7-2-1996

Investigating Vowel Duration as a Perceptual Cue to Voicing in the English of Native Spanish Speakers

Becky Jean George
Portland State University

THESIS APPROVAL

The abstract and thesis of Becky Jean George for the Master of Arts in TESOL were presented July 2, 1996, and accepted by the thesis committee and the department.

COMMITTEE APPROVALS: [redacted]

Beatrice Oshika, Chair

[redacted]

Thomas Dieterich

[redacted]

John Tetnowski
Representative of the Office of Graduate Studies

DEPARTMENT APPROVAL: [redacted]

Beatrice Oshika, Chair
Department of Applied Linguistics

* *

ACCEPTED FOR PORTLAND STATE UNIVERSITY BY THE LIBRARY

by [redacted] on 21 August 1996

ABSTRACT

An abstract of the thesis of Becky Jean George for the Master of Arts in TESOL presented July 2, 1996.

Title: Investigating Vowel Duration as a Perceptual Cue to Voicing in the English of Native Spanish Speakers.

Researchers in the cognitive sciences, and in particular those in acoustic phonetics, investigate the acoustic properties in the speech signal that enable listeners to perceive particular speech sounds. Temporal cues have been found to convey information about the linguistic content of an utterance. One acoustic characteristic that is particularly well documented in American English is the difference in vowel duration preceding voiced and voiceless consonants, which has been found to play a role in the perception of the voicing of postvocalic word-final consonants. Research on vowel duration and its role in the perception of the voicing distinction of the following consonant has primarily involved data from native English speakers.

The purpose of the present study was to investigate the vowel durations preceding word-final voiced and voiceless stops in the English production of four native Spanish speakers. This study sought to determine if differences in vowel duration are exhibited preceding voiced and voiceless consonants in the English production of the native Spanish speakers, and to determine if the vowel durations affected the perception of the voicing distinction of the postvocalic stop by four native English speakers.

A significant effect of voicing on the vowel durations in the English production of the native Spanish speakers was found. However, the degree of variation in the vowel lengths with respect to voicing was much less than the degree of difference exhibited in native English, and similar to the variation produced in native Spanish. The average mean difference in length with respect to the voicing of the following consonant was 17.8 msec. in the present study. In native English the mean difference between vowels preceding voiced and voiceless consonants ranges from 79 msec. to 92 msec. and in Spanish the average mean difference is 18 msec. Statistical analysis performed to quantify the contribution of vowel duration on the perception of the voicing distinction found only minimal affect. It was concluded that although the cue of vowel duration variation was present in the speech signal of this data, the listeners generally did not utilize it as a cue to the voicing distinction of the following stops.

INVESTIGATING VOWEL DURATION AS A PERCEPTUAL CUE TO VOICING IN THE ENGLISH OF NATIVE SPANISH SPEAKERS

by

BECKY JEAN GEORGE

A thesis submitted in partial fulfillment of the requirements for the degree of

MASTER OF ARTS
in
TESOL

Portland State University
1996

TABLE OF CONTENTS

ACKNOWLEDGMENTS iii

LIST OF TABLES iv

LIST OF FIGURES v

CHAPTER

I INTRODUCTION 1

Statement of the Problem 1
Background 2
Purpose of this Study 5
Research Questions 6
Definition of Terms 7

II REVIEW OF THE LITERATURE 9

Vowel Duration Preceding Voiced/Voiceless Consonants 9
Other Factors Influencing Vowel Duration 12
Speech Perception 16
Vowel Duration as a Perceptual Cue to Voicing 18
Vowel Duration Cross-Linguistically 26
Measurement and Segmentation 29
Summary 31

III RESEARCH DESIGN AND METHODOLOGY 34

Subjects 34
Stimuli 35
Data Collection 37
Instrumentation 38
Segmentation 39
Reliability of Segmentation 43
Phonetic Transcribers 43
Transcription Procedures 43
Reliability of Transcription 44
Data Analysis 45

IV RESULTS 46

Vowel Duration and the Voiced-Voiceless Influence 46
Inherent Vowel Duration 51
Reliability of Vowel Measurements 53
Transcription Analysis-Perception of the Stops 58
Reliability of Transcriptions 59
Effect of Vowel Duration on Perceived Voicing of Stops 63
Summary 66

V DISCUSSION AND CONCLUSIONS 68

Discussion and Conclusions 68
Implications 71
Limitations of the Study 72
Directions for Future Research 73

REFERENCES 75

APPENDIX A 79

ACKNOWLEDGMENTS

I would like to thank the following people for their support:

Members of the thesis committee, Beatrice Oshika, John Tetnowski, and Tom Dieterich. Thank you for your guidance, patience, helpful comments and encouragement. An additional thank you to John Tetnowski for allowing me to use the lab in the Speech and Hearing Department for the acoustical analysis in this study, and for the time spent teaching me about the equipment.

Transcribers for this study, Victoria Chung, Randy Evans, and Annette Murphy. Thank you for your support, interest, lively discussion and friendship along the way.

John and Jane George and Terri and Jerry Spinak. Thank you for your tremendous support and assistance at home. This was truly a joint effort.

Sam, Nick and especially Scott Spinak. Thank you for taking it all in stride. And Scott, for your unwavering encouragement, endless patience and for understanding why.

LIST OF TABLES

1. List of Voiced-Voiceless Oppositions Tested 36
2. Analysis of Variance for Vowel Duration - Effect of Voicing 47
3. Combined Mean Duration of Vowel in Msec. Before Voiced - Voiceless Stops 48
4. Mean Duration of Vowels in Msec. by Vowel Type 51
5. Vowel Durations at Time 1, Time 2 and Time X 54
6. Range of Vowel Measurements Between Time 1 and Time 2 55
7. Range of Vowel Measurements Between T1 and TX and T2 and TX 56
8. Perception of the Voiced - Voiceless Distinction by Transcribers 58
9. Kappa Values of Agreement Between Transcribers 61
10. Kappa Values of Agreement Between Transcribers and Target 62
11. Logistic Regression Analysis of Effects of Vowel Duration on Perception of Voicing by Transcriber 65
12. Logistic Regression Analysis of Effects of Vowel Duration on Perception of Voicing by Aggregate 66

LIST OF FIGURES

1. Waveform of the word pick 41
2. Waveform of the word mop 41
3. Waveform of the word fig 42
4. Mean vowel length variation by voicing of the following consonant 49
5. Correlation of vowel measurements between Time 1 and Time 2 55
6. Correlation of vowel measurements between Time 1 and Time X 57
7. Correlation of vowel measurements between Time 2 and Time X 57

CHAPTER I

INTRODUCTION

Humans engage in conversation and typically understand each other without much effort. Yet, the speech signal is a complex one that varies with regard to the speaker, the rate of speech, and even environmental noise. Given this variation, how is it that speech is consistently and reliably recognized? Researchers in the cognitive sciences, and in particular those in acoustic phonetics, have investigated the acoustic properties in the speech signal that enable listeners to perceive particular speech sounds. Temporal cues, such as the pattern of durations of individual phonetic segments, among other acoustic cues, have been found to convey information about the linguistic content of an utterance.

Statement of the Problem

One acoustic characteristic that is particularly well documented in American English is the difference in vowel duration preceding voiced and voiceless consonants (House, 1961; House & Fairbanks, 1953; Peterson & Lehiste, 1960). The length of the vowel in words ending in a voiced consonant is longer than the length of the vowel in words ending in the voiceless consonant counterpart. This effect is most pronounced in phrase-final position or in single syllable words (Klatt, 1976; Oller, 1973), and various researchers have found that the vowel duration can play a significant role in the perception of the voicing of postvocalic word-final consonants (Denes, 1955; Klatt, 1976; Malecot, 1970; Raphael, 1972).

Research on vowel duration and its role in the perception of the voicing distinction of the following consonant has primarily involved data from native English speakers. A few studies have investigated whether differences in vowel duration occur cross-linguistically. However, there is a lack of studies in the literature that have examined vowel duration in the English production of nonnative speakers. Do nonnative speakers of English exhibit the same degree of vowel duration differences preceding voiced and voiceless consonants that native English speakers exhibit? And how, if at all, might this affect perception of the target sounds?

This research investigates the vowel durations in the English production of native Spanish speakers (hereafter NSSs) and how the vowel durations may affect the perception of the voicing of the following consonants. Specifically, this research analyzes the vowel durations preceding target word-final voiced and voiceless stops in the English production of NSSs, and investigates the perception of the voicing distinction of the adjacent stops by native English speakers.

Background

The search for acoustic correlates to phonetic segments continues to be a live issue in speech perception and acoustic phonetics, and a current topic of inquiry. There is still much debate over which particular acoustic cues, or complex of cues, may signal any given phonetic segment. Researchers have investigated both spectral and temporal cues present in the speech wave in an attempt to determine the relative contribution of these various cues to a listener's perceptual judgments. Denes (1955) notes:

> In the past the tendency was to relate phonemic differences solely with spectral characteristics of the speech sound wave. It is recognized more and more that some nonspectral physical characteristics of the speech wave and context will also influence phonemic judgments. (p. 764)

The vowel duration differences in English preceding voiced and voiceless word final consonants are well documented in the literature, but the importance of these differences in vowel duration with regard to the perception of the voicing of the following consonant is debated. Many researchers have found the vowel duration to be a primary acoustic cue to the voicing distinction in the word-final or phrase-final position (Denes, 1955; Klatt, 1976; Malecot, 1970; Raphael, 1972). However Wardrip-Fruin (1982) argues, "A review of the literature indicates that, although vowel duration differences are very reliably produced, their role in perception is not as predictable" (p. 187).

More commonly, researchers have suggested that vowel duration is one acoustic cue among a number of different cues that contribute to the voicing distinction of the following consonant (Hillenbrand, Ingrisano, Smith & Flege, 1984; Hogan & Rozsypal, 1980; Wolf, 1978). However, Luce and Luce (1978) suggest that vowel duration cues may be the most consistently and reliably produced of the various cues that have been found to play significant roles in the perception of voicing of a postvocalic consonant. In addition, Wardrip-Fruin (1982) reports evidence of a trading relationship between the acoustic cues that signal the voicing of final stops. If one or more cues are missing, which is often the case in natural connected speech, Wardrip-Fruin suggests the remaining cue or cues will be given more weight in the listener's judgment of the voicing distinction. Whether the vowel durations are a primary or secondary cue to the voicing distinction in postvocalic, word-final consonants, it is nevertheless clear that this temporal correlate of voicing plays some role in the perception of voicing, especially when other cues are missing or ambiguous.

Taken together, the research suggests that native English speakers utilize the vowel duration cue in some capacity to judge voicing of final consonants. Investigating the English production of nonnative speakers of English to determine if this cue is present or lacking in the acoustic signal is important, since this may have some bearing on the perception of the voicing distinction of the postvocalic consonants by native English speakers. It is also germane to investigate whether the lengthening-before-voicing effect occurs in other languages, particularly Spanish.

A few studies have investigated the vowel duration differences exhibited cross-linguistically (Chen, 1970; Delattre, 1962; Zimmerman & Sapon, 1957). Chen reports that there is a tendency across languages for the vowel to become longer before voiced consonants and shorter before voiceless ones. But he notes that "...the voicing of the adjacent consonant influences its preceding vowel to different degrees in different languages" (p. 138). In fact, the research suggests that the vowel duration differences preceding voiced and voiceless consonants are much more pronounced in English than in other languages examined. For example, in his study Chen (1970) reported that the mean difference between the vowel durations before voiced and voiceless consonants in

English was 92 milliseconds (hereafter msec.), but in Spanish the mean difference reported by Zimmerman and Sapon (1957) was just 18 msec.

In light of this, it is useful to examine the English production of native Spanish speakers (NSSs) to determine if they utilize this temporal cue as native English speakers do. If NSSs do not exhibit the same degree of vowel duration differences as native English speakers do, it could result in perception errors or ambiguities by the listener with respect to the voicing of the following consonant.

Results of this type of research are useful in teaching pronunciation to nonnative speakers of English, or more specifically in the field of accent reduction. Accent reduction is important for many speakers of another language since the degree of accent a nonnative speaker exhibits is often a determining factor in the first impression conveyed by that speaker.

Purpose of This Study

This study has two major goals. One is to investigate the English productions of NSSs to determine if they exhibit the same degree of vowel duration differences, if any, preceding voiced and voiceless consonants that native English speakers exhibit. An examination of the vowel durations preceding voiced and voiceless consonants in the English production of NSSs may provide insight as to the strategies that these NSSs utilize to cue their listeners to the target sounds.

Another aim of this study is to examine how native English speakers perceive the postvocalic voiced and voiceless stops in the English production of NSSs. This research will attempt to determine if there is a correlation between the vowel durations produced and the perceived voicing of the target stops. This may provide insight as to whether the strategies that the NSSs employ to cue their listeners to the voicing distinction of the final consonants are effective. The information obtained for both the production and perception components of this study may be useful in ESL pedagogy, as well as in the field of accent reduction.

Research Questions

This study investigates both *production* and *perception* of speech by examining the English productions of NSSs and the perception of that speech by native speakers of English. In particular, this study focuses on the vowel durations produced and the following stop consonants and how they are perceived. The specific research questions addressed by this study are:

1. Is there a difference in the vowel durations preceding word-final voiced and voiceless stops in the English production of NSSs, and if so, what degree of difference is exhibited?
2. Is there a correlation between the vowel durations and the native English speakers' perception of the voiced/voiceless distinction of the stops in the English speech of NSSs?

Definition of Terms

The following terms used throughout this study are more specifically defined below:

Native English Speaker:	A speaker whose native language is English. (NES)
Native Spanish Speaker:	A speaker whose native language is Spanish. (NSS)
Nonnative Speaker:	A speaker whose native language is other than English.
Vowel Duration:	The length of a vowel (in msec.) as measured by standard criteria using digital acoustic analysis instrumentation.
Voiced:	A segment is said to be voiced if the vocal cords are vibrating during an articulation of the segment. The notion of voiced also includes a measure of voice onset time (VOT) for stops. VOT refers to the span of time between the release of stop closure when the vocal folds are open, to the onset of voicing, when the vocal folds start vibrating. A short VOT indicates a voiced stop because the vocal folds start vibrating before full release of the closure.
Voiceless:	A segment is voiceless if pronounced without vibration of the vocal cords. As with voiced stops, the notion of a voiceless stop includes a measure of VOT. A longer VOT is associated with voiceless stops. The voicing starts at

the moment of release of the closure, or later, which indicates an aspirated voiceless stop.

