



8-1990

Acoustic Characteristics of Vowels Produced by Men, Women and Children

Laura Arlene Getty
Western Michigan University

Follow this and additional works at: https://scholarworks.wmich.edu/masters_theses



Part of the Speech Pathology and Audiology Commons

Recommended Citation

Getty, Laura Arlene, "Acoustic Characteristics of Vowels Produced by Men, Women and Children" (1990). *Masters Theses*. 1042.

https://scholarworks.wmich.edu/masters_theses/1042

This Masters Thesis-Open Access is brought to you for free and open access by the Graduate College at ScholarWorks at WMU. It has been accepted for inclusion in Masters Theses by an authorized administrator of ScholarWorks at WMU. For more information, please contact wmu-scholarworks@wmich.edu.



ACOUSTIC CHARACTERISTICS OF VOWELS
PRODUCED BY MEN, WOMEN
AND CHILDREN

by

Laura Arlene Getty

A Thesis
Submitted to the
Faculty of The Graduate College
in partial fulfillment of the
requirements for the
Degree of Masters of Arts
Department of Speech Pathology and Audiology

Western Michigan University
Kalamazoo, Michigan
August 1990

ACOUSTIC CHARACTERISTICS OF VOWELS
PRODUCED BY MEN, WOMEN
AND CHILDREN

Laura Arlene Getty, M.A.

Western Michigan University, 1990

This investigation was a replication and extension of Peterson and Barney's (1952) research on the acoustics and perception of vowels. This study explored the acoustic characteristics of 12 vowels produced by 50 men, 50 women and 41 children. Speech samples were tape recorded as subjects read lists of words and isolated vowels. Speech samples were digitized. Speech waveforms, spectrograms and formant frequencies were examined and measured.

The average formant frequency values were similar but not identical to the values reported by Peterson and Barney (1952). The largest discrepancy between the two studies was the degree of overlap among vowel categories, which was much larger in the present study.

ACKNOWLEDGEMENTS

I wish to extend my appreciation and gratitude to the following people who contributed to this study.

First, I thank Dr. James M. Hillenbrand for making speech perception comprehensible and interesting. Working under his direction, I have gained an appreciation for the the field of speech science. His guidance, suggestions, sense of humor, extreme patience, words of encouragement and perceptive abilities were greatly appreciated throughout a project that gave new meaning to the phrase "research project."

I thank committee member Dr. Michael J. Clark, who gave so many hours to this project. I am fortunate to have had access to his critiquing skills as a phonotician and as a researcher. I thank committee member Dr. John M. Hanley whose skills as a researcher have been an asset to this project. His sincere words of encouragement were greatly appreciated.

I am indebted to Miss Kimberlee Boody for her long hours of dedication to this research project. I could not have completed such a task without her help, friendly advice, and quick wit on days when it was a true effort to work. Thanks also to Miss JoAnne Gladd, Mrs. Jenifer Smith and Miss Anne Wanzeck who helped me to achieve my

Acknowledgements--Continued

goal.

Thanks go to Mr. William Dawson who simultaneously solved several equipment problems, answered my questions, and pointed to the proverbial light at the end of the tunnel when I so often lost sight of it.

I extend thanks to the teachers and students at Parkwood-Upjohn Elementary School and The Gage School, Kalamazoo, Michigan, for working with me in the collection of speech samples. Also, thanks go to the men and women who provided speech samples to make this project possible.

I wish to thank Western Michigan University for contributing financial support to this project. I am also grateful to those who, in one way or another, helped me through this project. To list all names here is impossible, but each of you has my special thanks.

Finally, I will be forever grateful to my family for their constant emotional support, encouragement and laughter. I thank my parents, Samuel and Lucy, for giving me words, a hunger for education, and ambition, rather than silver platters, gold spoons and engraved invitations. I also thank my sisters, Susan, Elizabeth, Lois and Dorothy, for bringing me down to earth when I needed it; there's nothing like a sister!

Laura Arlene Getty

INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.



University Microfilms International
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600

Order Number 1341559

Acoustic characteristics of vowels produced by men, women and children

Getty, Laura Arlene, M.A.

Western Michigan University, 1990

U·M·I
300 N. Zeeb Rd.
Ann Arbor, MI 48106

TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	ii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
CHAPTER	
I. INTRODUCTION.....	1
History of Phonetic Perception.....	1
Phonetic Perception Research.....	2
Purpose of the Investigation.....	3
II. REVIEW OF RELATED LITERATURE.....	5
Considerations of Speech Perception.....	5
The Phonetic Perception Problem.....	5
Theories of Speech Perception.....	7
Characteristics Influencing Perception..	15
Formants and Vowel Identification.....	25
III. DESIGN.....	28
Limitations of Peterson and Barney.....	28
IV. METHODS.....	32
An Overview.....	32
Subject Participation.....	32
Talkers.....	32
Criteria for Subject Participation.....	33
Collection of the Data.....	34
Screening Procedure.....	34

Table of Contents--Continued

CHAPTER

	Determining Subject Eligibility.....	35
	Collection of the h-V-d Stimuli.....	37
	Acoustic Analysis.....	40
	Digitizing the Signals.....	40
	Formant Frequency Measurement.....	41
V.	RESULTS.....	45
	Presentation of the Data.....	45
	Formant Frequency Averages.....	45
	The Data in an F1-F2 Matrix.....	48
	Vowel Duration Data.....	50
VI.	DISCUSSION.....	52
	Comparison of the Investigations.....	52
	Vowel Duration.....	56
	Proposed Future Studies.....	57
APPENDICES		
A.	Peterson and Barney Listening Data.....	61
B.	Adolescent Voice Change.....	63
C.	Screening Procedure.....	65
D.	Reading Lists.....	69
E.	Original FTRACK Display of / ϵ /.....	82
F.	Edited FTRACK Display of / ϵ /.....	84
G.	Gray-Scale Spectrogram of / ϵ /.....	86
H.	Formants Prior to Interpolation.....	88

Table of Contents--Continued

APPENDICES

I. Formants Following Interpolation.....	90
J. Vowel Data of /i/, /I/ and /ɛ/.....	92
K. Vowel Data of /ɛ/, /æ/ and /ʌ/.....	94
L. Vowel Data of /ʌ/, /ɑ/ and /ɔ/.....	96
M. Vowel Data of /ɔ/, /U/ and /u/.....	98
N. Vowel Data of /u/, /U/ and /ʊ/.....	100
O. Original Productions of /ɛ/ and /æ/.....	102
P. Remeasured Productions of /ɛ/ and /æ/.....	104
Q. LPC Spectra of /æ/.....	106
R. The /æ/ and /ɛ/ Experiment.....	108
S. Consent Forms.....	110
BIBLIOGRAPHY.....	115

LIST OF TABLES

1. Formant Frequency Averages.....	46
2. Vowel Duration Averages.....	51

LIST OF FIGURES

1. Peterson and Barney (1952) Listening Data..... 62

CHAPTER I

INTRODUCTION

History of Phonetic Perception

Phonetic perception has been an area of interest to researchers for several years and an area in which many questions remain unanswered. Scientists studied this problem in the 18th and 19th centuries by generating artificial speech through models of the human respiratory system and vocal tract. It was not until the 20th century that Homer Dudley developed a speech synthesizer that synthesized continuous speech by using electronic circuits (Borden & Harris, 1984).

The sound spectrograph, a machine that gave an output which represented the frequencies of a signal across time, was invented at Bell Telephone Laboratories, Murry Hill, New Jersey, in the 1940s (Potter, Kopp & Green, 1947). The visible representation produced by the machine was called a spectrogram. The development of that system allowed investigators to formulate questions as to what acoustic cues were necessary to identify speech (Borden & Harris, 1984).

The reverse of the sound spectrograph, the Pattern Playback, was the vision of Franklin Cooper at Haskins

Laboratories, New York, New York. Speech pioneers such as Cooper, Liberman and Delattre directed their efforts toward developing a reading machine for the blind. They experimented by creating, or "painting," spectrograms, feeding them into the machine, and listening to the output. The team varied specific characteristics of the hand-painted spectrograms and were thus able to make inferences about the acoustic cues to phonetic categories. This method of studying speech and the related acoustic cues was most effective until the introduction of computer controlled synthesizers (Borden & Harris, 1984).

The introduction of computer synthesized speech in the 1950s has automated the field of speech perception. Computers have allowed speech scientists to view an utterance represented acoustically, play the utterance repeatedly and vary the signal more readily than in the past (Fant & Tatham, 1975).

Phonetic Perception Research

The specific focus of this study was to replicate and extend a previous experiment that explored the perception and acoustic characteristics of vowels. The previous study was conducted by Peterson and Barney (1952) at Bell Telephone Laboratories and is the most widely cited study of vowel perception. Peterson and Barney were interested in the spectral characteristics of vowels and the

relationship between a talker's intended vowel and a listener's perception of that same vowel. They found considerable acoustic variability of the spectra between the same vowel when spoken by men, women and children which they attributed, in part, to differences in vocal tract length. This finding of variability was not surprising since it was known that men tend to have relatively low formant frequencies, children have higher formant frequencies and women in between (Potter & Steinberg, 1950). This finding suggested that absolute formant frequencies probably do not play a primary role in the perception of vowels. Peterson and Barney also found overlap between vowel categories; vowels that were part of one vowel category occasionally were found in a region of acoustic space that was dominated by an adjacent vowel.

Purpose of the Investigation

The present study served two purposes. The first purpose will be to replicate and extend the speech production part of the Peterson and Barney (1952) study. This study will provide entire consonant-vowel-consonant (CVC) spectrograms and measurements of those data. The second purpose of the investigation will be the formation of an acoustic library for a series of resynthesis experiments to follow up on the findings of Hillenbrand and McMahon (1988) and Hillenbrand and Gayvert (1988).

These studies will be discussed later. The current study will investigate vowels, but the research problems associated with vowel identification might be similar for other classes of speech sounds. Spectrograms from the Peterson and Barney study are unpublished and, therefore, unavailable to other investigators. Therefore, this research was necessary to meet a need for the actual data.

CHAPTER II

REVIEW OF RELATED LITERATURE

Considerations of Speech Perception

The Phonetic Perception Problem

Several speech perception problems have been identified but remain unresolved. Klatt (1980) discussed four problems that play a role in the identification of phonemes: (1) invariance, (2) segmentation, (3) time normalization, and (4) talker normalization. Investigators have dealt with the problem of invariance, or how listeners identify context dependent cues, especially when the context is unknown. A phonetic segment varies depending on the particular phonetic environment. Delattre, Liberman and Cooper (1955, cited in Borden & Harris, 1984) asked listeners to identify two-formant formant CV syllables of /b/, /d/ and /g/. The researchers found that listeners identified the stimuli based on the changing formant pattern that occurred at the beginning of each syllable. For example, listeners differentiated /di/ from /du/. The former had two ascending formants and the latter had an ascending first formant and a descending second formant. Although the

formant transitions for /di/ and /du/ were quite different, listeners identified them as equally good examples of /d/. The auditory basis of this perceived similarity, however, is not clear.

Segmentation, a second problem discussed by Klatt (1980), involves the grouping of phonetic units for listener identification. In speech there are ambiguous segment boundaries when the articulators move asynchronously, which often occurs. Segmentation creates problems when duration measurements of a syllable are attempted since the cues that signal the beginning and ending of syllables, or phonetic segments, are often ambiguous. It is generally assumed that phonetic recognition involves some sort of pattern modeling process. However, given the ambiguities regarding syllable and segment boundaries, it is not clear how the time intervals over which the pattern matching occurs are determined.

Time normalization plays a role in the identification of vowels. The time normalization problem can be divided into two sub-problems. First, speech scientists need to learn how the auditory system ignores the irrelevant variations in segmental durations. Second, researchers need to learn how to incorporate durational information in segmental decision-making when durational perturbations due to syntax, semantics and stress are unknown. The

duration of the phonetic event can play a role in distinguishing between certain phonemes such as /æ/ - /ɛ/ and /s/ - /z/. Therefore, it becomes necessary to devise a method that will take into consideration the role that duration plays in the identification of phonemes when attempting to distinguish them.

Finally, talker normalization refers to the variability in the acoustic pattern associated with a particular speech sound from one individual to another. Talkers differ in length and shape of the vocal tract, fundamental frequency, speaking rate, dialect and other acoustic features related to the laryngeal source. Listeners have relatively little difficulty adjusting to inter-talker variation but the auditory basis of this ability is not well understood.

Theories of Speech Perception

Lloyd (cited in Miller, 1989) proposed the formant-ratio theory of vowel perception. He suggested that like vowel qualities could be associated with like ratios of the frequencies of their resonances. The theory developed following experimentation with synthetically produced vowels generated by exciting hand-blown glass tubes and by conducting studies of transposition done by slowing and speeding Ediphone recordings. The theory was also based on the frequency ratios in musical perception and on

differences in vowels naturally produced by men, women and children. Lloyd realized his formant-ratio theory was erroneous after Pipping (cited in Miller, 1989) noted that certain vowels have common ratios of second and first resonant frequencies ($F2/F1$) and that transposition only works over certain limits. Pipping also stated that some vowels appear to have one formant rather than two and that Lloyd's formulas for calculating the resonances of the tubes were in error.

Most recently, Miller (1989) proposed a three dimensional model ($\log F3 - \log F2$, $\log F2 - \log F1$, $\log F1 - \log F0$), entitled the auditory-perceptual theory, in an attempt to deal with the weaknesses of Lloyd's formant-ratio theory. Miller described a three stage process in which a waveform is transformed into a series of codes by the listener. In the first stage the speech waveform is transformed into the auditory-sensory dimension. Miller believed that the waveform is classified as either a glottal source spectrum, a burst-friction spectrum or a combination of both. The spectral envelope patterns are then represented as sensory responses within a phonetically relevant auditory-perceptual space.

Stage two of the process involves conversion of the sensory responses into one perceptual response which is also located within the same space. Miller (1989) stated that the perceptual response is a hypothetical construct

and is calculated by mathematically derived sensory-perceptual transformations. Such transformations are dependent upon histories, trajectories and dynamics of the sensory responses.

In the third stage of the process, segmentation and categorization mechanisms, which rely upon the dynamics of the sensory response, cause perceptual target zones to yield a series of codes which correspond to the allophones of the language. Miller (1989) concluded that the preliminary target zones can classify American English vowels with 93% accuracy.

More recently Strange (1989a) presented theories of vowel perception in an attempt to show how researchers have characterized vowels as articulatory and acoustic events. She described the "simple target" model of vowel perception. The theory considers the articulatory, acoustic and perceptual characterizations of the vowel as a unified concept known as the "vowel target." Articulatorily, vowel targets are canonical forms of the vowels. They are best represented by the production of a vowel sound in a static vocal tract position. Acoustically vowel targets are represented in a multidimensional space such as F1-F2 or F1-F2-F3.

Two problems arise when the perception of a simple target model of the vowel is considered. First, variation between talkers has been proven (Peterson & Barney, 1952),

thus rejecting the notion of the simple target model of production. Second, when a CVC syllable is produced at a normal to rapid rate of speech the vowel target is often not reached. This problem, also known as "target undershoot" may be affected by phonetic context, stress, rate of speech, and talker specific coarticulation strategies.

Strange (1989a) also described two vowel perception models which dominated the 1970s and 1980s. She referred to these as the "elaborated target" and "dynamic specification" models. Both models consider how to best represent vowels acoustically and perceptually. The elaborated target model focuses on vowels as static points in a multidimensional space. The dynamic specification model emphasizes the role of dynamic sources of information as critical. Such a model attempts to account for talker normalization (Syrdal & Gopal, 1986).

Gerstman (1968) sought to determine the minimum dimensionality and the minimum number of vowels required to classify an individual talker's vowels. He developed a computer algorithm for recognizing vowels which utilized a talker's extreme formant frequencies, usually /i/, /a/ and /u/, to scale all of the talker's vowels. The algorithm used these normalized values to classify the vowels obtained by Peterson and Barney (1952). He reported that the vowels were classified with 97.5% accuracy. He

concluded that all talkers have similar locations for all vowels within a self-normalized space.

Verbrugge, Strange, Shankweiler and Edman (1976) stated that Gerstman's algorithm was not necessarily a perceptual strategy, but rather a logically possible strategy. No evidence has been offered that has indicated that listeners perform the scaling of formants when identifying vowels.

Delattre, Liberman, Cooper and Gerstman (1952) and Fant (1973, both cited in Lieberman & Blumstein, 1988) conducted studies which suggested that vowels could be perceived correctly, for the most part, using only the first two formants. However, Delattre et al. noted that when only the first two formants were synthesized, an "effective second formant," or $F2'$, must be formed which would account for the shape of the spectrum. This altered second formant ($F2'$) would not be necessary if the higher formant frequencies were present since those frequencies contain that information. For example, Cooper, Delattre, Liberman, Borst and Gerstman (1952) synthesized /i/ at 270 Hz and 2700 Hz for $F1$ and $F2$, respectively, rather than 270 Hz and 2290 Hz; the latter figures being formant averages calculated by Peterson and Barney (1952).

Carlson, Fant and Granstrom (1975) found that the best two-formant synthetic Swedish vowels were different from naturally produced vowels and that such a difference

was systematic. For example, when synthesizing /i/ the second formant must be closer to the natural third formant than to the natural second formant. For the non-high vowels the second formant was synthesized between the natural second and third formants. When synthesizing a two-formant model for the back vowels, F2' was placed closer to the natural second formant. This research suggested that the third formant was more important to the perception of front vowels than to back vowels.

In a matching experiment, Carlson et al. (1975) simultaneously presented two signals and asked listeners to match them by adjusting a frequency setting. For each of the three phonetically trained listeners a four-formant stimulus was presented. The first formant of a two-formant stimulus was identical to F1 of the four-formant stimulus. Subjects were instructed to vary F2 of the two-formant stimulus to achieve a phonetic match between the two-formant stimulus and the four-formant stimulus. The investigators reported that subjects' setting of the variable F2 of the two-formant vowel (referred to as F2') was near the natural F2 in mid and back vowels. F2' was between the natural second and third formants for the front vowels excluding /i/ in which case it was in the region of F3 or higher.

