified as voiceless.

Liberman et al., (1956) showed that the distinction between stop, glide and semi-vowel was related to the duration of the F2TR. The authors showed that as duration of the F2TR increased, the /bɛ/ identification responses abruptly changed to /wɛ/ at durations greater than 40 msec, and the /wɛ/ identification responses in turn changed abruptly to /uɛ/ at approximately 200 msec. In the present study, subjects with cochlear hearing-impairments provided /bae/ identification responses only when the duration of the F2TR was short, i.e. 5-15 msec. As duration increased the identification responses changed to /dae/ and finally to

/gae/ . These findings imply, that, in addition to altered phoneme boundaries for the distinction among voiced stops, there may be alterations in the stop glide semi-vowel distinction as well. Thus, an experiment is proposed where: (1) the range of target frequencies should be extended beyond the range included in the present study, i.e. to values greater than 2234 Hz and less than 1312 Hz, and (2) the range of durations of the F2TR extended beyond 30 msec to at least 200 msec, in order to investigate the stop glide semi-vowel distinction in populations of subjects with sensorineural hearing-impairments.

Liberman et al., (1967), showed that the F1TR is a cue by which listeners infer manners of artiuclation. Rising F1TRs lead to identification of speech-like items as stops. Test items in which the F1TR is flat are identified as nasals. In the present study the F1TR rose to the steady-state frequency of F1. In addition, the subjects were restricted in their responses to a single choice among the stops /bae/ , /dae/ and /gae/ . However, two of the subjects with cochlear hearing-impairments remarked that on occassion, they believed a particular test item to be /mae/ or /nae/ . The phoneme /m/ and /n/ are nasals, but their place of articulation is identical to /b/ and /d/ . These subjective reports then, imply that there may be alterations in the detection of the F1TR in some persons with cochlear hearing-impairment, and that these alterations could lead to phoneme confusions. Thus, an experiment is proposed where the F2TR should be held constant, but the F1TR be varied in target frequency and duration, in order to investigate the stop-nasal distinction in subjects with cochlear hearing-impairments.

Single Unit Characteristics with Respect to Speech Stimuli

Kiang and Moxon (1974), investigated in single cochlear nerve fibers (cat), the low frequency tails of high frequency tuning curves. The authors were concerned with single unit correlates of the assymetric resonance curves of the cochlear partition (gradual slopes on the low frequency side; steep slopes on the high frequency side), and the eighth nerve correlates for human speech. Kiang and Moxon made the following statement: "Human speech sounds at normal speaking levels will elicit responses in the high CF units even when the speech sounds have negligible high frequency content." With reference to thier own data, the authors explained that the responses elicited by the phrase "shoo cat" were represented by a wide range of nerve fibers resulting in ". . . considerable spatial distributions of activity across the ensemble of auditory-nerve fibers." Analysis of fine-time resolution post-stimulus time histograms showed changes in gross discharge rates associated with phonetic components and phase locking for fundamental frequency and spectral peaks (formants) for the vowels. Thus, low CF units responded to the low frequency components of speech both with an increase in pulse rate density and with phase locking, while high CF units responded to the low frequency components of speech only with phase locking.

Kiang and Moxon (1974), suggested that both the time and frequency characteristics of the speech stimulus could be predicted from single unit responses. The authors proposed a multidimensional analysis, in the form of a neurogram to aid in the prediction of the time and frequency characteristics of the speech stimulus from single unit data. A neurogram could be considered as an analogue to the speech spectrogram where pulse rate density is plotted against characteristic frequency and time. According to the authors,

> . . . a neurogram would provide a comprehensive view of the time-varying spatial distribution of neural discharges. The role of the central processor would then be seen as one of analyzing spatio-temporal patterns of auditory-nerve activity.

Kiang and Moxon (1974) suggested that in cochlear pathology, specifically in lesions involving the basal turn of the cochlea, phase locking and rate of discharge of high CF units would be altered resulting in a deterioration of speech intelligibility. The addition of noise would result in a further deterioration of speech intelligibility since it has been shown previously that noise disrupts the rate of discharge in low CF units and the synchrony in high CF units.