Stop: A speech sound characterized by a complete obstruction of the vocal tract that is usually followed by an abrupt release of air. Also referred to as plosives.

Prepausal lengthening: Lengthening of a vowel in the syllable preceding a phrase boundary or a pause.

Digitization: The process of converting an analog continuous signal to a digital discrete form. A digital signal is represented as a sequence of numbers and can be stored in a digital computer.

Waveform: A graph showing the amplitude versus time function for a continuous signal such as the acoustic speech signal. Also referred to as an oscillogram or oscillographic display.

Sampling Rate: The number of times an analog signal is sampled per second during analog-to-digital conversion. The sampling occurs at periodically spaced points along the signal. A sampling rate of 10 kHz means that the analog signal is sampled 10,000 times per second. A sampling rate of twice the frequency range is required to completely represent a signal within that range. For example, to reconstruct a 5kHz range, a sampling rate of 10 kHz is necessary.

CHAPTER II

REVIEW OF THE LITERATURE

This chapter provides an overview of the research on vowel durations in American English, particularly in the environment preceding voiced and voiceless consonants. Factors other than voicing that influence vowel duration are discussed in order to provide a proper perspective for the current study. A section on speech perception is included to provide background for the discussion on research investigating the perceptual significance of vowel duration. In addition, research on the cross-linguistic validity of vowel duration is examined because of its relevance to this study involving nonnative speech.

Vowel Duration Preceding Voiced/Voiceless Consonants

A large body of acoustic research has established that in American English, the vowel nucleus is longer preceding a voiced consonant in word-final position than the vowel nucleus preceding its voiceless cognate. This phenomenon has been found across a wide variety of studies and phonetic environments.

House and Fairbanks (1953) reported that vowel duration varied systematically with the voicing of the following consonant. Their study compared the duration, fundamental frequency, and relative power of vowels. The stimuli in their analysis, read by ten subjects, consisted of 72 bisyllabic items consisting of 12 consonants, including the voiced and voiceless stops, fricatives and nasals, paired with six vowels. Each item had only one consonant, and was prefixed with the unstressed syllable [hΛ] to render items such as 'hupeep'. House and Fairbanks found that the consonant environment significantly influenced all the acoustical characteristics examined. However, when all three characteristics were compared, the effects of voicing were greatest. The vowels preceding voiced consonants were longer than those preceding the voiceless consonants in every case. When the values were pooled with respect to voicing, there was a statistically significant difference of .079 sec between the two means. In addition, the mean difference between the vowels lengths preceding stops, with respect to voicing, was 83.3 msec., and the mean difference overall, including all the consonants, was 79 msec. In a subsequent study, House (1961), confirmed that "The average duration of vowels varies as a function of the phonetic environment. The primary influence is contributed by the voicing characteristic of the following consonant" (p. 1175).

Denes (1955) examined the vowel durations preceding the fricatives in the words [jus] and [juz] produced by a number of speakers. The range of vowel durations preceding the voiceless and voiced final consonants support the claims of House and Fairbanks (1953) and House (1961). Malecot (1969) investigated vowels preceding word-final consonants in connected speech and found the duration differences to be consistent with the previously discussed studies as well. In addition, he discovered that the duration of vowels varied inversely with the amount of energy the subjects perceived was needed to articulate the consonant. Voiceless consonants were reported by subjects to require more energy to produce, and the vowels preceding them were consistently shorter than their voiced counterparts. Malecot reported vowel duration to be a significant indicator at .01 for consonant class in word-final position.

Peterson and Lehiste (1960) also found evidence that the voicing contrast of consonants influenced the preceding vowel duration. They recorded and acoustically analyzed two sets of data consisting of natural connected speech in order to examine the characteristics of vowel duration in English. The read data was embedded in a carrier sentence to provide for uniform stress and intonation pattern and included more than 130 minimal pairs of consonant-vowel-consonant (CVC) structure between both data sets. They compared the minimal pairs differing with respect to voicing in the final stop consonant, and found the average length of the vowel before voiceless stops to be only 66% of the average length of the vowel before voiced stops: a ratio of 2:3. Peterson and Lehiste report, "The durations of all syllable nuclei in English are significantly affected by the nature of the consonants that follow the syllable nuclei..." (p. 200).

Evidence in support of the conclusions by Peterson and Lehiste is provided in a study by Klatt (1973). Klatt analyzed 40 monosyllabic and 40 bisyllabic words spoken in a carrier sentence by three adult males. His findings for average vowel durations of the monosyllabic words also reveals a ratio of 2:3 preceding the voiceless and voiced consonants. The mean difference in the vowel lengths preceding voiced and voiceless consonants, averaged across the three speakers, was 66 msec. for the single syllable words and less for the bisyllabic words.

Zimmerman and Sapon's (1958) findings were similar to those already mentioned. Their English data consisted of read speech of monosyllabic word containing the vowel /i/ and ending in all possible consonants. On the basis of their findings, they concluded that the vowels preceding voiced consonants are longer that vowels preceding unvoiced consonants. The mean difference between the vowel durations, with all the consonants pooled by voicing, was 83.2 msec. A number of other studies, complied over three decades, report findings consistent with those discussed, with longer vowel durations preceding voiced consonants as opposed to the voiceless cognate (Chen, 1970; Halle, Hughes & Radley, 1957; Klatt, 1976; Wang, 1969).

The lengthening of vowels preceding a voiced word-final stop in English is well documented in the literature. However, there are several factors other than voicing that influence the vowel durations preceding stop consonants that any study investigating vowel duration must take into consideration. In order to provide a complete overview of vowel durations, and to provide context for the present study, the research regarding the influence of the factors other than voicing on vowel durations is reviewed in the following section.

Other Factors Influencing Vowel Duration

Many studies examining vowel duration in English have also investigated factors other than voicing that may influence the vowel lengthening effect before voiced consonants in English (Klatt, 1976; Luce & Luce, 1985; Oller, 1973; Umeda, 1975). Kent and Read (1992) report:

> Among the factors that influence vowel duration are: tense-lax (long-short) feature of the vowel, vowel height, syllable stress, speaking rate, voicing of a preceding or following consonant, place of articulation of a preceding or following consonant, and various syntactic or semantic factors such as utterance position or word familiarity. (p. 95).

The discussion in this section will be limited to those factors, aside from voicing of the following consonant, that are particularly relevant to the scope of the present study. These include (1) position in word; (2) sentence position (nonphrase-final versus phrase-final); and (3) inherent durations of individual vowels.

With regard to the vowel position in a word, Oller (1973) and Klatt (1976) both report that word-final syllables are somewhat longer in duration than non-final syllables, even in non-phrase-final positions. In addition, Oller's data indicates that single syllable vowel durations behave consistently like the vowels in word-final syllables with regard to the lengthening effect. Furthermore, Oller observed the final-syllable vowel lengthening in both stressed and unstressed syllables alike. He reported that 100 msec was the average difference between final-syllable vowels and non-final syllable vowels for both stressed and unstressed vowels. This vowel lengthening in word-final position is not to be confused with the lengthening effect of lexical stress patterns, but is rather considered a lengthening at word boundaries, which is a syntactic, not phonetic factor (Klatt 1976).

Umeda (1975) reports that in her study of connected speech, there is no significant difference between the durational behavior of the vowels in single syllable words and the stressed vowel in the final syllable of polysyllabic words in non-phrase final position. This is similar to Oller's (1973) findings. Umeda also found that the lengthening effect of the following consonant on the vowel durations in single syllable words and word-final syllables in running speech was greatly reduced when compared to the effect on phrase-final vowels. It was however detected, which supports the findings by Oller (1973) and Klatt (1976).

The position in the sentence or phrase of the vowel and following consonant is an additional factor that influences the vowel durations preceding voiced and voiceless consonants. Both Klatt (1976) and Umeda (1975) report that the vowel lengthening effect before voiced consonants becomes more pronounced in phrase-final position. This effect is also referred to as prepausal lengthening (Crystal & House, 1988; Umeda, 1975). Findings by Crystal and House (1988) support those by Klatt and Umeda. In their study with informal connected speech Crystal and House reported that "The average durations of vowels preceding prepausal word-final consonants are considerably greater than those of vowels preceding word-final consonants in general or, in particular, those preceding nonprepausal word-final consonants" (p. 1559).

The identity of the particular vowel is yet another factor that affects the vowel duration preceding voiced and voiceless stops. Different vowels have different inherent durations. Peterson and Lehiste (1960) computed the intrinsic durations of all the English vowels in stressed syllables and found that the four vowels [I, ɛ, Λ, U] are intrinsically shorter than the other English vowels. Therefore, in their analysis of the data, they compared vowel durations between minimal pairs with like vowels, such as *beat-bead* and *duck-dug*, rather than measuring all the vowel durations that occurred with a particular voiced or voiceless consonant. Otherwise, they noted, the results would have been weighted by the number of occurrences of intrinsically short or long vowels paired with the particular consonant.

Klatt (1976) also reports "Differences in inherent duration account for much of the variation in segmental timing in speech" (p. 1213). Not only do vowels have different inherent durations, Klatt notes, but inherent differences in consonant length are also observed for place of articulation. Crystal and House (1982) found that the vowel durations in their data of connected speech showed strong lengthening before voiced stops for the long (tense) vowels, but very minimal lengthening for the short (lax) vowels preceding voiced stops. They suggested that some unaccounted for parameters in connected speech modified the expected voicing influence on the vowel durations for the short vowels.

This discussion has been limited to only those factors that are particularly relevant to the present study. Other influences, such as vowel height, place of articulation of the postvocalic consonant, sentential stress and intonation all have been found to influence vowel duration as well, but consideration of these is beyond the scope of this study. Given that the vowel duration differences preceding voiced and voiceless consonants are reliably produced, and taking into consideration the various other factors that influence the vowel durations, it then becomes useful to ask what the perceptual importance of this acoustic cue may be to the voiced/voiceless distinction of the following consonant. In order to provide a context for the discussion on the role that vowel duration may play as an acoustic correlate to the voicing distinction of following consonants, a brief discussion on speech perception is provided.

Speech Perception

Speech perception involves recoding of the acoustic signal into a phonetic representation. The most straightforward view of speech perception would hold that a set or series of acoustic features corresponds to a particular phoneme in the language, and those phonemes are recognized by the listener directly in the speech wave and then recoded by some mechanism into words and phrases. This view was stated early on by Jakobson, Fant, and Halle (1952), who developed the universal distinctive-feature system by which phonemes are described in terms of acoustic distinctive features.

But the process of speech perception has been found to be much more complex than a simple mapping from acoustic features to phonetic segments. This is due in large part to the way speech is produced. Coarticulation, or the overlapping of adjacent articulations, complicates the recoding from acoustic to phonetic levels in two basic ways (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). It results in the *segmentation* and the *invariance* problems of the speech signal.

Segmentation refers to the parallel transmission of information for different phonemes in any given acoustic segment. This results in the overlapping of information for any given phonetic segment over that of adjacent or nearby segments. The invariance problem refers to the multiple and varied acoustic properties that identify a phonetic segment. There is no one single invariant property in the acoustic signal that corresponds uniquely to a given phonetic segment. To further complicate the process of perception, the acoustic properties used for phonetic identification will systematically vary depending on the linguistic context. On the basis of research with synthetic speech, Liberman, Cooper, Shankweiler, & Studdert-Kennedy (1967) concluded that:

> The acoustic cues for successive phonemes are intermixed in the sound stream to such an extent that definable segments of sound do not correspond to segments at the phoneme level. Moreover, the same phoneme is mostly commonly represented in different phonemic environments by sounds that are vastly different. There is, in short, a marked lack of correspondence between sound and perceived phoneme (p. 432).

This lack of one-to-one correspondence between attributes of the acoustic signal and the phonetic percept is well documented in the literature, and it has lead to a fundamental debate among researchers in speech perception about what necessary and sufficient cues exist in the speech signal and what kind of mechanism is required to recover those cues (Segalowitz & Gruber, 1977).

On one side of the debate, many researchers argue that some important cues for phoneme recognition do not occur in the acoustic signal at all, due to the complexities of the relation between the acoustic signal and the phonetic percept (Delattre, Liberman and Cooper, 1955; Liberman, Cooper, Harris, MacNeilage & Studdert-Kennedy, 1967; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Liberman & Mattingly, 1985). This viewpoint holds that the acoustic signal alone does not supply the sufficient and necessary information for speech perception, and hypothesizes that deeper cues exist in underlying, abstract structures.

In contrast to this view, other researchers focus on the surface aspects of the speech signal and argue that sufficient cues do exist in the acoustic signal which enable the listener to directly perceive speech (Blumstein & Stevens, 1979; Cole, 1977; Cole & Scott, 1974a; Cole & Scott, 1974b; Fant, 1967; Jakobson, Fant, & Halle, 1952). This viewpoint holds that although no single invariant acoustic feature signals the identity of any given phoneme, a *cluster* of features may form an invariant pattern for given phonemes in different contexts. Cole (1974) argues that speech perception involves the simultaneous identification of both invariant spectral cues for a given phoneme and context-conditioned cues that vary with the context. Various clusters of cues, according to Cole, interact to provide an invariant pattern for a given phoneme in a given context.

This theoretical debate in speech perception regarding whether the *necessary* and *sufficient* cues for perception actually exist in the acoustic signal is a fundamental one that continues to be investigated. Central to the debate is the issue of exactly which acoustic cues provide essential information to the listener for the identification of a phoneme. This study investigates just one cue, vowel duration, that may be used to signal a *distinctive feature* of a given phoneme in a given context; that of voicing. The following discussion of the research on the perceptual role of vowel duration as an acoustic correlate to the voicing distinction in postvocalic consonants reveals that there is even considerable debate regarding this one small piece of the much larger puzzle that is speech perception.

Vowel Duration as a Perceptual Cue to Voicing

Although the vowel duration differences with regard to the voicing of the following consonant are well established in English, the perceptual importance of this acoustic cue to the voicing distinction of the following consonant is debated by researchers. Numerous researchers maintain that vowel duration is a primary cue in word- final position, (Denes, 1955; Klatt, 1976; Malecot, 1977; Raphael, 1971) while others argue that it should be considered as just one of many cues that may be used by the listener to perceive speech (Hogan & Rozsypal, 1980; Port & Dalby, 1982; Slis & Cohen, 1969; Wardrip-Fruin, 1981; Wolf, 1977).

