To the Graduate Council:

I am submitting herewith a thesis written by Tracie Rice entitled "Effects of wide dynamic range compression on stop consonant place of articulation." I have examined the final copy of this thesis for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Master of Arts, with a major in Audiology.

Mark Hedrick

Mark Hedrick, Ph.D., Major Professor

We have read this thesis
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of the Graduate School

EFFECTS OF WIDE DYNAMIC RANGE COMPRESSION ON STOP CONSONANT PLACE OF ARTICULATION

A Thesis Presented for the Master of Arts Degree
The University of Tennessee, Knoxville

Tracie Rice
May 1999

DEDICATION

This thesis is dedicated to my dear friends and the audiology staff at the University of Tennessee-Knoxville. I am eternally grateful for the encouragement and support you all have given me. For my friends that have worked as hard as I have getting through the program at SSH I commend you for your efforts!!

ACKNOWLEDGMENTS

Thank you to my major professor, Dr. Mark Hedrick, who spent a great deal of time explaining all the aspects of speech perception related to this project. Thank you to my thesis committee for your time, effort and suggestions.

Thank you to my classmates for your undying support and encouragement. You guys have made this time in my life very special and I will never forget you.

Special thanks to my family back home and in Hilton Head for all the love and support you have given. I am eternally grateful for all the trips you made to Knoxville when you knew I just needed a hug and the phone conversations that kept me going.

Last but not least, thanks to my husband Merritt. Thank you for letting me talk for hours about speech perception and hearing aids and acting as if it were as important to you as it was to me. Thank you for always standing behind me and supporting my point of view, even when it may not have always been right.

TABLE OF CONTENTS

CHAPTER

I. Introduction

II. Review of Literature

III. Methods

IV. Results

V. Discussion

LIST OF FIGURES

FIGURE

ABSTRACT

The present study was conducted to determine if wide dynamic range compression alters listeners' perception of place of articulation when relative amplitude of a stop consonant and a vowel is manipulated synthetically and if so, does the Quantal Theory explain any differences found in listeners' perception of place of articulation.

The stimuli for this investigation were synthetic consonant-vowel (CV) syllables generated by a PC software version of Klatt's cascade/parallel formant synthesizer using a sampling rate of 10 kHz. Twenty-five subjects ranging in age from 22 to 35 years participated in this experiment.

Eight of the 25 subjects perceived more /p/ than /t/ for the processed condition. 7 of the 25 subjects perceived more /t/ than /p/ for the processed condition and can be explained by the Quantal Theory. Ten of the subjects had consistent changes between the unprocessed and processed stimuli. A plausible explanation for the 10 subjects that did not have significantly different perception of place of articulation is that perhaps those listeners' cue into formant transition changes that were not altered in this study.

CHAPTER I

Introduction

Background and Rationale

According to the Source-Filter Theory, speech sounds are produced by a multiplication of activity of the vocal folds with the resonances of the vocal tract shaped by movement of the articulators. The product of this interaction results in specific acoustic patterns for speech sounds (Fant, 1960). While vowels and resonant consonants such as nasals and liquids are characterized by free airflow and a definitive formant structure, other consonants such as stops, fricatives, and affricates are characterized by a restricted or obstructed airflow and have weak formant structures (Borden, Harris & Raphael, 1994). The current study will concentrate on the voiceless stops /p/ and /t/ paired with the vowel /ɑ/.

The essential articulatory feature of a stop consonant is a momentary blockage of the vocal tract. The blockage is formed by an articulatory occlusion occurring at either the bilabial, alveolar, or velar site. The articulatory blockage has a variable duration that varies from 50-100 ms. Following the blockage of air is the release of air occurring as the air pressure impounded behind the obstruction escapes, generally referred to as a burst. Typically, the burst is no longer than 5-40 ms in duration. The voicing feature for syllable-initial stops can be specified by a single number that gives the interval between the articulatory release of the stop and the onset of vocal fold vibrations. This time interval is called voice onset time (VOT). The VOT for voiceless consonants is a range from 25 - 100 ms. The voiceless consonant VOT is described as a range because there is no single value of VOT that will be used by all speakers or across all phonetic contexts (Kent and Reed, 1992).

Formants are defined as resonances of the vocal tract. In general, changes in vocal tract shape during speech are signaled acoustically by changes in the vocal tract resonances. These changes are referred to as formant transitions. It has been discovered that the acoustic changes have approximately the same duration as the underlying articulatory changes. Therefore, if the articulatory transition from a stop occlusion to vowel configuration takes 50 ms, the acoustic transition also has a duration of about 50 ms (Kent, 1992). Typically, the first three formant transitions are used most for classification of articulatory cues. These three formant cues are referred to as F1, F2, and F3 respectively. The F1 transition is considered to be a cue to manner of production and the F2 and F3 transitions may be cues to place of articulation.

In addition to formant transitions, the frequency spectrum of the burst is a salient cue for perception of stop consonant place of articulation. The bilabial stops /p b/ have a spectra that fall in amplitude as frequency is increased; the alveolar stops /t d/ have a spectra that rises in amplitude as frequency is increased; and the velar stops /k g/ have a compact spectrum with maximum amplitude in the mid-frequency region.

The Quantal Theory

Stevens and colleagues have conducted studies to determine how articulatory changes affect acoustic parameters of speech sounds. The Quantal Theory suggests there are certain regions that cause large changes in the acoustic perception when there are small changes in articulatory motion. This movement region where large changes occur could be considered, in some sense, as a threshold region such that as the acoustic parameter changes through this region the auditory response shifts from one type of pattern to another. For example, earlier studies have shown that the amplitude of the burst relative to the adjacent vowel in a particular frequency region affects listeners' perception of stop consonant place of articulation (Ohde & Stevens, 1983; Gravel & Ohde, 1983).

Stevens' Quantal Theory and the results of Ohde and Stevens (1983) predict that a higher burst amplitude in the F4-F5 frequency region relative to the adjacent vowel's amplitude in the same frequency region will yield an alveolar percept. Conversely, a lower burst relative to vowel amplitude in the F4-F5 frequency region will yield a labial percept. These predictions were borne out in the results of a recent study by Hedrick, Schulte, and Jesteadt (1995). Furthermore, this study and a companion study (Hedrick & Jesteadt, 1996) found that listeners with sensorineural hearing loss utilized this relative amplitude comparison more than formant transitions to make place of articulation judgments.

Relative Amplitude

Studies have found that the relative amplitude between the burst and the vowel in the F4-F5 frequency region along with overall presentation level can influence the perception of place of articulation (Ohde & Stevens, 1983; Gravel & Ohde, 1983). The amplitude of a major peak in the spectrum of the consonantal portion of a consonant/vowel syllable relative to a spectral peak in the same frequency region in the onset of the adjacent vocalic portion may be a cue for consonant perception (Hedrick et al., 1995).

Processing of the relative amplitude property for perception of place of articulation of stop consonants appears to be affected by the overall presentation level of the stimuli, in that higher presentation levels result in more alveolar than labial judgments (Gravel and Ohde, 1983).It is felt that when altering the characteristics of the amplitude and frequency spectrum of the consonant in relation to the vowel the accuracy of perception will be affected (Stevens & Blumstein, 1981).

In a study conducted in 1995 by Hedrick, Schulte, and Jesteadt, it was found that normal-hearing listeners yielded more alveolar responses with an increase in relative amplitude. An increase in relative amplitude can be described as increasing the energy of the consonant with respect to the vowel in the F4-F5 frequency region. It was also found that overall presentation level influences the processing of the relative amplitude cue for stops. This data supports the Quantal Theory for normal hearing listeners, in that small changes in acoustics near a threshold region caused shifts in perception of place of articulation.

Wide Dynamic Range Compression Amplification

Compression circuits can be described in terms of three primary characteristics: threshold or kneepoint of compression, compression ratio, and attack and release times (Tobin and Dempsey, 1997). Threshold of compression is the level of an incoming signal that is just intense enough to trigger the compression function. The compression ratio indicates the degree to which the signal will be compressed. The attack time is generally preferred to be very short in order to protect the wearer from high output levels. The release time should be longer than 50 ms in order to keep the wearer from hearing a "pumping sound" due to the circuit going in and out of compression too quickly (Tobin and Dempsey, 1997). Wide Dynamic Range compression circuits are generally used with individuals with more severe tolerance problems because of the low kneepoint with a low compression ratio that is activated for all but the very least intense signals.

Consonant Vowel Ratio

Consonant-vowel (CV) ratio is defined as the overall amplitude between a consonant and the following vowel. The amplitude of most consonant sounds is substantially lower than that of vowels, with the weakest consonant being as much as 30 dB lower than the strongest vowels (Dunn & White, 1940). Many amplification circuits alter the consonant-vowel ratio in hopes of providing better speech understanding. Compression amplification devices increase the amplitude of the consonant with respect to the vowel in hopes of providing the listener with better consonant recognition. The most positive benefit to increasing the CV ratio is that audibility is increased overall. This may help in the perception of sound in general; however, according to the concept of relative amplitude in the Quantal Theory, changing the CV ratio could result in errors of perception of place of articulation. For example when raising the amplitude of the /p/ to be more close to the amplitude of the vowel, the /p/ may be perceived as a /t/.

Thus it is plausible that Wide Dynamic Range Compression changes the normal relative spectral amplitude relations in natural speech between consonants and vowels.

Altering these spectral cues may in some cases alter consonant recognition. Because previous studies have shown that listeners with sensorineural hearing loss predominantly use relative amplitude to make place of articulation judgements, it is imperative to determine how Wide Dynamic Range Compression alters the relative amplitude cue. Therefore, the following questions will be the foci of this research:

1) Does Wide Dynamic Range Compression alter listeners' perception of place of articulation when relative amplitude is manipulated synthetically?

2) Does the Quantal Theory explain any differences found in listeners' perception of place of articulation?

CHAPTER II

Review of Literature

Background and Rationale

As stated previously, stop consonants are classified by the blockage of air followed by the sudden release of energy with constriction at either the bilabial, alveolar, or velar place of articulation. When a consonant is paired with a vowel the movement of the articulators change the resonance of the vocal tract which is called the formant transition. The formant transitions contribute to the perception of the place of articulation as well as the manner of articulation.

At the release of a stop consonant there is a sequence of acoustic events consisting of a brief burst of acoustic energy followed by a rapid change in the spectrum of the sound as the articulatory structures in the vocal tract move away from the constricted configuration used to produce the consonant. Certain properties of this acoustic pattern provide a listener with information concerning the place in the vocal tract where the consonant closure is made and the way in which the articulators are shaped to produce the constriction (Ohde & Stevens, 1983).

The Quantal Theory

When a parameter specifying the configuration or state of an articulatory structure is manipulated through a range of values, some acoustic parameter describing the resulting sound often changes in a non-monotonic fashion (Stevens, 1989). There appear to be ranges of the articulatory parameter motion for which there is very little change in the acoustic parameter and other ranges where the acoustic parameter is more sensitive to changes in articulation.

Experiments where the measures being manipulated are in a burst that precedes a synthetic vowel show that the initial consonant in the synthetic syllable is heard as a velar stop consonant for the bursts with a single compact spectral prominence and as a labial or alveolar consonant when the compactness is decreased (Hawkins & Stevens, 1987). The shift in identification for threshold from labial to alveolar occurs when the peak spectrum amplitude of the prominence is about 3-5 dB below the peak amplitude of the corresponding prominence in the adjacent vowel (Stevens, 1989).

Stevens (1989) found that for some articulatory and acoustic parameters the contrasts are inherently more salient than for others. Therefore, assuming the Quantal

Theory is correct, the change in amplitude ratio may cause perception of certain consonants to change and thus cause perceptual errors.

Relative Amplitude

Ohde and Stevens (1983) proposed two acoustic properties to account for the labial-alveolar distinction. One property is specified by the time course of the change in amplitude in the high frequency range (above 2500 Hz) in the few tens of ms following consonantal release, and the other is defined by the frequencies of spectral peaks associated with the second and third formants in relation to the first formant. Based on acoustic evidence (Zue, 1976; Ohde, 1979), these properties were hypothesized to have differential effects in the perception of place of articulation with the former more salient for voiceless stops and the latter more salient for voiced stops.

Hedrick et al. (1995) replicated earlier research in that for normal hearing listeners the salience of the relative amplitude cue for stop consonants with a relative amplitude increase yielded more alveolar responses. In addition, overall presentation level influences the processing of this cue for stops. Higher consonant burst amplitude than vowel amplitude in the F4 frequency region results in more alveolar than labial stop consonant responses (Hedrick et al., 1995). This study also found that hearing-impaired listeners selected more alveolar responses than listeners with normal hearing when tested at equivalent SPLs.

Gravel and Ohde (1983) also found that the processing of the relative amplitude property for perception of place of articulation of stop consonants appears to be affected by the overall presentation level of the stimuli, in that higher presentation levels result in more alveolar than labial judgments.