The present author views the paper just cited, as an extremely valuable one because: (1) a comparison of the speech spectrogam and neurogram would provide a means for determining the cochlear transfer function for speech (assuming correction for the middle ear transfer function), and (2) by comparison of the neurograms obtained (a) from subjects with normal hearing and (b) from subjects with sensorineural hearing-impairments and the speech spectrogram of the stimulus, the effects of cochlear pathology on peripheral auditory analysis for speech could be evaluated quantitatively. However, the present author believes that, at least in the early stages of research, computer synthesized speech items should be employed in lieu of real speech items as in Kiang and Moxon's experiment. In addition, there should be an orderly progression, in terms of the complexity of the signals from vowels to consonant-vowel syllables, with special emphasis on the F1TR and F2TR, and finally to words and phrases.

Summary

Gabors's application of the Heisenberg Uncertainty Principle to Acoustics states that frequency resolution is inversely proportional to signal duration ($\Delta f \times \Delta t \cong 1$). This relationship is not only applicable to mechanical or electronic analyzing devices, but it may be related to the auditory system as well. Thus, the auditory system, a dual time-frequency analyzer, may be subject to time resolution limitations imposed by its bandwidth, and bandwidth limitations imposed by its integration time.

Various psychoacoustic investigations including: (1) the difference limen for frequency, (2) the critical-band and (3) the spread of masking, suggest that frequency resolution deteriorates in the presence of cochlear pathology. Temporal integration experiments for subjects with sensorineural hearing-impairment, show the time constant of integration to be less than 200 msec, that is, shorter than normal. Thus, it appears that the peripheral auditory system, in cochlear pathology, shows finer time resolution capabilities at the expense of frequency resolution. The present author believes that changes in the time-frequency resolution of the auditory system, secondary to cochlear pathology, may be responsible for the deterioration of the speech processing abilities in persons handicapped by hearing-impairment of cochlear origin.

The purpose of the present investigation was to examine the trading relationship between time and frequency parameters of synthetic speech-like items and phoneme identification, for listeners with normal hearing and those with cochlear hearing-impairment.

It has been shown previously that the target frequency of the second-formant transition (F2TR) is a cue for phoneme identification with respect to places of articulation for listeners with normal hearing. The stimuli were prepared, under computer control of a parallel resonance speech synthesizer, at the Haskins Laboratories. The independent variables were the target frequency, in Hz, and duration, in msec, of the F2TR. The target frequency varied from 1312 Hz to 2234 Hz in thirteen (75 Hz) increments. For each target frequency the duration varied from 30 to 5 msec in six (5 msec) steps. Thus, a total of 78 stimuli were prepared. All other parameters of the test items were held constant. The dependent variables were: (1) the probability score for the identification responses /bae/ , /dae/ and /gae/ , (2) the frequency domain phoneme boundary values, in Hz, and (3) the time domain phoneme boundary values, in msec.

Five subjects with normal hearing and four subjects with audiometric test results suggestive of cochlear lesions, were chosen as subjects. Signals were presented monaurally at 30 dB SL via headphone.

The data indicated that for the subjects with normal hearing: (1) the frequency domain phoneme boundaries shifted to regions of higher frequency with decreasing duration of the F2TR, (2) the /d-b/ time domain boundaries occurred when target frequency exceeded 1541 Hz and decreased monotonically to a minimum at 1920 Hz, (3) for

$f \geq 1996$ Hz, /d-b/ and /g-d/ boundaries occurred, which monotonically increased with increasing target frequency, and (4) time domain boundaries became more abrupt with increasing target frequency, indicating improved time resolution with increasing frequency. The data were replotted in terms of rate of change of frequency of the F2TR (glide rate), measured in Hz/msec. Inspection of these graphs revealed that with increasing glide rate the identification responses varied systematically from /bae/ to /dae/ to /gae/ .

The subjects with cochlear hearing-impairment, in contrast to those with normal hearing, did not utilize target frequency of the F2TR for phoneme identification. However, the subjects with hearing-impairment showed distinct time domain phoneme boundaries at all target frequencies of the F2TR, i.e. 1312-2234 Hz. The subjects with cochlear pathology were, at comparable target frequencies, more sensitive to changes in the duration of the F2TR than were the subjects with normal hearing. In conclusion, it appeared, with respect to the identification of synthetic speech-like items, that the subjects with cochlear hearing-impairment were better time analyzers and poorer frequency analyzers than the subjects with normal hearing.