Disner (1980) studied vowel normalization procedures to determine which technique was most effective in

reducing "speaker-particular effects." She proposed a quantifiable evaluation procedure which she applied to the vowels of six different Germanic languages. The data had been normalized using four procedures (Gerstman, 1968; Harshman, 1970; Lobanov, 1971; and Nearey, 1977, all cited in Disner, 1980). A normalization procedure qualifies as ineffective if the quality judgments vary as a result of the data being normalized. Disner reported that Harshman's (1970) normalization procedure was most effective for cross-language and cross-dialect comparisons. Nearey's (1977) technique was most effective at reducing the scatter of the data. Disner reported that no one normalization procedure was effective for all of the languages she studied.

Syrdal and Gopal (1986) proposed a quantitative perceptual model of vowel recognition which was based on the spatial pattern of auditory excitation produced by American English vowels. The researchers performed bark-difference transformations $[b_3 - b_2$ (F3 in bark - F2 in bark), $b_2 - b_1$ and $b_1 - b_0]$ on the 1,520 measurements from Peterson and Barney (1952). Syrdal and Gopal stated that the bark difference model provided a connection between some acoustic characteristics of vowels and their phonetic quality. Linear discriminant analyses were also performed on the Peterson and Barney data. Syrdal and Gopal reported the higher classification accuracy for the bark-

transformed measurements than measurements represented in linear frequency. Furthermore, the investigators reported that the transformation reduced the acoustic differences between vowels as spoken by different talkers. Syrdal and Gopal concluded that their bark difference model of vowel perception provided a perceptually based quantitatively defined link between acoustic and phonetic features.

Hillenbrand and Gayvert (1988) compared several vowel classification models using a maximum likelihood classification technique. The Peterson and Barney (1952) vowel measurements of F0, F1, F2 and F3 provided the data for the study. Hillenbrand and Gayvert reported that error rates for the best models were more than twice as high as the error rates shown by human listeners. Classification models produced error rates of 12 to 13% which was more than twice as high as Peterson and Barney's 5.6% error rate for listeners. The investigators found no advantage of nonlinear transforms, such as bark differences and log differences, over linear frequency. This finding was in contrast to Syrdal and Gopal (1986). Hillenbrand and Gayvert hypothesized that the difference between the identification rate of vowel perception models and that shown by listeners was due to listeners using dynamic information as compared to static spectral information used by the acoustic classification algorithms.

Characteristics Influencing Vowel Perception

The exact information listeners rely upon to identify vowels remains undetermined, although researchers since the 1950s have offered suggestions as to dynamic characteristics of speech that might play a role. There has been controversy regarding whether the context in which a vowel is heard plays a role in the identification of the stimulus. Vowels have been studied in sentence context, consonantal context and in isolation.

Verbrugge et al. (1976) asked listeners to identify p-vowel-p (p-V-p) syllables which included all of the vowels Peterson and Barney (1952) studied excluding /ʒ/. The stimuli were produced by five men, five women, and five children and classified by 79 listeners. Listeners responded to the stimuli under mixed conditions in which all productions were heard in random order. Responses were also recorded for presentations of a segregated set in which only three talkers' productions were heard. Subjects made an average of 17% errors in the random order productions while an average error rate of 9.5% was reported for the segregated condition.

Strange, Verbrugge, Shankweiler and Edman (1976) reported that vowels could be identified more accurately in context than in isolation. Strange et al. utilized p-V-p syllables collected by Verbrugge et al. (1976). Strange et al. collected a corresponding set of isolated

vowels from five men, five women, and five children. The signals were identified by 30 listeners. Listeners responded to the isolated vowels under mixed and segregated conditions. Strange et al. reported an error rate of 42.6% for the mixed task and 31.2% for the segregated task. The results of the Verbrugge et al. (1976) and the Strange et al. (1976) studies indicated that while talker variation did contribute to error rates, the presence or absence of context was most significant.

Diehl, McCusker and Chapman (1981) reported low error rates for synthetic, isolated vowels. They conducted two experiments; the first to determine identification rates of synthetic vowels, and the second to determine identification rates for natural speech stimuli. In the first experiment the Peterson and Barney (1952) vowels, excluding /ɜ/, were synthesized using the average adult male and female frequencies. Vowels were synthesized at 240 msec durations, 300 msec durations and 200 msec durations bounded by 30 msec transitions representing a b-V-b context. The stimuli were classified by 14 listeners who responded on paper to the presentations. Diehl et al. reported error rates of 24.1%, 22.6% and 28.8% for short isolated vowels, long isolated vowel and the b-V-b syllables, respectively. Diehl et al. noted that these identification rates were the opposite of the pattern typically observed for natural speech.

In their second experiment, Diehl et al. tape recorded the same vowels as they were produced by subjects. Two groups of 13 served as listeners. One group responded to the presentations on an answer sheet composed of $\bar{E}\bar{E}$, \check{I} , \check{E} , AE, AH, AW, UH, $\check{O}\check{O}$, $\bar{O}\bar{O}$. The other group marked their responses on a b-V-b word list (e.g., beeb, bib, beb, etc.). Subjects who responded to the word list showed more errors with the exception of the medial vowel condition which indicated a 0.4% decrease in error rate. Listeners using the vowel answer sheet identified 15% of the short isolated vowels, 15.8% of the long isolated vowels, and 18.2% of the vowels of the b-V-b syllables. Listeners using the b-V-b answer sheets had corresponding error rates of 17.7%, 18.6% and 17.8%.

Diehl et al. (1981) noted that the results of the two experiments differed in important ways. First, there was no significant advantage for the isolated vowels. Second, error rates for both answer sheets in experiment two were smaller than the error rates reported for experiment one. Diehl et al. suggested that the reduced error rates may be attributed to natural durational differences and to longer interstimulus intervals.

Assmann, Nearey and Hogan (1982) reported that isolated vowels could be identified accurately, and the way in which listeners record their responses to a listening task can affect performance. Assmann et al.

conducted two experiments in which they reported that listeners perceived the stimuli accurately but their response was not in accord with their perception. Listeners heard the stimuli, which were the 10 words studied by Peterson and Barney (1952), and then recorded their responses on paper. As they recorded their response on paper they were asked to say the word into a microphone. Assmann et al. stated that for 14 of the 18 listeners at least half of the vowels misidentified on the written task were accurately identified in the spoken task. These data suggested that the manner in which subjects record their responses to a listening task might affect the error rate.

In a follow-up experiment Assmann, Nearey and Hogan (1982) hypothesized that responses to spelling would show higher identification rates than keyword responses. Seven listeners were given a list of words containing the 10 words studied by Peterson and Barney (1952). An additional seven listeners were given a list of words showing the same vowels but within a p-V-p context. Subjects were asked to cross off the corresponding word that they heard over a loudspeaker. The stimuli heard by the 14 subjects were the p-V-p words. Assmann et al. reported that subjects using the h-V-d lists made an average of 7.68% errors while listeners using the p-V-p lists made an average of 3.86% errors. Assmann et al.

concluded that the way in which a subject documented the stimulus can play a significant role in the error rate.

Strange (1989b) reported on vowel perception from altered contexts after being collected in sentence context. Three experiments were conducted in which she studied the contribution of target information available in vowels, intrinsic duration related to syllable length and dynamic spectral information of onsets and offsets of syllables. She reported that silent-center syllables were perceived relatively accurately. (Silent-center syllables were defined as those in which the vowel portion of a CVC syllable was attenuated but the initial and final transitional sections were preserved in the temporal relationship in which they were produced.) Strange reported that error rates increased slightly when silent-center syllables were neutralized by equating the time between the initial and final transition of segments of the vowel. Strange also reported that both silent-center syllables and vowel portions of syllables were identified no less accurately than in unmodified conditions.

In addition to context, duration has been addressed by investigators of phonetic perception. Peterson and Lehiste (1960) reported data on the duration of vowels. They studied CVC words collected in sentence context. They found that syllables containing /I/, /E/, /A/, and /U/ were relatively short in duration, ranging between 180

and 200 msec. Syllables containing the vowels /i/, /a/ and /u/ ranged between 240 and 260 msec in duration. The vowel /æ/ averaged 330 msec. Other vowels such as /eI/, /a/, /ɔ/ and /oU/ as well as diphthongs /aI/, /aU/ and /ɔ I/ had relatively longer durations. The durations of these vowels depended on the rate of speech, the degree of stress placed on the words and the position of the word within an utterance. Context of neighboring phonemes also played a role in the duration of a vowel. Peterson and Lehiste reported that listeners were sensitive to these differing durations and that listeners identified the vowels based on intrinsic duration.

Assman et al. (1982) conducted an experiment in which they gated, or blocked out, a portion of the waveform so that it would not be heard by the listener. Thus, they eliminated duration and diphthongization. Next, they presented the gated stimuli in blocked trials, in which only one talker's presentations were heard, and mixed conditions, in which several talkers' presentations were randomized, to a group of phonetically trained listeners. Their overall error rate for the blocked condition was 9.50% and the mixed condition yielded an error rate of 13.75%. Based on previous research Assmann, Nearey and Hogan reported that although gated vowels were identified with less accuracy than full presentations of the vowels, the error rate for the gated-mixed condition was

relatively low.

Nearey and Assmann (1986) windowed nucleus and offglide sections from isolated vowels which were played for listeners to identify. Two sections from each vowel were presented which included natural order (nucleus-offglide), repeated nucleus (nucleus-nucleus) and reverse nucleus (offglide-nucleus). Listeners heard the full presentations of the isolated vowels in a practice session. Nearey and Assmann reported an error rate of 14% for the natural ordered stimuli and 13% for the full, or unmodified, stimuli. The researchers reported error rates of 32% for repeated nucleus and 38% for reverse presentations. From these data, Nearey and Assmann concluded that the offglide portion of the stimuli played an important role in identification. The increased error rate in reverse order presentation suggested that temporal ordering of the two stimuli was necessary in vowel identification. They further stated that vowel-inherent dynamic characteristics provide information in the identification of the isolated Canadian English vowels they studied.

In addition to context, researchers have also considered speaking rate. Rate, still another dynamic characteristic of vowels, must be considered when listeners are asked to identify speech since increasing the speaking rate tends to neutralize vowels. Lindblom

(1963) reported that in British and American varieties of English, there is a tendency for the lesser stressed vowels to approach each other and to approximate schwa. For example, the second formant of /i/ would decrease while the second formant of /u/ would increase. In turn the listener has to compensate for these shifts by using contextual cues of the utterance.

Tiffany (1959) demonstrated that the vowel diagram becomes smaller when an isolated vowel is unstressed instead of stressed. He noted that vowels tend to move toward the center of the diagram to a more neutral position. Tiffany proposed that a large vowel diagram would result in higher intelligibility. He proposed that the tongue should be higher and farther forward for high vowels and lower and more back for lower vowels. He did not investigate placement in relation to intelligibility in his study.

Gay (1978) collected three types of speech samples from four adult talkers to investigate the effects of speaking rate on the attainment of vowel targets. Gay also investigated the relative time and movement toward the vowel targets. Stimuli consisted of CVC and CVCVC utterances produced at slow and fast rates. The vowels studied were the same vowels studied by Peterson and Barney (1952), excluding /ɜ/. Gay reported that the acoustic targets of vowels as a function of rate showed

little variability, usually falling well within the range of measurement error. Measurements were taken at the midpoint of the first, second and third formant frequencies. Gay noted that the speed of movement from the consonant into the vowel did not change; rather, the duration of the vowel was lessened when talkers spoke at a faster rate.

Experiments using natural speech and synthetic speech have been employed in vowel perception research. Dynamic information within each signal must be accounted for in the study of natural speech since speech is a complex waveform with temporal and spectral variations. Research has shown that human listeners can correctly identify natural speech stimuli when presented in a variety of ways. Perhaps not as obvious is the fact that synthetically generated speech stimuli can also contain sufficient acoustic information to be correctly identified.

Miller (1953) experimented with a series of synthetically produced vowels in which three parameters, fundamental frequency, amplitude, and the number of formants necessary for identification, were altered. Nine listeners classified these vowels. Miller found that there were general areas in F1-F2 space in which listeners classified the two-formant synthetic vowels. Miller also found that when the formant frequencies were fixed and

amplitudes were varied, the varying amplitudes played a sometimes critical role in the identification of the stimuli. Lastly, the addition of a third formant resulted in a higher identification rate by listeners. These conclusions suggested that it was possible for listeners to classify synthetically altered vowels fairly accurately.

Hillenbrand and McMahon (1988) investigated listeners' ability to identify vowels based exclusively on static spectral information. Steady state versions of the vowels produced in the Peterson and Barney (1952) study were generated. Twelve listeners, all of whom had some phonetics training, classified the stimuli. Hillenbrand and McMahon reported an overall error rate of 24.8% which was more than four times greater than the error rate of the Peterson and Barney study. The results of Hillenbrand and McMahon indicated the importance of dynamic information to the identification of a signal. Therefore, Hillenbrand and Gayvert (1988) hypothesized that the addition of dynamic information to a signal would decrease the error rate.

Hillenbrand and Gayvert (1988) reported that error rates were significantly reduced when minimal dynamic information was provided to the listeners. They used the same stimuli that were used in the Hillenbrand and McMahon (1988) study except that Hillenbrand and Gayvert added a

$\pm 25\%$ pitch contour to the stimuli. The introduction of a pitch contour resulted in a small but highly significant 2.1% reduction in error rate when compared to stimuli with flat pitch contours.

Formants and Vowel Identification

The most complete set of data on the acoustics and perception of vowels was the result a study conducted at Bell Telephone Laboratories by Peterson and Barney (1952). They investigated how people produce and perceive vowels. The study was conducted in two parts; part one involved the productions of vowels and part two involved the identification of vowels by listeners.

The first part of the study involved 76 talkers (33 men, 28 women and 15 children) who recorded a set of 10 vowels in an h-V-d context. Each talker made two recordings of the following list of words: "heed," "hid," "head," "had," "hod," "hawed," "hood," "who'd," "hud" and "heard." Peterson and Barney used narrow-band amplitude sections from a sound spectrograph to obtain fundamental frequency and formant measurements of the 1,520 utterances. They measured the steady state of the vowel which they defined as the portion "following the influence of the [h] and preceding the influence of the [d] during which a practically steady state is reached" (p. 177).

Part II of Peterson and Barney's study was a

listening experiment that was conducted in an auditorium. The recorded words from part one were played over a loudspeaker to a group of 70 listeners. They identified each of the stimuli by marking the corresponding word on a piece of paper. There were eight listening sessions in which the listeners identified 200 vowels per session. At the end of the eighth listening session each listener had categorized the corpus of 1,520 words that were recorded from part one. The data from the listening portion of the study were organized in the form of a confusion matrix (see Appendix A).

Peterson and Barney (1952) reported that listeners categorized some vowels more accurately than others. For example, [i], [ɜ], [æ] and [u] had overall error rates of less than 4%. They also noted that [I] and [ɛ] were often substituted for one another, which they attributed to the dialect of the talker. Tokens of /ɑ/ and /ɔ/ were misidentified on 13% of the trials. This finding is also related, at least in part, to dialect since the /ɑ/ - /ɔ/ distinction is poorly maintained or not maintained at all in some dialects of American English. The overall error rate for these vowels in the listening study was 5.6%.

The reason that the Peterson and Barney (1952) study gained wide recognition was apparent when the listening portion of the study was compared to the acoustic analysis results. The investigators reported that, although there

was great variability in the production of a particular vowel, the listeners identified the intended vowels with a very high degree of accuracy. Across the vowel categories, listeners' identification of the vowels agreed with the intended vowels of the speaker on 94.4% of the utterances. The results of the Peterson and Barney study suggested that listeners do not rely on absolute formant frequencies to correctly identify vowels.

CHAPTER III

DESIGN

This investigation was a replication and extension of Peterson and Barney's (1952) research. This study explored the acoustic characteristics of 12 vowels produced by 50 men, 50 women and 41 children. Speech samples were tape recorded as subjects read a list of h-V-d words and a list of isolated vowels. Speech samples were digitized. Speech waveforms, spectrograms and formant frequencies were examined and measured. Attempts were made to address several limitations of the original study.

Limitations of Peterson and Barney

Peterson and Barney (1952) included 33 men, 28 women and 15 children in their study. Their subject groups were unbalanced. The current experiment included 50 men, 50 women and 41 children in an attempt to balance the number of talkers per group. (Recordings have been made of nine additional children. The data are currently being analyzed and the results will eventually be added to the data base.) Although the number of children was not equal to the men and women groups, the ratio of children to

adults was smaller in the present study.

In addition to increasing the number of children in the experiment more information was reported for these subjects in the present experiment. Specific numbers of boys and girls were noted in this experiment, unlike data reported by Peterson and Barney (1952). Perhaps the most relevant characteristic that was noted was the ages of the children. The age ranges for the children were restricted so that pitch change was unlikely to be an interfering variable. A child who has experienced an adolescent voice change might produce adult-like formant values. Peterson and Barney could not make comparisons between the boys and girls since they considered the children as one subject group.

A screening procedure was developed in collaboration with a phonetician to control for dialect. Peterson and Barney (1952) did not report the use of any type of screening measure for dialect. Their study included talkers who were born outside the United States and who learned another language before learning English. Peterson and Barney also included talkers from the "Middle Atlantic area" as well as those who spoke General American English. Any difficulty that a listener might have had distinguishing between vowels might have been due to different dialects that were included in the study. The present experiment included only those subjects who spoke

English as a native language and who spoke a General American dialect. Once each subject's dialect was determined, the talker read the list of h-V-d words and isolated vowels.

Talkers in this investigation read a list of 16 h-V-d words. Unlike the subjects in the original study, each talker also produced the corresponding isolated vowel. Peterson and Barney (1952) recorded 76 talkers producing 10 vowels in an h-V-d context. The equipment used to analyze the speech samples from the 141 talkers was much different from the sound spectrograph that Peterson and Barney utilized.

The recording equipment utilized in this experiment was more technologically advanced. Speech samples were recorded using high quality audio equipment. The samples were viewed in several different forms through the use of custom designed computer programs.