CV Ratio

Consonant/Vowel (CV) ratio is known to be an important factor in speech intelligibility. Research on speech acoustics has shown that the intensity of most consonant sounds is substantially lower than that of vowels, with the weakest consonant being as much as 30 dB lower than the strongest vowels (Dunn & White, 1940). Many of the options in amplification deal with altering the consonant-vowel ratio in hopes of providing better speech understanding. Compression amplification devices increase the intensity of the consonant with respect to the vowel in hopes of providing the listener with better consonant recognition.

Studies have not concluded a high relation of positive benefit occurring with respect to increasing the consonant amplitude. Results reported by Freyman and Nerbonne (1989) suggested that the major benefits of CV ratio enhancement, increasing the overall amplitude of the consonant with relation to the vowel, relate to consonant audibility; that is, listeners' performance was determined more by absolute levels of consonant segments than by CV ratio, and that CV ratio would be more important at lower signal levels. It was also noted that CV ratio modification influenced recognition cues for fricatives but not stops (Freyman & Nerbonne, 1989). Another study conducted by Montgomery and Edge (1988) found that the consonant amplification loses its beneficial effect as the overall level approximates PB max. Unfortunately, this is near the level where hearing aids would typically be set, so selective enhancement of CV ratio appears to be of limited practical value at least for the moderately hearing-impaired subjects.

It has been a concern that modification of CV ratio will adversely affect the amplitude envelope, thereby causing problems for consonant recognition. Boothroyd, Springer, Smith and Schulman (1988), Moore and Glasberg (1986), and Plomp (1988) have expressed concern that amplitude compression using short attack and release times could distort waveform envelope cues and reduce speech recognition performance in hearing-impaired individuals. Boothroyd et al. (1988) found that performance of 8 of 9 profoundly hearing-impaired listeners was actually poorer for single-channel amplitude compressed speech than for uncompressed speech. They suggested that amplitude compression may often be inadvisable for profoundly hearing-impaired persons, who probably rely more heavily than normal-hearing listeners on the gross amplitude and temporal features of the waveform. Research conducted by Hedrick et al. (1995) suggested that hearing aids with short time constants will increase the amplitude of low-level speech segments relative to high-level segments. Hearing aids such as this might boost the overall amplitude of a low-level consonant, thereby altering the natural relative amplitude values between consonant and vowel. This may lead to perceptual errors in determining place of articulation.

Many experiments have found that place of articulation information was more vulnerable to spectral degradation than were manner or voicing cues (Balakrishnan et al, (1996); Van

Tasell et al. (1987); ter Keurs et al. (1992); and Moore and Glasberg (1993)). Balakrishnan et al. (1996) found some notable exceptions where manner of articulation errors were seen. Some of the errors include affricate and glide for nasal errors, as well as /v/ substitutions for /b/ and /w/. Studies conducted by Kennedy et al. (1998) found that nasal consonants benefitted least from adjustment to the CV intensity ratio.

The most recent research conducted in the area of CV ratio was conducted by Balakrishnan et al. (1996), and Kennedy et al. (1998). In both experiments it was concluded that when CV ratios were altered, substantial changes in the recognition of many individual consonants occurred. Balakrishnan et al. (1996) found that there were a few sounds, notable the weak voiceless fricatives, that were affected negatively by increasing their amplitudes relative to the natural amplitude. It was also found that intelligibility of glides and sibilant fricatives increased with increasing CV ratio, nasals and weak fricatives were best at low CV ratios, and stops and affricates were generally independent of CV ratios (Balakrishnan et al, 1996). Both Kennedy et al. (1998) and Balakrishnan (1996) found that where CV ratio was important, optimum CV ratios were often in the vicinity of the natural CV ratio. Kennedy (1998) found an overall increase in scores when CV ratios were altered but were a result of the individualized adjustments made for each subject for each CV combination. Without such individual adjustment, it is possible that some CV ratio changes may be detrimental to accurate consonant perception.

In terms of signal processing and improved understanding by hearing-impaired listeners, it is important to determine the strengths and weaknesses involved in CV ratio adjustments. If adjusting the CV ratio alters perception of the relative amplitude cue, then it is possible that a wide dynamic range compression circuit may alter the cue that listeners with hearing loss give the most perceptual weight (Hedrick et al., 1995; Hedrick & Jesteadt, 1996).

CHAPTER III

Methods

Subjects

Twenty-five subjects ranging in age from 22 to 35 years participated in this experiment. Inclusion criteria were: (i) hearing sensitivity less than or equal to 15 dB HL (ANSI S3.6-1989) for 250-8000 Hz in the right ear; (ii) no evidence of abnormality of the pinna or ear canal; and (iii) normal tympanogram. There was one experimental session for each listener and all listeners were unpaid volunteers.

Stimuli

The stimuli for this investigation were synthetic consonant-vowel (CV) syllables generated by a PC software version of Klatt's cascade/parallel formant synthesizer (Klatt, 1980) using a sampling rate of 10 kHz. Nine stimuli ranging from -12 dB to +12 dB relative amplitude values in 3 dB steps were synthesized. A positive relative amplitude value means that the amplitude of the consonant burst is greater than the amplitude of vowel onset in the F4-F5 frequency region. A negative relative amplitude value means that the amplitude of the consonant burst is less than the amplitude of vowel onset in the F4-F5 frequency region.

The consonantal portion of the CV syllables was 60 ms, and the vocalic portion of the syllables was 200 ms. Each CV was initiated by a 25 ms burst of frication noise. Aspiration noise was initiated 10 ms after burst onset and remained on until 5 ms after vocalic onset. The aspiration noise began at a low level and reached a maximum amplitude in 20 ms and sharply fell during its last 15 ms. Voicing amplitude gradually rose in 40 ms to an overall peak level and remained there for 160 ms.

Formant amplitude parameters A4 and A5 of the Klatt synthesizer were manipulated to yield a particular relative amplitude value. The relative amplitude values were measured using linear predictive coding (LPC) spectra of the stop burst and vowel onset. The LPC spectra were made using 15 coefficients and a 25.6 ms full Hamming window. By manipulating the A4 and A5 amplitude parameters, the peak of the burst spectra was compared to the peak of the vowel onset spectra at F4 to make the peak of the burst either -12, -9, -6, -3, 0, +3, +6, +9, or +12 dB relative to the peak of the vowel onset spectra.

For each stimulus, the formant frequency onset value for F1 began at 400 Hz and in 40 ms (from just before burst termination to initiation of voicing) rose to the vowel steady-state value of 700 Hz. There were no formant transitions for F2 and F3; in this way, the transition cue was neutral for the labial/alveolar contrast. Thus, the formant frequency onset values and the vowel steady-state formant frequency values for F2, F3, F4, and F5 were constant. Thus, the relative amplitude manipulation theoretically was the only acoustic cue provided for listeners to make a determination of place of stop consonant articulation. Steady-state vowel formant frequency values were F1=700 Hz, F2=1220 Hz, F3=2600 Hz, F4=3500 Hz, and F5=4200 Hz. F0 began at 130 Hz at voicing onset and declinated to 100 Hz at voicing offset. Appendix A shows the synthetic parameter values used to create the stimuli.

Following processing by the hearing aid, the stimuli were analyzed acoustically to determine if either the formant transition or relative amplitude cues were altered acoustically. This was done by a comparison of unprocessed and processed signal spectrograms and frequency spectra.

Recording system

The major comparison to be made in the current study is between perception of the original stimuli (or unprocessed stimuli) and the stimuli recorded through a compression hearing aid (or processed stimuli). To this end, the stimuli were recorded through a hearing aid having a wide dynamic range compression circuit (Unitron BTE IKON K). The hearing aid was checked to insure that it performed according to ANSI specifications. Appendix B presents the ANSI summary of the electroacoustic specifications of the hearing aid. The hearing aid potentiometers were set accordingly: threshold knee-point at normal, high-cut tone control at normal (which resulted in increased output for frequencies above 2 kHz), and the low-cut tone control at h (which resulted in reduced output in the low frequencies below 2.5 kHz). According to ANSI specification, the attack time of the hearing aid was 10 ms, and the release time 536 ms. The compression ratio for the hearing aid was 2:1.

Following stimulus generation, the stimuli were recorded on a DAT recorder (Sony 2000 ES), routed to an audiometer (Madsen OB 822), and sent to a loudspeaker (Infiniti 2000.6) inside a double-wall sound booth (IAC). The peak level of the stimulus emanating from the loudspeaker was 80 dB SPL, as measured using a Larson-Davis 800B sound level meter located 3 meters from the loudspeaker. The unprocessed signal was then recorded by a ½ inch pressure microphone (Larson-Davis) located at the end of a Zwislocki coupler. The coupler and microphone were located 3 meters from the loudspeaker cone. The output of the microphone was fed to a preamplifier (Larson-Davis PRM 900B), an amplifier (Larson-Davis 2200C), routed to a preamplifier outside the booth (Tucker-Davis MA2), sent to an anti-aliasing filter (Tucker-Davis FT6, set a 20 kHz low-pass), routed through a high-pass filter set at 75 Hz (Tucker Davis PF1), sent to an A/D (Tucker-Davis DD1) and digitized at a sampling rate of 44.1 kHz in a software speech analysis package (CSRE version 4.5) using a Compaq 2000 586 PC. These stimulus files were then presented to the listeners. The processed stimuli had the same recording apparatus, except that the BTE hearing aid was coupled to the Zwislocki coupler via insertion of the bore of the full-shell earmold into the coupler opening.

Procedure

The presentation of stimuli to the subjects was implemented using interactive signal generation and control software (CSRE version 4.5, with a Compaq 2000 586 PC). The stimuli were output using a Tucker-Davis DD1 D/A converter, low-pass filtered at 4.9 kHz (Tucker-Davis PF1), sent to a final attenuator (Tucker-Davis PA4), routed to a headphone buffer (Tucker-Davis HB), and sent to ER-3A insert headphones having a roll-off of approximately 10 dB per octave above 4 kHz. The headphones were located inside the IAC booth.

Generation of random orderings and online data collection were performed using the interactive software. Subjects were instructed to identify the consonant perceived by selecting the appropriate symbol ("P" or "T") displayed on a computer monitor via a mouse. All subjects were given a criterion test using continua endpoints. Subjects had to classify the most /p/ - like and the most /t/ - like stimuli with at least 80% accuracy before inclusion into the study. The criterion test and data collection were presented at a level of 80 dB peak SPL through the insert earphones, as measured using a Larson-Davis 800B sound level meter with a one-inch microphone and a 2cc coupler. Both sets of stimuli (9 unprocessed and 9 processed) were presented in 10 random orders, with the order of the sets counterbalanced across listeners.

CHAPTER IV

Results

Acoustic Analysis

Figures 1-6 shows the effects of processing by the hearing aid on the waveform and spectrographic representation of the speech stimuli. In these waveforms the amplitude of the consonant portion of the processed stimuli is large relative to the vowel. Figures 1 and 2 are the most /p/ like stimuli, unprocessed and processed. The criteria for the /p/ stimuli is a lower amplitude for the burst comparted to the vowel for the F4-F5 frequency region. In comparing these figures, the waveform of the processed stimuli shows a larger amplitude of the consonant relative to the vowel, along with a frequency damped appearance for the initial burst. The spectrograph of Figure 2 shows increased amplification in the higher frequency region for the processed stimuli. Figures 3 and 4 are the most neutral stimuli, unprocessed and processed. These two figures again show the larger amplitude of the consonant in relation to the vowel, the frequency damped shape of the consonant burst and the increased amplification of the higher frequencies for the processed stimuli. Figures 5 and 6 are the most /t/