Appendix

Appendix A

Probability of Response for Hearing-Impaired Subject: J.W.

Target Frequency F2TR	Probability of Response											
	Duration of the second-formant transition											
	30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	.90	.10	.00	.90	.10	.00	1.0	.00	.00	1.0	.00	.00
1386	.90	.10	.00	.80	.10	.00	.90	.10	.00	.90	.10	.00
1465	.80	.10	.10	.70	.30	.00	.80	.20	.00	.90	.10	.00
1541	.30	.30	.40	.60	.30	.10	.70	.20	.10	.80	.20	.00
1620	.10	.30	.60	.40	.40	.20	.80	.20	.00	.90	.10	.00
1695	.00	.40	.60	.30	.50	.20	.10	.50	.40	.90	.10	.00
1772	.00	.10	.90	.10	.60	.30	.10	.50	.40	.70	.30	.00
1845	.00	.10	.90	.00	.70	.30	.00	.40	.60	.50	.50	.00
1920	.00	.10	.90	.00	.30	.70	.00	.50	.50	.30	.50	.20
1996	.00	.00	1.00	.00	.30	.70	.00	.50	.50	.30	.60	.10
2078	.00	.00	1.00	.00	.50	.50	.00	.50	.50	.20	.60	.20
2156	.00	.00	.90	.00	.30	.70	.10	.10	.80	.30	.60	.10
2234	.00	.00	1.00	.00	.10	.90	.00	.20	.80	.30	.50	.20

Probability of Response for Hearing-Impaired Subject: J.W.

Target Frequency F2TR	Probability of Response					
	Duration of the second-formant transition					
	10 msec			5 msec		
	b	d	g	b	d	g
1312	.90	.10	.00	.90	.10	.00
1386	.90	.10	.00	1.00	.00	.00
1465	.80	.20	.00	1.00	.00	.00
1541	1.00	.00	.00	.90	.10	.00
1620	.90	.10	.00	1.00	.00	.00
1695	.90	.10	.00	1.00	.00	.00
1772	-	-	-	.90	.10	.00
1845	.90	.10	.00	1.00	.00	.00
1920	.90	.00	.10	.80	.10	.10
1996	.90	.10	.00	.90	.10	.00
2078	.90	.10	.00	.50	.50	.00
2156	.70	.20	.10	.80	.10	.10
2234	.90	.10	.00	.60	.30	.10

Appendix B

Probability of Response for Hearing-Impaired Subject: J.K.L.

Target Frequency F2TR	Probability of Response											
	Duration of the second-formant transition											
	30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	.00	.00	1.00	.30	.60	1.0	.60	.40	.00	.60	.40	.00
1386	.00	.00	1.00	.20	.70	.10	.20	.70	.10	.60	.40	.00
1465	.00	.00	1.00	.40	.50	.10	.30	.70	.00	.70	.30	.00
1541	.00	.00	1.00	.10	.90	.00	.20	.80	.00	.20	.80	.00
1620	.00	.00	1.00	.00	.10	.90	.20	.80	.00	.70	.30	.00
1695	.00	.00	1.00	.20	.60	.20	.10	.80	.10	.30	.70	.00
1772	.00	.00	1.00	.20	.80	.00	.20	.70	.10	.30	.70	.00
1845	.00	.00	1.00	.20	.70	.10	.00	1.0	.00	.50	.50	.00
1920	.00	.00	1.00	.10	.90	.00	.20	.80	.00	.80	.20	.00
1996	.00	.10	.90	.00	.70	.30	.60	.30	.10	.60	.40	.00
2078	.00	.00	1.00	.00	.80	.20	.00	.70	.30	.30	.70	.00
2156	.00	.00	1.00	.00	.70	.30	.10	.70	.20	.30	.70	.00
2234	.00	.00	1.00	.00	.50	.50	.10	.90	.00	.50	.50	.00

Probability of Response for Hearing-Impaired Subject: J.K.L.