The experimenters in this investigation made measurements of formant frequency patterns for the entire vowel. This was in contrast to the single measurement per vowel that was reported in the original study. Measurements of duration and bandwidth were also made in addition to formant frequency and formant level. Peterson and Barney reported formant levels and formant frequencies. Additional measurements that were made in this investigation, but were not reported in the original

investigation, were fourth formant values. The fourth formant was not measurable for all vowels but for those utterances in which there was a fourth formant, a value was noted.

Procedures were used to assess measurements of reliability. Two experimenters independently made judgments of the steady state portion and duration of each vowel.

CHAPTER IV

METHODS

An Overview

The present study was a replication and extension of the phonetic perception research conducted by Peterson and Barney (1952). Number of talkers, data collection procedures, analysis methods and vowels studied were modified to: (a) balance the number of talkers per group, (b) create a screening procedure to establish criteria for subject participation, (c) incorporate more sophisticated analysis techniques, and (d) investigate the acoustic characteristics of two additional vowels. Incorporation of modern recording equipment and custom designed computer software made it possible to replicate the study with more technologically advanced equipment than was available to Peterson and Barney in the 1950s. Many methodological procedures for conducting the study remained the same.

Subject Participation

Talkers

A total of 141 talkers were involved: 50 men, 50 women and 41 children. Men and women ranged in age from

19 years to 50 years of age. The average age of the men was 23 years, seven months and the average age of the women was 26 years, 11 months. Men and women subjects were current or previous students or faculty at Western Michigan University. The majority of the women were speech and language pathology undergraduate and graduate students. Many of the men were students in the business and engineering departments at Western Michigan University. The 41 children consisted of 19 girls and 22 boys. Boys ranged in age from 10 years, two months to 11 years, 11 months with an average age of 10 years, 10 months. Girls ranged in age from 10 years, one month to 11 years, 11 months with an average age of 10 years, nine months. The children lived in the Kalamazoo, Michigan area and attended one of two elementary schools where the recordings were made. The pre-teen age ranges for the children were selected in order to minimize the influence of adolescent pitch change (see Appendix B).

Criteria for Subject Participation

Criteria for subject selection included the following: (a) no speech, language, voice or hearing disorders, (b) no current respiratory infections or allergy problems, (c) English as a native language, with a General American dialect which included the ability to distinguish between /a/ and /ɔ/, (d) within the age range

established for the respective subject group, and (e) a distinction between /a/ and /ɔ/.

Collection of the Data

Screening Procedure

A screening procedure was conducted with each subject before recordings were made. Potential subjects were first briefly interviewed to collect background information (see Appendix C). During this interview, the potential subject was engaged in an informal conversation in order to obtain a five- to seven-minute spontaneous speech sample on topics that arose in conversation. This portion of the screening was recorded using Sony HF 60 cassette tapes and an Audiotronics model 148 B cassette tape recorder.

An audio recording was also made of each potential subject reading a short passage constructed in collaboration with a phonetician (see Appendix C). All potential subjects read the passage and recorded samples which were later used to aid in determining the talker's dialect. Particular attention was paid to the distinction between /a/ and /ɔ/. The 128 word passage was constructed based on four /a/ -/ɔ/ word pairs as well as four additional words containing either the vowel /a/ or /ɔ/. One of the words, the proper name "Don," was repeated three times. The recording was later evaluated by a

phonetician.

Each potential subject was then asked to judge whether a randomized list of minimal pairs and homonyms sounded the same or different when pronounced by the subject (see Appendix C). Both minimal pairs and homonyms were randomized to reduce the possibility that the subject would be aware of the /a/ - /ɔ/ distinction that the experimenter was investigating. The potential subject marked either "S" if the word pair sounded the "SAME" or "D" if the word pair sounded "DIFFERENT."

The experimenter and the subject then reviewed the completed task. The experimenter pronounced one of the words from each word pair that the person marked as "DIFFERENT." The potential subject was asked to put a checkmark by the word that was pronounced. The experimenter and the potential subject continued this process for all of the word pairs marked "DIFFERENT." The experimenter and the subject then reversed roles; the experimenter marked a word and the subject pronounced it.

Determining Subject Eligibility

The experimenter reviewed the responses to questions regarding background data to determine subject eligibility. Informal assessments of speech, language, voice, and dialect were made by the experimenter during the collection of this information. For other criterion

areas, the experimenter relied on the subject's report. Subjects might have received speech therapy in the past but not within three years. Subjects might have reported an allergy, flu or cold but not one that affected them at the time of recording according to subject report. Instances of other languages spoken in the home where a subject was raised were noted as were the languages. English as a native language and General American dialect were mandatory.

A phonetician listened to the audio tapes recorded for each subject and eliminated those who did not reliably distinguish between /a/ and /ɔ/, or who showed any systematic departure from General American English.

A phonetician evaluated the spontaneous speech sample and the reading passage to determine whether a potential subject maintained the /a/ - /ɔ/ distinction. The phonetician noted all clearly produced words containing /a/ in one column and clearly produced /ɔ/ words in another column for both the spontaneous speech sample and the reading passage. Words that were produced with the vowel approximating neither the /a/ nor the /ɔ/ were written in a third column. The phonetician then made the final evaluation as to whether the person maintained a /a/ - /ɔ/ distinction clearly, marginally or not at all.

The word differentiation task also was used as a means of assessing the /a/ - /ɔ/ contrast. A subject was

required to have an overall score of at least 92% for the word differentiation task to be considered eligible.

In summary, participants were considered ineligible if they: (a) did not meet the criteria established for each of the screening questions, (b) were unable to produce the h-V-d words and isolated vowels, (c) did not speak a General American dialect, or (d) did not make a distinction between /ɹ/ and /ɔ/.

In order to select the appropriate number of subjects, more than 141 subjects were interviewed. Sixty-eight children were interviewed to gather the 41 children who were included in the study. The experimenter interviewed 130 adults in order to gather samples from 100 adults. The most common reasons for rejecting subjects were difficulty performing the task and inappropriate dialect for the investigation.

Collection of the h-V-d Stimuli

Following completion of the screening tasks, each subject was instructed to read a list containing the 16 h-V-d words. The following words were read aloud by each participant: "hoyed," "hide," "hewed," "howed," "heed," "hid," "head," "had," "hod," "hawed," "hood," "who'd," "hud," "heard," "hoed" and "hayed." This list of words was identical to Peterson and Barney (1952) except for the addition of "hoed" and "hayed" and the words "hoyed,"

"hide," "hewed" and "howed" which were also produced by each subject but were not analyzed in this study.

One of 12 different randomizations of the word list was read by each participant (see Appendix D). The 16 words were listed on a page followed by the same list in reverse order. Three of the 16 words were added to the end of the list in order to account for downward inflections that commonly occur when reading word lists. The last three words of each recording were not analyzed. Next to each word on the list was a familiar word as well as the corresponding vowel. For example,

"hayed"	"ay" as in "bay"
"hid"	"ih" as in "bit"

Each subject read aloud the list of 16 words to demonstrate understanding of the task before the recording was made. If a subject did not understand the orthography, the experimenter provided a spoken model of the word. The participant also read each of the vowels in isolation. Subjects were not excluded from the study or asked to articulate words in a different manner based on the experimenter's judgment of correct pronunciation. When both the experimenter and the subject agreed that the subject had an understanding of the pronunciation of the words and vowels, the recording was made.

During the recording process all subjects were asked to read both the list of words and the list of isolated

vowels more than one time in the event that misreadings of the orthography occurred. In general, the children were not able to read the lists more than three times. They often became confused with the pronunciation of words with repetitions of the task or appeared to be uninterested in the task. Most adult subjects read the list of words and the list of isolated vowels three to four times.

Adult subjects were seated in an Industrial Acoustics Company (IAC) booth model 401 to make the recordings. A Sure 570-S dynamic microphone was placed approximately 10 centimeters from the subject's mouth. A Sony PCM-F1 digital recorder (16 bits, 44.1 kHz sample rate) and a Canon VR40 VCR were used to record the signals. The same equipment was used for both the adults and the children with the exception of the IAC booth.

Since children's speech samples were recorded in an elementary school, an IAC booth was not available. Instead, the screening procedure was conducted in an art supply room and the h-V-d word readings were recorded in an adjacent art supply room. The recording of the h-V-d words occurred in a 15-by-5-foot room. The child was seated at the furthest point from the hallway door and the recording equipment was positioned in the middle of the room.

Acoustic Analysis

Digitizing the Signals

Word productions from 141 subjects were digitized. Each subject provided 16 words for a corpus of 1,692 words. The words studied were analyzed using computer programs custom designed for a PDP 11/73 computer system.

The recordings were digitized with a program called AUDED (Prall & Hillenbrand, 1981). The signals were sampled at a rate of 16 kHz with 12 bits of amplitude resolution and were digitized one word at a time. A low pass input filter was set at 7200 kHz. Sixteen words per person were digitized, although only 12 words were included in this study. Record-level gain was adjusted so that peak amplitude was at least 80% of the $\pm 10V$ dynamic range of the A/D with no peak clipping. The waveform of each signal was displayed on a computer graphics terminal. In cases where the talker did not release the final consonant, caution was exercised not to eliminate the /d/. Each signal was windowed to include the entire /h/, the vowel, and the entire /d/. During the digitizing and windowing process the signal was played over a loudspeaker so that the experimenter was able to listen to the signal before and after it was windowed.

After digitization, all 16 signals were played to verify recording fidelity and to insure that all signals

were windowed correctly. Signals were redigitized if necessary, then all 16 signals were played a second time and verified by two experimenters. Once the fidelity of the 16 words was confirmed, the starting and ending times of the vowel were measured for a computer generated ossillographic display.

Formant Frequency Measurement

Formant-frequency analysis began with the calculation of linear-predictive coding (LPC) spectra every 6.4 msec for each signal. A 20-pole LPC model was used with a 32 msec window size for the men and a 64 msec window size for the women and children. The frequencies, amplitudes, and bandwidth of the first seven peaks were then extracted from the LPC spectrum files. Files containing the LPC peak data then served as the input to a formant editor called FTRACK (Smith, 1989). This program allowed the experimenter to determine which of the LPC peaks corresponded to F1-F3 and, for cases in which the fourth formant was visible, F4.

Examples of original and edited FTRACK displays are shown in Figures 2 and 3, respectively (see Appendix E). In addition to the FTRACK peak display, the experimenter also viewed the corresponding wide-band FFT grey-scale spectrogram produced on a Macintosh computer by the GW Instruments MacSpeech Lab II package (see Figure 4 in

Appendix E). The first step in formant editing involved windowing the LPC peak data to include only the vocalic portion of the syllable. The starting and ending times of the vowel were measured from the grey-scale spectrogram using criteria described by Peterson and Lehiste (1960). After windowing, an automatic formant editing algorithm was applied to the peak data. The automatic editing algorithm assigned peaks to five formant slots using very simple logic based on maximum departures from median values for each of the formant slots. In many cases further hand editing was required. The hand editing involved positioning a cursor over a particular LPC peak and striking a key that would instruct FTRACK to either: (a) assign the peak to a particular formant number, or (b) eliminate the peak. Editing decisions were based on general constraints regarding formant continuity and close examination of the grey-scale spectrogram.

In many cases, general knowledge of acoustic phonetics played a role in the editing process. For example, editing decisions could be influenced by the experimenter's knowledge of the close proximity of F2 and F3 for vowels such as /i/ and /ɜ/, the close proximity of F1 and F2 for vowels such as /a/ and /u/, and so on. Quite often considerations such as these led the experimenter to conclude that a formant merger had occurred. For example, the FTRACK display might show a

single sequence of peaks for /i/ corresponding to F2 and F3. In these cases, the LPC analysis was recomputed with different parameters until the merged formants separated. A trial-and-error approach was used, with the most common manipulation involving increases in the number of LPC poles.

It was not uncommon for formant contours to show "holes"; that is, a frame or sequence of frames in which a peak was not present, with the peak re-emerging later in the signal. These "holes" were filled in later using linear interpolation to estimate the formant frequencies, bandwidths, and amplitudes based on corresponding values immediately prior to and following the disappearance of the spectral peak. Figures 5 and 6 (see Appendix F) show examples of a formant trace prior to and following interpolation.

In order to produce a data set comparable to Peterson and Barney (1952), a measurement was made for each syllable of the "steady state" portion of the vowel. This measurement was made separately by two experimenters based on visual inspection of both the edited FTRACK display and the grey-scale spectrogram. In accordance the relatively vague guidelines described by Peterson and Barney, the experimenters were instructed to locate the point "following the influence of the [h] and preceding the influence of the [d] during which a practically steady

state is reached" (p. 177). All of the data plotted in Figures 7 through 11 (see Appendices J through N) were based on formant measurements sampled at the steadiest portion, as judged by one experimenter, of the vowel. However, there were some signals that were found to have no identifiable steady state, that is, anywhere from one to all of the first three formants were found to be changing continuously. In those cases the measurement was taken at vowel onset.

CHAPTER V

RESULTS

Presentation of the Data

Formant Frequency Averages

Averages were calculated for the first, second and third formants for each group of talkers. The averages were compiled in table form which included the averages reported by Peterson and Barney (1952) for comparative purposes (see Table 1). Fundamental frequency data have been collected and will be reported at a later time so that comparisons between the original study and this investigation will be possible. The men showed the lowest averaged formant frequencies for all the vowels and the girls showed the highest averages. Formant frequency averages of the women were between the men and children. The women talkers' averages were often close to the averages for the boys and girls which, unlike Peterson and Barney, were calculated separately.

The averages for boys and girls were separately noted in an attempt to learn about similarities and differences between these two groups. In general, girls showed higher average first, second and third formant frequencies across

Table 1
Formant Frequency Averages

		Vowel											
		/ɪ/	/I/	/ɛ/	/æ/	/a/	/ɔ/						
Formant Frequencies (Hz)													
F1	M	340	(270)*	445	(390)	583	(530)	584	(660)	748	(730)	841	(570)
	W	432	(310)	508	(430)	705	(610)	745	(860)	943	(850)	785	(590)
	B	456	(370)	542	(530)	709	(690)	726	(1010)	977	(1030)	821	(680)
	G	458		529		717		774		1051		904	
F2	M	2293	(2290)	1922	(1990)	1758	(1840)	1867	(1720)	1289	(1090)	1020	(840)
	W	2757	(2790)	2207	(2480)	2012	(2330)	2072	(2050)	1573	(1220)	1248	(920)
	B	2981		2422	(2730)	2170	(2610)	2264	(2320)	1681	(1370)	1328	(1060)
	G	3087	(3200)	2507		2218		2292		1747		1538	
F3	M	2967	(3010)	2616	(2550)	2562	(2480)	2531	(2410)	2484	(2240)	2491	(2410)
	W	3410	(3310)	3027	(3070)	2946	(2990)	2882	(2860)	2815	(2810)	2803	(2710)
	B	3619	(3730)	3318	(3600)	3248	(3570)	3094	(3320)	2874	(3170)	2859	(3180)
	G	3649		3421		3331		3143		3008		3047	

*The values in parentheses are from Peterson and Barney (1952).

Table 1--Continued

		Vowel										
		/u/	/u/	/u/	/ʌ/	/ɜ/	/o/	/e/				
Formant Frequency (Hz)												
F1	M	469	(440)	380	(300)	613	(640)	466	(490)	477		445
	W	523	(470)	459	(370)	746	(760)	514	(500)	521		473
	B	577	(560)	495	(430)	698	(850)	565	(580)	574		520
	G	576		489		735		588		553		496
F2	M	1129	(1020)	993	(870)	1182	(1190)	1351	(1350)	894		2099
	W	1342	(1160)	1116	(950)	1501	(1400)	1578	(1640)	1013		2583
	B	1590	(1410)	1331	(1170)	1565	(1590)	1728	(1820)	1102		2681
	G	1638		1464		1599		1756		1107		2781
F3	M	2399	(2240)	2297	(2240)	2522	(2390)	1736	(1590)	2433		2664
	W	2809	(2680)	2738	(2670)	2873	(2780)	1959	(1960)	2814		3071
	B	2966	(3310)	2937	(3260)	3052	(3360)	2107	(2160)	2959		3299
	G	3151		2993		3096		2216		3058		3326

all vowel categories as compared to the boys. For 31 of the 36 comparisons shown in Table 1, the average formant frequency value was higher for the girls than the boys. However, the averages for boys usually fell close to the averages for the girls and in some cases were nearly identical to them.

The Data in an F1-F2 Matrix

The data resulting from this investigation are plotted in a series of five figures (see Appendices J through N). Initially, the 10 vowels were plotted in one figure modeled after the Peterson and Barney (1952) data. However, it was not possible to visually distinguish each of the 10 different vowel categories due to the lack of separation of the data. Therefore, only three vowels were plotted on one graph. One vowel category from the previous figure was plotted on the succeeding figure to serve as a reference point. Each graph will be considered separately.

Figure 7 (see Appendix J), plottings of /i/, /I/ and /E/, shows the greatest amount of separation between vowel categories. The least amount of overlap occurred between these vowel categories. Some overlap was to be expected even among the best defined vowels. The vowel /E/ shows some outlying data points; that is, they appear more distant from the center of the F1-F2 /E/ region than do

the majority of the /ɛ/ vowels.

Figure 8 (see Appendix K) shows the vowels /ɛ/, /æ/ and /ʌ/. The vowel /ʌ/ was fairly distinct from /æ/ and /ɛ/, but the adjacent vowels /ɛ/ and /æ/ show a large degree of overlap.

Figure 9 (see Appendix L) shows F1 and F2 values for the vowels /ʌ/, /ɑ/, and /ɔ/. Little separation of the three vowels is shown. The men talkers' utterances for /ʌ/ and /ɔ/ appear to share the same space on the graph. The rest of the utterances for these three vowels appear to be intermingled, the exception being a few productions of /ʌ/ whose first and second formants were below approximately 600 Hz and above approximately 1,500 Hz, respectively.

Examination of Figure 10 (see Appendix M) revealed some distinction between the vowels /ɔ/, /U/ and /u/. The vowels /u/ and /U/ appear to share nearly the same region on the graph with the exception of the men's productions of /u/. The vowel /ɔ/ was seen more easily since most of the utterances did not overlap with the other vowels on this particular graph.