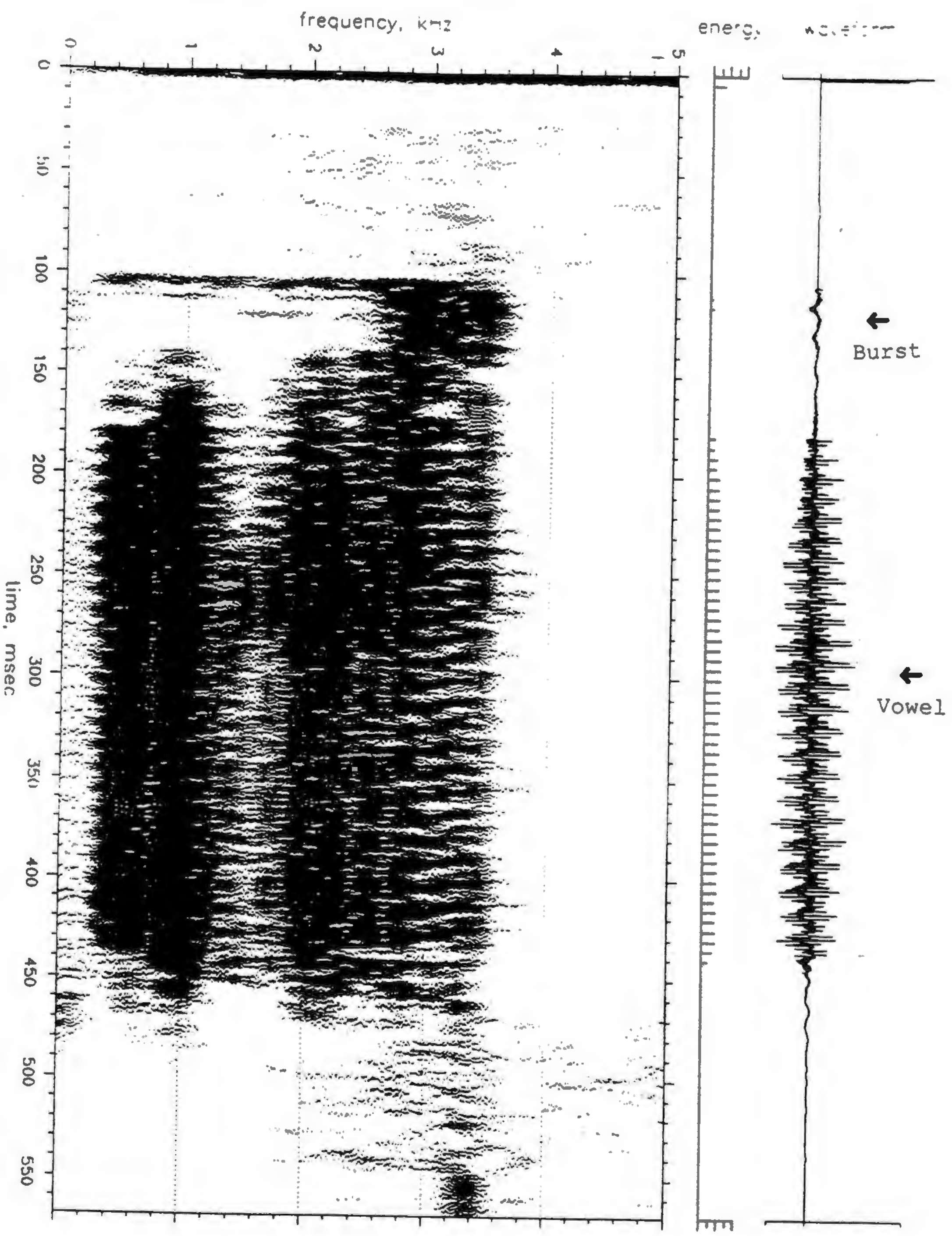


Figure 1. Most /p/ like stimulus - unprocessed

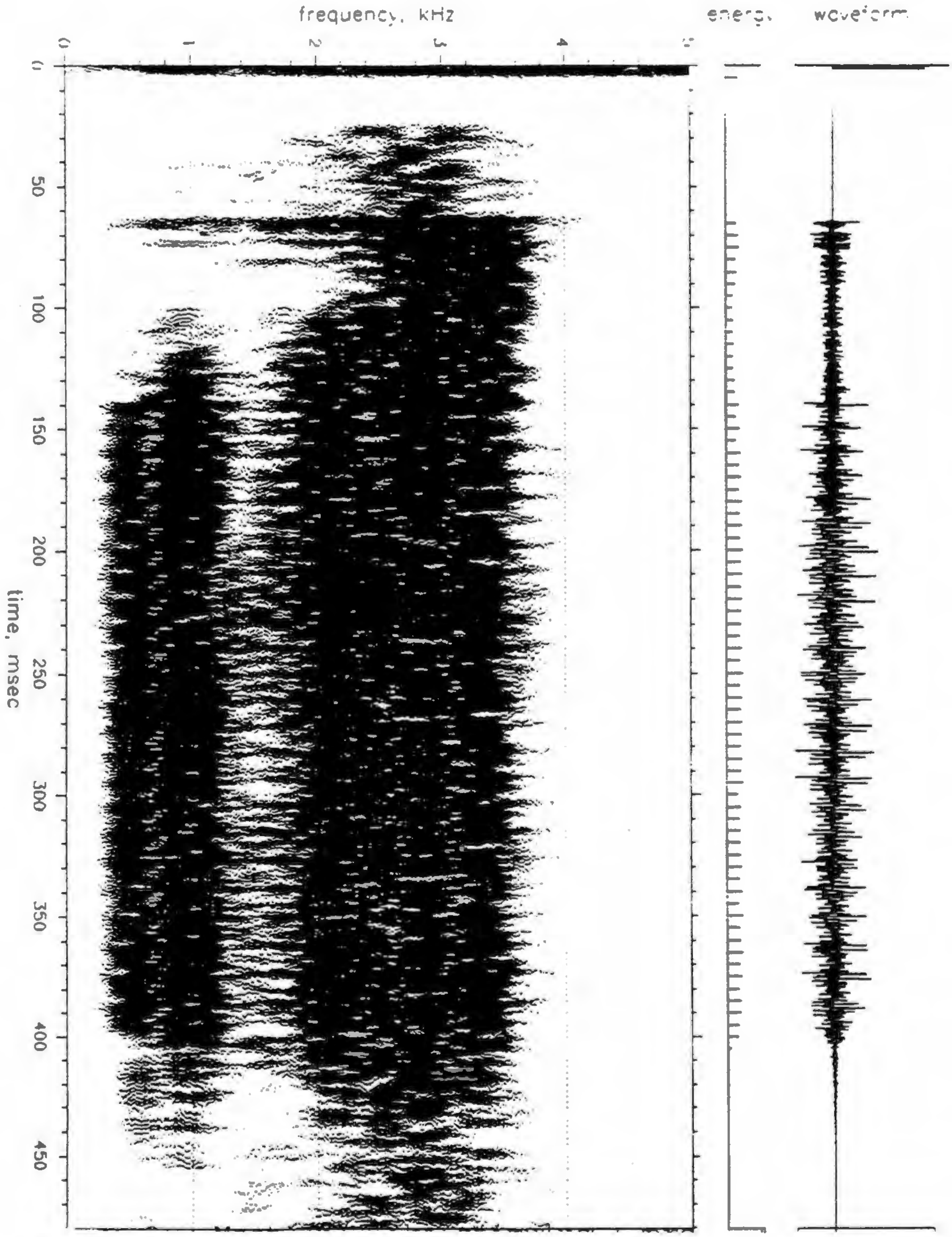


Figure 2. Most /p/ like stimulus - processed

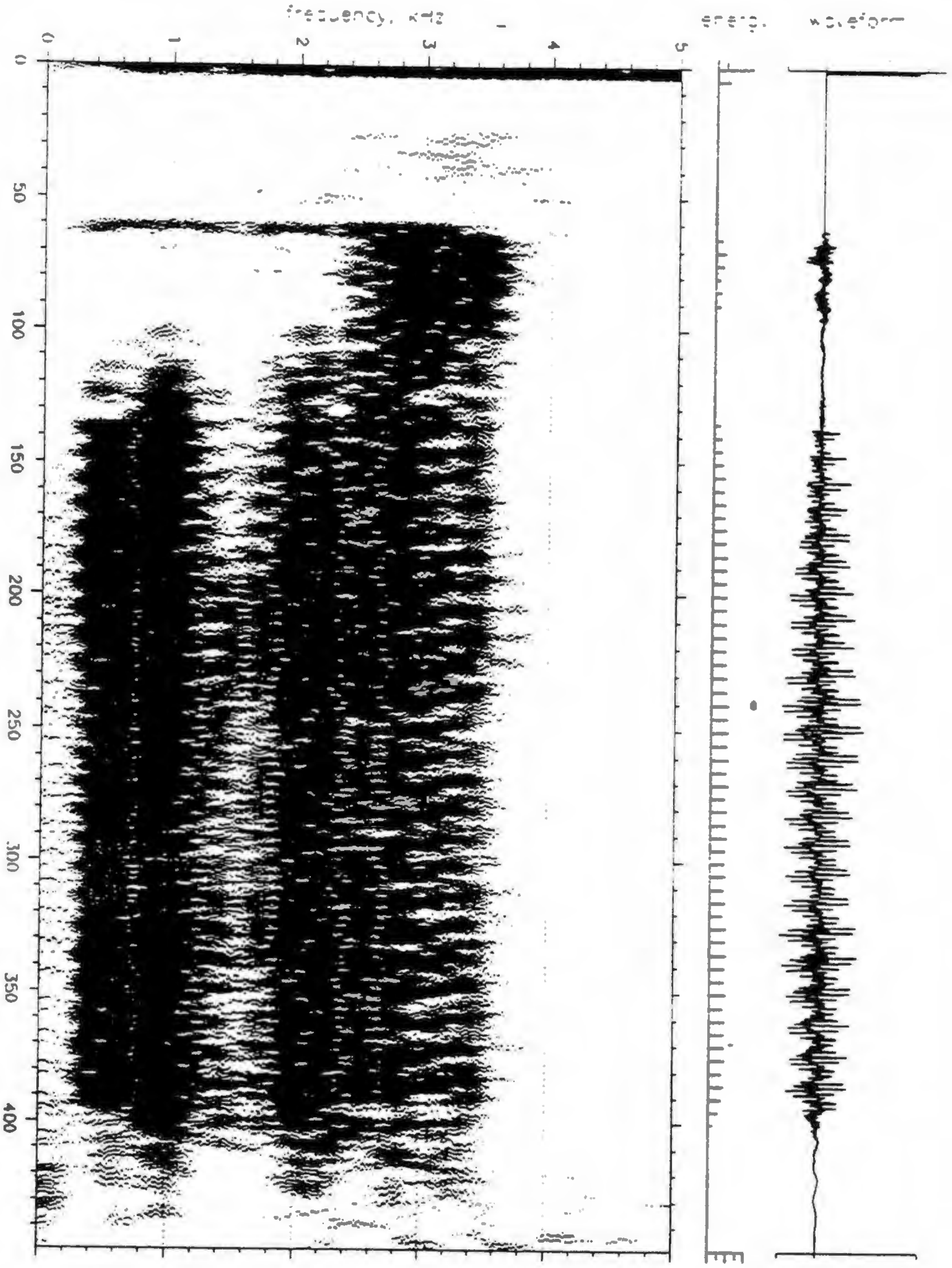


Figure 3. Most neutral stimulus - unprocessed

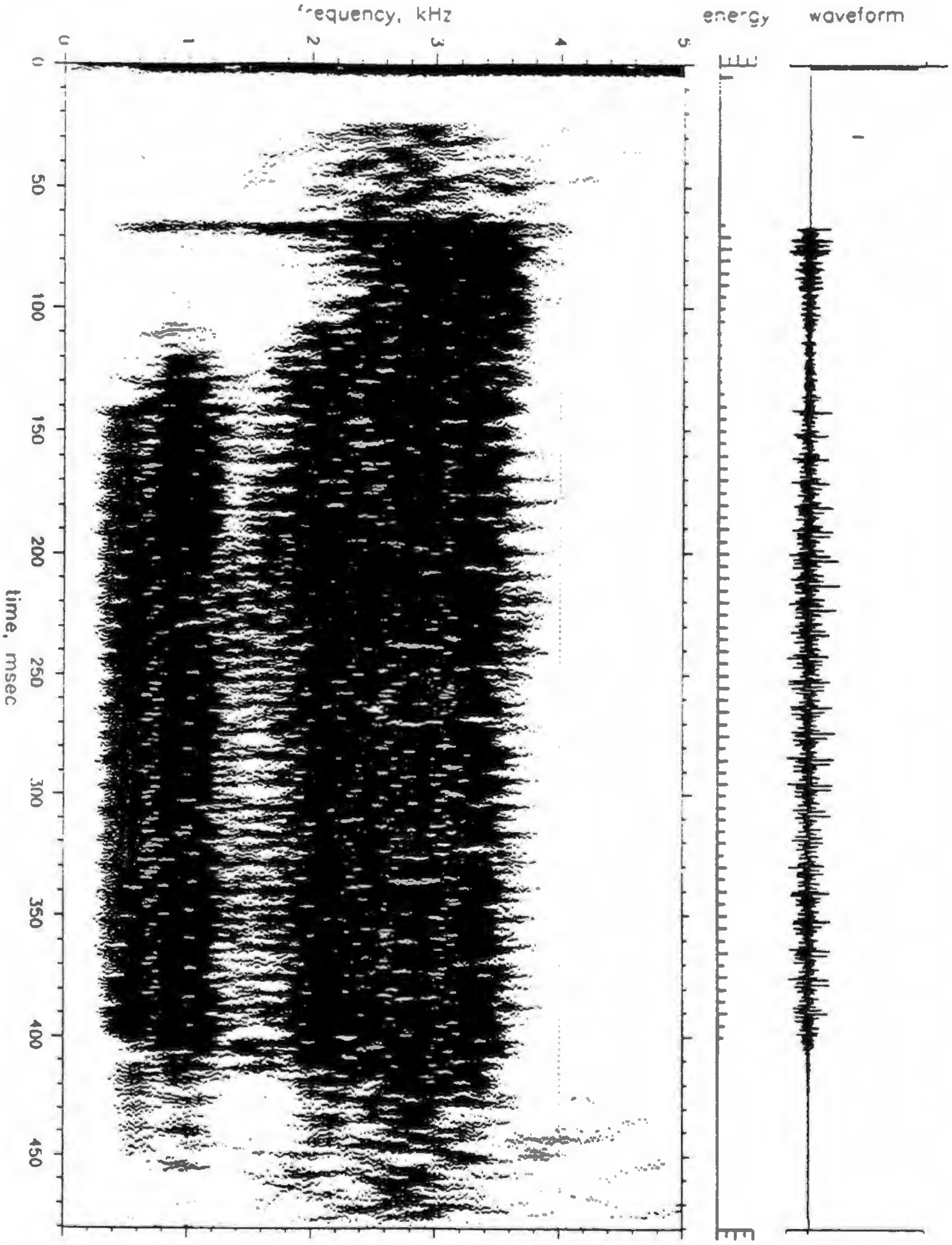


Figure 4. Most neutral stimulus - processed

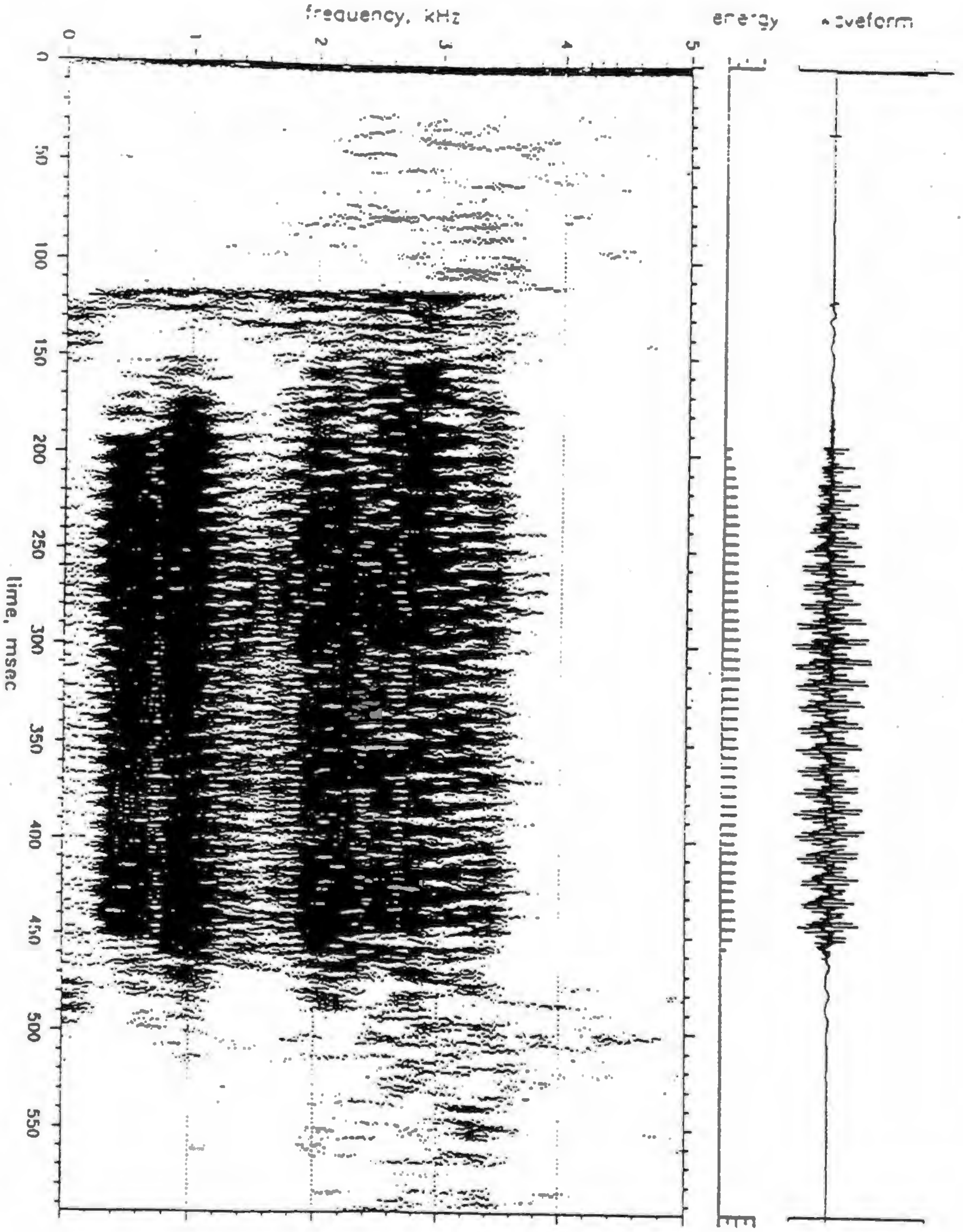


Figure 5. Most /t/ like stimulus - unprocessed

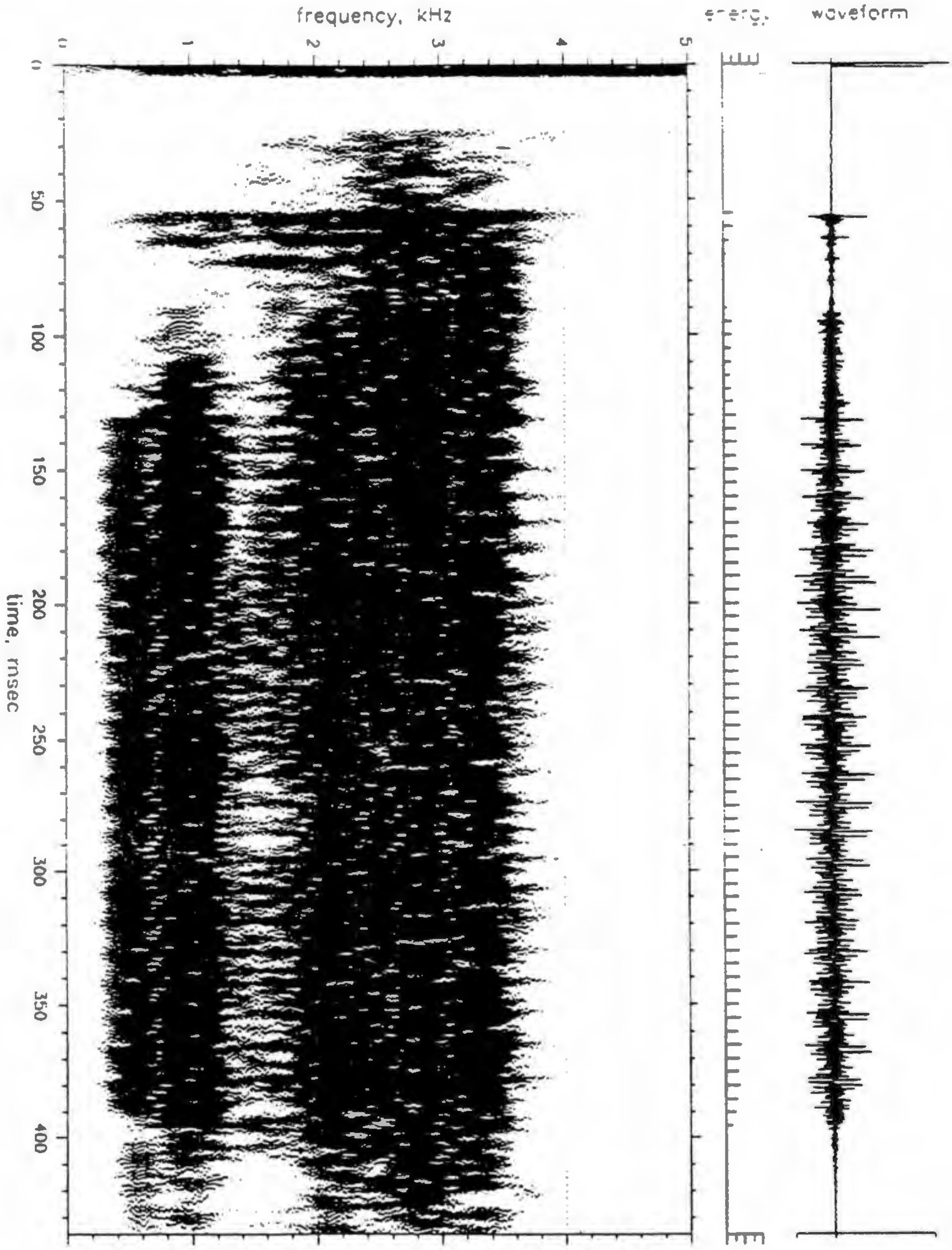


Figure 6. Most /t/ like stimulus - processed

like stimuli, unprocessed and processed. In comparing these figures, the same pattern as described above is noted for the processed stimuli. Also note the frequency damped waveform of the processed /t/ stimuli in Figure 6.

As a further comparison, Figures 7 and 8 show waveform and spectrograms obtained from naturally produced /p/ and a naturally produced /t/ respectively. Note the frequency damped shape of the initial burst of the /p/ production in comparison to the /t/ production. Thus it is evident that the naturally produced /p/ burst has a frequency damped shape and the naturally produced /t/ does not.

LPC Spectra

Figures 9-14 show the LPC spectra of the stimuli both unprocessed and processed. Figures 9 and 10 show the most /p/ like stimuli for both consonant and the vowel. Note that the consonant has a smaller overall amplitude than the vowel. In comparing these two figures one can see the increase in the amplitude of both the consonant and the vowel after being processed by the hearing aid. In particular, the level of the burst is relatively closer to that of the vowel for the processed stimulus. Figures 11 and 12 show the most neutral stimuli for both conditions.