Denes (1955) was one of the first to investigate the effect of vowel duration on the perception of voicing. In his study, 33 listeners were asked to discriminate the final consonant in the synthetically produced minimal pair [jus, juz], in which the vowel and consonant durations were varied. Denes found that the final consonant was perceived as voiced more often as the vowel duration lengthened and the consonant duration became shorter. Therefore, he suggested that the effect on perception for word-final /s/ and /z/ was a function of the ratio of the duration of the vowel and following consonant. He states:

> The duration of the vowels and of the final consonants have a definite and consistent influence on the perception of "voicing." The effect of the duration of the vowel on the perception of "voicing" is not independent of the duration of the consonant, and vice versa. (p. 763)

Further experimental evidence on the importance of vowel duration to perception of the voicing distinction is supplied by Raphael (1972). Raphael investigated the effect of varying the vowel duration before synthetic word final stops and fricatives in minimal pairs. A series of voiced consonants was produced, accompanied by varying vowel lengths, and then an identical voiceless series was produced by eliminating the final 50-msec of the first formant transition.

Raphael (1972) reported that regardless of the first formant transition cues, the final consonants were perceived as voiceless when preceded by shorter vowels and as voiced when preceded by longer vowels. He concluded "The preceding vowel duration is a sufficient (*and for the types of stimuli employed here*, a necessary) cue to the perception of the voicing characteristic of a word-final stop, fricative, or cluster" (p. 1301). Further, he concedes that vowel duration is not the only potential cue to the voicing distinction, but argues that it is the most consistently present of all the cues, and therefore may be a primary one. He argues that cues such as voicing during stop closure is inconsistent, and that the stop release burst is often absent in American English.

Malecot (1970) also argues that vowel duration has become a major cue for voicing of consonants in word final position, although other cues may be available.

Some of the other possible cues he mentions include voicing during closure, closure duration, formant transition, and stop release burst. He maintains however, that in real speech some of these cues are missing. As does Raphael (1972), Malecot argues that voicing during closure is frequently absent and the closure duration is typically not present when the consonant is in initial position or unreleased in final position.

In a review of the durational effects that are known to convey linguistic information, Klatt (1976) notes that a meaningful difference in vowel duration due to a postvocalic consonant is only seen in phrase-final environments. In fact, Klatt claims that "A large difference in vowel duration is only seen in phrase-final environments, so it is only in these cases that the durational cue has primary importance" (p. 1219). He notes, though, that it is common for English speakers to devoice postvocalic voiced fricatives, which makes the voicing during closure cue unavailable. However, Klatt maintains that the variation in vowel duration is more regular, and suggests that as a result, English may be changing towards using the vowel duration, or a ratio of the vowel duration and following consonant (Denes 1955) as the primary cue for the voicing distinction.

Klatt (1976), Malecot (1970) and Raphael (1972) all agree that vowel duration as a perceptual cue to voicing in word-final position may be a primary cue because it is so consistently and reliably produced, unlike several of the other possible acoustic cues that have been investigated. They do not deny that other cues may be important to perception, only that because of its consistency, vowel duration may be utilized more by the listener to judge the voicing distinction in postvocalic word-final consonants.

In a more recent study, Luce and Luce (1985) investigated the use of vowel and closure duration as correlates of voicing of word-final stops in connected speech. Their study controlled for a number of factors known to influence duration, including: (1) sentence position (phrase-final versus nonphrase-final); (2) phonetic context; (3) place of articulation of the stops; and (4)inherent duration of the vowel preceding the stop. Their intent was to determine whether vowel duration (Raphael, 1972), closure duration (Port & Dalby, 1982), or the consonant/vowel ratio (Denes 1955) was more robust in signaling the voicing distinction in various environments.

Their research revealed longer vowel durations for words (single syllable) in phrase-final position than in non-phrase final position, thus supporting the findings of previous studies (Crystal & House, 1982; Klatt, 1976; Oller, 1973; Umeda, 1975). In addition, they found support for the well documented lengthening-before-voicing effect (House & Fairbanks, 1953; Peterson & Lehiste, 1960). The average vowel duration for test words ending in voiced stops was 55 ms longer than those ending in voiceless stops. And finally, the vowel durations preceding voiced and voiceless stops were significantly affected by the inherent vowel duration. They found the vowel duration differences to be largest for the long vowel /i/ and smallest for the short vowel /I/, similar to the findings of Crystal and House (1982). Of the three possible voicing correlates Luce and Luce examined, they found vowel duration to be the most consistent cue to the voicing distinction. The closure durations of voiced and voiceless stops were unreliable and, as a result, the consonant/vowel ratio was not as reliable as the vowel duration cue either. They report, "The finding of primary importance was that the durations of vowels preceding word-final stops most consistently distinguished voicing across the various manipulations employed" (p. 1955).

Numerous studies, however, have found vowel duration to be superseded by other spectral cues, or to be part of a complex of cues, without primary status. A study by Wolf (1977) investigated the role of various acoustic characteristics of final stop consonants in the perception of voicing, including formant transitions, release burst, closure duration, vowel formant frequencies and vowel duration. Six vowel-consonant (VC) syllables ending in the stop consonants were recorded, with the stop released, and the tape was then cut at six different intervals, creating six new stimuli with various amounts of the stop and vowel transition deleted. Subjects judged whether each syllable ended in a voiced or voiceless stop. Wolf discovered that even when the subjects did not hear a final consonant (because of deletion) in the stimulus, they responded differently to the voiced and voiceless data. She concluded this could be due to either the vowel duration or the vowel formant frequencies. When the vowel duration was controlled in a second experiment, the response of the subjects was still different for the voiced and voiceless stimuli, so Wolf postulated that the first formant and/or second formant frequency in the first 50 msec. of the syllables served as a perceptually significant cue to voicing.

On the basis of these results, Wolf proposed two types of acoustic cues as perceptually important for the voicing contrast in final stops. Offset cues, in the vicinity of closure, include the first formant transition, closure duration and the stop release burst. Vowel cues include the duration of the vowel and the vowel formant frequencies.

Hogan and Rozsypal (1980) conducted a vowel gating experiment in an effort to determine the category boundaries at which the perception of a consonant would cross over to its cognate. They varied the length of the vowels preceding word-final consonants at different intervals. In a listener identification test, out of the 24 stimuli used, there were nine instances in which the category boundary was obtained and 15 in which it was not reached. They concluded "It is apparent that factors other than vowel duration alone are involved in the recognition of the voicing opposition" (p. 1768). Hogan and Rozsypal proposed that a complex of acoustic characteristics are important to the perception of voicing in word final position, including: vowel duration, voicing during closure, closure duration, and burst duration.

Wardrip-Fruin (1981) argues that vowel duration differences have gained unwarranted status as a perceptual cue, and that they are neither necessary nor adequate to the voicing distinction. She attempted to determine the status of preceding vowel duration relative to other cues to the voicing distinction by deleting or expanding portions of the vowel nucleus at different intervals and comparing it to deletions of the final transitions. She concluded that the location of the deletions were markedly more important than the amount deleted for the perception of voicing.

Further, she reports that the voicing during closure had more effect on the voicing percept than did the change in the vowel durations.

Hillenbrand, Ingrisano, Smith and Flege (1984) used computer editing techniques and acoustic measurements to examine single syllable, CVC words in order to gain information about the acoustic bases of the listeners' judgments pertaining to the voiced/voiceless distinction. They were unable to identify a single acoustic cue, or any combination of cues, that clearly explained the listeners' judgments of the voicing distinction.

Port and Dalby (1982) suggest that it is the consonant/vowel ratio (Denes, 1955), rather than the duration of either factor independently, that serves as a primary cue for English voicing in syllable-final position. They argue that the consonant/vowel ratio acts as a phonetic unit that can be directly extracted from the speech wave by the listener, and that the ratio provides a way to reduce the number of independent variables in the speech wave that must be measured and combined by the perceptual system.

In summary, the experiments by numerous researchers (Denes, 1955; Klatt, 1976; Luce and Luce, 1985; Malecot, 1970; Raphael, 1972) suggest that vowel duration may be a primary cue to the voiced/voiceless distinction for word-final consonants. On the other hand, Wardrip-Fruin (1981) argues that the temporal cues are secondary to the spectral cues such as voicing during closure. Still others insist that a complex of cues, not a single cue such as vowel duration, is necessary to signal the voicing distinction (Hogan & Rozypal, 1980; Port & Dalby, 1982; Wolf, 1977).

In short, a bewildering variety of acoustic cues has been shown to affect voicing judgment of postvocalic word-final consonants. However, a recurring theme emerges in the literature reviewed. The researchers who argue that vowel duration is a primary cue to the voicing distinction do not dispute that other cues can be powerful indicators of voicing as well. Rather, they point out that the vowel duration cue is the most *consistent* and *reliable* cue present in the speech signal. This is an important point. And those researchers who argue vowel duration is one of a complex of cues that signal the voicing distinction include the vowel duration cue as an important part of the complex. Therefore, overall, the research suggests that vowel duration is indeed an important and significant cue to the voicing distinction of postvocalic consonants, whether primary in relation to other cues or not. It is useful, then, to investigate whether this vowel duration cue is exhibited by nonnative speakers of English, since it is apparent that native English speakers do utilize this cue in some capacity to judge voicing of postvocalic word-final consonants.

Therefore, it is germane to consider whether vowel duration differences preceding voiced and voiceless consonants occur cross-linguistically. Is this lengthening-before-voicing effect a universal phenomenon? If so, then it is reasonable to expect that nonnative speakers of English might exhibit the same vowel duration differences as native English speakers. However, if it is not found across different languages, or varies greatly among languages, then it is reasonable to expect that this acoustic cue might be lacking in the English production of nonnative speakers of English.

Vowel Duration Cross-Linguistically

Several researchers have investigated whether the vowel duration differences preceding the voiced and voiceless consonants seen in English are present in other languages (Chen, 1970; Slis & Cohen, 1969; Zimmerman & Sapon, 1958). The goal of this line of inquiry is to ascertain if the vowel lengthening before voiced consonants is a learned speech habit specific to certain linguistic structures, or whether it is an inherent physiological feature.

An early study by Zimmerman and Sapon (1958) examined the influence of a following consonant on the vowel duration in both English and Spanish. The data consisted of two word lists read and recorded by Spanish and English speakers. The Spanish list contained 90 bisyllabic words using five vowels with stress on the first syllable, followed by all possible consonants. Items such as *pato, pavo,* and *paso* were used. The English list consisted of 38 monosyllabic words with the vowel /i/, followed by all possible consonants, including items such as *neat, need,* and *niece.* Zimmerman and Sapon reported findings of the English data to be in agreement with previous studies (House & Fairbanks, 1954; Peterson & Lehiste 1960); longer vowels preceding voiced consonants than preceding the voiceless counterpart. The same phenomenon was found in the Spanish data. However, a major difference was reported in the mean difference of the vowel durations in each language. In English, the mean difference between the vowels preceding voiced consonants and those preceding voiceless consonants was 83.2 msec. In Spanish it was just 18.2 msec. In addition, the range of vowel duration in the two languages was very different. In

Spanish, the range was 36.1 msec. and in English it was 136.0 msec. Zimmerman and Sapon concluded that while in English the vowel length may constitute an important feature in the perception of voicing, there is a lack of corresponding significance of the feature in Spanish due to the small increase in vowel length before voiced consonants and to the small overall range of vowel length found in their Spanish data.

These findings are criticized by Delattre (1962) on several grounds. First, Delattre argues that the voiced/voiceless contrast practically does not exist in Spanish in final or intervocalic position, where it can affect the length of a preceding vowel. He states that the voiced cognates of final or intervocalic voiceless stops are all fricativated. In addition, Delattre notes that Zimmerman and Sapon compared monosyllabic English items to bisyllabic Spanish items. This, he argues, is not an equivalent comparison. The vowel duration influence of a medial consonant in a bisyllable is much less pronounced than that of a final consonant in a monosyllable (Klatt, 1976; Umeda, 1975). Also, Delattre claims that Zimmerman and Sapon included *all* the consonants in the calculation of the mean difference between vowel durations before voiced and voiceless consonants. Factoring out the stops from the consonants, he argues, would be more meaningful. He reanalyzed the data and found a more significant vowel duration difference preceding the voiced and voiceless stops of 33 msec. for the Spanish data. This is still markedly less than what Zimmerman and Sapon, or other studies (House & Fairbanks, 1954; Peterson & Lehiste, 1960) found for English. Finally, Delattre reports that it is not known whether the vowel length differences are due to voicing or to fricativation.

Slis and Cohen (1969) investigated vowel durations in Dutch as potential correlates of the voicing distinction of the following consonants. Their data consisted of two syllable words of the structure /bVC∂/, in which the V stands for a vowel and the C for a stop or fricative. They found that if the consonants differed by voicing only, the vowel preceding the voiced consonant was always longer than the vowel preceding the voiceless cognate. The mean difference was 30 msec.

The vowel duration lengthening was investigated in other languages by Chen (1970). He examined how vowel length is influenced by a following voiced or voiceless consonant in French, Russian, Korean and English. A native speaker read a list of minimal or near-minimal pairs for each language. The data was recorded and acoustically analyzed. In all four languages Chen found the vowels to be longer before a voiced consonant than before a voiceless consonant.

Chen reviewed the literature and supplemented his data to include Spanish, Norwegian, and German. He calculated ratios of vowel duration before voiced consonants to vowel duration before voiceless consonants for all the languages. The ratios varied greatly across languages and he noted that the voicing of the consonant influences the preceding vowel to different degrees in different languages. For example in English, the ratio is large at .61; the vowel is reduced by 40% preceding a voiceless consonant. But in French, the ratio is .87; the vowel is reduced by only

13% preceding a voiceless consonant. In Spanish, the ratio is .86 (Zimmerman & Sapon, 1958). Chen concludes:

> (a) it is presumably a *language-universal* phenomenon that vowel duration varies as a function of the voicing of the following consonant, and (b) the extent, however, to which an adjacent voiced or voiceless consonant affects its preceding vowel durationwise is determined by the *language-specific* phonological structure. (p. 139).

In light of Chen's findings it would appear that the large degree of vowel duration differences preceding voiced and voiceless consonants exhibited in English is a learned trait, or language-specific. Therefore, it is useful to investigate the speech of nonnative speakers of English to determine if they exhibit the same degree of vowel duration differences as native speakers, or if they exhibit less, as in other languages. If this trait has not been learned, or acquired, by nonnative speakers of English, and it is utilized by native English speakers to judge voicing of postvocalic consonants, as has been established, then perception errors may result.