Target Frequency F2TR	Probability of Response					
	Duration of the second-formant transition					
	10 msec			5 msec		
	b	d	g	b	d	g
1312	.70	.30	.00	.80	.20	.00
1386	.70	.30	.00	.70	.30	.00
1465	.50	.50	.00	.50	.50	.00
1541	.80	.20	.00	.50	.50	.00
1620	1.00	.00	.00	.90	.10	.00
1695	.50	.50	.00	.60	.40	.00
1772	-	-	-	.50	.50	.00
1845	.50	.40		.70	.30	.00
1920	.80	.20	.00	.70	.30	.00
1996	.70	.30	.00	.50	.50	.00
2078	.90	.10	.00	.60	.40	.00
2156	.70	.20	.10	.80	.20	.00
2234	.60	.30	.10	.80	.10	.10

Appendix C

Probability of Response for Hearing-Impaired Subject: S.B.

Target Frequency F2TR	Probability of Response											
	Duration of the second-formant transition											
	30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	.00	.00	1.00	.50	.50	.00	.80	.20	.00	.40	.60	.00
1386	.00	.00	1.00	.00	.60	.40	.00	.80	.20	.20	.80	.00
1465	.00	.00	1.00	.00	.50	.50	.00	.70	.30	.00	.80	.20
1541	.00	.00	1.00	.30	.70	.00	.50	.50	.00	.50	.50	.00
1620	.00	.00	1.00	.00	.10	.90	.20	.80	.00	.30	.70	.00
1695	.00	.10	.90	.40	.60	.00	.40	.60	.00	.00	1.0	.00
1772	.00	.50	.50	.00	.90	.10	.20	.80	.00	.10	.90	.00
1845	.00	.80	.20	.30	.70	.00	.70	.20	.10	.00	.90	.10
1920	.10	.90	.00	1.0	.00	.00	.50	.50	.00	.50	.50	.00
1996	.00	.50	.50	.30	.70	.00	.70	.30	.00	.00	.80	.20
2078	.00	.20	.80	.30	.70	.00	.70	.30	.00	.60	.40	.00
2156	.00	.10	.90	.00	.30	.70	.40	.60	.00	.80	.20	.00
2234	.00	.40	.60	.10	.70	.20	.10	.90	.00	.30	.70	.00

Probability of Response for Hearing-Impaired Subject: S.B.

Target Frequency F2TR	Probability of Response					
	Duration of the second-formant transition					
	10 msec			5 msec		
	b	d	g	b	d	g
1312	.20	.80	.00	.10	.80	.10
1386	.50	.50	.00	.60	.20	.10
1465	.70	.30	.00	.80	.20	.00
1541	.80	.20	.00	.20	.70	.10
1620	.70	.30	.00	1.00	.00	.00
1695	.30	.70	.00	.50	.50	.00
1772	-	-	-	.80	.20	.00
1845	.00	1.00	.00	.00	1.00	.00
1920	.60	.40	.00	.60	.40	.00
1996	.20	.70	.10	.60	.30	.10
2078	.40	.50	.10	.70	.30	.00
2156	.90	.10	.00	1.00	.00	.00
2234	.70	.30	.00	.90	.10	.00

Appendix D

Probability of Response for Hearing-Impaired Subject: S.S.

Target Frequency F2TR	Probability of Response											
	Duration of the second-formant transition											
	30 msec			25 msec			20 msec			15 msec		
	b	d	g	b	d	g	b	d	g	b	d	g
1312	.00	.10	.90	.50	.50	.00	.60	.40	.00	.20	.80	.00
1386	.00	.00	1.00	.30	.20	.50	.20	.70	.10	.40	.60	.00
1465	.00	.30	.70	.20	.40	.40	.40	.60	.00	.50	.50	.00
1541	.20	.10	.70	.20	.80	.00	.50	.50	.00	.50	.40	.10
1620	.30	.50	.20	.10	.60	.30	.30	.50	.20	.20	.80	.00
1695	.30	.70	.00	.20	.80	.00	.60	.40	.00	.70	.30	.00
1772	.10	.70	.20	.10	.80	.10	.20	.70	.10	.20	.80	.00
1845	.00	.50	.50	.40	.40	.20	.20	.60	.20	.30	.70	.00
1920	.00	.90	.10	.30	.50	.20	.50	.50	.00	.40	.50	.10
1996	.20	.40	.40	.20	.50	.30	.30	.70	.00	.50	.50	.00
2078	.00	.30	.70	.10	.30	.60	.40	.50	.10	.50	.50	.00
2156	.00	.20	.80	.00	.40	.60	.40	.60	.00	.40	.60	.00
2234	.00	.20	.80	.10	.50	.40	.40	.60	.00	.60	.40	.00

Probability of Response for Hearing-Impaired Subject: S.S.