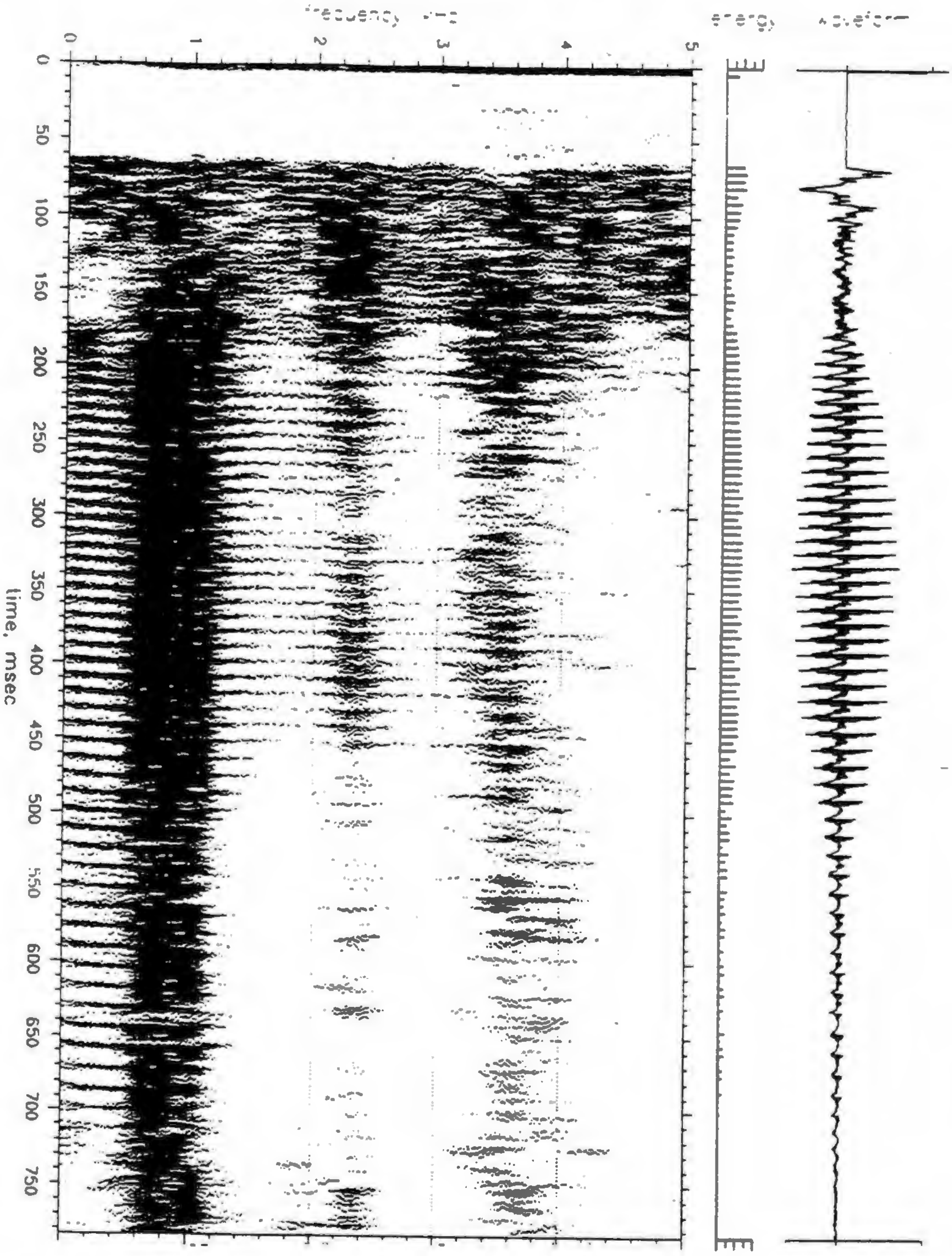


Figure 7. Naturally produced /p/

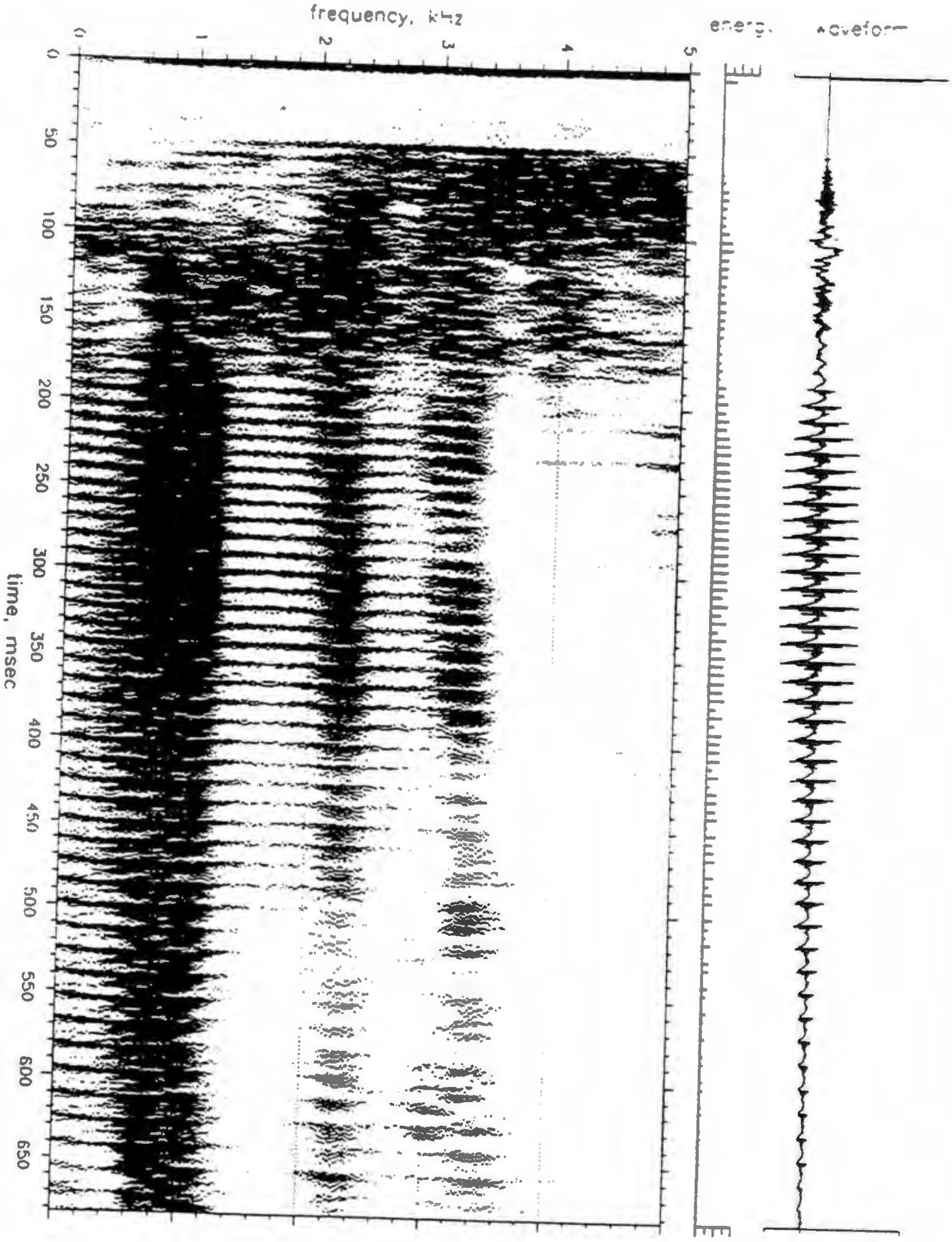


Figure 8. Naturally produced /t/

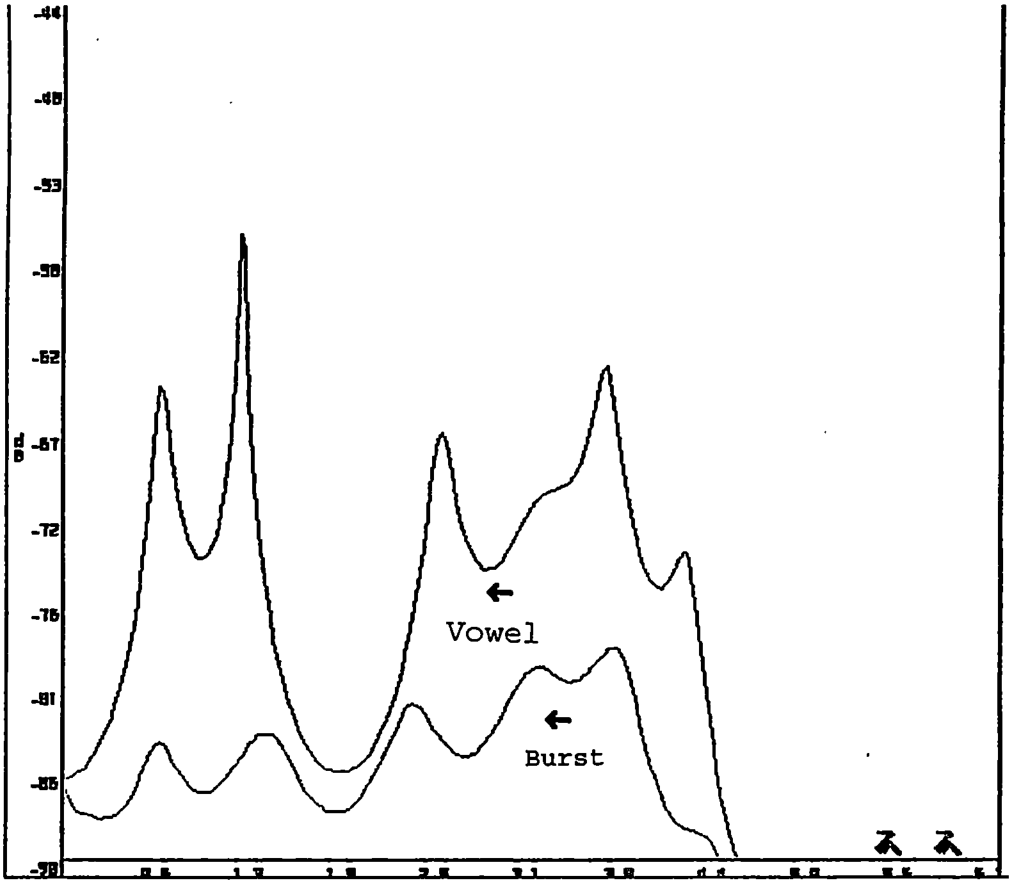


Figure 9. Burst and vowel onset for most /p/ like stimulus - unprocessed

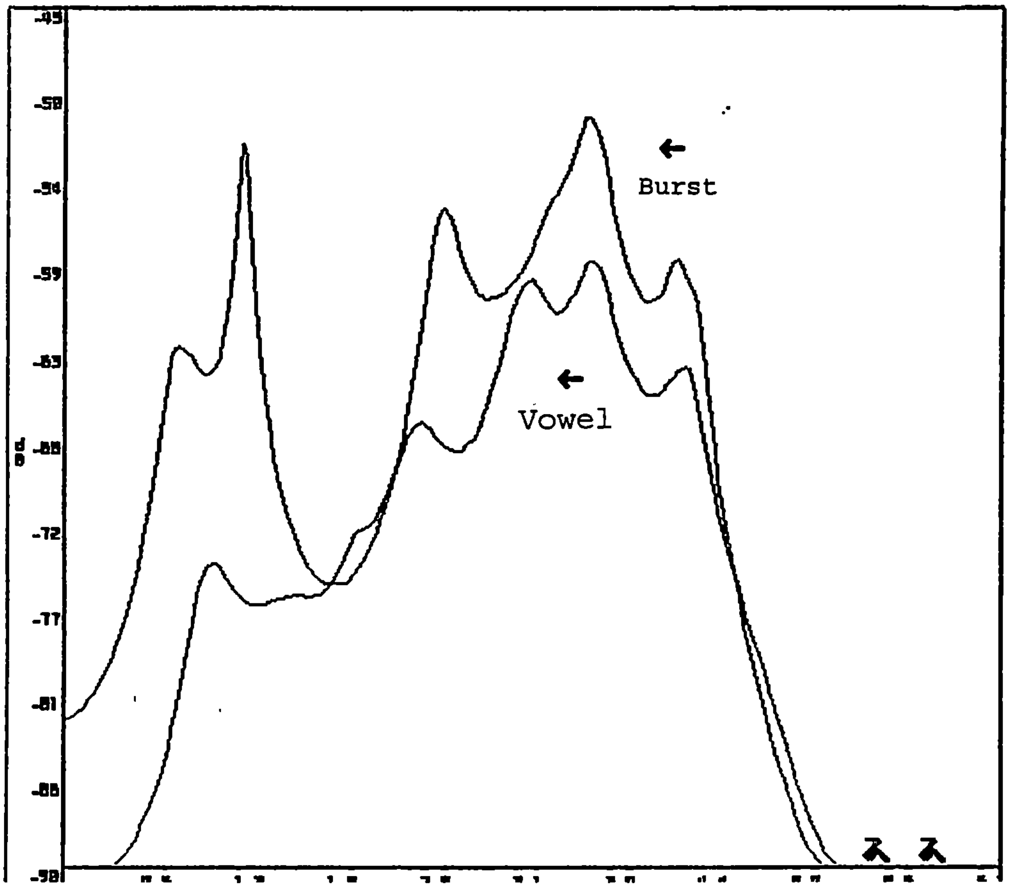


Figure 10. Bust and vowel onset for most /p/ like stimulus - processed

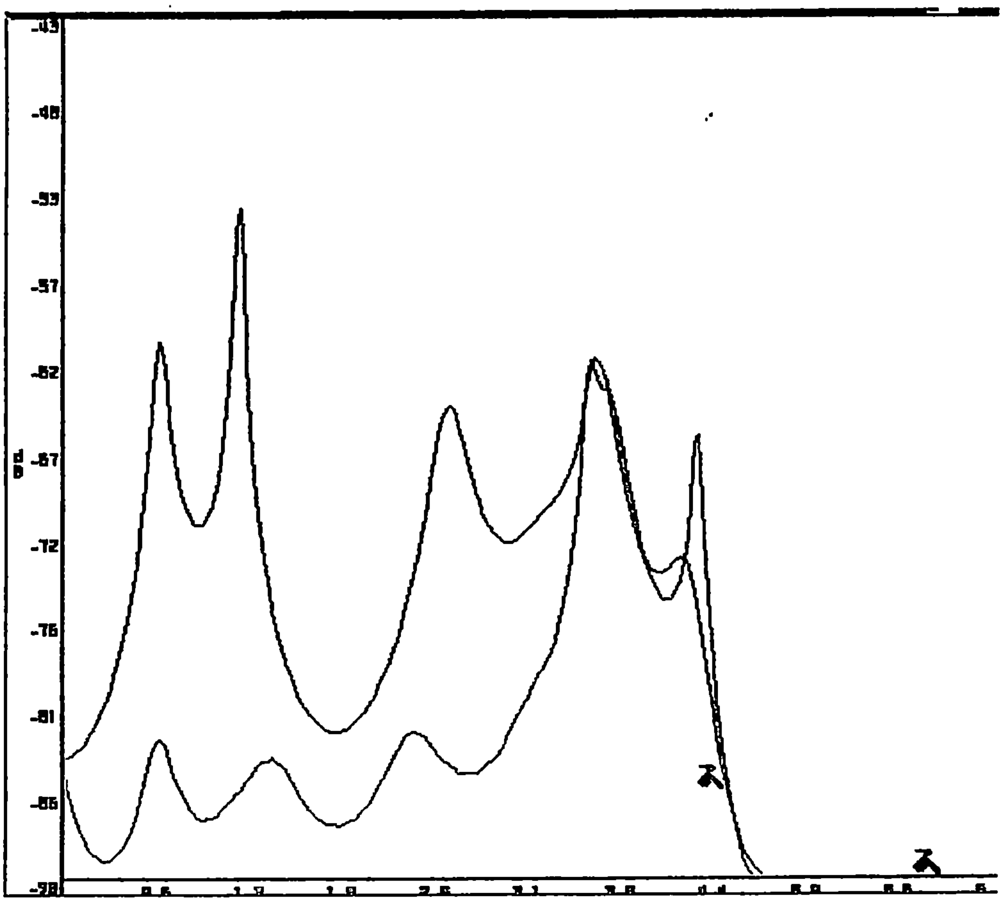


Figure 11. Burst and vowel onset for most neutral stimulus - unprocessed.