Yet another related line of inquiry, that is beyond the scope of this study, is the question as to the basis or cause of this apparently universal phenomenon of lengthening before voicing. Numerous researchers have investigated or commented on the possible intrinsic factors of vowel length variation (Belasco, 1953; Lisker, 1957; Malecot, 1970; Raphael, 1975; Scharf, 1964; Slis & Cohen, 1969; Wang & Fillmore, 1961; Zimmerman & Sapon, 1958). Further discussion of the intrinsic or extrinsic factors involved in the vowel length variation is, however, beyond the scope of this study.

Measurement and Segmentation

Central to the entire issue of vowel duration is the notion of measurement, which involves the segmentation of the speech signal. For the results of any research to be meaningful, the criteria for segmentation to determine measurements must be clearly stipulated and adhered to. Standards in the literature for the segmentation of vowels will be briefly discussed in order to provide precedent for the measurements in this study.

Segmentation has long been a major problem in speech analysis. As outlined in the previous section on speech perception, the information for different phonemes in any given acoustic segment overlaps that of adjacent phonemes, making the delineation of the ending of one segment and the beginning of another a very difficult task in many cases. Peterson and Lehiste (1960) state:

> There are many instances in which the cues signaling the beginning and the end of a syllable nucleus are relatively unambiguous, but there are many other instances where it is very difficult to specify the point of segmentation. (p. 694).

Therefore, most researchers have aimed for consistency in their measurements, since complete accuracy is elusive, by describing the major cues that are used in the segmentation basic to the measurement.

Segmentation particularly relevant to this study is the marking of the onset and offset of the vowel followed by voiced and voiceless stops. In general, vowel duration is measured from the beginning of *vocalic* periodic energy to an end or marked decrease in amplitude of vocalic periodic energy (Luce & Luce, 1985) when using a waveform display. When using a spectrogram display the vowel duration is measured by onset and offset of both voicing and formant structure (House, 1961; Peterson & Lehiste, 1960). Some researchers include the aspiration, and the following frication, after an initial voiceless stop as part of the vowel measurement (Peterson & Lehiste, 1960; Umeda, 1975), while others regard it as part of the preceding stop and do not include it in the vowel measurement (Chen, 1970; House, 1961).

At the vowel offset, with voiced stops, it can be more difficult to determine where the vowel ends and the stop begins, because full voicing often continues through the transition period and release (Peterson & Lehiste, 1960). Generally, in this case, a marked decrease in amplitude of the periodic energy is used as a cue to the onset of the following stop (Luce & Luce, 1985; Peterson & Lehiste, 1960). Other consonants, especially /l/, /r/, /w/ and /y/ present particular difficulty in segmentation (Peterson & Lehiste, 1960). These will not be discussed in depth here, since none were used in this study. Clearly, the segmentation of the speech signal in acoustic analysis involves some human judgment, especially when the beginning or end of a segment is ambiguous. For this reason, conventions for segmentation and consistency in adhering to them is important in measurement of the speech signal. The cues and procedures used for segmentation in this study will be fully outlined in the next chapter.

Summary

This chapter has provided a review of the literature on vowel duration differences preceding voiced and voiceless consonants exhibited in English. The significance of vowel length as a perceptual cue to the voicing distinction of the following consonant varies with the phonetic environment and with the presence or absence of other spectral and temporal acoustic cues. It may be that at times vowel duration serves as a primary cue to voicing, and at other times it functions as part of a cluster of acoustic cues (Malecot, 1970; Klatt, 1976; Hogan & Rozsypal, 1980). Nevertheless, it is apparent that native English speakers utilize the vowel duration cue to judge the voicing distinction in postvocalic consonants, and it is therefore useful to determine if this cue is present or lacking in the English production of nonnative speakers of English, since it may affect perception.

Factors other than voicing that influence the vowel durations were discussed to provide a more complete context for the present study and its design. Issues related to the present study within the broader context of speech perception were also examined, as well as the research on the cross-linguistic validity of vowel duration differences. The vowel duration differences preceding voiced and voiceless consonants also occur in various other languages examined, but to different degrees in different languages (Chen 1970). Compared to other languages investigated, the variation in vowel durations are much larger in English, and most notably minimized in Spanish (Zimmerman & Sapon, 1957). This provides a rationale to investigate the vowel durations preceding voiced and voiceless consonants in the English production of native Spanish speakers (NSSs) as a possible source of perceptual errors. The issue as to whether the vowel duration differences are language-specific or intrinsic to the human speech mechanism is a topic for further research.

Segmentation and measurement of the speech signal is a central issue in research involving acoustic analysis. Segmentation procedures in the literature were discussed to provide precedent for the segmentation in this study. Researchers investigating acoustic properties in the speech signal typically utilize acoustic analysis tools that enable extremely accurate measurements of the data. In particular, digital instrumentation allows the speech signal to be sampled with very high precision, making segmentation conventions and consistency crucial.

CHAPTER III

RESEARCH DESIGN AND METHODOLOGY

In this study vowel durations preceding word-final voiced and voiceless stop consonants were examined from a small corpus of connected English speech recorded by four native Spanish speakers. Acoustic analysis techniques were used to determine if the vowel durations differed before voiced and voiceless stops. In addition, this study involved phonetic transcription by four native English speakers of the test words ending in postvocalic word-final stops generated in the data. The purpose of the phonetic transcription was to determine if there was a correlation between the voicing of the stops that the native English speakers perceived and the duration of the vowels produced. Statistical analyses were applied to both the results of the vowel measurements and of the phonetic transcriptions.

Subjects

Four native Spanish speakers from Mexico drawn from the adult ESL program at Oregon Coast Community College served as subjects in this study. All the subjects are young male adults living and working in the United States and have obtained functional fluency in English as determined by their class placement. Prior to arrival in the US none of the subjects had exposure to or instruction in English. The subjects had no apparent disorder of speech, hearing, or language. The subjects were recorded individually in a single session lasting approximately 20 minutes. They were paid for their participation.

Stimuli

The stimuli for this study consisted of 48 single syllable words of CVC structure. The stimuli formed 24 minimal pairs, differing only by the voicing of the final consonant, and ending in all possible voiced and voiceless stop consonants. (The exception to this are the pairs *fib-sip* and *fig-sick*. In order to construct minimally contrastive pairs of authentic English words, it was necessary to form near-minimal pairs for these vowel and stop combinations). Two intrinsically short vowels /I/ and /Λ/ and two intrinsically long vowels /æ/ and /a/ (Crystal & House, 1982; Peterson & Lehiste, 1960) were each combined twice with the following voiced/voiceless stop pairs: /b/ and /p/, /d/ and /t/ and /g/ and /k/. A combination of both long and short vowels were chosen to examine the differences in inherent vowel duration along with any duration differences due to the voicing of the following consonant.

Initial consonants for the minimal pairs that would make the segmentation of the onset of the vowel as unambiguous as possible were used. For this reason, /h/, the fricatives /s/ and /f/, and the nasals /m/ and /n/ were used as often as possible. The /h/ is considered neutral, with minimal effect on adjacent sounds (House & Fairbanks, 1953), and the fricatives and nasals in initial position provide distinct cues for segmentation of the vowel boundary (Peterson & Lehiste, 1960). In all other cases, stop consonants were used. The test words are listed in Table 1.

TABLE 1
List of Voiced-Voiceless Oppositions Tested

Vowels	Final Stop Consonant /b-p/	/d-t/	/g-k/
	dib	hid	pig
/I/	dip	hit	pick
	fib	kid	fig
	sip	kit	sick
	pub	bud	bug
/Λ/	pup	but	buck
	cub	mud	dug
	cup	mutt	duck
	nab	had	hag
/æ/	nap	hat	hack
	cab	sad	bag
	cap	sat	back
	mob	cod	hog
/a/	mop	cot	hock
	sob	nod	dog
	sop	not	dock

All the test words were spoken in the carrier sentence, "Say _______again". The carrier sentence was used to control several factors. A phrase, rather than isolated words, was used in order to produce speaking rates similar to conversational speech (Klatt, 1973), since speaking rate can influence vowel duration (Kent & Read, 1992; Klatt, 1976). The test words were positioned in the center of the phrase to avoid effects due to prepausal lengthening of the vowel in phrase-final position (Klatt, 1973; Oller, 1973, Umeda, 1975). In addition, the same carrier sentence for all the test words helps control the stress and intonation patterns which can affect vowel durations (Klatt, 1976; Peterson & Lehiste, 1960). The word "again" was used at the end of the phrase to encourage the release of the preceding stop.

Data Collection

The stimuli were recorded in a sound treated room using a Shure SM58 microphone connected to a Sony A7 digital audio recorder at 44.1 kHz sampling rate. The subjects were seated approximately 6-8 inches from the microphone. Each subject read two repetitions of each word, randomly ordered, spoken in the same sentence frame. This resulted in 96 test utterances for each subject (48 test words times two repetitions) resulting in a total of 384 test utterances for the study. The words were randomized in the list to avoid practice effects that might occur if the words were in unvarying order (Peterson & Barney, 1952). Each subject received the same randomization of the test words.

At the beginning of the session, the subjects were given time to review the list and they were allowed to ask the investigator about the meaning or pronunciation of any unfamiliar words. They then read a short practice list of sentences to familiarize themselves with the materials and to allow for adjustment of the recording levels. For the recording, the subjects were instructed to read each sentence in as natural a manner as possible. At the end of the recording each subject was asked to repeat any words that he may have stumbled on, or mispronounced, as determined by the investigator. A stumble was defined as several false starts on a word before finishing the phrase. A word was considered mispronounced only if it included a gross substitution, insertion or deletion of the initial consonant or vowel when compared to the target. Because the subjects were nonnative English speakers, it was expected that there would be some variation from the English target in pronunciation. No attempt was made by the investigator to judge the final consonants. Two subjects made only one error, and two subjects made four errors. The repeated words were substituted for the errors and included in the analysis.

Instrumentation

For computer-based acoustic analysis, the recorded data was delivered via a Sony A7 digital audio tape player through a low-pass filter set at 20 kHz. The signal was digitized through a 2 channel, 16 bit A/D (analog-to digital) and D/A (digital-to-analog) converter to a Gateway 2000 local bus computer system with a 486 processor. The digitized speech signal was displayed as oscillographic representations, at a sample rate of 40 kHz, using the Computerized Speech Research Environment (CSRE) computer program designed for speech analysis (Jamieson, 1993). An oscillogram display was chosen because "Records of the waveform and the intensity provide a good way of studying variations in length" (Ladefoged, 1993, p. 188) The oscillogram is commonly used in acoustic research for measuring duration. Although not as complete an acoustic record of speech as the spectrogram, the oscillogram was sufficient for purposes of this study.

Using auditory playback and visual cues, the CVC test words were identified on the waveform display within the carrier phrase. The vicinity of the onset and offset of the test word vowels were then magnified for closer inspection, and the beginning and end of the vowels were marked with the cursor. The CSRE program automatically calculated the duration of the vowel in milliseconds (msec.) by measuring the distance between the cursors and displayed it on the screen. The oscillographic display was then cut to include only the test word (not the entire phrase), and each edited waveform was assigned a filename and saved to computer disk.

Segmentation

As mentioned in the last chapter, a basic problem in the measurement of the duration of vowels is that of segmentation. Successive speech sounds overlap one another in the speech signal due to coarticulation, which involves transition between the articulators. These points of transition between the end of one sound segment and the beginning of the next can make it difficult to specify the segment boundaries. In some instances, the onset or offset of the vowel is relatively unambiguous. But in other instances, the location of the segment boundaries is more arbitrary and must be defined by convention. A description follows of the major cues that were used in the segmentation basic to the measurements in this study.

Vowel duration was measured from the first negative peak at the onset of vocalic periodic energy to the last negative peak at the offset of vocalic periodic energy. Other sounds, such as nasals and final voiced stops represent periodic energy on the waveform as well, so *vocalic* periodicity is specified. The vowel onset and offset cues varied with the environment. The general conventions for segmentation used, with regard to the different consonants involved, are outlined below.

1. Initial and Final Stops

The release of the voiceless initial stops is followed by a period of frication and then aspiration on the waveform (Peterson & Lehiste, 1960). In the voiced initial stops the period of aspiration is absent, but the period of frication following the release is usually more prominent than in the voiceless stops. The frication and aspiration were considered part of the stop, and were not included in the vowel measurement (Chen, 1970). The onset was marked at the first negative peak following the aspiration in the voiceless stops and following the frication in the voiced stops.

The final stops presented more difficulty, especially the voiced ones. With the final voiceless stops, there was generally an abrupt change in the amplitude and periodicity on the waveform due to the cessation of voicing, and the vowel offset was marked at this location. However, the voiced stops often exhibited full voicing and periodicity, and in addition to that, these subjects often fricated the final voiced stops. In Spanish, intervocalic and final voiced stops are always "fricativated" (Delattre, 1962); the fricated stops are allophones of a stop, not contrastive. This occurred most often with the voiced velar stop /g/. In these instances, the transition between the vowel and the voiced stop would be much longer and gradual, with no abrupt or marked decrease in the amplitude of the waveform. The boundary was determined by zooming in to determine changes in the peaks, amplitude, and patterns of the waveform and the vowel offset was marked at these locations. Figure 1 shows the vowel segmentation of an utterance with a voiceless initial and final stop.

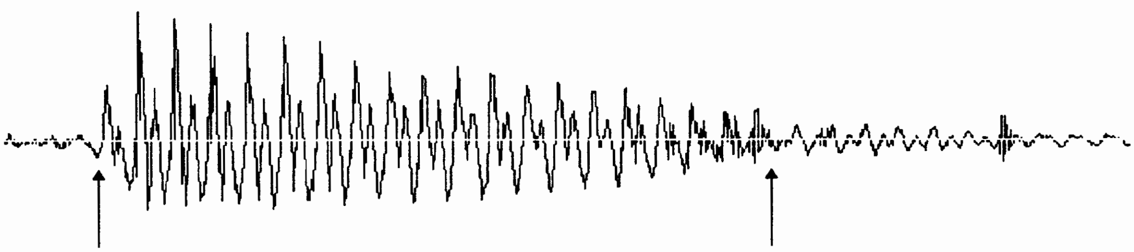

Figure 1. Waveform of the word *pick*. Vowel onset is unambiguously marked by beginning of periodicity. Vowel offset is indicated by changes in the shape, complexity, and amplitude of the waveform.