Target Frequency F2TR	Probability of Response					
	Duration of the second-formant transition					
	10 msec			5 msec		
	b	d	g	b	d	g
1312	.20	.80	.00	.70	.30	.00
1386	.30	.70	.00	.30	.70	.00
1465	.60	.40	.00	.80	.20	.00
1541	.70	.30	.00	.70	.30	.00
1620	.60	.40	.00	.70	.30	.00
1695	.20	.80	.00	.00	.90	.10
1772	-	-	-	.90	.10	.00
1845	.50	.50	.00	.70	.20	.10
1920	.70	.30	.00	.70	.30	.00
1996	.60	.40	.00	.90	.10	.00
2078	.60	.40	.00	.80	.20	.00
2156	.80	.20	.00	.90	.10	.00
2234	.70	.30	.00	.80	.20	.00

REFERENCES

Abbs, J.H. and Sussman, H.M. "Neurophysiological Feature Detectors and Speech Perception: a Discussion of Theoretical Implications," Journal of Speech and Hearing Research, 14, 1, 23-36, (1971)

Butler, R.A. and Albrite, J.P. "The Pitch-Dscrimination Function of the Pathological Ear," Archives of Otolaryngology, 63, 411, (1956)

Cherry, C. On Human Communication: a Review and a Criticism. Second edition, (MIT Press, Cambridge) 1966

Dallos, P. and Olsen, W.O."Integration of Energy at Threshold with Gradual Rise-Fall Tone Pips," Journal of the Acoustical Society of America, 36, 743-751, (1964)

Danaher, E.M., Osberger, M.J. and Pickett, J.M. "Discrimination of Formant Frequency Transitions in Synthetic Vowels," Journal of Speech and Hearing Research, 16, 439-451 (1973)

Danaher, E.M and Pickett, J.M. "Some Masking Effects Produced by Low-Frequency Vowel Formants in Persons with Sensorineural Hearing Loss," Journal of Speech and Hearing Research, 18, 2, 261-271, (1975)

DeBoer, E. and Bouwmeester, J. "Critical Bands and Sensorineural Hearing Loss," Audiology, 13, 236-259 (1974)

Delattre, P.C., Liberman, A.M. and Cooper, F.S. "Acoustic Loci and Transitional Cues for Consonants," Journal of the Acoustical Society of America, 27, 4, 769-773 (1955)

Deutsch, L. "The Threshold of the Stapedius Reflex for Pure Tone and Noise Stimuli," Acta Otolaryngologica, 74, 248-251 (1972)

DiCarlo, L.M. "Some Relationships Between Frequency Discrimination and Speech Reception Performance," Journal of Auditory Research, 2, 37-49 (1962)

Fant, G. "Auditory Patterns of Speech," in Models for the Perception of Speech and Visual Form, ed. Wathen-Dunn. Cambridge: The MIT Press (1967)

Flottorp, G., Djupesland, G. and Winther, F. "The Acoustic Stapedius Reflex in Relation to Critical Bandwidth," Journal of the Acoustical Society of America, 49, 457-461 (1971)

Fulton, R.T. and Waryas, A. "Frequency Differentials," in Auditory Stimulus Response Control, ed. by Robert T. Fulton, (University Park Press, Baltimore), 1974

Gengel, R.W. and Watson, C.S. "Temporal Integration: I. Clinical Implications of a Laboratory Study. II. Additional Data from Hearing-Impaired Subjects," Journal of Speech and Hearing Disorders, 36, 2, (1971)

Gengel, R.W. "Temporal Effects in Frequency Discrimination by Hearing-Impaired Listeners," Journal of the Acoustical Society of America, 54, 1, 11-15, (1973)