Figure 11 (see Appendix N), which includes vowels /u/, /U/ and /ʊ/, showed almost no separation other than the small cluster of the vowel /u/ whose first formant was below approximately 400 Hz and whose second formant was below approximately 1,400 Hz.

Vowel Duration Data

Average vowel duration data by talker groups is presented in Table 2. All productions, which included productions by men, women, boys, and girls, of each vowel also were averaged. Averages for boys and girls were calculated separately but were found to be nearly identical. The average vowel durations of the women were similar to the averages for the children. The computed averages for the men were markedly lower than those of the other three groups of talkers and the relative difference was rather consistent across vowel categories.

Although hypotheses have been generated and tested, a logical conclusion cannot be drawn to account for the high degree of scatter of the data. At the present time tests are being conducted in an attempt to account for the low degree of separability of the data as compared to Peterson and Barney (1952).

Table 2
Vowel Duration Averages

	Vowel											
	/i/	/I/	/ɛ/	/æ/	/a/	/ɔ/	/U/	/u/	/ʌ/	/ɜ/	/o/	/e/
Vowel Duration Averages in msec												
M	256	205	204	285	272	286	201	248	202	274	277	279
W	342	255	261	348	341	366	259	329	244	338	342	340
B	327	262	254	346	329	331	268	289	260	319	319	318
G	330	274	261	349	334	353	280	327	251	320	338	338
All Talkers	308	240	240	326	314	331	241	294	233	310	315	315

CHAPTER VI

DISCUSSION

Comparison of the Investigations

Probably the most informative comparison that could be made between the Peterson and Barney (1952) study and this investigation was between the F1-F2 displays of the data. The Peterson and Barney display depicting talkers' productions showed a fairly high degree of separation between the vowel categories. Although there was overlap between vowel categories in the original study, it was relatively small when compared to this investigation. The data from this experiment showed a much higher degree of scatter and overlap.

The vowel categories /i/ and /I/ from this investigation show separation similar to that reported by Peterson and Barney (1952). These two vowel categories do not share the same region of the F1-F2 display as the same two vowels from the Peterson and Barney study. While the two vowel categories were fairly well separated, other categories showed considerable overlap.

Peterson and Barney (1952) reported overlapping areas between /ɜ/ and /ɛ/, /ɜ/ and /ʊ/, /ʊ/ and /u/ and /ɑ/ and /ɔ/. Peterson and Barney noted that /ɜ/ might be easily

distinguished from the other vowel categories if the third formant frequency was used. As noted earlier, this investigation showed a shared vowel space between the vowel categories /U/ and /u/. In fact, the only separable portion of these two categories was displayed in the men talker's productions of /u/. The vowel categories /Q/ and /ɔ/ of this experiment share somewhat the same region on the F1-F2 graph. Close examination of the display revealed that /Q/ was the higher of the two vowels, following the same general pattern of the Peterson and Barney data.

One possibility that was considered was that the discrepancy in category separability was based on differences in the times at which spectral measurements were made. As in the original study, measurements were reported at "steady-state" points; that is, at the point in the syllable that was judged to be maximally stable. It is possible that the criteria determining the steady-state point may have differed sufficiently across the two studies to account for a substantial amount of the difference in category separability.

An attempt was made to obtain an approximate estimate of the amount of the discrepancy that might be attributable to differences in the judgment of steady-state points. This was done by remeasuring a subset of the data using a method of choosing steady-state points

that was purposely designed to provide maximal agreement with Peterson and Barney (1952). The signals chosen for remeasurement were tokens of /æ/ and /ɛ/ produced by the women. These signals were chosen because formant mergers between F1 and F2 were not likely for these vowels.

FTRACK displays were reexamined for each of these signals while simultaneously examining an F1-F2 plot of the Peterson and Barney (1952) data. Steady-state points were chosen to ensure the closest possible agreement with Peterson and Barney, even in cases where there was little acoustic justification for the decision. It was reasoned that if separability improved significantly with the new set of steady-state points, a large portion of the discrepancy might be attributable to differences in the judgement of steady-state points. However, if separability failed to improve significantly, other explanations for the differences in category separability would have to be sought.

Figure 12 (see Appendix O) shows the original /æ/ and /ɛ/ measurements from the adult female talkers and Figure 13 (see Appendix P) shows the remeasured tokens. It can be seen that category separability improved only very slightly as a result of the redefinition of steady-state points. This finding suggests that different decision criteria regarding the judgment of steady-state points probably played a relatively minor role in

accounting for the discrepancy in category separability between Peterson and Barney (1952) and the present study.

Another possibility that was considered was that a different method of LPC analysis might produce better separability. The approach that was taken in the present study followed Miller's (1989) method of using a relatively large number of LPC poles. This method produces a fine-grain spectrum with a tendency to produce spurious peaks. Figure 14 (see Appendix Q) shows LPC spectra from an /æ/ produced by a woman with the number of poles varied between 12 and 30.

The general strategy that was followed in this study was to begin with a fine-grain spectrum using 24 LPC poles and to use subsequent FTRACK editing to remove spurious peaks. Although there was no *a priori* reason to believe that this technique was flawed, a subset of the signals was remeasured using a coarser LPC analysis. Again, the /æ/ and /ɛ/ ' vowels produced by the women were used as test signals. The LPC analysis used 10 poles; the number of poles was increased only if F2-F3 mergers occurred. In these cases, the number of poles was increased by two until F2 and F3 separated. The results of these analyses, shown in Figure 15 (see Appendix R), indicate little or no improvement in category separability as a result of the change in analysis technique.

Vowel Duration

As mentioned previously, the vowel durations for the women, boys, and girls were similar. Vowel durations for the men were significantly shorter for all vowel categories. All subjects were instructed to read the list of words in a relaxed manner.

The vowel duration measurements from this investigation were compared to the results of Peterson and Lehiste (1960) in which the average duration of vowels was calculated for five talkers. It is unknown whether the five talkers in the Peterson and Lehiste study were men, women, or children or a combination of the groups but the researchers noted that the talkers were "of the same general dialect" (p. 694). Comparing the results of the Peterson and Lehiste data and the men talkers' data from this investigation indicated very similar averages for almost all vowel categories. There appeared to be no logical reason why the average durations for the women were more similar to the children than to the men. This difference might be further investigated in the future.

At the current time no logical explanations can be given to account for the data being dissimilar to Peterson and Barney (1952). Studies involving duration measurements might be conducted to account for the difference between the men talkers and the remaining three groups. Many studies remain to be conducted with this

data set. A follow-up study is currently under way which involves remeasurement using: (a) smooth Fourier spectra (not LPC)--this will eliminate any questions regarding the number of LPC poles, and (b) "steady state" measurements derived automatically by a computer algorithm. The algorithm will locate the point of minimum velocity in the movement of a pointer through Miller's auditory-perceptual space ($\log F_3 - \log F_2$, $\log F_2 - \log F_1$, $\log F_1 - \log F_0$).

Proposed Future Studies

One aspect of vowel research that can be further explored as a result of the data collected from this study is vowel duration. As noted earlier, the average durations for the women were similar to the children's average duration times. It is unclear why the men and women should have differed in vowel durations. The average duration for each vowel category should be further investigated since it would seem logical that adult vowel durations would be similar while boys' and girls' duration times would be similar (Smith, 1978). Such a study might involve determining whether the same difference exists in the spontaneous speech samples recorded during the screening interview.

Productions of 16 isolated vowels were also collected from each subject during this investigation. These recordings might serve as a data base for future studies

of naturally produced isolated vowels. Comparisons can be made between the durations of isolated vowels and vowels produced within an h-V-d context. Researchers have begun to deal with issues concerning context versus isolated vowels. Diehl et al. (1981) stated that vowels produced in isolation might have less durational variation across vowel categories when compared to vowels produced within consonantal context. They showed that vowels produced in context were not necessarily identified with greater accuracy than vowels produced in isolation.

Several perception experiments investigating the comparison of natural and synthetic isolated vowel productions might be conducted utilizing the signals collected from this study. Researchers might generate synthetic stimuli based on the formant frequency measurements of the natural productions of isolated vowels. Any combination, including the four diphthongs /aI/, /oi/, /yu/ and /au/, of the 16 recorded vowels may be explored. These studies could lead to a better understanding of the differences between synthetic, isolated vowels and natural vowel productions of men, women and children. Although researchers might use the data from this investigation in several ways, the focus of future experiments might be, at least in part, on perception studies.

Perhaps the next most logical phase for the data

collected in this investigation would be a perception experiment modeled after the listening study in the original experiment. Peterson and Barney's (1952) listening study indicated that a high percentage of the stimuli from the first part of the study were classified with accuracy. A formal listening study of the stimuli from the present experiment has not been conducted. However, informal listening suggests that the great majority of the signals could be identified as the vowel that was intended by the talker. Since the results of this investigation were dissimilar to Peterson and Barney's findings it would be of primary importance to determine how well listeners perceived these vowels.

Other perception studies might include asking listeners to classify the 10 vowels produced by one talker or asking listeners to classify the 10 vowels produced by 10 different talkers to determine which method proves most accurate. Results of such a study might enable researchers to learn whether listeners normalize to an individual talker's characteristics.

Still another perception experiment that would be of interest to vowel perception researchers would be the classification accuracy of the vowels /o/ and /e/. An experiment should be conducted to determine how well listeners are able to classify these vowels.

Replication and extension of the Peterson and Barney

(1952) study was necessary to measure the pattern of spectral change for a set of vowels similar to those studied in the original experiment. This study, in which the formant frequency pattern was measured for the entire vowel portion, will provide researchers with more information pertaining to each particular vowel. Also, having access to the entire utterance might allow speech scientists to manipulate the whole vowel segment as well as the two consonants on either side of the vowel. Investigators might begin to better understand what information listeners rely upon to accurately identify vowels of a given language by manipulating and analyzing selected portions of the utterance.

Appendix A
Peterson and Barney Listening Data

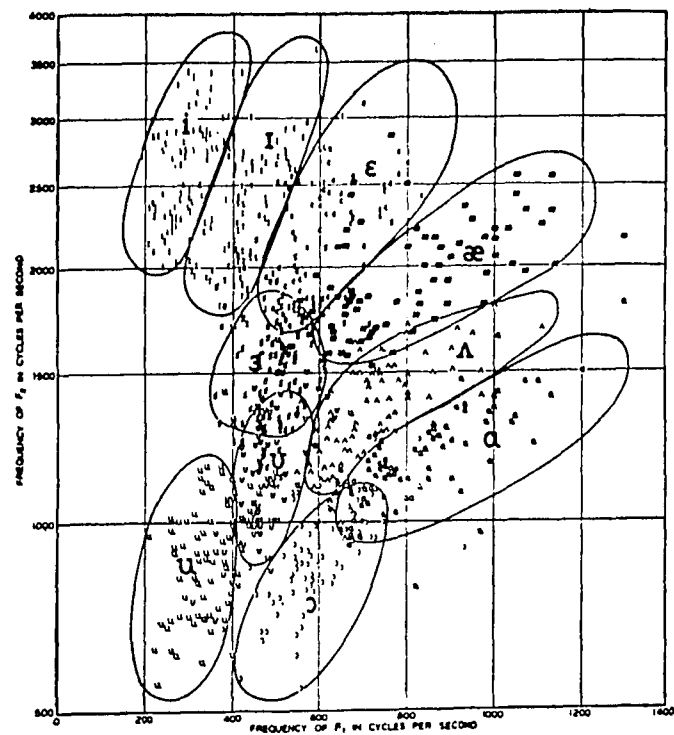


Figure 1. Peterson and Barney (1952) Listening Data. Reproduced from Peterson and Barney (1952), p. 182.

Appendix B
Adolescent Voice Change

Adolescent Voice Change

During adolescence an average pitch change of one octave occurs in a male's voice and a female's pitch lowers approximately three to four semitones (Boone and McFarlane, 1988). Bennett (1983) conducted a longitudinal study of fundamental frequency in which 25 7-, 8- to 11-year-old children were involved. She found that none of the 10- or 11-year-old boys evidenced any change in their voices since their fundamental frequencies were above 195 Hz. Bennett reported that the group of girls she studied over a period of time possessed child-like voices since the group average for fundamental frequency was 221 Hz. Bennett's (1983) findings were in agreement with a cross-sectional study by Hollien, Malcic and Hollien (1965) who concluded that the 10-year-old boys they studied possessed child-like voices.

Appendix C
Screening Procedure

BACKGROUND INFORMATION -- INTERVIEW**NAME:****AGE: BIRTHDATE: YEARS: MONTHS:****SEX:****GRADE (FOR CHILDREN):****LOCATION BORN AND RAISED:****NATIVE LANGUAGE:****OTHER LANGUAGES SPOKEN IN THE HOME WHILE GROWING UP:****SPEECH AND LANGUAGE PROBLEMS:****HEARING PROBLEMS:****LEARNING PROBLEMS:****PRESENT RESPIRATORY INFECTIONS (INCLUDING FLU AND COLDS):****ALLERGIES:**

THE CAMPING TRIP

DON AND SARAH, WHO LIVED IN THE CITY, WENT CAMPING ONE WEEKEND WITH THEIR DOG, ROVER. WHEN THEY ARRIVED AT THEIR DESTINATION DON DISCOVERED THAT HE HAD FORGOTTEN TO BRING THEIR TENT AND COTS. HE SAWED SOME WOOD TO MAKE A FIRE AND THEY EVENTUALLY FELL ASLEEP ON THE WET SOD. THEY WOKE UP AT DAWN THE NEXT MORNING AND SARAH REALIZED SHE HAD CAUGHT A BAD COLD. DON LOOKED AROUND FOR SOMETHING TO EAT, BUT ALL HE COULD FIND WAS WATER AND SOME DRIED-UP BEAN PODS ON A SOYBEAN BUSH. ROVER TURNED UP HIS NOSE AT THESE, WHINED, AND PAWED AT THE CAR DOOR AS IF HE WANTED TO GO HOME. DON AND SARAH AND ROVER GOT INTO THE CAR AND DROVE BACK TO THE CITY.

SUBJECT #: _____

Do the words _____ and _____ sound 1. EXACTLY THE
SAME (S)
or
2. DIFFERENT
when you say them?

_____SUM	SOME	_____THERE	THEIR
_____LIST	LEAST	_____HOCK	HAWK
_____THROUGH	THREW	_____AUNT	ANT
_____COT	CAUGHT	_____TEN	TIN
_____DON	DAWN	_____POOL	PULL
_____LAUGHED	LEFT	_____RIGHT	WRITE
_____SHOOT	SHUT	_____TOT	TAUGHT
_____MEET	MEAT	_____OTTO	AUTO
_____BODY	BAWDY	_____TOCK	TALK
_____SOD	SAWED	_____BUY	BYE
_____FOUR	FOR	_____COLLAR	CALLER
_____HURT	HUT	_____PAWED	POD
_____WOOD	WORD	_____CAN	CANE

Appendix D
Reading Lists

file: PBLIST.1N

"hayed"	"ay" (as in "bait")
"hood"	"oo" (as in "book")
"hod"	"ah" (as in "cot")
"hide"	"eye" (as in "eye")
"heard"	"er" (as in "bird")
"head"	"eh" (as in "bet")
"hoyed"	"oy" (as in "boy")
"who'd"	"oo" (as in "boot")
"howd"	"ow" (as in "ouch")
"hid"	"ih" (as in "bit")
"hawed"	"aw" (as in "caught")
"hewed"	"you" (as in "you")
"heed"	"ee" (as in "beet")
"hoed"	"o" (as in "coat")
"HUD"	"uh" (as in "but")
"had"	"a" (as in "bat")
"had"	"a" (as in "bat")
"HUD"	"uh" (as in "but")
"hoed"	"o" (as in "coat")
"heed"	"ee" (as in "beet")
"hewed"	"you" (as in "you")
"hawed"	"aw" (as in "caught")
"hid"	"ih" (as in "bit")
"howd"	"ow" (as in "ouch")
"who'd"	"oo" (as in "boot")
"hoyed"	"oy" (as in "boy")
"head"	"eh" (as in "bet")
"heard"	"er" (as in "bird")
"hide"	"eye" (as in "eye")
"hod"	"ah" (as in "cot")
"hood"	"oo" (as in "book")
"hayed"	"ay" (as in "bait")
"who'd"	"oo" (as in "boot")
"howd"	"ow" (as in "ouch")
"hid"	"ih" (as in "bit")

file: PBLIST.2N

"hod"	"ah"	(as in "cot")
"head"	"eh"	(as in "bet")
"hoyed"	"oy"	(as in "boy")
"heed"	"ee"	(as in "beet")
"hayed"	"ay"	(as in "bait")
"hoed"	"o"	(as in "coat")
"howd"	"ow"	(as in "ouch")
"who'd"	"oo"	(as in "boot")
"hid"	"ih"	(as in "bit")
"had"	"a"	(as in "bat")
"hide"	"eye"	(as in "eye")
"hood"	"oo"	(as in "book")
"hewed"	"you"	(as in "you")
"HUD"	"uh"	(as in "but")
"hawed"	"aw"	(as in "caught")
"heard"	"er"	(as in "bird")
"heard"	"er"	(as in "bird")
"hawed"	"aw"	(as in "caught")
"HUD"	"uh"	(as in "but")
"hewed"	"you"	(as in "you")
"hood"	"oo"	(as in "book")
"hide"	"eye"	(as in "eye")
"had"	"a"	(as in "bat")
"hid"	"ih"	(as in "bit")
"who'd"	"oo"	(as in "boot")
"howd"	"ow"	(as in "ouch")
"hoed"	"o"	(as in "coat")
"hayed"	"ay"	(as in "bait")
"heed"	"ee"	(as in "beet")
"hoyed"	"oy"	(as in "boy")
"head"	"eh"	(as in "bet")
"hod"	"ah"	(as in "cot")
"who'd"	"oo"	(as in "boot")
"hid"	"ih"	(as in "bit")
"had"	"a"	(as in "bat")