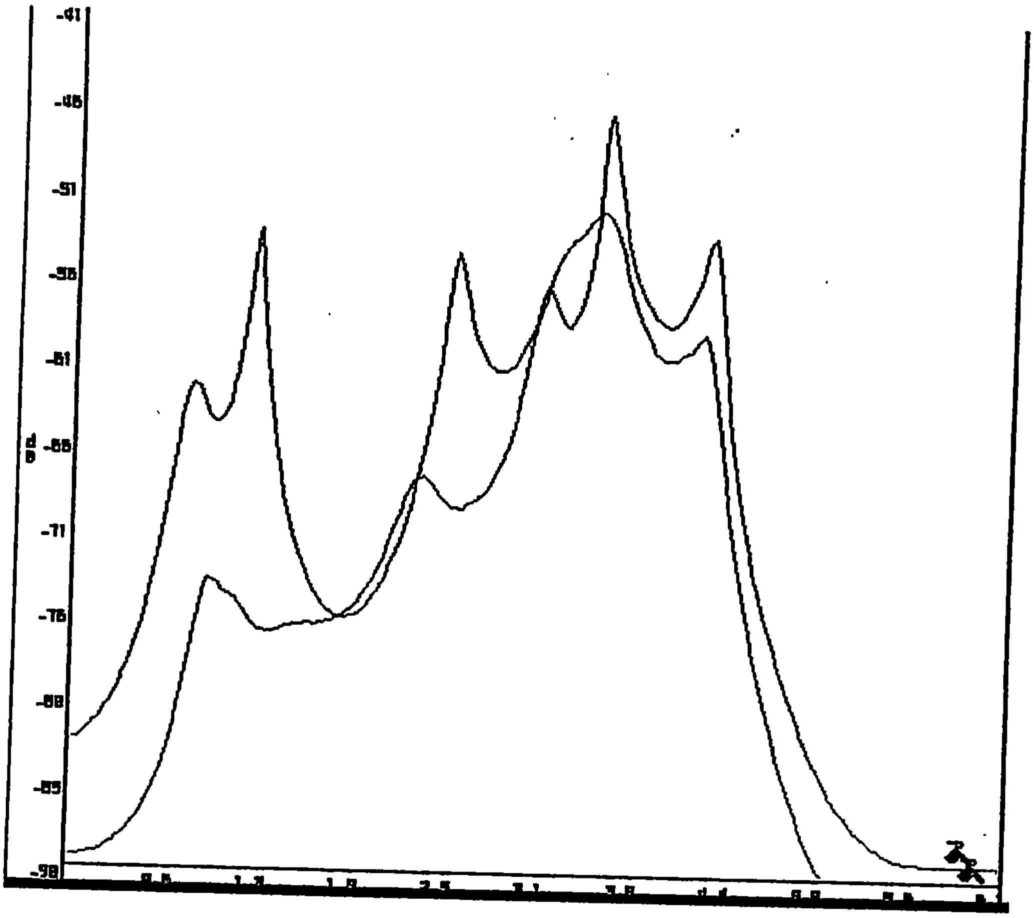


Figure 12. Burst and vowel onset for most neutral stimulus - processed

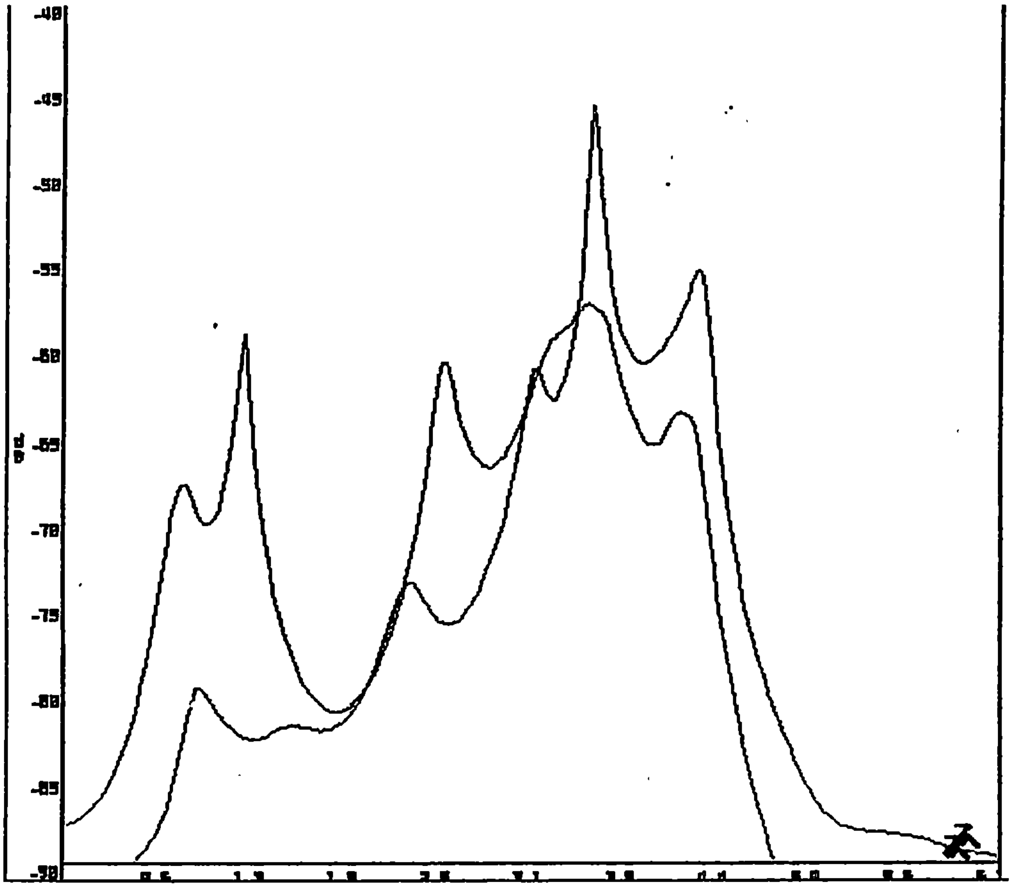


Figure 13. Burst and vowel onset for most /t/ like stimulus - unprocessed

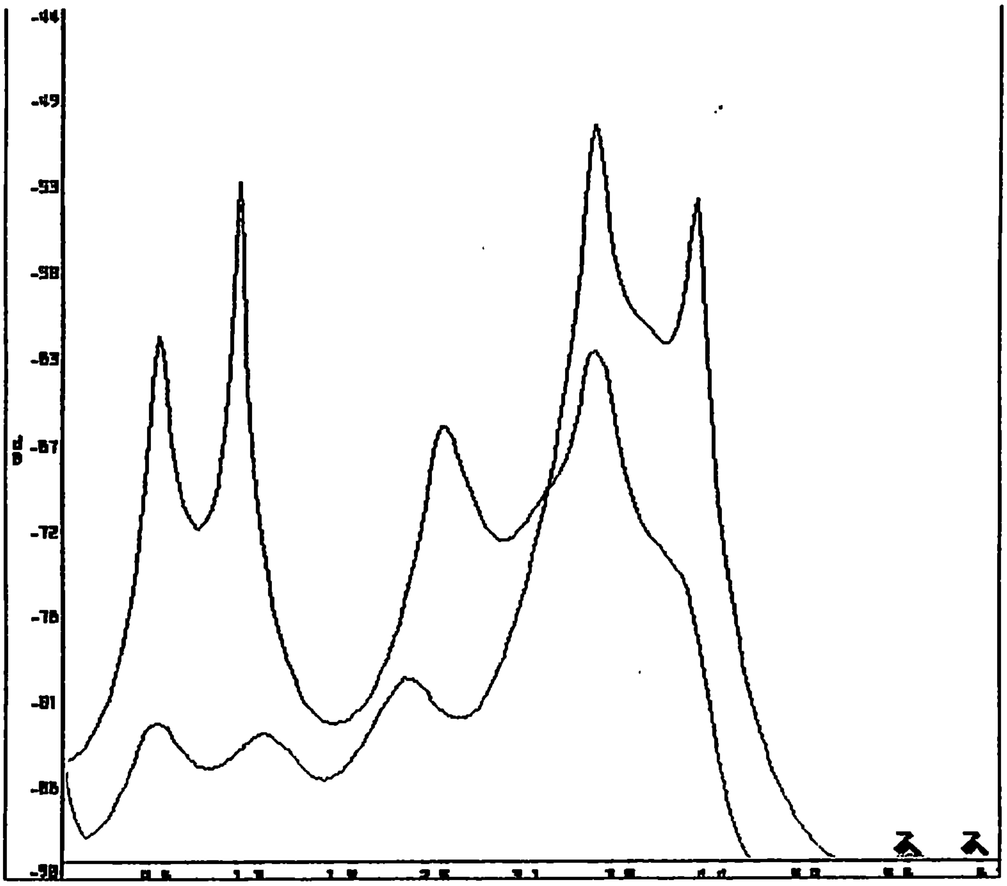


Figure 14. Burst and vowel onset for most /t like stimulus - processed

Again the processed stimuli shows increases in the amplitude for both the consonant and the vowel for formants 1-5. In fact, with the processed stimulus, the burst is now higher in amplitude than the vowel in the high frequency region. Figures 13 and 14 show the most /t/ like stimuli both unprocessed and processed. A comparison of the unprocessed stimuli to the processed stimuli clearly shows increases in the amplitude for the consonant and the vowel for formants 1-5.

Perceptual Results

For 15 of the 25 subjects there was a statistically significant difference in perception of stop consonant place of articulation. For each listener, a t-test (see Table 1) was performed to determine if there was difference between the psychometric functions obtained via processed and unprocessed stimuli. There were 3 trends in the data: 10 of the 25 subjects did not show any difference in perception of stop consonant place of articulation after being processed by the hearing aid. 8 of the 25 subjects showed more /t/ than /p/ and the remaining 7 showed more /p/ than /t/. Examples of these trends are shown in Figures 15-17. Figure 15 shows a subject that chose more /p/ responses under the

Table 1. T test showing significance of change for each subject

SUBJECT NUMBER	t value	Sig. (2-tailed)
1	-4.024	.004*
2	-4.585	.002*
3	5.587	.001*
4	-6.174	.000*
5	4.209	.003*
6	-.839	.426
7	-1.941	.088
8	-1.512	.169
9	1.540	.162
10	3.221	.012*
11	2.949	.018*
12	5.246	.001*
13	1.663	.135
14	-.873	.408
15	-3.200	.013*
16	-2.786	.024*
17	.665	.525
18	8.232	.000*
19	5.030	.001*
20	-.430	.679
21	1.706	.126
22	-1.809	.108
23	-2.800	.023*
24	3.833	.005*
25	-5.349	.001*

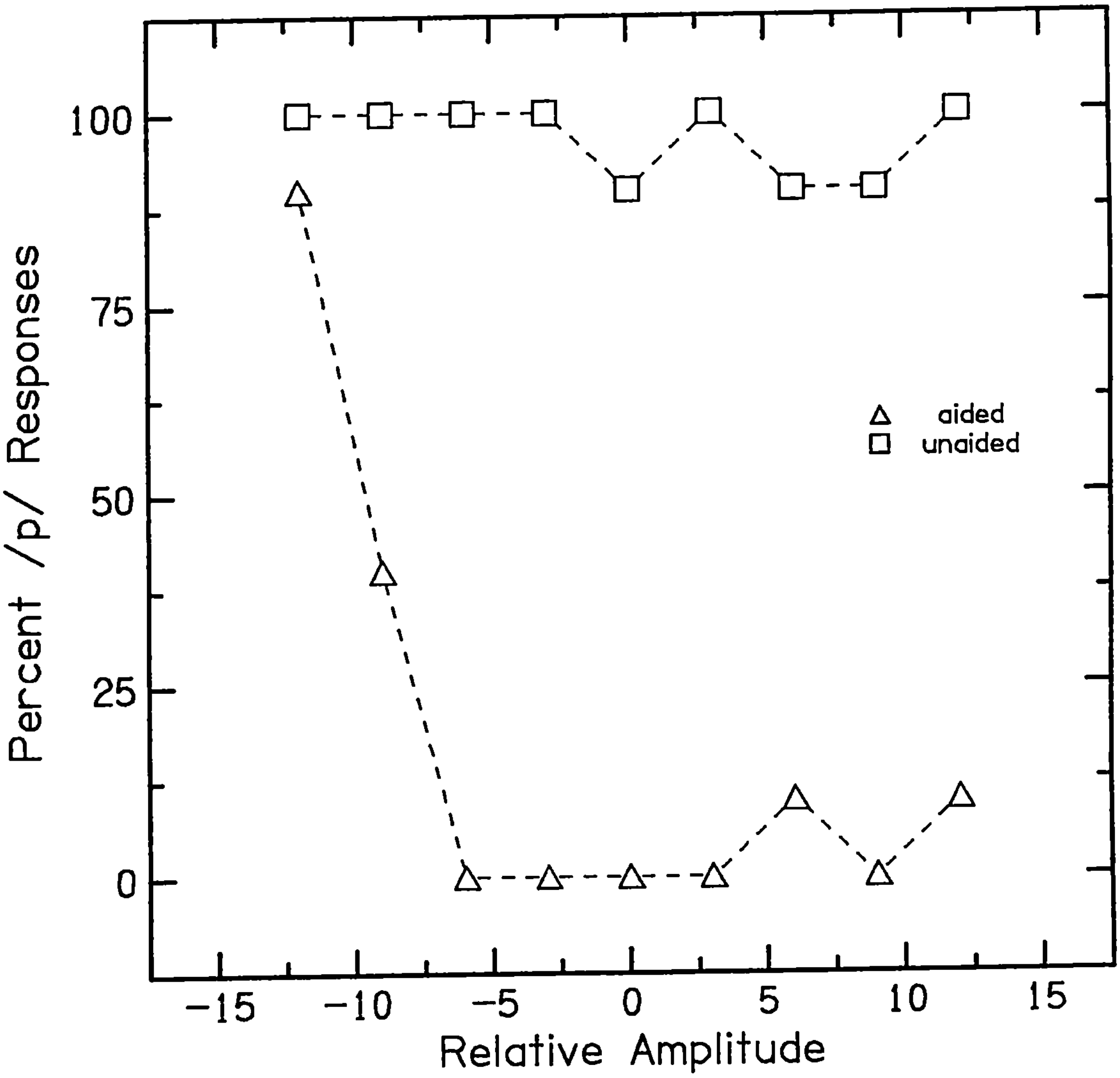


Figure 15. More /p/ responses chosen under processed stimuli as a function of relative amplitude - subject #18