Greenwood, D.N. "Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane," Journal of the Acoustical Society of America, 33, 1344-1356, (1961)

Harris, J.D. "Pitch Discrimination," Journal of the Acoustical Society of America, 24, 750-755, (1952)

Harris, J.D. Introductory statement for section entitled, "Pitch Modulation and Pitch Memory," in Forty Germinal Papers in Human Hearing, ed. by J.D. Harris, (Journal of Auditory Research, Groton, Conn.), 1969

Jerger, J.F. "Development of Synthetic Sentence Identification (SSI) as a Tool for Speech Audiometry," in Speech Audiometry, ed. by C. Rojskjaer (Second Danavox Symposium, Odense, Denmark), 1970

Jerger, J.F., Tillman, T.W. and Peterson, J.L. "Masking by Octave Bands of Noise in Normal and Impaired Ears," Journal of the Acoustical Society of America, 32, 385-390, (1960)

Jerger, J.F. and Jerger, S. " Psychoacoustic Comparison of Cochlear and VIIIth Nerve Disorders," Journal of Speech and Hearing Research, 10, 659-688, (1967)

Jerger, J.F., Burnery, P., Maudlin, L. and Crump, B. "Predicting Hearing Loss from the Acoustic Reflex," Journal of Speech and Hearing Disorders, 39, 11-22, (1974)

Kiang, N.Y.-s. and Moxon, E.C. "Tails of Tuning Curves of Auditory Nerve Fibers," Journal of the Acoustical Society of America, 55, 3, 620-630, (1974) Koenig, W., Dunn, H.K. and Lacey, L.Y. "The Sound Spectrograph," Journal of the Acoustical Society of America, 17, 19-49, (1946)

Liberman, A.M., Delattre, P.C., Cooper, F.S. and Gerstman, L.J. "The Role of Consonant Vowel Transitions in the Perception of the Stop and Nasal Consonants," Psychological Monographs, 68, 8, (Whole No. 379), (1954)

Liberman, A.M., Delattre, P.C., Gerstman, L.J. and Cooper, F.S. "Tempo of Frequency Change as a Cue for Distinguishing Classes of Speech Sounds," Journal of Experimental Psychology, 52, 2, 127-137, (1956)

Liberman, A.M., Harris, K.S., Hoffman, H.S. and Griffith, B.C. "The Discrimination of Speech Sounds Within and Across Phoneme Boundaries," Journal of Experimental Psychology, 54, 5, 358-368, (1957)

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. "Perception of the Speech Code," Psychological Review, 74, 6, 431-461, (1967)

Liberman, A.M., Cooper, F.S., Harris, K.S., MacNeilage, P.F. and Studdert-Kennedy, M. "Some Observations on a Model for Speech Perception," in Models for the Perception of Speech and Visual Form ed. by W. Wathen-Dunn. Cambridge: MIT Press, 1967

Licklider, J.C.R. "Basic Correlates of the Auditory Stimulus," in Handbook of Experimental Psychology, ed. by S.S. Stevens, New York: John Wiley and Sons Inc. 1951

Lisker, L. and Abramson, A.S. "A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements," Word, 20, 384-422, (1964)

Littler, T.S. The Physics of the Ear, New York: Pergamon Press, 1965

Martin, E.S., Pickett, J.M. and Colten, S. "Discrimination of Vowel Formant Transitions by Listeners with Severe Sensorineural Hearing Loss," in Speech Communication Ability and Profound Deafness, G. Fant Ed. (A.G. Bell Association for the Deaf, Washington, D.C.), 1972

Martin, E.S. and Pickett, J.M. "Sensorineural Hearing Loss and Upward Spread of Masking," Journal of Speech and Hearing Research, 13, 2, 426-437, (1970) Martin, F.N. and Wofford, M.J. "Temporal Summation of Brief Tones in Normal and Cochlear-Impaired Ears," Journal of Auditory Research, 10, 82-86, (1970)

Moore, B.C.J. "Frequency Difference Limens for Short Duration Tones, "Journal of the Acoustical Society of America, 54, 3, 610-619, (1973)