file: PBLIST.3N

"HUD"	"uh" (as in "but")
"hood"	"oo" (as in "book")
"hoed"	"o" (as in "coat")
"head"	"eh" (as in "bet")
"hoyed"	"oy" (as in "boy")
"hewed"	"you" (as in "you")
"heed"	"ee" (as in "beet")
"hawed"	"aw" (as in "caught")
"had"	"a" (as in "bat")
"hod"	"ah" (as in "cot")
"who'd"	"oo" (as in "boot")
"hid"	"ih" (as in "bit")
"hayed"	"ay" (as in "bait")
"hide"	"eye" (as in "eye")
"heard"	"er" (as in "bird")
"howd"	"ow" (as in "ouch")
"howd"	"ow" (as in "ouch")
"heard"	"er" (as in "bird")
"hide"	"eye" (as in "eye")
"hayed"	"ay" (as in "bait")
"hid"	"ih" (as in "bit")
"who'd"	"oo" (as in "boot")
"hod"	"ah" (as in "cot")
"had"	"a" (as in "bat")
"hawed"	"aw" (as in "caught")
"heed"	"ee" (as in "beet")
"hewed"	"you" (as in "you")
"hoyed"	"oy" (as in "boy")
"head"	"eh" (as in "bet")
"hoed"	"o" (as in "coat")
"hood"	"oo" (as in "book")
"HUD"	"uh" (as in "but")
"hawed"	"aw" (as in "caught")
"had"	"a" (as in "bat")
"hod"	"ah" (as in "cot")

file: PBLIST.4N

"hewed"	"you" (as in "you")
"head"	"eh" (as in "bet")
"hide"	"eye" (as in "eye")
"hood"	"oo" (as in "book")
"hoyed"	"oy" (as in "boy")
"who'd"	"oo" (as in "boot")
"hid"	"ih" (as in "bit")
"hawed"	"aw" (as in "caught")
"hod"	"ah" (as in "cot")
"heard"	"er" (as in "bird")
"howd"	"ow" (as in "ouch")
"heed"	"ee" (as in "beet")
"had"	"a" (as in "bat")
"HUD"	"uh" (as in "but")
"hayed"	"ay" (as in "bait")
"hoed"	"o" (as in "coat")
"hoed"	"o" (as in "coat")
"hayed"	"ay" (as in "bait")
"HUD"	"uh" (as in "but")
"had"	"a" (as in "bat")
"heed"	"ee" (as in "beet")
"howd"	"ow" (as in "ouch")
"heard"	"er" (as in "bird")
"hod"	"ah" (as in "cot")
"hawed"	"aw" (as in "caught")
"hid"	"ih" (as in "bit")
"who'd"	"oo" (as in "boot")
"hoyed"	"oy" (as in "boy")
"hood"	"oo" (as in "book")
"hide"	"eye" (as in "eye")
"head"	"eh" (as in "bet")
"hewed"	"you" (as in "you")
"hawed"	"aw" (as in "caught")
"hod"	"ah" (as in "cot")
"heard"	"er" (as in "bird")

file: PBLIST.5N

"had"	"a" (as in "bat")
"heed"	"ee" (as in "beet")
"HUD"	"uh" (as in "but")
"hoyed"	"oy" (as in "boy")
"hood"	"oo" (as in "book")
"hid"	"ih" (as in "bit")
"howed"	"ow" (as in "ouch")
"hewed"	"you" (as in "you")
"hide"	"eye" (as in "eye")
"hod"	"ah" (as in "cot")
"who'd"	"oo" (as in "boot")
"hayed"	"ay" (as in "bait")
"heard"	"er" (as in "bird")
"head"	"eh" (as in "bet")
"hawed"	"aw" (as in "caught")
"hoed"	"o" (as in "coat")
"hoed"	"o" (as in "coat")
"hawed"	"aw" (as in "caught")
"head"	"eh" (as in "bet")
"heard"	"er" (as in "bird")
"hayed"	"ay" (as in "bait")
"who'd"	"oo" (as in "boot")
"hod"	"ah" (as in "cot")
"hide"	"eye" (as in "eye")
"hewed"	"you" (as in "you")
"howd"	"ow" (as in "ouch")
"hid"	"ih" (as in "bit")
"hood"	"oo" (as in "book")
"hoyed"	"oy" (as in "boy")
"HUD"	"uh" (as in "but")
"heed"	"ee" (as in "beet")
"had"	"a" (as in "bat")
"hewed"	"you" (as in "you")
"hide"	"eye" (as in "eye")
"hod"	"ah" (as in "cot")

file: PBLIST.6N

"hod"	"ah" (as in "cot")
"hawed"	"aw" (as in "caught")
"heed"	"ee" (as in "beet")
"hewed"	"you" (as in "you")
"hoyed"	"oy" (as in "boy")
"head"	"eh" (as in "bet")
"hid"	"ih" (as in "bit")
"heard"	"er" (as in "bird")
"hoed"	"o" (as in "coat")
"had"	"a" (as in "bat")
"hood"	"oo" (as in "book")
"hide"	"eye" (as in "eye")
"hayed"	"ay" (as in "bait")
"howd"	"ow" (as in "ouch")
"HUD"	"uh" (as in "but")
"who'd"	"oo" (as in "boot")
"who'd"	"oo" (as in "boot")
"HUD"	"uh" (as in "but")
"howd"	"ow" (as in "ouch")
"hayed"	"ay" (as in "bait")
"hide"	"eye" (as in "eye")
"hood"	"oo" (as in "book")
"had"	"a" (as in "bat")
"hoed"	"o" (as in "coat")
"heard"	"er" (as in "bird")
"hid"	"ih" (as in "bit")
"head"	"eh" (as in "bet")
"hoyed"	"oy" (as in "boy")
"hewed"	"you" (as in "you")
"heed"	"ee" (as in "beet")
"hawed"	"aw" (as in "caught")
"hod"	"ah" (as in "cot")
"heard"	"er" (as in "bird")
"hoed"	"o" (as in "coat")
"had"	"a" (as in "bat")

file: PBLIST.7N

"had"	"a" (as in "bat")
"head"	"eh" (as in "bet")
"heard"	"er" (as in "bird")
"HUD"	"uh" (as in "but")
"hood"	"oo" (as in "book")
"hid"	"ih" (as in "bit")
"howd"	"ow" (as in "ouch")
"hide"	"eye" (as in "eye")
"hoed"	"o" (as in "coat")
"hayed"	"ay" (as in "bait")
"hawed"	"aw" (as in "caught")
"heed"	"ee" (as in "beet")
"hod"	"ah" (as in "cot")
"who'd"	"oo" (as in "boot")
"hewed"	"you" (as in "you")
"hoyed"	"oy" (as in "boy")
"hoyed"	"oy" (as in "boy")
"hewed"	"you" (as in "you")
"who'd"	"oo" (as in "boot")
"hod"	"ah" (as in "cot")
"heed"	"ee" (as in "beet")
"hawed"	"aw" (as in "caught")
"hayed"	"ay" (as in "bait")
"hoed"	"o" (as in "coat")
"hide"	"eye" (as in "eye")
"howd"	"ow" (as in "ouch")
"hid"	"ih" (as in "bit")
"hood"	"oo" (as in "book")
"HUD"	"uh" (as in "but")
"heard"	"er" (as in "bird")
"head"	"eh" (as in "bet")
"had"	"a" (as in "bat")
"hide"	"eye" (as in "eye")
"hoed"	"o" (as in "coat")
"hayed"	"ay" (as in "bait")

file: PBLIST.8N

"had"	"a"	(as in "bat")
"head"	"eh"	(as in "bet")
"heard"	"er"	(as in "bird")
"heed"	"ee"	(as in "beet")
"hayed"	"ay"	(as in "bait")
"hood"	"oo"	(as in "book")
"HUD"	"uh"	(as in "but")
"hewed"	"you"	(as in "you")
"hawed"	"aw"	(as in "caught")
"howd"	"ow"	(as in "ouch")
"who'd"	"o"	(as in "boot")
"hod"	"ah"	(as in "cot")
"hoed"	"o"	(as in "coat")
"hid"	"ih"	(as in "bit")
"hoyed"	"oy"	(as in "boy")
"hide"	"eye"	(as in "eye")
"hide"	"eye"	(as in "eye")
"hoyed"	"oy"	(as in "boy")
"hid"	"ih"	(as in "bit")
"hoed"	"o"	(as in "coat")
"hod"	"ah"	(as in "cot")
"who'd"	"oo"	(as in "boot")
"howd"	"ow"	(as in "ouch")
"hawed"	"aw"	(as in "caught")
"hewed"	"you"	(as in "you")
"HUD"	"uh"	(as in "but")
"hood"	"oo"	(as in "book")
"hayed"	"ay"	(as in "bait")
"heed"	"ee"	(as in "beet")
"heard"	"er"	(as in "bird")
"head"	"eh"	(as in "bet")
"had"	"a"	(as in "bat")
"hewed"	"you"	(as in "you")
"hawed"	"aw"	(as in "caught")
"howd"	"ow"	(as in "ouch")

file: PBLIST.9N

"hid"	"ih"	(as in "bit")
"hayed"	"ay"	(as in "bait")
"heard"	"er"	(as in "bird")
"hoyed"	"oy"	(as in "boy")
"howd"	"ow"	(as in "ouch")
"head"	"eh"	(as in "bet")
"had"	"a"	(as in "bat")
"HUD"	"uh"	(as in "but")
"who'd"	"oo"	(as in "boot")
"hod"	"ah"	(as in "cot")
"heed"	"ee"	(as in "beet")
"hoed"	"o"	(as in "coat")
"hood"	"oo"	(as in "book")
"hide"	"eye"	(as in "eye")
"hewed"	"you"	(as in "you")
"hawed"	"aw"	(as in "caught")
"hawed"	"aw"	(as in "caught")
"hewed"	"you"	(as in "you")
"hide"	"eye"	(as in "eye")
"hood"	"oo"	(as in "book")
"hoed"	"o"	(as in "coat")
"heed"	"ee"	(as in "beet")
"hod"	"ah"	(as in "cot")
"who'd"	"oo"	(as in "boot")
"HUD"	"uh"	(as in "but")
"had"	"a"	(as in "bat")
"head"	"eh"	(as in "bet")
"howd"	"ow"	(as in "ouch")
"hoyed"	"oy"	(as in "boy")
"heard"	"er"	(as in "bird")
"hayed"	"ay"	(as in "bait")
"hid"	"ih"	(as in "bit")
"HUD"	"uh"	(as in "but")
"who'd"	"oo"	(as in "boot")
"hod"	"ah"	(as in "cot")

file: PBLIST.10N

"hid"	"ih"	(as in "bit")
"hoed"	"o"	(as in "coat")
"hood"	"oo"	(as in "book")
"who'd"	"oo"	(as in "boot")
"hoyed"	"oy"	(as in "boy")
"hide"	"eye"	(as in "eye")
"heard"	"er"	(as in "bird")
"hayed"	"ay"	(as in "bait")
"howd"	"ow"	(as in "ouch")
"head"	"eh"	(as in "bet")
"heed"	"ee"	(as in "beet")
"hawed"	"aw"	(as in "caught")
"hewed"	"you"	(as in "you")
"HUD"	"uh"	(as in "but")
"hod"	"ah"	(as in "cot")
"had"	"a"	(as in "bat")
"had"	"a"	(as in "bat")
"hod"	"ah"	(as in "cot")
"HUD"	"uh"	(as in "but")
"hewed"	"you"	(as in "you")
"hawed"	"aw"	(as in "caught")
"heed"	"ee"	(as in "beet")
"head"	"eh"	(as in "bet")
"howd"	"ow"	(as in "ouch")
"hayed"	"ay"	(as in "bait")
"heard"	"er"	(as in "bird")
"hide"	"eye"	(as in "eye")
"hoyed"	"oy"	(as in "boy")
"who'd"	"oo"	(as in "boot")
"hood"	"oo"	(as in "book")
"hoed"	"o"	(as in "coat")
"hid"	"ih"	(as in "bit")
"hayed"	"ay"	(as in "bait")
"howd"	"ow"	(as in "ouch")
"head"	"eh"	(as in "bet")

file: PBLIST.11N

"hod"	"ah" (as in "cot")
"hood"	"oo" (as in "book")
"HUD"	"uh" (as in "but")
"hewed"	"you" (as in "you")
"had"	"a" (as in "bat")
"heard"	"er" (as in "bird")
"hoed"	"o" (as in "coat")
"hid"	"ih" (as in "bit")
"hoyed"	"oy" (as in "boy")
"who'd"	"oo" (as in "boot")
"heed"	"ee" (as in "beet")
"hide"	"eye" (as in "eye")
"hawed"	"aw" (as in "caught")
"howd"	"ow" (as in "ouch")
"hayed"	"ay" (as in "bait")
"head"	"eh" (as in "bet")
"head"	"eh" (as in "bet")
"hayed"	"ay" (as in "bait")
"howd"	"ow" (as in "ouch")
"hawed"	"aw" (as in "caught")
"hide"	"eye" (as in "eye")
"heed"	"ee" (as in "beet")
"who'd"	"oo" (as in "boot")
"hoyed"	"oy" (as in "boy")
"hid"	"ih" (as in "bit")
"hoed"	"o" (as in "coat")
"heard"	"er" (as in "bird")
"had"	"a" (as in "bat")
"hewed"	"you" (as in "you")
"HUD"	"uh" (as in "but")
"hood"	"oo" (as in "book")
"hod"	"ah" (as in "cot")
"hid"	"ih" (as in "bit")
"hoyed"	"oy" (as in "boy")
"who'd"	"oo" (as in "boot")

file: PBLIST.12N

"hoed"	"o" (as in "coat")
"head"	"eh" (as in "bet")
"hawed"	"aw" (as in "caught")
"heard"	"er" (as in "bird")
"hayed"	"ay" (as in "bait")
"howd"	"ow" (as in "ouch")
"hid"	"ih" (as in "bit")
"hoyed"	"oy" (as in "boy")
"heed"	"ee" (as in "beet")
"who'd"	"oo" (as in "boot")
"hewed"	"you" (as in "you")
"HUD"	"uh" (as in "but")
"hod"	"ah" (as in "cot")
"had"	"a" (as in "bat")
"hood"	"oo" (as in "book")
"hide"	"eye" (as in "eye")
"hide"	"eye" (as in "eye")
"hood"	"oo" (as in "book")
"had"	"a" (as in "bat")
"hod"	"ah" (as in "cot")
"HUD"	"uh" (as in "but")
"hewed"	"you" (as in "you")
"who'd"	"oo" (as in "boot")
"heed"	"ee" (as in "beet")
"hoyed"	"oy" (as in "boy")
"hid"	"ih" (as in "bit")
"howd"	"ow" (as in "ouch")
"hayed"	"ay" (as in "bait")
"heard"	"er" (as in "bird")
"hawed"	"aw" (as in "caught")
"head"	"eh" (as in "bet")
"hoed"	"o" (as in "coat")
"hoyed"	"oy" (as in "boy")
"heed"	"ee" (as in "beet")
"who'd"	"oo" (as in "boot")

Appendix E
Original FTRACK Display of ϵ

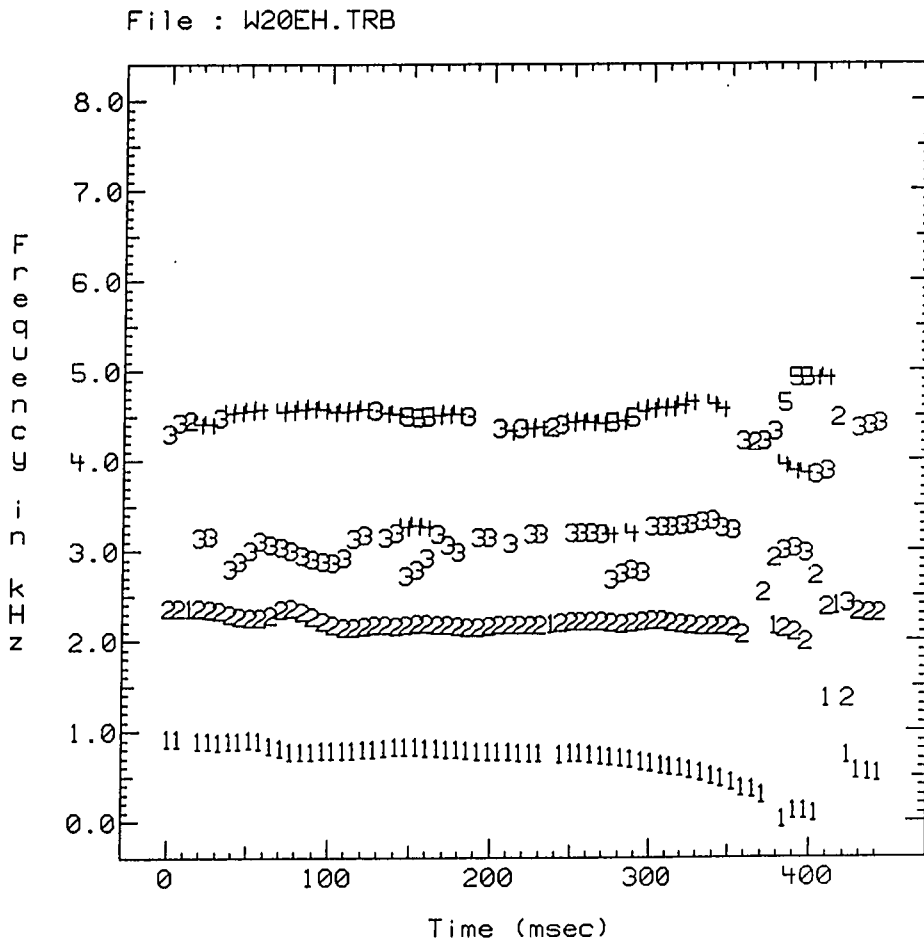


Figure 2. Original FTRACK Display of /E/.