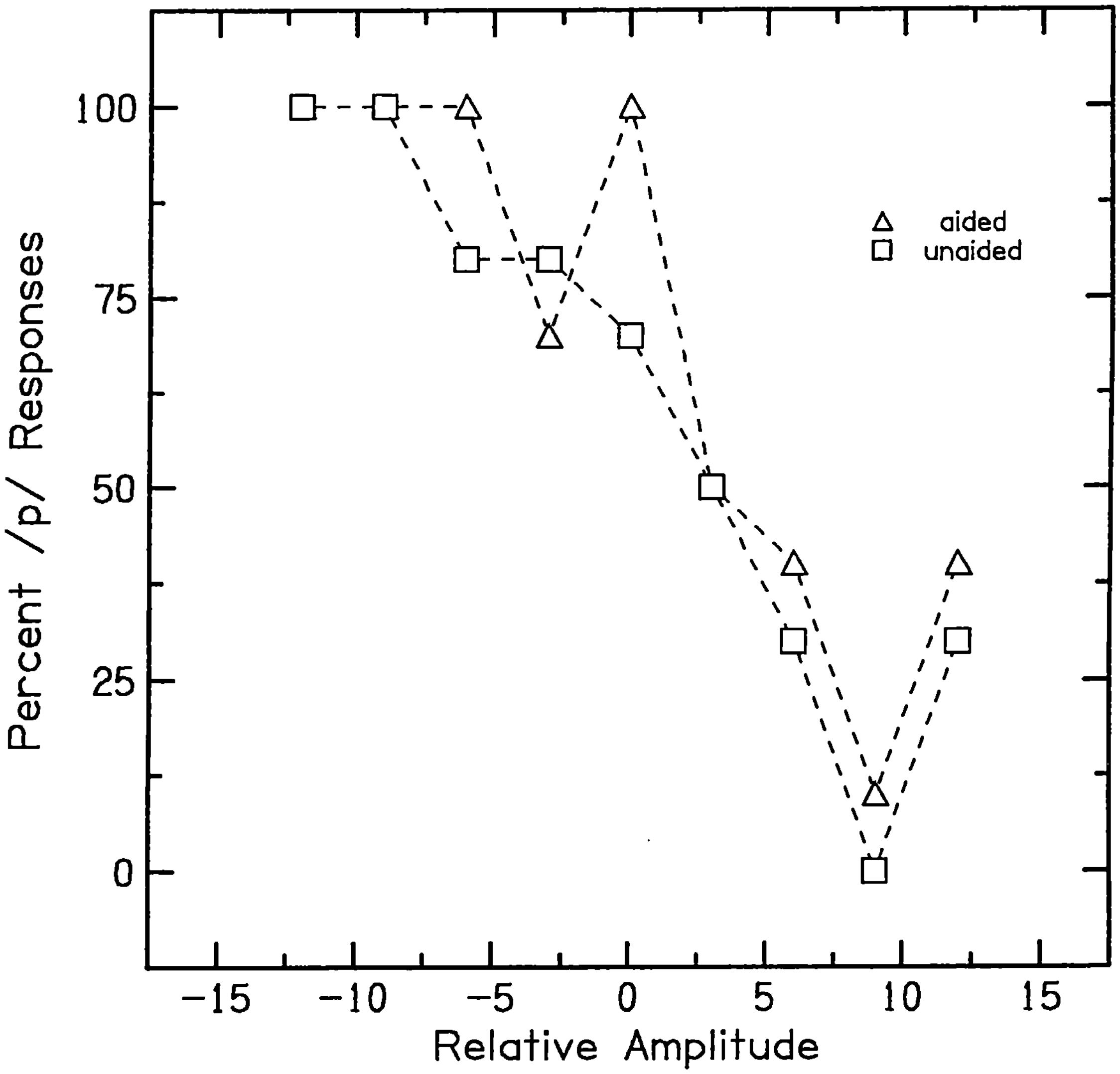


Figure 16. Similar changes in perception between processed and unprocessed stimuli as a function of relative amplitude - subject #7

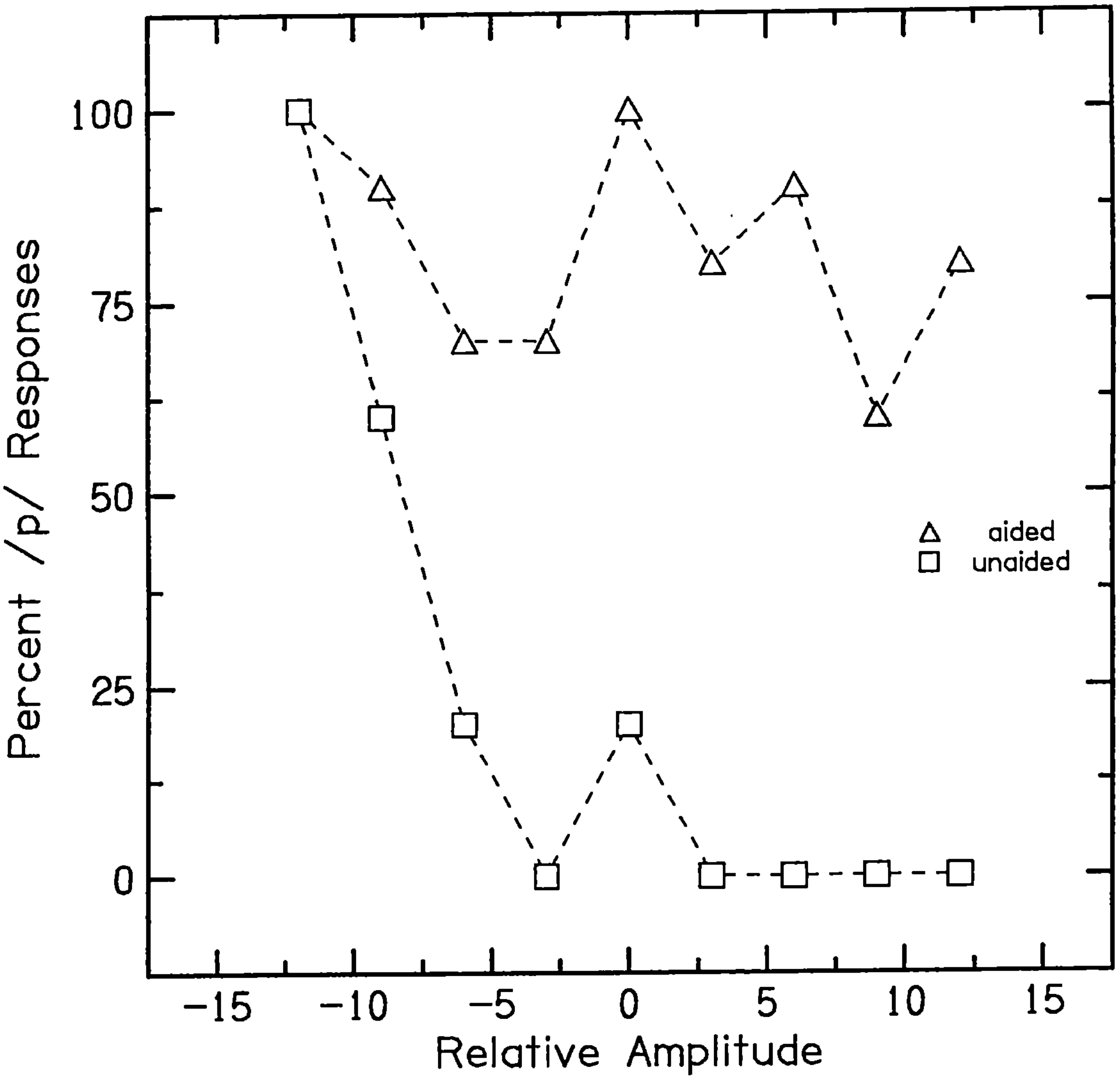


Figure 17. More /t/ responses chosen under processed stimuli as a function of relative amplitude - subject #4

processed condition. Figure 16 shows a subject that perceived no change in stimuli after being processed. Figure 17 shows a subject that perceived more /t/ like sounds under the processed condition.

CHAPTER V

Discussion

This research intended to answer the following questions:

1) Does Wide Dynamic Range Compression alter listeners' perception of place of articulation when relative amplitude is manipulated synthetically?

2) Does the Quantal Theory explain any differences found in listeners' perception of place of articulation?

Specifically, would listeners' perceive more /t/ than /p/ when listening to the processed stimuli. Steven's Quantal Theory explains the 8 of the 25 subjects who chose more /t/ than /p/ responses after being processed by the hearing aid (Ohde & Stevens, 1983; Gravel & Ohde, 1983). Therefore, Stevens' Quantal Theory only explains 31% of the perceptual differences. The three trends that were discovered by this research indicate that there are three groups of listeners that use different cues to discriminate speech.

The processed stimuli shows that wide dynamic range compression increases the amplitude of the consonant relative to the vowel, exaggerates the shape of the consonant to a frequency damped shape, and increases the amplification in the higher frequencies.

Plausible explanations for the three trends found in this project are as follows:

1) Subjects which perceived no difference between the unaided and the aided stimuli perhaps were listeners who focus more on formant transition; therefore, any changes in burst amplitudes were not attended to.

2) Subjects who chose more /t/ than /p/ would follow the hypothesis made in our study based on Steven's Quantal Theory which states that certain changes in amplitude ratio may cause perception of certain consonants to change and thus cause perceptual errors.

3) Persons who chose more /p/ than /t/ responses may have cued in to the frequency damped shape of the consonant which is similar to a naturally produced /p/ (Ladefoged, 1975).

The above explanations are described from research that has shown individual differences in the perceptual weights given to various cues in an acoustical signal (Christensen and Humes, 1996, 1997). Stevens' Quantal Theory partially explains our results, in that the listeners who chose more /t/ than /p/ cued in to the relative amplitude changes. Steven's Quantal Theory suggests there are certain regions that cause large changes in the acoustic perception when there are small changes in articulatory motion. This being the case, the changes caused by the hearing aid resulted in changes in a particular region that a listener used to perceive the consonant.

In looking at research similar to this project, one finds that results indicate that a manipulation that would tamper with the amplitude of the burst relative to the vowel might further confuse the cues leading to correct discrimination of the syllable (Hedrick, Schulte & Jesteadt, 1995). Sammeth, Dorman and Stearns (1999) recently conducted a study to determine the effects of consonant/vowel (CV) ratio on the effects of perception of place of articulation in the presence of background noise for hearing impaired individuals. Although Sammeth, Dorman and Stearns (1999) did not use synthetic speech stimuli nor did they process the speech through a hearing aid, they found that altering the CV ratio degraded the perception of place of articulation for stop consonants.

Clinical and Theoretical Implications

In looking at research by Hedrick, Schulte and Jesteadt (1995)and Sammeth, Dorman and Stearns (1999)one would expect that hearing impaired individuals would have more perceptual problems when listening to speech that is being processed by an amplification device that alters CV ratio. An important area would be to explore various hearing loss configurations to determine if there are differences between groups and to see if the three trends carry over from the normal hearing to the hearing impaired group. If this is the case then perhaps the cues that are used to discriminate speech could be discovered early and help with aural rehabilitation.

It is also important to begin exploring how perception differs when not only a single syllable is being altered but the entire conversation is being processed. There are many patterns and cues that listeners' use to recognize English phonemes while being presented in a conversation; therefore, it is important to determine the effects of processing for a string of speech stimuli.

Future studies should include projects looking at various consonants and vowels to compare the differences in perception of place of articulation. Perhaps stop consonants only show a small portion of changes that take place under amplified conditions. Freyman and Nerbonne (1989) reported, for example, that recognition and CV ratio were negatively correlated for /f/ and /θ/, but positively correlated for /s/ and /ʃ/. Balakrishnan, Freyman, Chiang, Nerbonne, and Shea (1996) found that for different phonemes there may be certain CV ratios that are more optimal for correct perception of place of articulation.

Other research could include looking at various types of amplification circuits. One may find that digital amplification devices provide more accurate cues to correctly discriminate, thus being another cause to focus more on turning to new technological devices.