Miller, G.A. and Nicely, P.E. "An Analysis of Perceptual Confusion Among Some English Consonants, "Journal of the Acoustical Society of America, 27, 338-352, (1955)

Nabelek, I. and Hirsh, I.J. "On the Discrimination of Frequency Transitions," Journal of the Acoustical Society of America, 45, 6, 1510-1519, (1969)

Olsen, W.O., Rose, D.E. and Noffsinger, D. "Brief-Tone Audiometry with Normal, Cochlear and Eighth Nerve Tumor Patients," Archives of Otolaryngology, 99, 185-189, (1974)

Plomp, R. and Bouman, M.A. "Relation Between Hearing Threshold of Pure Tones as a Function of Duration," Journal of the Acoustical Society of America, 31, 749-758, (1959)

Pollack, I. "Detection of Rate of Change of Auditory Frequency," Journal of Experimental Psychology, 77, 4, 535-541, (1968)

Pisoni, D.B. On the Nature of Categorical Perception, unpublished doctoral dissertation, University of Michigan (Psycholinguistics), 1971

Rosenblith, W.A. and Stevens, K.N. "On the DL for Frequency," Journal of the Acoustical Society of America, 25, 980-985, (1953)

Rittmanic, P.A. "Pure-Tone Masking by Narrow Noise Bands in Normal and Impaired Ears," Journal of Auditory Research, 2, 287-304, (1962)

Scharf, B. and Hellman, P. "Model of Loudness Summation Applied to Impaired Ears," Journal of the Acoustical Society of America, 40, 71-78, (1966)

Shower, E.G. and Biddulph, R. "Differential Pitch Sensitivity of the Ear," Journal of the Acoustical Society of America, 3, 275-287, (1931) Simon G. "The Critical Bandwidth Level in Recruiting Ears and its Relation to Temporal Summation," Journal of Auditory Research, 3, 109-119, (1963)

Speaks, C. and Jerger, J. "Method for Measurement of Speech Identification," Journal of Speech and Hearing Research, 8, 185-194, (1965)

Stevens, K.N and Halle, M. "Remarks on Analysis-by-Synthesis and Distinctive Features," in Models for the Perception of Speech and Visual Form, ed. by Wathen-Dunn, Cambridege, Mass: The MIT Press, 1967

Stevens, K.N. "Segments, Features and Analysis-by-Synthesis," in Language by Ear and by Eye, ed. by James F. Kavanagh and I.G. Mattingly, Cambridge, Mass: MIT Press, 1972

Stevens, K.N. and House, A.S. "Speech Perception," in Foundations of Modern Auditory Theory, ed. by J.V. Tobias, Vol. II (Academic Press, New York), 1972

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S. and Cooper, F.S. "Motor Theory of Speech Perception: A Reply to Lane's Critical Review," Psychological Review, 77, 3, 234-249, (1970)

Turnbull, W.W. "Pitch Discrimination as a Function of Tonal Duration," Journal of Experimental Psychology, 34, 302, (1944)

Wang, M.D. and Bilger, R.C. "Consonant Confusion in Noise: A Study of Perceptual Features," Journal of the Acoustical Society of America, 54, 5, 1248-1266, (1973)

Wright, H.N. "Clinical Measurement of Temporal Auditory Summation," Journal of Speech and Hearing Research, 11, 109-129, (1968a.)

Wright, H.N. "The Effect of Sensorineural Hearing Loss on Threshold Duration Functions," Journal of Speech and Hearing Research, 11, 842-852, (1968 b.)

Zwicker, E. "Masking and Psychological Excitation as Consequences of the Ear's Frequency Analysis," in Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G.F. Smoorenburg Eds. (A.W. Sijthoff, Leiden), 1970

Zwicker, E., Flottorp, F. and Stevens S.S. "Critical Band Width in Loudness Summation," Journal of the Acoustical Society of America, 29, 548-557, (1957) Zwislocki, J.J. "Analysis of Some Auditory Characteristics," in Handbook of Mathematical Psychology, R.D. Luce, R.R. Bush and E. Galanter (Eds.) New York, John Wiley and Sons Inc., 1965

Zwislocki, J.J. "Theory of Temporal Auditory Summation," Journal of the Acoustical Society of America, 32, 1046-1060, (1960)