Appendix F
Edited FTRACK Display of / ϵ /

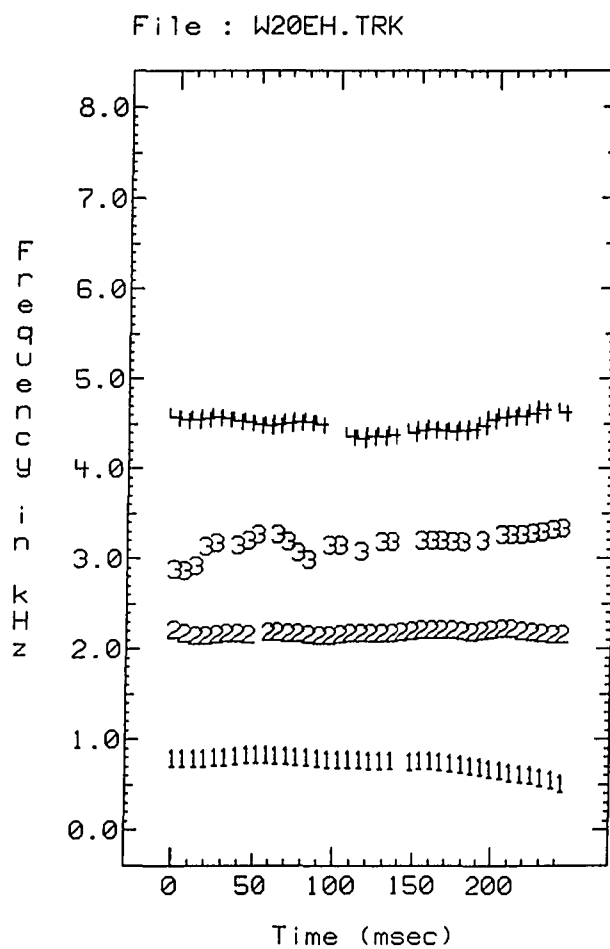


Figure 3. Edited FTRACK Display of /E/.

Appendix G
Gray-Scale Spectrogram of /ɛ/

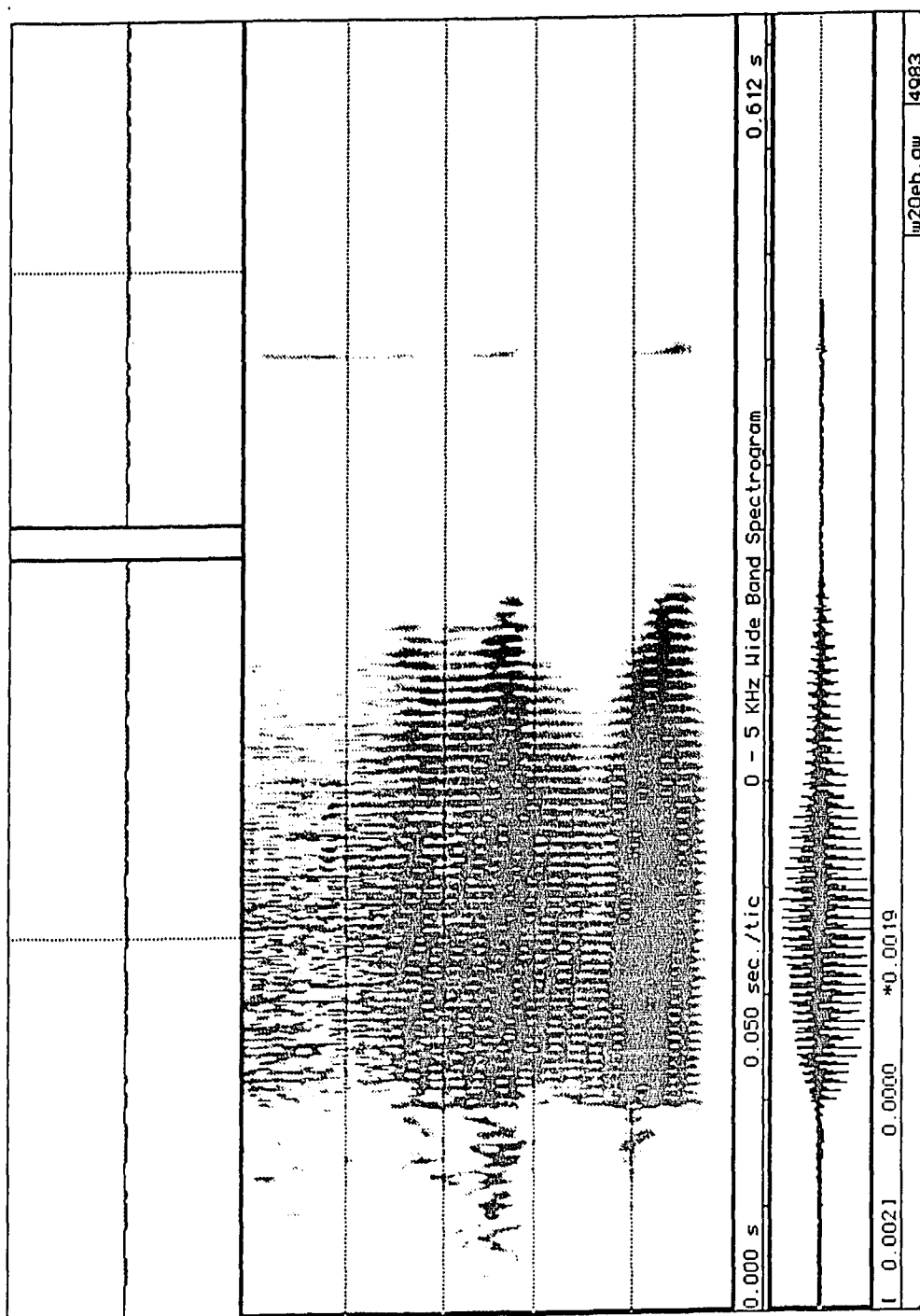


Figure 4. Gray-Scale Spectrogram of /ε/.

Appendix H
Formants Prior to Interpolation

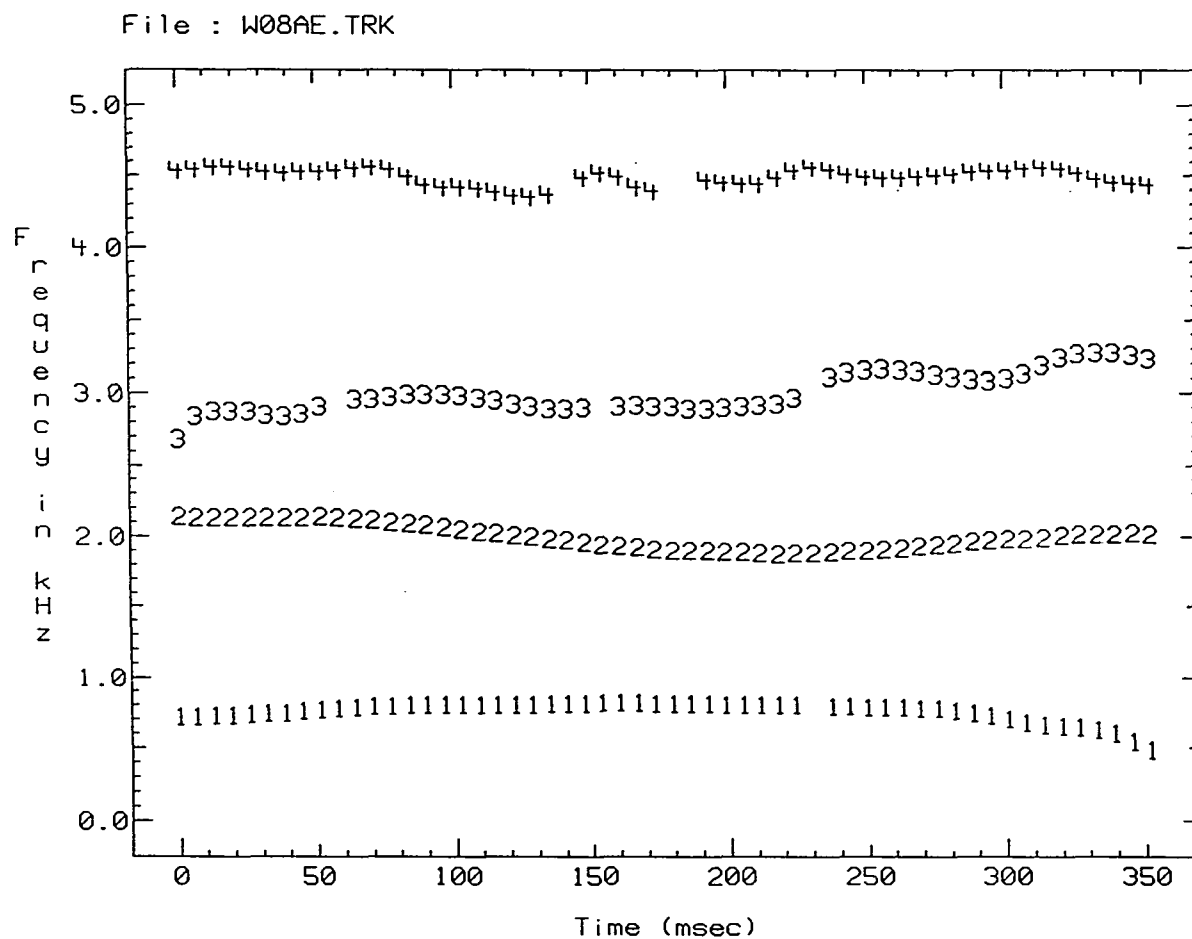


Figure 5. Formants Prior to Interpolation.

Appendix I
Formants Following Interpolation

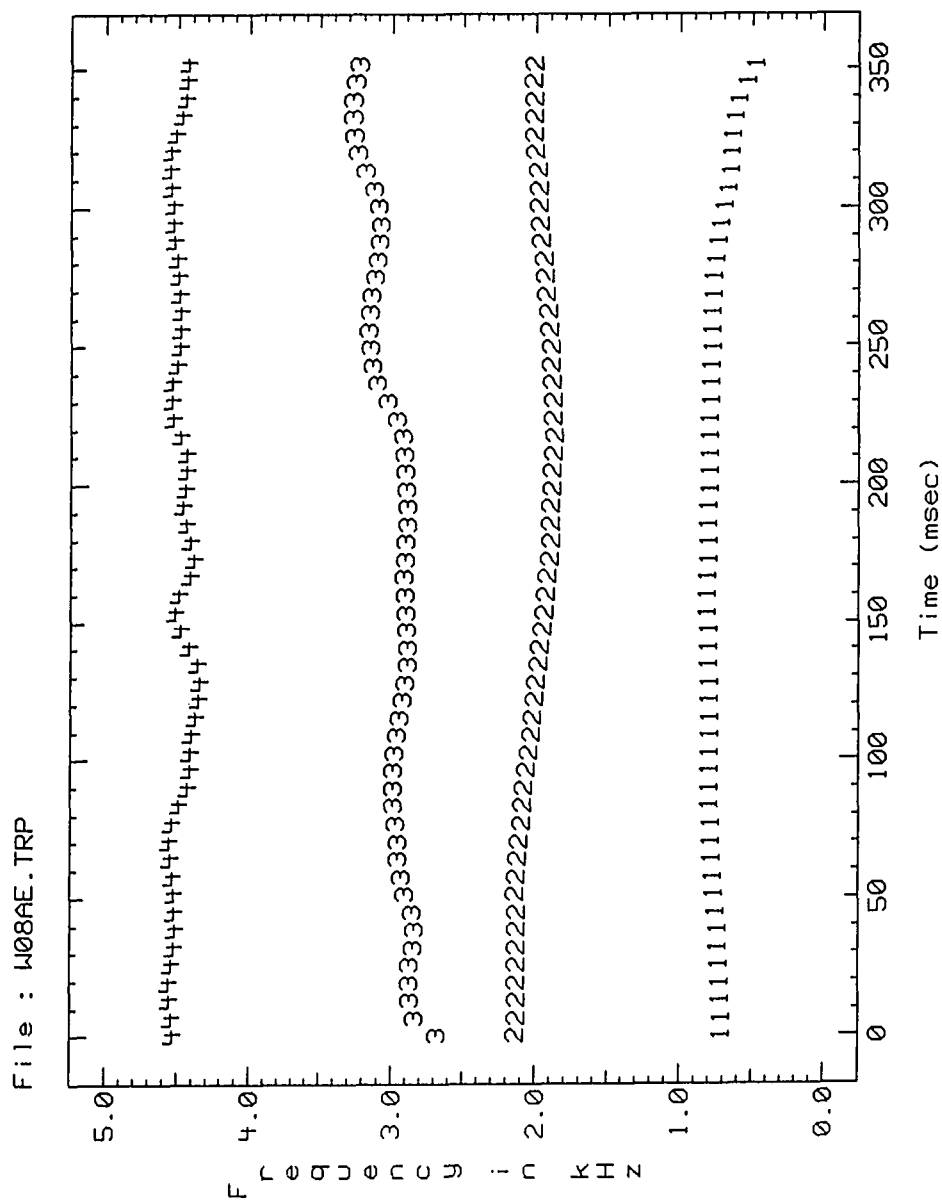


Figure 6. Formants Following Interpolation.

Appendix J
Vowel Data of /i/, /I/ and /ɛ/

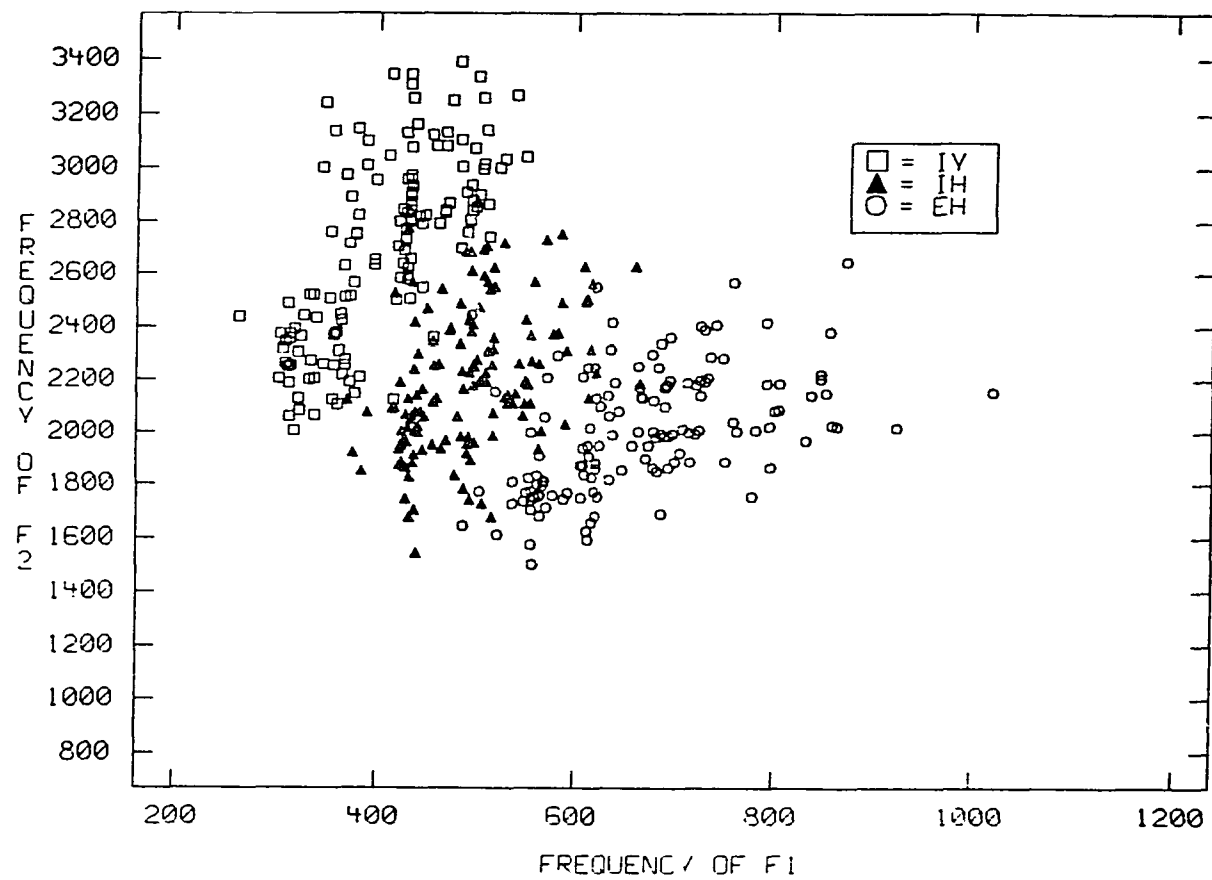


Figure 7. Vowel Data of /i/, /I/ and /E/.