Indeed it seems a wide dynamic range compression hearing aid causes changes in relative amplitude of the consonant in relation to the vowel. This being the case, a person being fit with an amplification device may need strategies to relearn cues in order to correctly discriminate speech stimuli.

REFERENCES

REFERENCES

ANSI (**1987**). ANSI S3.22-1987, "Specifications for hearing aids," (American National Standards Institute, New York).

ANSI (**1989**). ANSI S3.6-1989, "Specifications for audiometers," (American National Standards Institute, New York).

Balakrishnan, U., Freyman, R. L., Chiang, Y. C, Nerbonne, G. P., and Shea, K. J. (**1996**). "Consonant recognition for spectrally degraded speech as a function of consonant-vowel intensity ratio," J. Acoust. Soc. Am. **99**, 3758-3769.

Boothroyd, A., Springer, N., Smith, L., and Schulman, J. (**1988**). "Amplitude compression and profound hearing loss," J. Speech Hear. Res. **31**, 362-376.

Borden, G., Harris, K., and Raphael, L. (**1994**). Speech science primer: physiology, acoustics, and pereption of speech. Baltimore: Williams & Wilkens.

Christensen, L., and Humes, L. (**1996**). "Identification of multidimensional complex sounds having parallel dimension structure," J. Acoust. Soc. Am. **99**, 2307-2315.

Christensen, L., and Humes, L. (**1997**). "Identification of multidimensional stimuli containing speech cues and the effects of training," J. Acoust. Soc. Am. **102**, 2297-2310.

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

Fant, G. (**1960**). Acoustic theory of speech production. The Hague: Mouton.

Freyman, R. L., Nerbonne, G. P. (**1989**). "The importance of consonant-vowel intensity ratio in the intelligibility of voiceless consonants," J. Speech Hear. Res. **32**, 415-426.

Gravel, J., and Ohde, R. (**1983**). "Perception of stop place of articulation: Effects of stimulus amplitude" Am. Speech Lang. Hear. Assoc, **25**, 101.

Hawkins, S., and Stevens, K. N., (**1987**). "Perceptual and acoustical analysis of velar stop consonants." Perception and Psychophysics **32**, 409-418.

Hedrick, M. S., and Ohde, W. (**1993**). "Effect of relative amplitude of frication on perception of place of articulation," J. Acoust. Soc. Am. **94**, 2005-2026.

Hedrick, M. S., Schulte, L., and Jesteadt, W. (**1995**). "Effect of relative amplitude of frication on perception of place of articulation." J. Acoust. Soc. Am. **98**, 1292-1303.

Kennedy, E., Levitt, H., Neuman, A. C., and Weiss, M. (**1998**). "Consonant-vowel intensity ratios for maximizing consonant recognition by hearing-impaired listeners." J. Acoust. Soc. Am. **103**, 1098-1114.

Kent. R. D., and Read, C., (**1992**). The acoustic analysis of speech. Singular: McNaughton and Gunn.

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

Ladefoged, P. (**1975**). A course in phonetics. New York: Harcourt Brack Jovanovich, INC.

Montgomery, A. A., and Edge, R. A. (**1988**). "Evaluation of two speech enhancement techniques to improve intelligibility for hearing-impaired adults," J. Speech Hear. Res. **31**, 386-393.

Moore B. C. J., and Glasberg, B. R. (**1986**). "A comparison of two-channel and single channel compression hearing aids," Audiology **25**, 210-226.

Ohde, R. N. (**1979**). "Acoustic characteristics of stops in CV syllables," paper presented at the convention of the American Speech-Language-Hearing Association in Atlanta, Georgia.

Ohde, R. N., and Stevens, K. N. (**1983**). "Effect of burst amplitude on the perception of stop consonant place of articulation," J. Acoust. Soc. Am. **74**, 706-714.

Plomp, R. (**1988**). "The negative effect of amplitude compression in multi-channel hearing aids in the light of the modulation transfer function," J. Acoust. Soc. Am. **83**, 2322-2327.

Stevens, K. N., and Blumstein, S. E., (**1981**). "The search for invariant acoustic correlates of phonetic features," Perspectives in the study of speech. Hillsdale: Erlbaum.

Stevens, K. N. (**1985**). "On the quantal nature of speech," J. Phonetics **17**, 3-45.

ter Keurs, M., Festen, J. M., and Plomp, R. (**1992**). "Effect of spectral envelope smearing on speech perception II," J. Acoust. Soc. Am. **93**, 1547-1552.

Tobin, H., and Dempsey, J. (**1997**). "Practical hearing aid selection and fitting." Rehabilitation Research and Development Service. Baltimore.

Van Tasell, D.J., Soli, S. D., Kirby, V. M., and Widin, G. P. (**1987**). "Speech waveform envelope cues for consonant recognition," J. Acoust. Soc. Am. **82**, 1152-1161.

Zue, V. (**1976**). "Acoustic characteristics of stop consonants: A controlled study, unpublished Sc. D. thesis, M.I.T. (A revised version was published as M.I.T., Lincoln Laboratory Technical Representative NO.523).

APPENDICES

APPENDIX A

PARAMETER TIME AND VALUE FOR FORMANT SYNTHESIZER

The following table corresponds to the most /p/ like stimulus. The table illustrates how a given parameter was varied over time. The following is a list of control parameters for the software formant synthesizer: AV=Amplitude of voicing (dB), AF= Amplitude of frication (dB), AH=Amplitude of aspiration (dB), F0=Fundamental frequency of voicing (Hz), F1=First formant frequency (Hz), F2=Second formant frequency (Hz), F3=Third formant frequency (Hz), F4=Fourth formant frequency, A4=Fourth formant amplitude (dB), A5=Fifth formant amplitude (dB), B1=first formant bandwidth (Hz), B2=Second formant bandwidth (Hz), B3=Third formant bandwidth (Hz), F5=Fifth formant frequency (Hz).

Time (ms)	AV	AF	AH	F0	F1	F2	F3	F4	A4	A5	B1	B2	B3	F5
0	0	65	0	0	400	1220	2600	3500	10	10	300	150	220	4200
10	-	-	20	-	-	-	-	-	-	-	-	-	-	-
20	-	55	40	-	400	1220	2600	-	-	-	300	150	220	-
25	-	0	50	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50	-	-	50	-	-	-	-	-	10	10	-	-	-	-
55	0	-	-	0	-	-	-	-	-	-	-	-	-	-
60	40	-	0	130	700	1220	2600	-	0	0	130	70	160	-
260	55	0	0	100	700	1220	2600	3500	0	-	130	70	160	4200

The following table corresponds to the most neutral stimulus. The table illustrates how a given parameter was varied over time. The following is a list of control parameters for the software formant synthesizer: AV=Amplitude of voicing (dB), AF= Amplitude of frication (dB), AH=Amplitude of aspiration (dB), F0=Fundamental frequency of voicing (Hz), F1=First formant frequency (Hz), F2=Second formant frequency (Hz), F3=Third formant frequency (Hz), F4=Fourth formant frequency, A4=Fourth formant amplitude (dB), A5=Fifth formant amplitude (dB), B1=first formant bandwidth (Hz), B2=Second formant bandwidth (Hz), B3=Third formant bandwidth (Hz), F5=Fifth formant frequency (Hz).

Time (ms)	AV	AF	AH	F0	F1	F2	F3	F4	A4	A5	B1	B2	B3	F5
0	0	65	0	0	400	1220	2600	3500	32	32	300	150	220	4200
10	-	-	20	-	-	-	-	-	-	-	-	-	-	-
20	-	55	40	-	400	1220	2600	-	-	-	300	150	220	-
25	-	0	50	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50	-	-	50	-	-	-	-	-	32	32	-	-	-	-
55	0	-	-	0	-	-	-	-	-	-	-	-	-	-
60	40	-	0	130	700	1220	2600	-	0	0	130	70	160	-
260	260	0	0	100	700	1220	2600	3500	-	-	130	70	160	4200

The following table corresponds to the most /t/ like stimulus. The table illustrates how a given parameter was varied over time. The following is a list of control parameters for the software formant synthesizer: AV=Amplitude of voicing (dB), AF= Amplitude of frication (dB), AH=Amplitude of aspiration (dB), F0=Fundamental frequency of voicing (Hz), F1=First formant frequency (Hz), F2=Second formant frequency (Hz), F3=Third formant frequency (Hz), F4=Fourth formant frequency, A4=Fourth formant amplitude (dB), A5=Fifth formant amplitude (dB), B1=first formant bandwidth (Hz), B2=Second formant bandwidth (Hz), B3=Third formant bandwidth (Hz), F5=Fifth formant frequency (Hz).

Time (ms)	AV	AF	AH	F0	F1	F2	F3	F4	A4	A5	B1	B2	B3	F5
0	0	65	0	0	400	1220	2600	3500	43	43	300	150	220	4200
10	-	-	20	-	-	-	-	-	-	-	-	-	-	-
20	-	55	40	-	400	1220	2600	-	-	-	300	150	220	-
25	-	0	50	-	-	-	-	-	-	-	-	-	-	-
30	-	-	-	-	-	-	-	-	-	-	-	-	-	-
40	-	-	-	-	-	-	-	-	-	-	-	-	-	-
50	-	-	50	-	-	-	-	-	43	43	-	-	-	-
55	0	-	-	0	-	-	-	-	-	-	-	-	-	-
60	40	-	0	130	700	1220	2600	-	0	0	130	70	160	-
260	55	0	0	100	700	1220	2600	3500	-	-	130	70	160	4200

APPENDIX B

ANSI S3.22-1987 STANDARD FOR HEARING AID

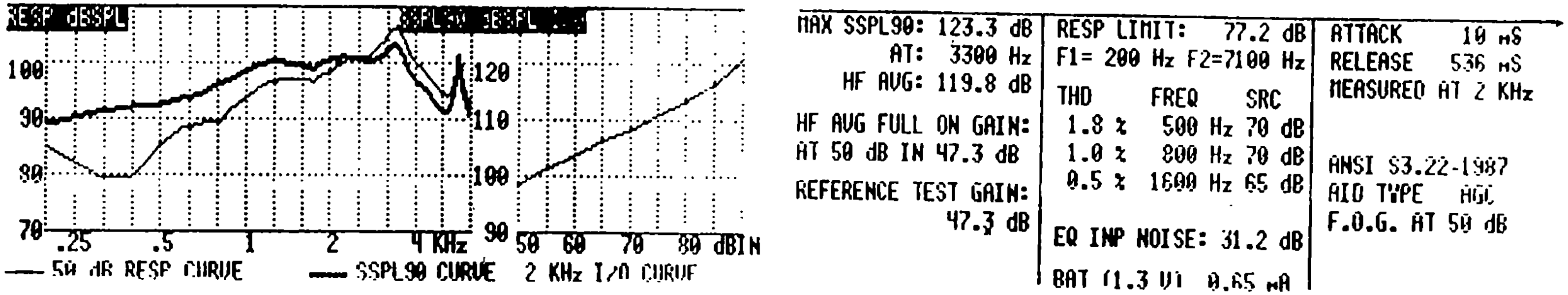
ANSI Technical Data

Unitron IKON K

Frequency Range		
L (Low-cut Tone)	-N-	430-6800 Hz
	-H-	600-7000 Hz
H (High-cut Tone)	-N-	430-6800 Hz
	-H-	270-6300 Hz
Reference Test Gain		48 dB
HF-Avg. Gain 50 dB in		48 dB
HF-SSPL90 Output		121 dB
Typical Battery Life		285 h
Current Drain at RTP		.8 mA
Output with Inductive Input		
at 1000 Hz		100 dB
Equivalent Input Noise		
typical 26 dB		<29 dB
Typical Harmonic Distortion at RTP		
500 Hz	not applicable	
800 Hz	typical 3%	<6%
1600 Hz	typical 2%	<6%
Attack Time		<30 ms
Release Time		700 ms
Compression Ratio		2:1

ANSI S3.22-1987

Electroacoustic Analysis of Unitron IKON K



```
MAX SSPL90: 123.3 dB   RESP LIMIT:     77.2 dB   ATTACK        10 mS
       AT:  3300 Hz    F1= 200 Hz F2=7100 Hz     RELEASE      536 mS
   HF AVG: 119.8 dB                              MEASURED AT 2 KHz
                       THD      FREQ     SRC
HF AVG FULL ON GAIN:    1.8 %    500 Hz 70 dB
AT 50 dB IN 47.3 dB     1.0 %    800 Hz 70 dB    ANSI S3.22-1987
REFERENCE TEST GAIN:    0.5 %   1600 Hz 65 dB    AID TYPE   AGC
             47.3 dB                             F.O.G. AT 50 dB
                       EQ INP NOISE: 31.2 dB

                       BAT (1.3 V)  0.65 mA
```

VITA

Tracie Rice was born in Franklin, North Carolina on March 4, 1976. She received her Bachelor of Science degree with a double major in Communication Disorders and Sociology from Western Carolina University in Cullowhee, North Carolina. In May 1999, she received a Master of Arts degree in Audiology from the University of Tennessee - Knoxville. She will complete a Clinical Fellowship Year following graduation.