2. Initial Nasals

The location of the vowel onset with initial nasals was generally very unambiguous. Although nasals represent periodicity on the waveform, the pattern of the periodicity is very different from that of vowels. The vocalic waveform is much more complex, and higher in amplitude. The vowel onset was marked at the first negative peak following the completion of the nasal pattern cycle, where the vowel pattern began. Figure 2 shows the segmentation of the vowel with an initial nasal and voiceless final stop.

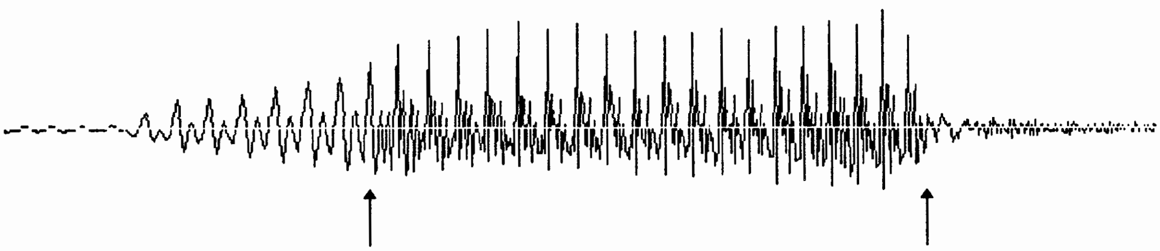

Figure 2. Waveform of the word *mop*. Vowel onset is marked by the change in the complexity of the waveform. Vowel offset is indicated by abrupt change in the waveform amplitude and periodicity.

3. Initial Fricatives and /h/

The beginning of the vowel after an initial voiceless fricative and after an initial /h/ was easily recognized. On the waveform, the onset of the vowel was accompanied by an abrupt increase in amplitude and onset of voicing, which looked completely different from the preceding frication. The first negative peak at this point was easily determined. Figure 3 shows the segmentation of the vowel with a voiceless initial fricative and a voiced final stop.

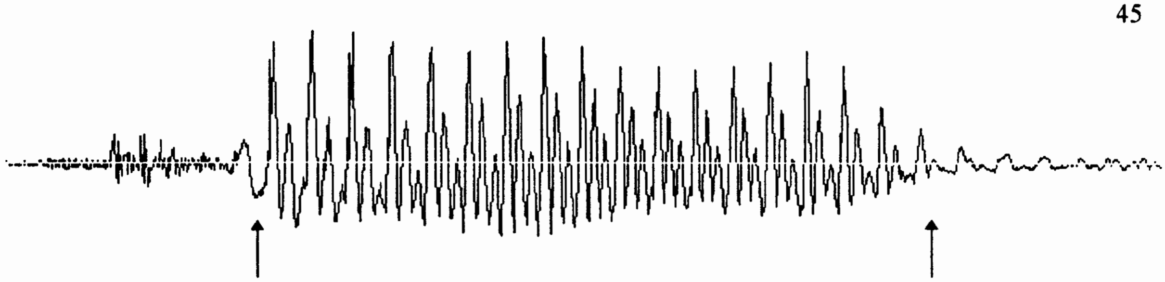

Figure 3. Waveform of the word *fig*. Vowel onset and offset are indicated by the abrupt changes in both the pattern and amplitude of the waveform.

Every attempt was made to be consistent in following the segmentation conventions as outlined when measuring the vowel durations in this study. Reliability scores discussed in the following chapter indicate to what extent the measurements are consistent.

Reliability of Segmentation

In order to provide a measure of consistency for the vowel durations obtained, both intra-rater and inter-rater reliability of the measurements were examined. A representative sample of the vowels, consisting of 10% of the data, was measured a second time after a one week interval to provide for intra-rater reliability. To provide for inter-rater reliability, the same representative sample of vowels was measured by another person trained in acoustic analysis. All of the 384 edited waveforms (96 for each subject times 4 subjects) were saved as files on disk during the original analysis and these were used for the additional measurements. Ten files from each subject's data were randomly chosen for the representative sample set, resulting in a total of 40 measurements for comparison. The reliability results are discussed in the next chapter.

Phonetic Transcribers

Four native speakers of English, including the investigator, served as transcribers for this study. All are graduate students at Portland State University and all had some training in phonetic transcription.

Transcription Procedure

The transcribers were provided with an analog tape recording of the data, and a symbol set for transcribing (Ladefoged, 1993). Transcribers used their own equipment to listen to the data, and could playback utterances as often as they wished, making changes to the transcription until they were satisfied. The data was transcribed in the same order by all transcribers with respect to the speakers, that is, speakers W, X, Y, Z in that order.

Broad phonetic transcription was used based on the International Phonetic Alphabet (IPA). Ladefoged (1993) defines broad transcription as "...a transcription that uses a simple set of symbols" (p. 41). The transcribers were instructed to choose from the symbol set provided, which included the vowels, stops, fricatives, approximants, and nasals. Diacritics were not included in the symbol set. In addition, the transcribers were free to comment on any portion of the text, but to make a forced choice if they had difficulty deciding, for example, between a voiced or voiceless stop. Of the four transcribers, only the investigator knew the focus of the study.

Reliability of Transcription

As with the segmentation results, both intra-rater and inter-rater reliability of the results were examined for the transcriptions. To measure the intra-rater reliability of the transcriptions, a representative sample of the utterances was transcribed a second time after a two week interval. The representative sample consisted of 10% of the test words for each subject, resulting in 40 utterances transcribed for comparison. Only the perceived voicing of the word-final stops was compared for reliability. The inter-rater reliability of the transcriptions was examined by comparing the results of the four individual transcribers. Again, only the perceived voicing of the word-final stops were compared for reliability. The measures of reliability obtained are discussed in the next chapter.

Data Analysis

Statistical analyses were performed on both the vowel durations obtained and the transcriptions of the perceived voicing of the word-final stops. The vowel duration measurements were categorized with respect to the voicing of the following target stop. Each measurement was labeled as preceding a voiced (1) or voiceless (0) stop, resulting in a total of 384 vowel durations (96 test words times 4 subjects) for analysis. An analysis of variance (ANOVA) procedure was performed for each vowel type to determine if the vowel durations differed significantly with respect to the voicing of the following consonant. The level of significance was set at .05, or a 95% level of confidence, for all statistical tests.

For analysis of the perceptual results, the perceived voicing of the word-final stops by each transcriber was categorized. Each stop was labeled as voiced (1) or voiceless (0) with respect to the transcribers' labels. A logistic regression was performed by vowel type for each transcriber and then by vowel type using an aggregate. This analysis determined if the vowel duration affected the perceived voicing of the postvocalic stops for individual transcribers, or if the vowel lengths affected the transcribers' perceptions as a group.

CHAPTER IV

RESULTS

This chapter reports the results of the vowel measurements, the transcriptions and the relationship between the duration measurements and the transcriptions, in that order. The first sections in this chapter address the questions regarding the differences in vowel duration produced by the subjects, the inherent vowel durations exhibited, and the reliability of the vowel measurements. The next sections report how the transcribers perceived the voicing of the stops in the data, and examine the reliability of the results. The final section addresses the question of the effect of vowel duration in this data on the perception of the transcribers with respect to the voicing distinction of the stops.

Vowel Duration and the Voiced-Voiceless Influence

A significant difference in the vowel durations was found preceding the voiced and voiceless stops as a function of the voicing distinction, for every vowel type ($p < 0.05$). These findings thus replicated the well documented evidence that vowels are longer preceding voiced consonants than preceding voiceless consonants (Chen, 1970; House & Fairbanks, 1953; Peterson & Lehiste, 1960). The analyses of variance were performed by vowel type, since any differences in inherent vowel lengths could confound the results (Peterson & Lehiste, 1960). Each subject produced 24 tokens of each of the four vowels, resulting in 96 cases for each vowel type. The findings of the analyses of variance are summarized in Table 2.

TABLE 2
Analysis of Variance for Vowel Duration - Effect of Voicing

Vowel Type	Source	Degrees of Freedom	Sum of Squares	F Ratio	P Value
/I/	Voicing	1	8715.37	25.39	.000*
/Λ/	Voicing	1	6303.42	11.09	.001*
/æ/	Voicing	1	6306.66	10.88	.001*
/a/	Voicing	1	9178.72	15.43	.000*

*$p < 0.05$

The subjects in this study, therefore, did exhibit significantly different vowel durations preceding the voiced and voiceless stops, as do native English speakers. However, the *degree* of difference in the vowel durations produced by these subjects was very different from the degree of difference in vowel durations exhibited by native English speakers. In this data, the mean differences in length with respect to the voicing of the following consonant ranged from 16.2 msec. for the vowels /Λ/ and /æ/ to 19.5 msec. for /a/. The mean durations of the vowels preceding voiced and voiceless stops, for the four subjects combined, are listed in Table 3 by vowel type. For each vowel type investigated, 48 cases occurred before a voiceless stop and 48 cases occurred before a voiced stop.

TABLE 3
Combined Mean Duration of Vowels in Msec. Before Voiced-Voiceless Stops

Vowel Type	Number	Before Voiceless Stops	Standard Deviation	Before Voiced Stops	Standard Deviation	Mean Difference	Ratio
/I/	96	128.8	36.5	147.9	31.5	19.1	.87
/Λ/	96	171.4	30.2	187.6	32.4	16.2	.91
/æ/	96	167.3	36.0	183.5	39.6	16.2	.91
/a/	96	154.5	34.7	174.0	29.7	19.5	.89
All	398	155.5		173.3		17.8	.90

For all vowel types combined, the average vowel duration preceding a voiceless stop is 155.5 msec. and preceding a voiced stop is 173.3 msec. This represents an average mean difference of 17.8 msec. and an average ratio between vowel preceding voiced and voiceless stops of 0.90. Therefore, the differences in vowel duration exhibited by the subjects in this study are much less than those normally exhibited in English (Chen, 1970; House & Fairbank, 1953; Peterson & Lehiste, 1960; Zimmerman & Sapon, 1958). In English Chen (1970) reported a mean difference of 92 msec. between vowels preceding voiced and voiceless consonants. House and Fairbanks (1953) found a mean difference of 79 msec., and

Zimmerman and Sapon (1958) noted a similar mean difference of 83.3 msec. Peterson and Lehiste (1960) reported the ratio of vowel before voiceless consonant to vowel before voiced consonant was approximately 0.66. These findings indicate that the difference in vowel duration produced in English by native speakers is much greater than that produced by the subjects in this study.

However, the difference in vowel duration found in this study is similar to the difference in vowel durations exhibited in Spanish (Chen, 1970; Zimmerman and Sapon, 1958). Zimmerman and Sapon reported a mean difference in vowel durations of 18 msec. in Spanish preceding voiced and voiceless consonants, and Chen reported a vowel ratio difference of 0.86 for Spanish, based on the findings of Zimmerman and Sapon. The findings in this study, of a mean difference in vowel durations of 17.8 msec. preceding voiced and voiceless stops, and an average ratio between vowel durations of 0.90 is consistent with what was found in Spanish. Since the subjects in this study are native Spanish speakers (NSSs), these findings are not surprising. A comparison of the difference in vowel durations before voiced and voiceless consonants in English from Chen (1960), in Spanish from Zimmerman and Sapon (1957) and in English by NSSs from this study are shown in Figure 4.

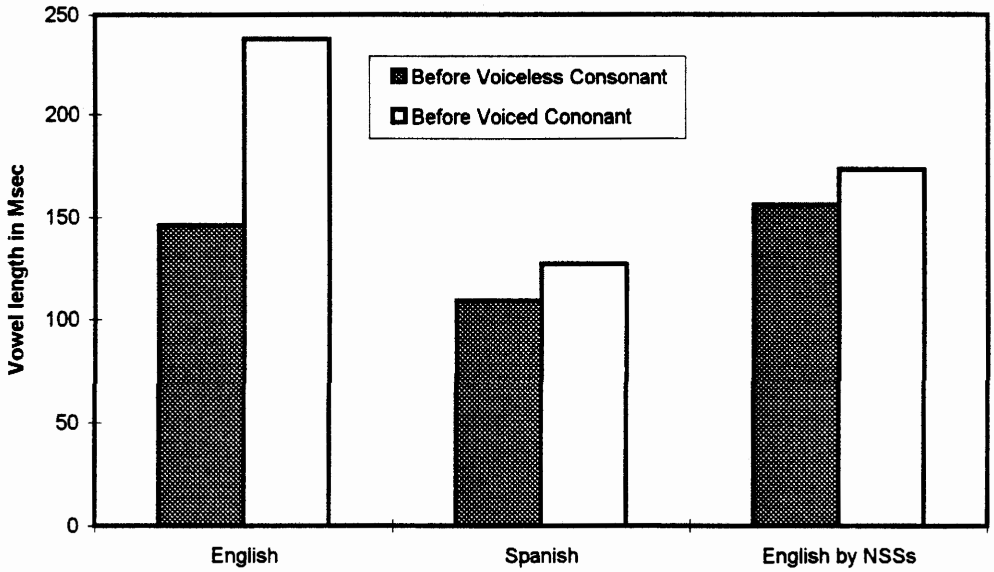

Figure 4. Mean vowel length variation by voicing of the following consonant.

The differences in vowel duration produced in English by the NSSs preceding the voiced and voiceless stops are clearly similar to the differences in vowel durations exhibited in Spanish. However, it is interesting to note that the mean lengths of the vowels preceding voiceless consonants in the English by NSSs (155.5 msec.) are similar to those in English (146 msec.), and are much longer than those in Spanish (109 msec.). The vowels preceding voiced consonants in the English by the NSSs are also markedly longer than in Spanish, but much shorter than those produced in English by native speakers. It is possible that the native Spanish speakers have learned to lengthen their vowels overall in English, but have not learned to increase their vowel durations even more when preceding voiced consonants.

Another explanation is that the vowel durations in the test words in the data were affected by "word prominence" (Umeda, 1975). It is possible that the subjects lengthened their vowels overall because of the importance of the word in the carrier sentence; the test words obviously carried the information load in the phrase. Even though the subjects were instructed to produce the carrier phrase as naturally as possible, they were still aware that the test words were the important words in the sentence since only the test words differed in each utterance. It is also likely that the subject lengthened the vowels in any words that were unfamiliar to them. Umeda (1975) reports "Unpredictable or important words take more exaggerated acoustic attributes than more predictable or less important words. The vowel duration may be included among attributes which are affected by this factor" (p. 436).