Appendix K

Vowel Data of /ɛ/, /æ/ and /ʌ/

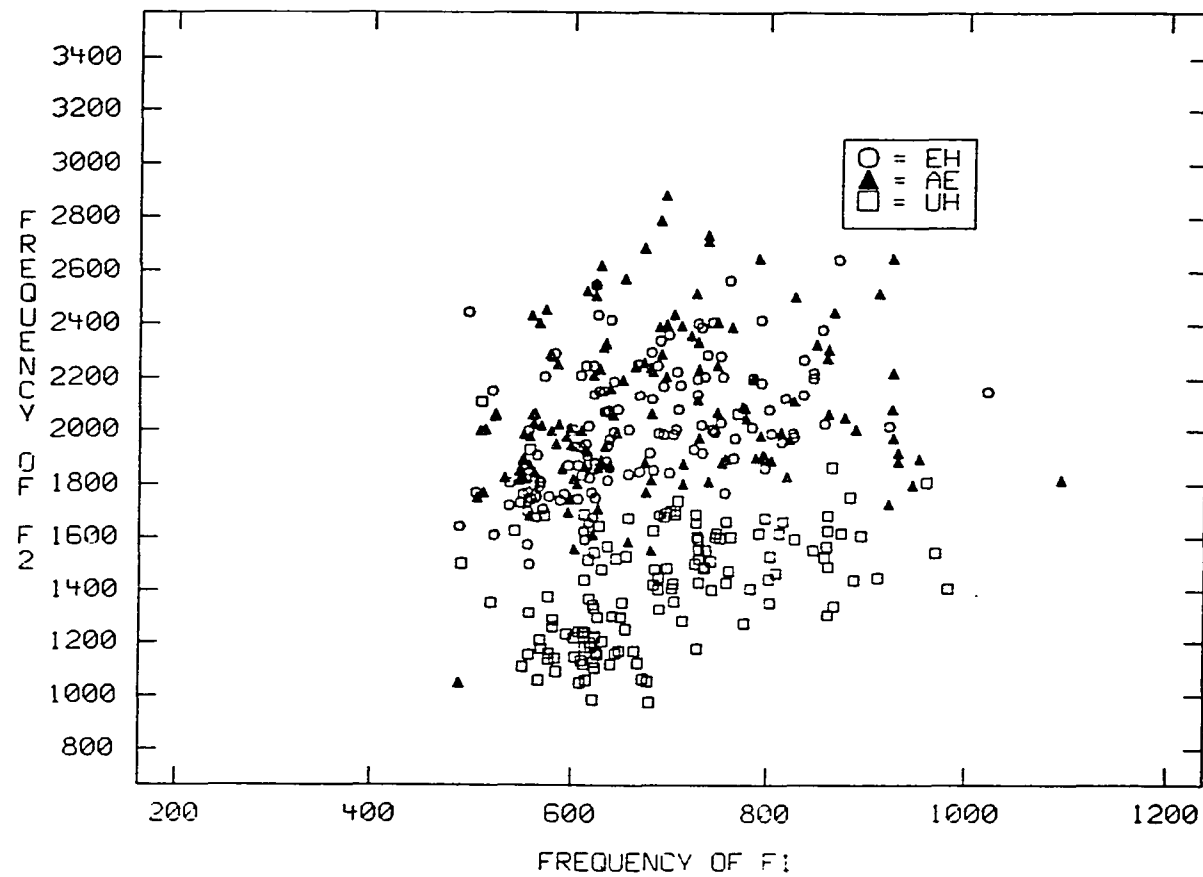


Figure 8. Vowel Data of /e/, /æ/ and /a/.

Appendix L

Vowel Data of /ʌ/, /ɑ/ and /ɔ/

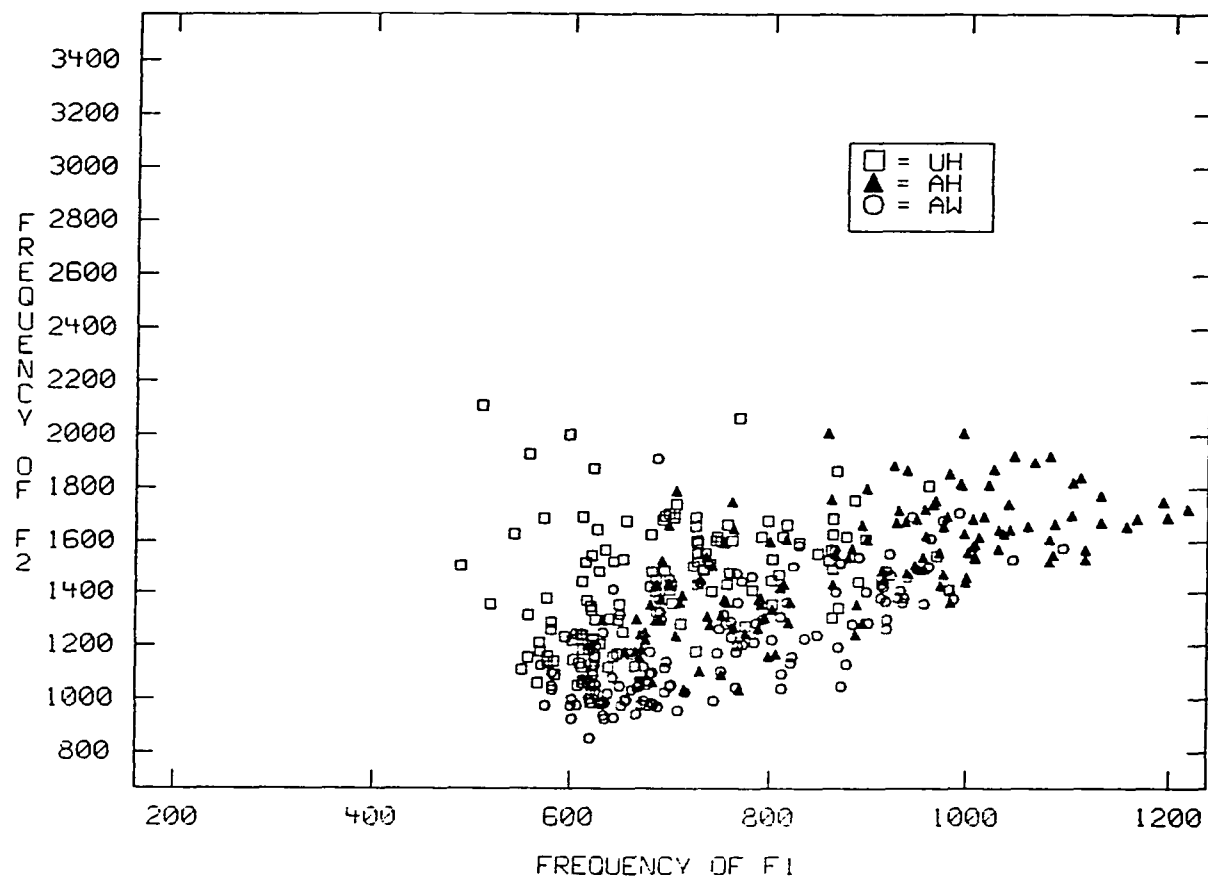


Figure 9. Vowel Data of /ʌ/, /a/ and /ɔ/.

Appendix M
Vowel Data of /ɔ/, /U/ and /u/

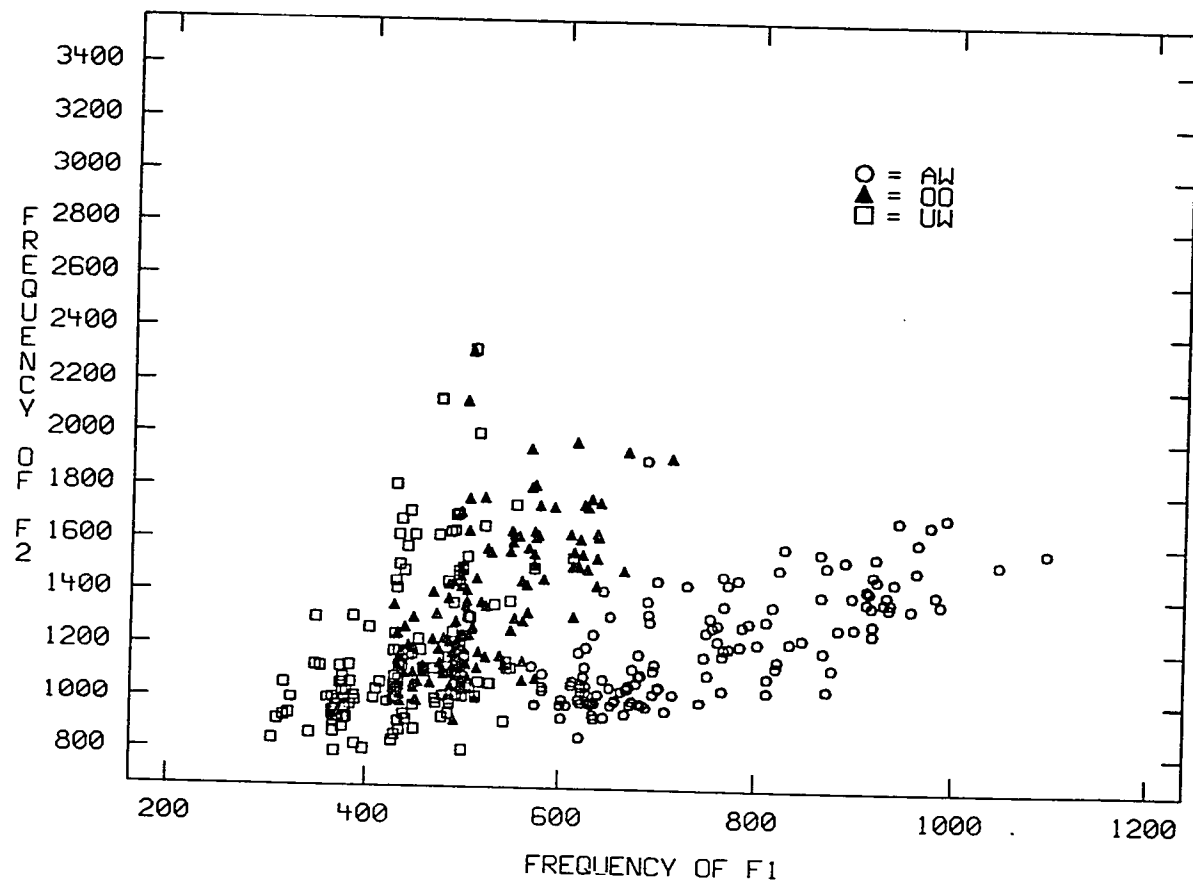


Figure 10. Vowel Data of /ɔ/, /U/ and /u/.

Appendix N

Vowel Data of /u/, /U/ and /ʊ/

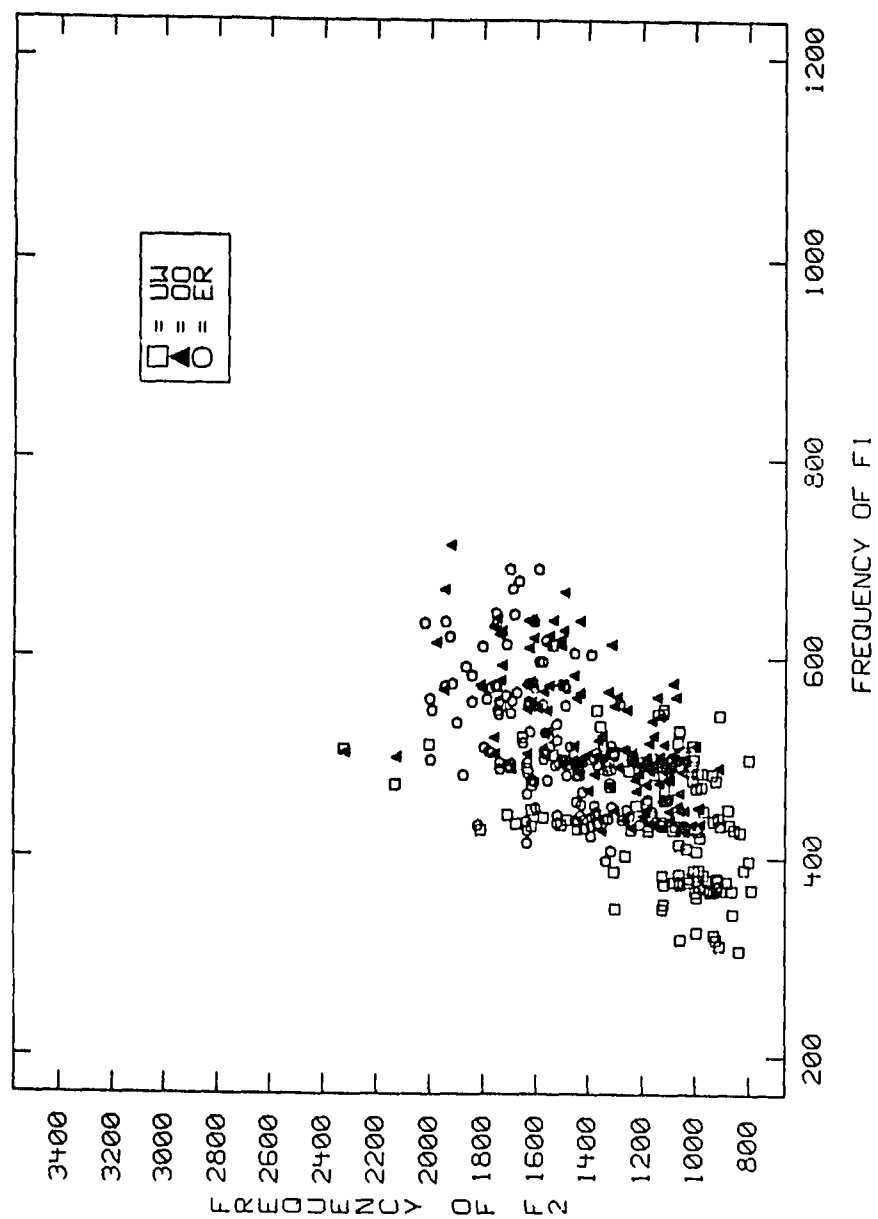


Figure 11. Vowel Data of /u/, /U/, and /ɜ/.

Appendix O
Original Productions of /ɛ/ and /æ/

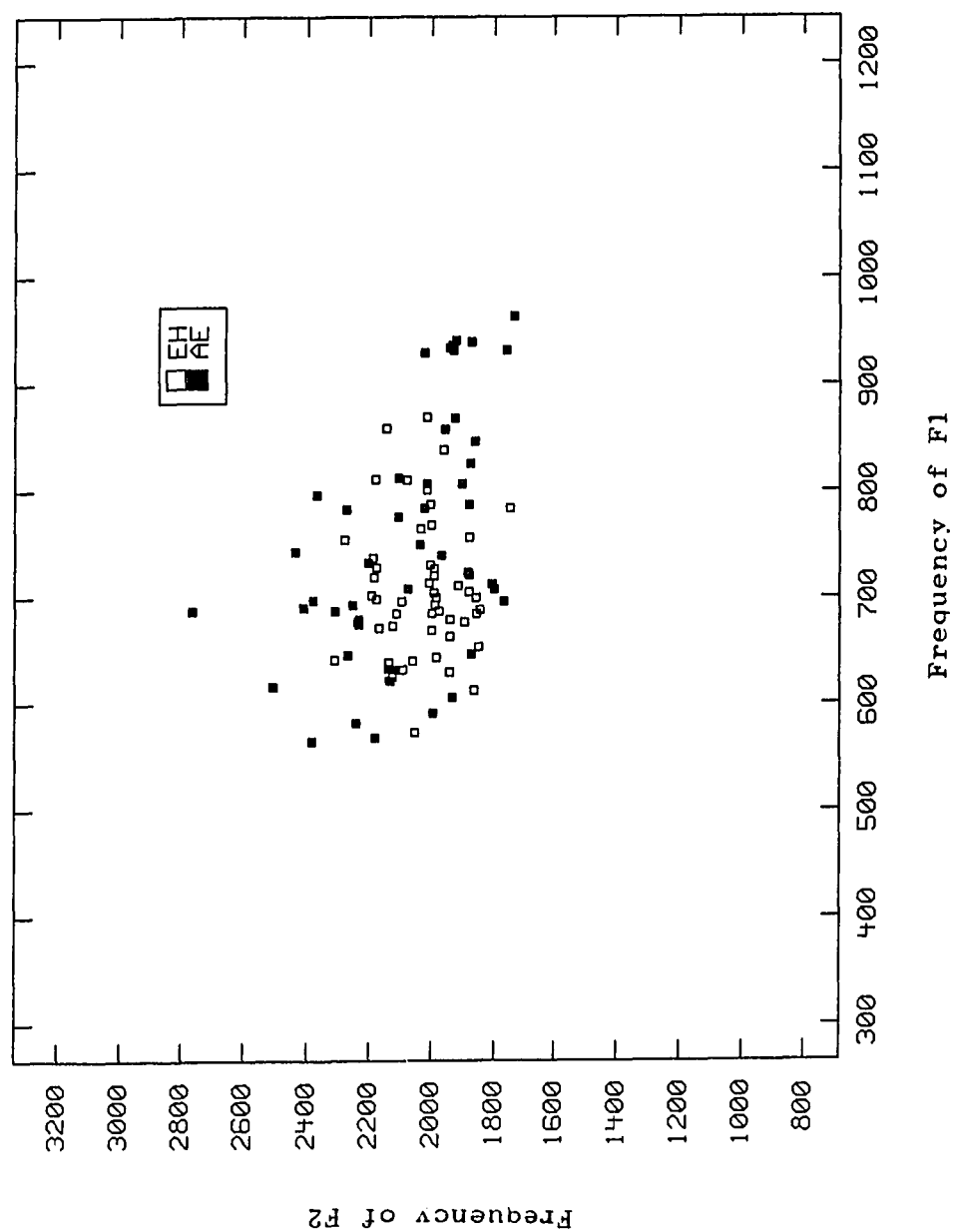


Figure 12. Original Productions of /ε/ and /æ/.