Inherent Vowel Duration

In addition to investigating the differences in vowel lengths with respect to voicing in the data, the inherent vowel durations were also examined. The analysis of variance provided comparisons of the mean duration of each vowel type that was produced by individual subjects, as well as averages for all the subjects combined. The mean durations calculated are the average duration of the vowels measured from minimal pairs differing in the voicing of the final stop consonant. This assumes that the voiced consonant has a lengthening influence on a vowel that is comparable to the shortening influence of a voiceless consonant (Peterson & Lehiste, 1960). Table 4 presents the mean durations and standard deviations found for each vowel type in the data.

TABLE 4
Mean Duration of Vowels in Msec. by Vowel Type

Vowel Type	Mean for 4 Speakers n=96	Subject 1 n=24	Subject 2 n=24	Subject 3 n=24	Subject 4 n=24
/I/	138.3 (35.3)	125.5 (15.9)	180.2 (22.0)	143.0 (26.8)	104.8 (21.9)
/ʌ/	179.6 (32.2)	161.4 (21.8)	214.4 (20.0)	168.6 (34.8)	173.9 (20.8)
/æ/	175.4 (38.5)	163.6 (16.3)	225.0 (28.8)	164.4 (31.7)	148.5 (20.0)
/a/	164.2 (33.6)	152.5 (24.1)	199.2 (20.7)	158.1 (27.0)	147.3 (34.0)

Note. Standard deviation in parentheses.

The information in Table 4 reveals that the mean duration of the vowels varied by subject, which is to be expected (Peterson & Barney, 1952). Subject 2 consistently produced longer vowels for every vowel type, and Subject 4 generally produced the shortest vowels. Also, the average durations in Table 4 indicate that for every subject, the /I/ vowel was inherently shorter than the other vowels This is consistent with findings that the /I/ vowel is one of the intrinsically shorter vowels in English (Peterson & Lehiste, 1960). However, the shortest vowel, /I/, did not result in the smallest duration difference due to voicing, as has been found in English (Crystal & House, 1982; Luce & Luce, 1985). The mean difference in vowel length for /I/ preceding voiced and voiceless stops was 19.1 msec. and for both the vowels

/Λ/ and /æ/, which are longer, the mean duration difference due to voicing was just 16.2 msec. (see Table 3).

In addition, for three out of the four subjects, the central vowel /Λ/ was inherently longer than the other vowels. This also is not consistent with findings from English data that report the vowel /Λ/ to be intrinsically shorter than both /æ/ and /a/ (Peterson & Lehiste, 1960). It is notable that this vowel does not occur in Spanish. One possible explanation is the length of the vowel /Λ/ was affected by the fact that the vowel was less familiar to the subjects, and therefore it took on exaggerated prominence and was lengthened. Umeda (1975) discusses this effect for vowels within unfamiliar words in connected speech. The other vowel in the data that does not occur in Spanish, /æ/, was the second longest vowel produced for each subject. Notably, the two vowels that do not occur in Spanish were the two longest vowels produced by each subject.

Reliability of Vowel Measurements

Both *intra*-rater and *inter*-rater reliability of the vowel measurements were evaluated. To provide for intra-rater reliability, the vowel durations in approximately 10% of the utterances were measured a second time using the files stored on disk. A representative sample of 10 utterances from each subject's data were randomly chosen, resulting in a total of 40 measurements for comparison. There was approximately a week long interval between the original and second measurements. To provide for inter-rater reliability, another person trained in acoustic analysis procedures measured the same representative sample of 40 vowels for comparison with both the original measurements and the second measurements. The vowel durations obtained from the original measurement (Time 1), from the second measurement (Time 2) and from another person, (Time X) are presented in Table 5.

TABLE 5
Vowel Durations at Time 1, Time 2 and Time X

Subject	Time 1	Time 2	Time X	Subject	Time 1	Time 2	Time X
1	179.8	185.8	180.3	**3**	113.8	113.6	117.8
	170.7	172.9	177.8		182.5	170.3	179.9
	105.8	106.4	105.7		119.5	123.2	119.6
	153.7	158.2	153.6		160.3	163.6	169.0
	144.4	171.0	169.4		137.4	132.1	131.2
	164.8	166.0	165.1		137.5	142.4	139.2
	183.4	200.8	183.7		170.9	170.6	167.4
	108.7	118.7	124.3		160.3	146.7	151.6
	167.0	167.7	171.1		160.7	167.4	151.2
	129.5	135.2	144.8		131.9	131.7	127.0
2	221.8	219.1	219.9	**4**	158.9	170.9	162.1
	204.5	205.7	200.3		144.2	149.1	139.2
	210.7	210.9	211.1		116.4	136.9	119.5
	190.1	188.9	180.0		154.9	155.9	154.6
	233.0	233.1	227.9		196.8	196.6	190.2
	203.8	216.8	197.1		155.9	157.4	155.5
	198.6	202.5	194.9		110.3	109.5	112.0
	216.3	216.1	215.7		1170.3	117.3	113.9
	182.9	198.0	183.1		139.3	137.7	140.2
	198.6	198.2	196.8		100.6	100.2	91.2

To evaluate the intra-rater reliability of the results, the range of differences between the first two sets of vowel measurements (T1 and T2) were examined. Three different ranges of proximity between the measurements were calculated. The ranges are given in Table 6. For example, within each group of 10 utterances measured for each subject, 8 vowel measurements were within 10 msec. of the first values obtained. The variation in measurements therefore, was evenly distributed across the data.

TABLE 6
Range of Vowel Measurements Between Time 1 and Time 2

Number Compared	Range of Measurements	Number Within Range	Percentage of Total
40	Within 5 Msec.	27	68%
40	Within 10 Msec.	32	80%
40	Within 15 Msec.	37	93%

In order to determine a score for the degree of consistency between the vowel measurements obtained at Time 1 and Time 2, a Pearson product-moment correlation value was calculated between the two sets of vowel durations. The values for a Pearson coefficient range from -1.0 to 1.0, with perfect positive correlation represented by a score of 1. A high correlation score ($r = 0.9795$) was obtained between the original vowel measurements and the second measurements. Figure 5 illustrates the correlation between the two sets of measurements.

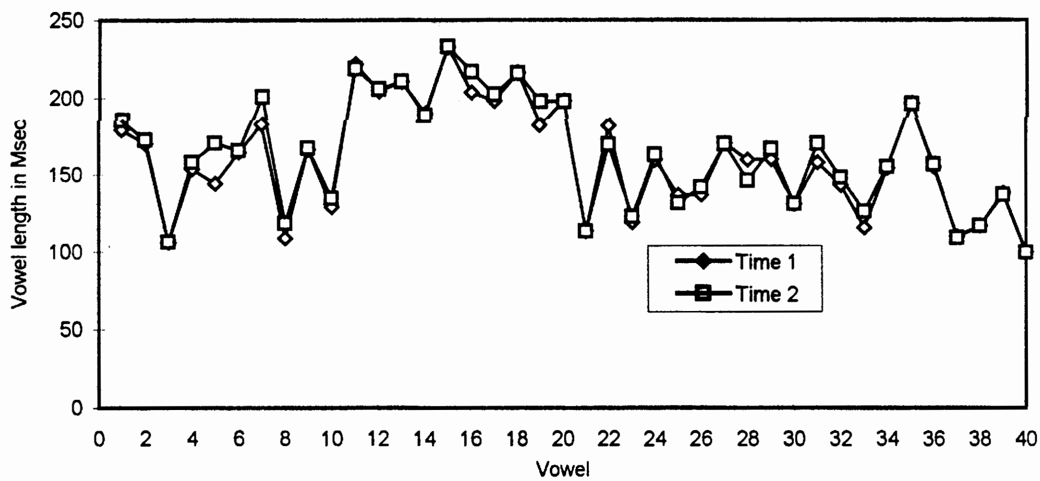

Figure 5. Correlation of vowel measurements between Time 1 and Time 2.

For inter-rater reliability, the results of the measurements at Time X were compared to both the original measurements (Time 1) and the second measurements (Time 2) obtained. Again, the range of differences between the three sets of data were examined. The results between Time 1 and Time X were compared, as well as the results between Time 2 and Time X. The range of the measurements are listed in Table 7.

TABLE 7
Range of Vowel Measurements Between T1 and TX and T2 and TX

Measurements Compared	Number Compared	Range of Measurements	Number Within Range	Percentage of Total
T1 and TX	40	Within 5 Msec.	27	68%
		Within 10 Msec.	36	90%
		Within 15 Msec.	37	93%
T2 and TX	40	Within 5 Msec.	22	55%
		Within 10 Msec.	36	90%
		Within 15 Msec.	37	93%

To determine the degree of correlation between the sets of measurements compared in Table 7, Pearson product-moment correlation scores were calculated. High correlation values were obtained between both Time 1 and Time X ($r = 0.9804$), and between Time 2 and Time X ($r = 0.9819$). These values are consistent with the correlation found between Time 1 and Time 2 ($r = 0.9795$). Figure 6 illustrates the correlation between Time 1 and Time X and figure 7 illustrates the correlation between Time 2 and Time X.

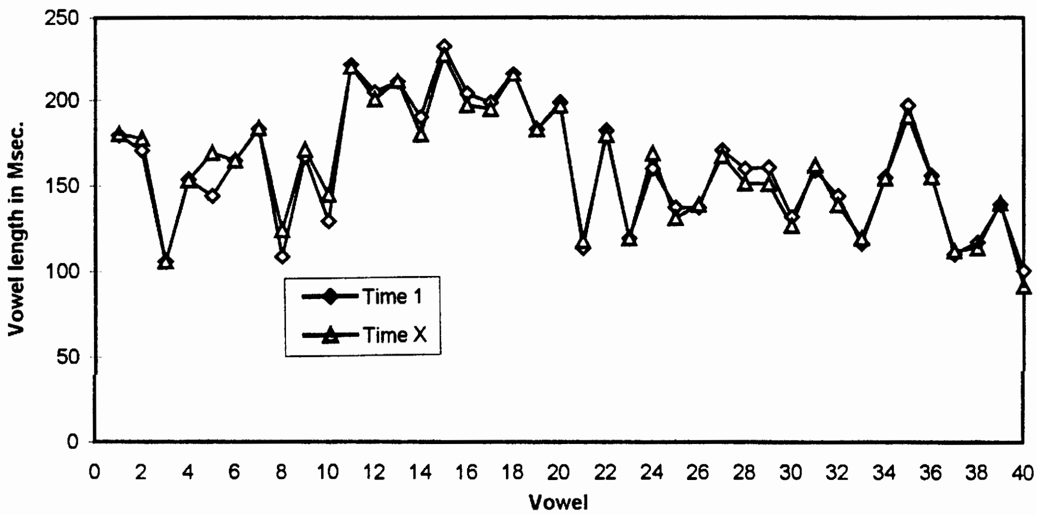

Figure 6. Correlation of vowel measurements between Time 1 and Time X.

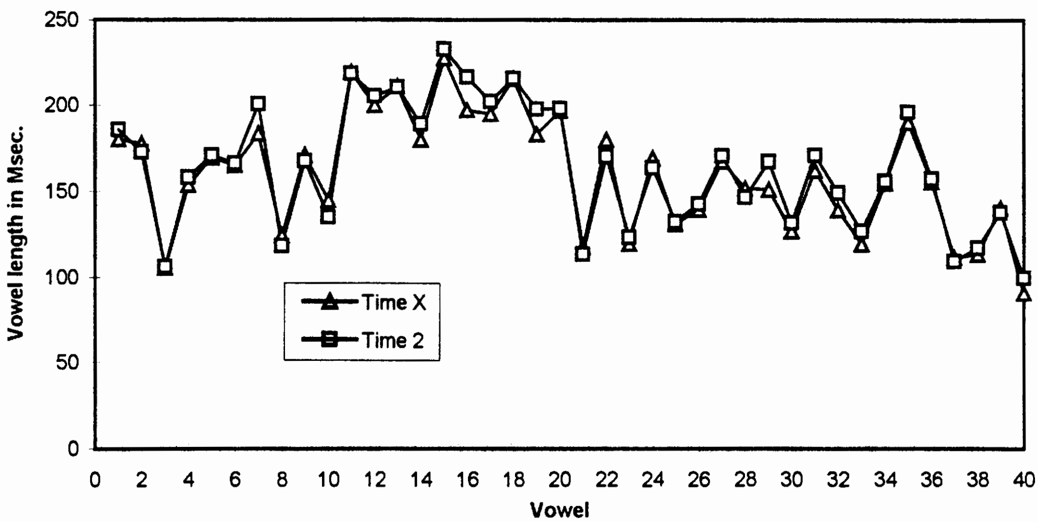

Figure 7. Correlation of vowel measurements between Time 2 and Time X.

Both the intra-rater and inter-rater reliability found for the vowel duration measurements in this study indicate a high degree of consistency in the segmentation procedures used.

Transcription Analysis-Perception of the Stops

The results of how the transcribers perceived the voicing of the final stops is summarized in Table 8. There were 384 stops in the data, and 192, or 50% were voiced and 50% were voiceless, according to the target.

TABLE 8
Perception of the Voiced-Voiceless Distinction by Transcribers

Transcriber	# Perceived as Voiceless	% of Total	# Perceived as Voiced	% of Total
1	218	57%	166	43%
2	227	59%	157	41%
3	264	69%	120	31%
4	240	63%	144	37%

It is evident from the information in Table 8 that all the transcribers perceived voiceless stops more often than voiced stops, even though there were an equal amount of each in the data. The percentages range from Transcriber 3 who heard 69% of the stops as voiceless to Transcriber 1 who heard 57% of the stops as voiceless. On average, the transcribers perceived 62% of the stops as voiceless and 38% of the stops as voiced. When the vowel durations produced in the data are considered, it is interesting to note that the mean duration before the target voiceless stops is 155.5 msec. and before the target voiced stops it is 173.3 msec. This is very similar to the range of vowel durations produced in English before voiceless consonants. Chen (1970) reports a mean vowel duration of 146 msec. before voiceless consonants. House and Fairbanks (1958, cited in Chen, 1970) found a mean vowel duration of 174 msec. before voiceless consonants and Peterson and Lehiste (1960, cited in Chen, 1970) found a mean duration before voiceless consonants of 197 msec.