Appendix P
Remeasured Productions of /ɛ/ and /æ/

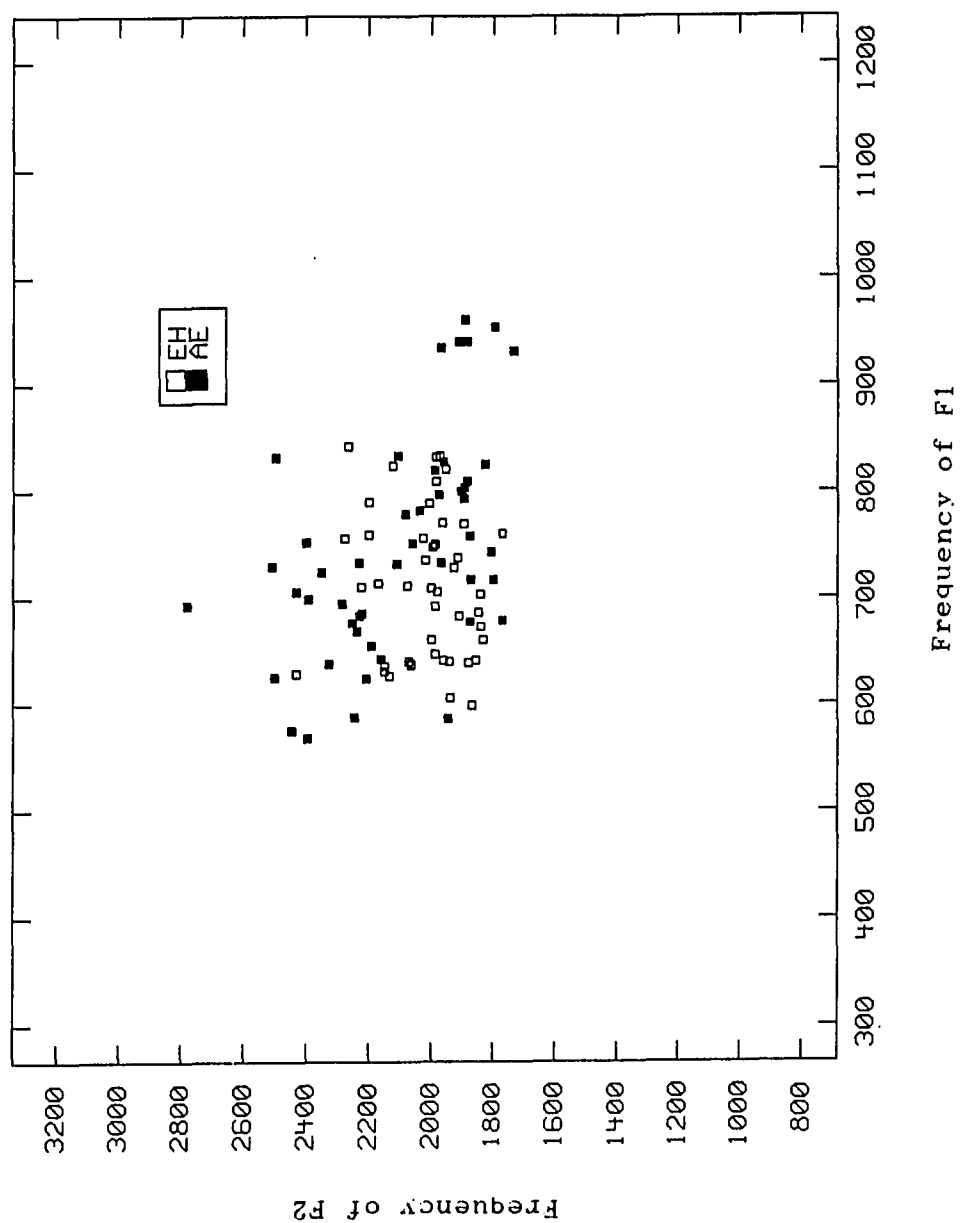
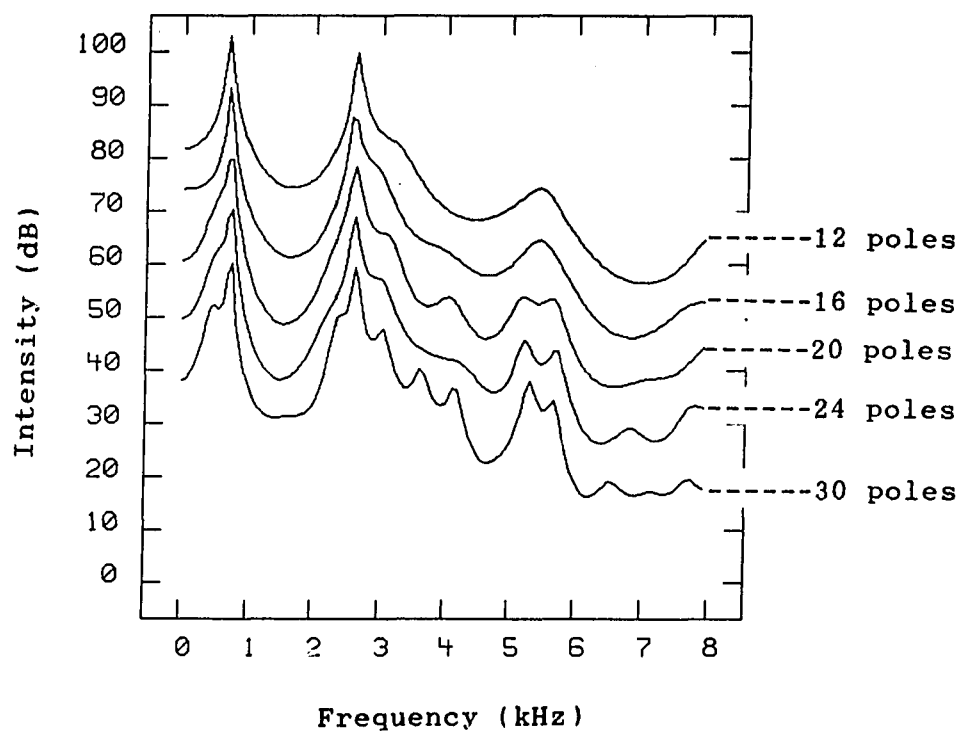


Figure 13. Remeasured Productions of /ε/ and /æ/.

Appendix Q
LPC Spectra of /æ/



** The spectra are offset from one another by 10 dB to facilitate comparisons; there are no overall energy differences

Figure 14. LPC Spectra of /æ/.

Appendix R
The /æ/ and /ɛ/ Experiment

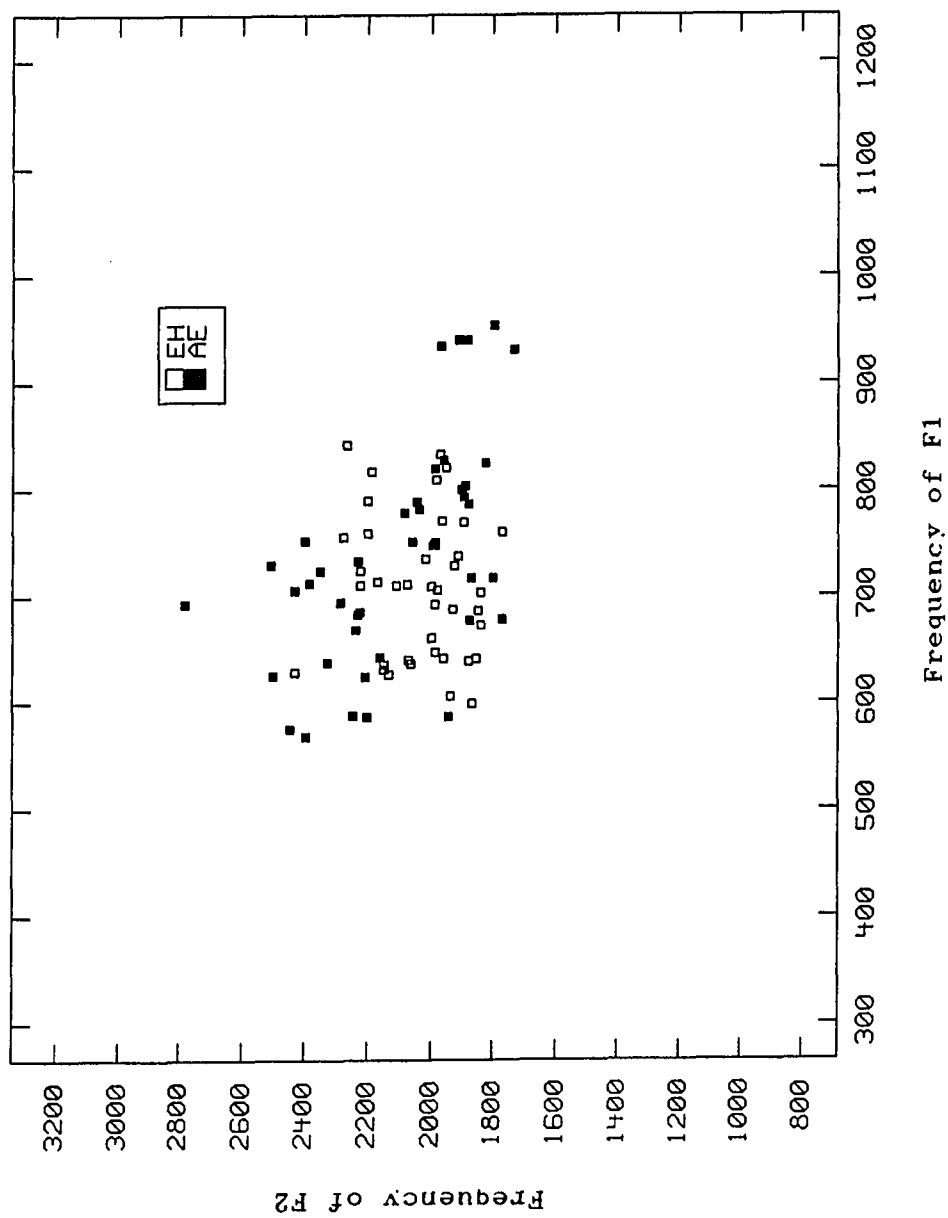


Figure 15. The /æ/ and /ɛ/ Experiment.

Appendix S
Consent Forms

Human Subjects Institutional Review Board

Kalamazoo, Michigan 49008-3899

WESTERN MICHIGAN UNIVERSITY

Date: October 4, 1989

To: James Hillenbrand

From: Mary Anne Bunde, Chair *Mary Anne Bunde*

This letter will serve as confirmation that your renewal application, "Evaluation of Auditory Models of Vowel Perception", submitted on 8-11-89, has been approved by the HSIRB. The conditions and duration of this approval are specified in the Policies of Western Michigan University. You may now begin to implement the research as described in the approval application. You must seek reapproval for any change in this design.

The Board wishes you success in the pursuit of your research goals.

Consent Form
Recording of Speech Samples

The purpose of this study is to learn something about how people understand and interpret speech sounds. The research will be applied to a number of practical problems, including the improvement of speech synthesis techniques, and the development of automatic speech recognition systems.

If you should choose to participate, your job will be to speak into a microphone and provide samples of speech which will be recorded on audio tape. You will be asked either to produce short words, a sustained vowel, or possibly a series of short sentences. You may also be asked to produce speech samples at a variety of different pitches. The acoustic characteristics of the speech samples will then be analyzed with the aid of a computer. In some cases, the speech samples may be played to a group of listeners who will make certain judgments about the speech (e.g., identifying the speech sound, or judging the similarity of two sounds). The recording session will last approximately 20 minutes.

In any written reports of this research, you will be identified by subject number only. Your name will not be made available to anyone other than project personnel. Although you will not directly benefit from this study, the information that is gained could lead to better speech synthesis and speech recognition techniques, and to the improvement of communication aids for the nonvocal. If you should consent to participate in this study, you may withdraw that consent at any time without penalty. Please feel free to ask questions now or at any time during the recording session. If you have questions at any time after the experiment, contact Dr. James Hillenbrand at 387-8066.

Signature of Investigator

Signature of Research Subject

Date

Department of Speech Pathology and Audiology
Charles VanRice Language, Speech and Hearing Clinic

Kalamazoo, Michigan 49008-3825
616 387-8045

WESTERN MICHIGAN UNIVERSITY

Dear Parent or Guardian,

December 20, 1989

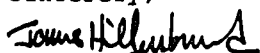
We are attempting to learn more about how people produce and understand speech. The particular focus of our research project has to do with understanding the similarities and differences between the speech of adults and young children.

The information we are interested in obtaining from your child will take approximately 20 minutes. During his/her participation, your child will be asked a set of identification questions (name, age, birthdate, etc.). The child will then be engaged in informal conversation for 3 to 5 minutes. An example of such a conversation may begin "Tell me what you did over summer vacation." The interview and conversation will be recorded on cassette tape to be heard only by persons directly involved with this project. Your child will then read a list of 32 words which will be recorded on tape. The tape will later be analyzed by a computer which will make a variety of acoustic measurements from the speech sample. This procedure involves no known hazards or risks.


All of the information gathered from your child will be held in confidence and at no point will the information be associated with your child by name. If at any time you should choose not to allow your child to continue participation in this project, you are free to withdraw your consent. Should you have any questions regarding this project or this consent, please feel free to call Dr. James Hillenbrand, advisor of this research, at Western Michigan University at 387-8066 between 9 a.m. and 5 p.m. Monday through Friday. You may also feel free to contact Laura Getty, who will be collecting the information from your child, at 387-8067, from 9 a.m. to 5 p.m. Monday through Friday.

If you are willing to allow your child to participate in this project, please fill out the attached sheet and give it to your child to give to his/her classroom teacher. Thank you for your cooperation.

Sincerely,



James Hillenbrand
Associate Professor
Department of Speech Pathology and Audiology



Laura Getty
Graduate Student
Department of Speech Pathology and Audiology

Graduate Programs Accredited by Educational Standards Board, American Speech-Language-Hearing Association

Department of Speech Pathology and Audiology
Charles VanRiper Language Speech and Hearing Clinic

Kalamazoo, Michigan 49008-3825
616 387-8045

WESTERN MICHIGAN UNIVERSITY

I, _____, hereby give my permission for
my child _____, to participate in the
project described in the attached letter. I have read and
understand this statement and I have had all my questions
answered.

Signature: _____ Date: _____

Graduate Programs Accredited by Educational Standards Board, American Speech-Language-Hearing Association

BIBLIOGRAPHY

- Assmann, P. F., Nearey, T. M., & Hogan, J. T. (1982). Vowel identification: Orthographic, perceptual and acoustic aspects. The Journal of the Acoustical Society of America, 71, 975-989.
- Bennett, S. (1983). A 3-year longitudinal study of school-aged children's fundamental frequencies. Journal of Speech and Hearing Research, 26, 137-142.
- Bernstein-Ratner, N. (1985). Dissociations between vowel durations and formant frequency characteristics. Journal of Speech and Hearing Research, 28, 255-264.
- Boone, D. R., & McFarlane, S. C. (1988). The voice and voice therapy (4th ed.). Englewood, NJ: Prentice-Hall.
- Borden, G. J., & Harris, K. S. (1984). Speech science primer: Physiology, acoustics, and perception of speech (2nd ed.). Baltimore, MD: Williams and Wilkins.
- Carlson, R., Fant, G., & Granstrom, B. (1975). Two-formant models, pitch and vowel perception. In G. Fant & M. A. A. Tatham (Eds.), Auditory analysis and perception of speech (pp. 55-82). London: Academic Press.
- Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. (1952). Some experiments on the perception of synthetic speech sounds. The Journal of the Acoustical Society of America, 24, 597-606.
- Diehl, R. L., McCusker, S. B. & Chapman, L. S. (1981). Perceiving vowels in isolation and in consonantal context. The Journal of the Acoustical Society of America, 69, 581-593.
- Disner, S. F. (1980). Evaluation of vowel normalization procedures. The Journal of the Acoustical Society of America, 67, 253-261.
- Fant, G. & Tatham, M. A. A. (1975). Auditory analysis and perception of speech. London: Academic Press.
- Gay, T. (1978). Effect of speaking rate on vowel formant movements. The Journal of the Acoustical Society of America, 63, 223-230.

- Gerstman, L. J. (1968). Classification of self-normalized vowels. IEEE Transactions on Audio and Electroacoustics, AU-16, 78-80.
- Grundner, T.M. (1986). Informed consent: A tutorial. Baltimore, MD: National Health Publishing.
- Hillenbrand, J. M. & Gayvert, R. T. (1988). Speaker-independent vowel classification based on fundamental and formant frequencies. (Northeast Artificial Intelligence Consortium, Technical paper 8726.) Rochester, NY: Author.
- Hillenbrand, J. & McMahon, B. J. (1988). The role of static spectral properties in vowel identification. (Northeast Artificial Intelligence Consortium, Technical paper 8728.) Rochester, NY: Author.
- Hollien, H., Malcic, E., & Hollien, B. (1965). Adolescent voice change in southern white males. Speech Monographs, 32, 87-90.
- Klatt, D. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), Perception and production of fluent speech (pp. 243-288). Hillsdale, NJ: Erlbaum.
- Lieberman, P. & Blumstein, S. (1988). Speech physiology, speech perception and acoustic phonetics. New York: Cambridge University Press.
- Lindblom, B. (1963). Spectrographic study of vowel reduction. The Journal of the Acoustical Society of America, 35, 1773-1781.
- Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. The Journal of the Acoustical Society of America, 85, 2114-2133.
- Miller, R. L. (1953). Auditory tests with synthetic vowels. The Journal of the Acoustical Society of America, 25, 114-121.
- Nearey, T. M. & Assmann, P. F. (1986). Modeling the role of inherent spectral change in vowel identification. The Journal of the Acoustical Society of America, 80, 1297-1308.
- Peterson, G. E. & Barney, H. L. (1952). Control methods used in a study of the vowels. The Journal of the Acoustical Society of America, 24, 175-184.

- Peterson, G. & Lehiste, I. (1960). Duration of syllable nuclei in English. The Journal of the Acoustical Society of America, 32, 693-703.
- Potter, R., Kopp, G. & Green, H. (1947). Visible speech. New York: Van Nostrand.
- Potter, R.K. & Steinberg, J.C. (1950). Toward the specification of speech. The Journal of the Acoustical Society of America, 22, 807-820.
- Prall, C. W. & Hillenbrand, J. M. (1981). AUDED: A general purpose waveform editor [computer program]. Evanston, IL: Northwestern University.
- Smith, B. L. (1978). Temporal aspects of English speech production: A developmental perspective. Journal of Phonetics, 6, 37-67.
- Smith, S. (1989). FTRACK: A formant frequency editor [computer program]. Birmingham: University of Alabama at Birmingham.
- Strange, W. (1989a). Evolving theories of vowel perception. The Journal of the Acoustical Society of America, 85, 2081-2087.
- Strange, W. (1989b). Dynamic specification of coarticulated vowels spoken in sentence context. The Journal of the Acoustical Society of America, 85, 2135-2153.
- Strange, W., Verbrugge, R. R., Shankweiler, D. P. & Edman, T. R. (1976). Consonant environment specifies vowel identity. The Journal of the Acoustical Society of America, 60, 312-224.
- Syrdal, A. K. & Gopal, H. S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. The Journal of the Acoustical Society of America, 79, 1086-1100.
- Tiffany, W. R. (1959). Nonrandom sources of variation in vowel quality. Journal of Speech and Hearing Research, 2, 305-317.
- Verbrugge, R. R., Strange, W., Shankweiler, D. P. & Edman, T. R. (1976). What information enables a listener to map a talker's vowel space? The Journal of the Acoustical Society of America, 60, 198-212.