That is, the mean vowel durations produced in this study before both target voiced *and* voiceless stops fall within the range of mean vowel durations produced before only the voiceless consonants in English. The mean vowel durations produced before voiced consonants in English are much greater (Chen, 1970; House & Fairbanks, 1958; Peterson & Lehiste, 1960). Therefore, it may be that the transcribers heard voiceless stops more often because the vowel lengths produced were within the range of those normally found in English preceding a voiceless consonant rather than a voiced one. The transcribers may have utilized the vowel duration cue to signal the majority of the stops as voiceless, especially if the voicing was ambiguous. It is notable that all the transcribers exhibited the same tendency to hear voiceless stops more often than voiced stops.

Reliability of Transcriptions

As with the segmentation results, both *intra*-rater and *inter*-rater reliability of the transcriptions were examined. To provide for intra-rater reliability, a representative sample of the utterances was transcribed a second time by the investigator. There was a two week interval between transcriptions. The representative sample consisted of 10% of the utterances for each subject, resulting in 40 utterances transcribed twice for comparison. Only the perceived voicing of the word-final stops were compared for reliability.

Of the 40 utterances transcribed the second time, 37 out of 40, or 93%, agreed with the first transcription on the voicing of the word-final stop. All three disagreements were in the data of one subject. In addition, to determine a measure of agreement between the two transcriptions a Cohen's Kappa value was calculated. Cohen's Kappa is similar to the Pearson's product-moment correlation, but is used for categorical information. Kappa values range from 0 to 1 where 1 represents complete agreement and 0 represents chance agreement. The categories compared were voiced (1) or voiceless (0) as perceived by the transcriber. A Kappa measure of agreement of 0.827 was obtained between the first and second transcriptions.

To measure the inter-rater reliability of the results, the transcriptions were compared among the four individual transcribers. Only the perceived voicing of the word-final stops were compared, and Kappa values between each pair of transcribers were calculated by vowel type and also with the vowels combined. The range of agreement between all transcribers by vowel type showed no consistent pattern.

Pairwise comparisons between transcribers also showed no consistent pattern, except that Transcriber 4 had lower agreement overall with the other transcribers. Table 9 gives a summary of the Kappa values obtained between each pair of transcribers. The investigator is represented by T1.

TABLE 9
Kappa Values of Agreement Obtained between Transcribers

Between Transcribers	Kappa Value of Agreement
T1 and T2	.51405
T1 and T3	.55007
T1 and T4	.36388
T2 and T3	.40182
T2 and T4	.20873
T3 and T4	.49425

Overall, the inter-rater reliability among all pairs of transcribers was low. The highest agreement obtained was .55007 between Transcriber 1 and Transcriber 3. The lowest agreement obtained was .20873 between Transcriber 2 and Transcriber 4. Confusion over the word-final stops in the test utterances was expected and these results are consistent with that expectation. There is little agreement between any pair of transcribers as to the voicing distinction of the word-final stops. Some of the variance in perception among the transcribers could be due to the different equipment that each transcriber used to listen to the data. However, other research on labeler agreement (Cole, Oshika, Noel, Lander & Fanty, 1994; Lander, Oshika, Cole & Fanty, 1995) suggests that "there is no single "correct" transcription of an utterance" (p. 3). Cole et al. (1994) and Lander et al. (1995) investigated transcriber agreement involving expert transcribers using waveform and spectrogram analysis tools for the labeling of the segments. For native English speakers transcribing fluent English, they obtained 89% and 83% agreement respectively. In the study presented here, the transcribers were transcribing accented English speech, and had no analytic tools for assistance.

Kappa values for transcriber agreement with respect to the voicing of the target was also determined. Table 10 presents the results of the comparisons.

TABLE 10
Kappa Values of Agreement Between Transcribers and Target

Transcriber with Target	Kappa Value of Agreement
Transcriber 1	0.552
Transcriber 2	0.526
Transcriber 3	0.323
Transcriber 4	0.229

As shown in Table 10, Transcribers 1 and 2 agreed with the target to an extent similar to their agreement with each other (see Table 9). Transcribers 3 and 4 have much lower agreement with the target, and, in general, had similarly low agreement with the other transcribers and with each other. That is, the values of agreement between each transcriber and the target are consistent with the values obtained between the transcribers. This indicates that the perception of the voicing distinction reflects actual choices and not merely convergence on "expected" target values.

Effect of Vowel Duration on Perceived Voicing of Stops

The results reported in the first part of this chapter established that the subjects in this study varied the length of their vowels with respect to the voicing of the following stop. The second part of this chapter established how the transcribers perceived the voicing of the word-final stops. In this section, the question regarding the effect of the vowel durations on the perception of the voicing distinction is addressed.

A logistic regression analysis was performed to determine if the differences in the vowel durations had a significant effect on the perceived voicing of the stops. This analysis was used to quantify the contribution of the vowel duration as an independent variable to the variation in the perception of the voicing of the stops as the dependent variable. Because of the variability of the perceived voicing of the stops among transcribers, it was decided to perform the analysis for individual transcribers by vowel type first. In effect, this determined if any one of the transcribers utilized vowel duration as a cue to the voicing distinction of the stops. In addition to the analysis with respect to each individual transcriber, the four transcribers were treated in the aggregate, using the target voicing as a tie breaker when the transcribers were split evenly on the voicing distinction. This determined any general trends of the group with respect to the effects that vowel duration had on the perceived voicing of the stops.

For the individual transcribers, the vowel duration was found to be a significant variable in the perception of voicing for two transcribers with two vowels ($p < .05$). The data indicated a significant effect of vowel duration on the perceived voicing for Transcriber 1 for vowel /I/, ($p < .0006$), and for Transcriber 4 for vowel /æ/, ($p < .0143$). That is, for Transcriber 1 and vowel /I/, every 10 msec. increase in vowel length increased the probability of perceiving the stop as voiced by approximately 70%. The contribution of the vowel duration for Transcriber 4 and vowel /æ/ was considerably lower, yet statistically significant. In this case, every 10 msec. increase in vowel length increased the probability of Transcriber 4 perceiving the stop as voiced by approximately 20%.

However, for 14 of the 16 transcriber/vowel combinations (4 transcribers times 4 vowels) no significant effect of vowel duration on voicing was found. In other words, approximately 88% of the time, the perceived voicing of the stops was not affected by the vowel durations produced. This would seem to indicate that, generally, the transcribers were not using the vowel duration differences as an acoustic cue to the voicing distinction of the following stops. The results of the logistic regression are presented in Table 11.

TABLE 11
Logistic Regression Analysis of Effects of Vowel Duration on Perception of Voicing by Transcriber

Transcriber/ Vowel		Independent Variable	B	Standard Deviation	Exp(B)
T1	1	Vowel Msec	.0540	.0158	1.0555*
	2		-.0051	.0077	.9949
	3		.0037	.0068	1.0037
	4		-.0116	.0083	.9005
T2	1		.0115	.0075	1.0116
	2		.0058	.0077	1.0058
	3		.0084	.0071	1.0084
	4		.0041	.0079	1.0042
T3	1		.0114	.0078	1.0115
	2		.0066	.0072	1.0067
	3		.0014	.0058	1.0014
	4		.0058	.0087	1.0058
T4	1		.0126	.0072	1.0127
	2		.0179	.0073	1.0181*
	3		.0015	.0056	1.0015
	4		.0053	.0070	1.0053

Note: Vowel types are 1-/I/, 2-/Λ/, 3-/æ/, 4-/a/, * $p < .05$

The logistic regression performed with the aggregate had similar results. Vowel duration, as an independent variable, did not have a significant effect on the perceived voicing of the stops for the group, for every vowel type. Table 12 gives the findings for the analysis by aggregate.

TABLE 12
Logistic Regression Analysis of Effects of Vowel Duration on Perception of Voicing by Aggregate

Vowel Type	Independent V	B	Standard Error	Exp(B)
/I/	Vowel Msec	.0190	.0100	1.0192
/ʌ/		-.0019	.0081	.9981
/æ/		.0056	.0074	1.0056
/a/		.0053	.0096	1.0053

Summary

The main finding regarding the vowels produced in this data was a significant effect of voicing on the vowel durations. However, the *degree* of variation in the vowel lengths with respect to voicing was much less than the degree of difference exhibited in native English. A mean difference of 17.8 msec. was found between vowels preceding voiced and voiceless stops in the English production of these native Spanish speakers. In native English, the mean difference in vowel length ranges from 79 msec. to 100 msec. (House & Fairbanks, 1953; Peterson & Lehiste, 1960, cited in Chen, 1970). Furthermore, the mean difference in vowel length exhibited in this data was very similar to the mean difference of 18 msec. found in Spanish preceding voiced and voiceless consonants (Zimmerman & Sapon, 1958).

The results of the transcriptions revealed that all transcribers perceived more voiceless stops than voiced stops. On average, the transcribers perceived 62% of the stops as voiceless and 38% as voiced. It is notable that the durations of all the vowels produced in the data were within the range of the mean durations of vowels preceding voiceless stops in native English. In addition there was fairly low agreement as to the perceived voicing of the stops among transcribers, ranging from approximately 20% to 55%, but the variability was consistent between transcribers and between transcribers and the target.

A logistic regression performed by transcriber and by vowel type to quantify the contribution of vowel duration on the perception of the voicing distinction found a significant effect for only two transcriber/vowel combinations. For the other transcriber/vowel combinations, no effect of vowel duration on perception of voicing was found. From the statistical analysis, it appears that although the cue of vowel duration variation was present in the speech signal of this data, the listeners generally did not utilize it as a cue to the voicing distinction of the following stops.

CHAPTER V

DISCUSSION AND CONCLUSIONS

This chapter begins with a discussion of the results and conclusions to be drawn from the findings. A discussion of the implications of the results is presented next, followed by an overview of the limitations of the study. Finally, some suggestions for future research are offered.

Discussion and Conclusions

This study had two major goals. One was to investigate whether a well known acoustic cue in English that is used by listeners to make perceptual judgments is also present in the English production of native Spanish speakers. The other goal was to investigate what effect this same acoustic cue, if present, has on the perceptual judgments of native English speakers. The specific research questions addressed by this study were:

1. Is there a difference in the vowel durations preceding word-final voiced and voiceless stops in the English production of NSSs, and if so, what degree of difference is exhibited?

2. Is there a correlation between the vowel durations and the native English speakers' perceptions of the voiced/voiceless distinction of the stops in the English speech of NSSs?

An analysis of variance showed that the vowel durations in this data do vary significantly with respect to the voicing of the following consonant. These findings thus replicate the well documented evidence that vowels are longer preceding voiced consonants than preceding voiceless consonants (Chen, 1970; House & Fairbanks, 1953; Peterson & Lehiste, 1960). However, the degree to which the vowel durations differ is quite disparate from the degree of variation exhibited by native English speakers. The native Spanish speakers in this study vary the length of their vowels an average of just 17.8 msec before voiced and voiceless stops, while an average range in variation of 79 msec. to 100 msec. is reported for native English speakers (House and Fairbanks, 1953; Peterson & Lehiste, 1960, cited in Chen, 1970).

These findings seem to support those of Chen (1970) with regard to the vowel duration differences preceding voiced and voiceless consonants as a language universal phenomenon, but one that varies with the language-specific phonological structure. The variation in vowel duration in this data with regard to voicing is much more like the mean difference of 18 msec. found in Spanish (Zimmerman & Sapon, 1958) than the variation normally produced in English. A possible explanation is that the subjects have not learned the degree of vowel duration differences that are language-specific to English, and thus they produce the degree of difference that is specific to their native language of Spanish as a result of language transfer.

Once it was established that the differences in vowel duration are produced in the data, the affect that this cue has on the perception of the transcribers was investigated. A logistic regression analysis was used to determine if there is a correlation between the differences in vowel duration exhibited by the native Spanish speakers and the perceived voicing of the stops by the native English speakers. The results of the analysis indicate that, with the exception of two transcribers with two vowels, the vowel durations do not significantly affect the perceived voicing of the stops by the transcribers. In short, although the acoustic cue is present, the transcribers, *in general,* did not use it to judge the voicing of the stops. These findings differ with those that report the vowel duration cue to have a significant effect on the perception of voicing for native English data (Denes, 1955; Luce & Luce, 1985; Malecot, 1977; Raphael, 1971).

One explanation for these findings is that the subjects do not exhibit a large enough difference in their vowel durations to make it a meaningful or useful acoustic cue to the native English speakers. Wang (1959) states:

> The perception of speech in its everyday form involves at least two sets of variables: the physical information present in the acoustical wave and the linguistic code with which the listener interprets the physical information. (p. 66).

Although statistically, minimal significant effect of vowel duration on perception of voicing was found, the results of the transcriptions reveal an interesting trend among the transcribers. All the transcribers perceived voiceless stops more often than voiced stops, even though there were an equal amount of each in the data. On average, the transcribers perceived 62% of the stops as voiceless and 38% as voiced. When the vowel durations produced in the data are considered, it is interesting to note that the mean vowel durations in this data before both voiced and voiceless stops fall within the range of mean vowel durations produced before the voiceless consonants in native English. These findings indicate that the vowel durations may have some bearing on the transcribers' perceptions of the voicing distinction. Further investigation, however, was beyond the scope of this study.

Implications

The findings of this study may be applied in several contexts. For instance, in the broad context of speech perception research, this study provides another example of variation in vowel duration produced with respect to the voicing of the following consonant. Therefore, in a small way, the findings add to the body of knowledge concerning acoustic cues in the speech signal and possible language universals.

In a more specific context, the results of this study may be applied in the field of English as a Second Language (ESL) pedagogy. In particular, the findings of this study can be incorporated into the teaching of English pronunciation. The results indicate that although native Spanish speakers may vary their vowel durations with respect to voicing, the *degree* of vowel duration differences may carry over, or transfer, from the native language. This may be true for other language groups as well. Variation in vowel duration is an acoustic characteristic of English that can be taught, which may improve pronunciation for ESL students.

In another specific context, acoustic analysis of the speech of nonnative English speakers is also useful in the field of accent reduction. This study provides specific acoustic information regarding the English production of NSSs. The identification of acoustic properties in the speech of nonnative speakers of English that are similar to or different from native English provides information which can be used to reduce perceived accents. Accent reduction is important for many nonnative speakers of English since the degree of accent produced by a speaker is often a determining factor in the first impression conveyed by that speaker.

Limitations of the Study

This study investigated only one acoustic cue that has been found to signal the voicing distinction of following consonants in English. Other cues are present in the speech signal that contribute to the voicing distinction as well. However, to make this topic more manageable, this study was limited to the acoustic analysis of the vowel durations only.

Other limitations of this study include the number of subjects used and the phonetic context investigated. For both the speech production and perception, only four subjects were used. Although a small number of subjects is common in this field of research, it limits the generalization of the results for a larger population. Also, nonnative English speakers from only one language background are represented, which limits the applicability of the results to only native Spanish speakers. This could be resolved by investigating various languages and including several subjects from each language group.

The phonetic context of this study is very specific. The variation in vowel duration examined was limited to only four vowels and to only the stop consonants, in order to make the study more tractable. However, it is well known that variation in vowel duration occurs across a much broader spectrum of phonetic environments. Further studies could examine the vowel duration variation preceding fricatives and nasals for example. In addition, the corpus generated for this research was relatively small, and involved only read data. Therefore, the findings of this research may not apply to other contexts, such as spontaneous conversational speech. However, in a comparison between measurements of various studies involving nonsense syllables, read discourse and spontaneous speech, Klatt (1976) reports that the "similarities are greater than the differences" (p. 1209).

Finally, the transcriptions were not done in a controlled environment, which may have contributed to the variability in the results of the perceived voicing by transcribers. Each transcriber used different mechanical equipment to listen and transcribe. This could be resolved by requiring all transcribers to transcribe in the same setting on the same equipment.

Directions for Future Research

There are many issues in the area of speech production and perception, related to the English produced by nonnative English speakers, that are open for future research. This section discusses only a few possibilities that closely pertain to this study.

A possible continuation of this study could involve digitally editing the vowel lengths to determine if changing the degree of difference in the vowel durations to more closely resemble that of native English would change the results of the perceptual tests. For example, it would be interesting to discover if lengthening the vowels before the voiced consonants increased the percentage of stops perceived as voiced, or if vowel duration was found to have a more significant effect on the voicing distinction if the degree of difference preceding voiced and voiceless stops was increased. Another approach might involve equalizing the vowel lengths to determine if the disagreement or confusion among the transcribers as to the voicing distinction increased due to the lack of the vowel duration cue.

Another related study could involve an attempt to determine at what point the vowel length cue begins to affect the perception of the voicing. The vowel durations preceding both the voiced and voiceless stops could be equalized and then incrementally increased to determine at what point the transcribers begin to use the vowel length as a cue to the voicing distinction. In other words, at what vowel length do the transcribers begin to hear voiced rather than voiceless stops?

Yet another approach might be to take only those perceptions that all the transcribers agree on and determine if the vowel duration plays a more significant role in the perception of the voicing distinction in this subset of data.

The notion of voicing and what that consists of was not explored in this study. However, voice onset time, in addition to the vibration of the vocal cords, contributes to the notion of voicing of a stop. The voice onset time refers to the period of voicelessness following the release of the stop before voicing starts (Ladefoged, 1993). The length of the voice onset time has a bearing on whether a stop is perceived as voiced or voiceless: the shorter the voice onset time, the more fully voiced a stop is perceived. Another possible continuation of this study could include an acoustic analysis of the voice onset times of the stops in the data, to determine if there is some kind of correlation between the voicing of the stops with respect to voice onset time, and the perception of the voicing by the transcribers. It would be interesting to investigate whether the voice onset time was a cue that the transcribers utilized to perceive the voicing of the stops rather than the vowel duration differences in this data.

Further research investigating many aspects of the English production of speakers of various different language background could provide information useful in ESL pedagogy and in the field of accent reduction. For example, related to this study, an investigation of whether nonnative speakers learn to use the vowel duration cue more like native English speakers over time could be conducted by comparing the present findings with a similar investigation involving more advanced nonnative speakers of English.

And finally, an open area for research is the question regarding the cause of the vowel length variation. There is much debate in the literature as to what inherent physiological basis might exist that causes the vowel duration differences (Belasco, 1953; Chen, 1970; Wang & Fillmore, 1961). If this phenomenon is language-universal, as it seems to be, what is the source of the variation?

REFERENCES

Belasco, S. (1953). The influence of force of articulation of consonants on vowel
duration. J. Acoust. Soc. Am., 25, 1015-1016.

Blumstein, S. & Stevens, K. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. J. Acoust. Soc. Am. 67 (2) 648-662.

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. Phonetica, 22, 129-159.

Cole, R. (1977). Invariant features and feature detectors: Some developmental implications. In S. Segalowitz & F. Gruber (Eds.), Language development and neurological theory (pp. 319-354). New York: Academic.

Cole, R. & Scott, B. (1974a). Toward a theory of speech perception. Psychological Review, 81, (4) 348-374.

Cole, R. & Scott, B. (1974b). The phantom in the phoneme: Invariant cues for stop consonants. Perception and Psychophysics, 15 (1) 101-107.

Cole, R., Oshika, B., Noel, M., Lander, T. & Fanty, M. (1994). Labeler agreement in phonetic labeling of continuous speech. Proceedings of the ICSLP-94, USA, Sept., 1994.

Crystal, T. & House, A. (1982). Segmental durations in connected speech signals: Preliminary results. J. Acoust. Soc. Am., 72 (3) 705-716.

Crystal, T. & House, A. (1988). Segmental durations in connected-speech signals: Current results. J. Acoust. Soc. Am., 83 (4) 1553-1573.

Delattre, P. (1962). Some factors of vowel duration and their cross-linguistic validity. J. Acoust. Soc. Am., 34, 1141-1143.

Delattre, P., Liberman, A., & Cooper, F. (1955). Acoustic Loci and transitional cues for consonants. J. Acoust. Soc. Am., 27 (4) 769-773.

Denes, P. (1955). Effect of duration on the perception of voicing. J. Acoust. Soc. Am., 27 (4) 761-764.

Fant, G. (1967). Auditory patterns of speech. In W. Wathen-Dunn (Ed.), Models for the perception of speech and visual form. (pp. 111-125). Cambridge: MIT Press.

Halle, M., Hughes, G., & Radley, J. (1957). Acoustic properties of stop consonants. J. Acoust. Soc. Am., 29 (1) 107-116.

Hillenbrand, J., Ingrisano, D., Smith, B. & Flege, J. (1984). Perception of the voiced-voiceless contrast in syllable-final stops. J. Acoust. Soc. Am., 76 (1) 18-26.

Hogan, J. & Rozsypal, A. (1980). Evaluation of vowel duration as a cue for the voicing distinction in the following word-final consonant. J. Acoust. Soc. Am. 67 (5) 1764-1771.

House, A. (1961). On vowel duration in English. J. Acoust. Soc. Am., 33 (9) 1174-1178.

House, A. & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. J. Acoust. Soc. Am., 25 (1) 105-113.

Jakobson, R., Fant, G. & Halle, M. (1952). Preliminaries to speech analysis: The distinctive features and their correlates. (Tech. Re. 13) Cambridge: MIT Press.

Jamieson, D. G. (1993). Computerized Speech Research Environment (version 4.2) [Computer software]. London, Ontario: AVAAZ Innovations Inc.

Kent, R. & Read, C. (1992). The acoustic analysis of speech. San Diego: Singular Publishing.

Klatt, D. (1973). Interaction between two factors that influence vowel duration. J. Acoust. Soc. Am., 54 (4) 1102-1104.

Klatt, D. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. J. Acoust. Soc. Am., 59 (5) 1208-1221.

Kopp, G. & Green, H. (1946). Basic phonetic principles of visible speech. J. Acoust. Soc. Am., 18 (1) 74-89.

Ladefoged, P. (1993). A course in phonetics. Fort Worth: Harcourt Brace.

Lander, T., Oshika, B., Cole, R. & Fanty, M. (1995, August). Multi-language speech database: Creation and phonetic labeling agreement. Proceedings of the International Congress of Phonetic Sciences, Stockholm, Sweden.

Liberman, A., Cooper, F., Harris, K., MacNeilage, P. & Studdert-Kennedy, M. (1967). Some observations on a model for speech perception. In W. Wathen-Dunn (Ed.), Models for the perception of speech and visual form. (pp. 68-87). Cambridge: MIT Press.

Liberman, A., Cooper, F., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74 (6) 431-461.

Liberman, A., & Mattingly, I. (1985) The motor theory of speech perception revised. Cognition, 21, 1-36.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. Language, 33 (1) 42-49.

Luce, P. & Charles-Luce, J. (1985). Contextual effects on vowel duration, closure duration, and the consonant/vowel ratio in speech production. J. Acoust. Soc. Am., 78 (6) 1949-1957.

Malecot, A. (1970). The lenis-fortis opposition: Its physiological parameters. J. Acoust. Soc. Am., 47, 1588-1592.

Miller, G. (1956). The perception of speech. In M. Halle (Ed.), For Roman Jakobson, Essays on the occasion of his sixtieth birthday. (pp. 353-360). The Hague: Mouton.

Oller, D. (1973). The effect of position in utterance on speech segment duration in English, J. Acoust. Soc. Am., 54 (5) 1235-1247.

Peterson, G. & Barney, H. (1952). Control methods used in a study of the vowels. J. Acoust. Soc. Am., 24 (2) 175-184.

Peterson, G. & Lehiste I. (1960). Duration of syllable nuclei in English. J. Acoust. Soc. Am., 32 (6) 693-703.

Port, R. & Dalby J. (1982). Consonant/vowel ratio as a cue for voicing in English. Perception & Psychophysics, 32 (2) 141-152.

Raphael, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. J. Acoust. Soc. Am., 51 (4) 1296-1303.

Raphael, L. (1975). The physiological control of durational differences between vowels preceding voiced and voiceless consonants in English. Journal of Phonetics, 3, 25-33.

Segalowitz, S. & Gruber, F. (1977). Speech perception. In S. Segalowitz & F. Gruber (Eds.), Language development and neurological theory. (pp. 313-317). New York: Academic.

Sharf, D. (1964). Vowel duration in whispered and in normal speech. Language and Speech, 7 (2) 89-97.

Slis, I. & Cohen, A. (1969). On the complex regulating the voiced-voiceless distinction I. Language and Speech, 12 (2) 80-102.

Umeda, N. (1975). Vowel duration in American English. J. Acoust. Soc. Am., 58 (20) 434-445.

Wang, W. S-Y. (1959). Transition and release as perceptual cues for final plosives. Journal of Speech and Hearing Research, 2 (1) 66-73.

Wang, W. S-Y & Fillmore, C. (1961). Intrinsic cues and consonant perception. Journal of Speech and Hearing Research, 4 (2) 130-136.

Wardrip-Fruin, C. (1982). On the status of temporal cues to phonetic categories: Preceding vowel duration as a cue to voicing in final stop consonants. J. Acoust. Soc. Am., 71 (1) 187-195.

Wolf, C. (1978). Voicing cues in English final stops. Journal of Phonetics, 6, 299-309.

Zimmerman, S. & Sapon, S. (1957). Note on vowel duration seen cross-linguistically. J. Acoust. Soc. Am., 30, 152-153.

APPENDIX A

Note: T = Transcriber

Subject 1 Data					
Word	Vowel (ms)	Voicing T1	Voicing T2	Voicing T3	Voicing T4
cab	165.4	1	1	0	0
dug	196.3	1	1	0	0
bag	169.5	1	1	0	0
sob	127.3	1	1	0	0
pick	102	0	0	0	1
mob	127.5	0	0	0	0
bud	138.9	0	1	0	0
hid	116.2	0	1	0	0
hog	171.7	1	1	1	1
nap	135.8	0	0	0	0
duck	179.2	0	0	0	0
sad	163.4	1	1	0	0
dip	147.6	0	0	0	0
hock	155.5	0	0	0	0
sip	104	0	0	0	0
hat	136.1	0	1	0	0
fig	129.2	1	1	1	1
hag	170.1	0	1	0	0
mop	126	0	0	0	0
bug	191.8	1	1	1	1
dib	140.7	1	1	0	0
dock	162.6	0	0	0	0
kid	116.3	0	0	0	0
not	157.3	0	1	0	0
cub	161	0	1	0	0
hack	160.7	0	0	0	0
nod	165	0	0	0	0
cap	165.1	0	1	0	0
pub	159.6	1	1	0	0
cup	165.1	0	0	0	0
pig	114.5	0	1	0	1
sick	107.9	0	0	0	0
buck	170	0	0	0	0
hit	123.9	0	0	0	0
mutt	144.8	0	0	0	0
cod	157.2	0	1	0	0
pup	140.9	0	0	0	0
back	174.3	0	0	0	0
fib	132	1	1	0	0

Subject 1 Data					
Word	**Vowel (ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
sat	153.8	0	0	0	0
but	162.7	0	0	0	0
dog	182.5	1	1	1	1
kit	147.4	0	0	0	0
had	161.9	0	1	0	0
mud	181.6	1	0	0	0
cot	157.1	0	0	0	0
nab	172.1	1	1	0	0
sop	141.7	1	1	0	0
cab	164.7	1	1	0	0
dug	128.1	0	1	1	0
bag	144.1	1	1	0	1
sob	167.7	1	1	0	0
pick	106.4	0	0	0	0
mob	171	0	1	0	0
bud	180.3	1	1	0	0
hid	116.2	0	1	0	0
hog	152.5	1	1	1	1
nap	143.5	0	0	0	0
duck	111.8	0	0	0	0
sad	185.8	1	1	1	0
dip	150.3	0	1	0	0
hock	70.7	0	0	0	0
sip	111	0	0	0	0
hat	158.2	0	1	0	0
fig	119	1	1	1	0
hag	200.8	1	1	0	0
mop	141.9	0	0	0	0
bug	175.6	1	1	1	0
dib	150.7	0	1	0	0
dock	183.8	0	1	1	0
kid	135.2	1	1	0	0
not	169	0	1	0	0
cub	152.1	0	0	0	0
hack	164.6	0	0	0	0
nod	172.9	0	1	0	0
cap	149.3	0	0	0	0
pub	173.4	1	1	0	0
cup	142.4	0	0	0	0
pig	119	0	0	0	1
sick	118.7	0	0	0	0
buck	193.4	0	0	0	1

Subject 1 Data					
Word	**Vowel (ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
hit	144.2	0	1	0	0
mutt	130.5	0	1	0	0
cod	168.7	0	1	0	0
pup	166	0	1	0	0
back	196.2	0	0	0	1
fib	115.4	1	1	1	0
sat	164.7	1	0	1	0
but	161.5	0	0	0	0
dog	142.7	1	1	1	1
kit	143.9	0	1	0	0
had	152.3	0	1	0	0
mud	167	1	1	0	0
cot	152.2	0	1	0	0
nab	175.1	1	1	0	0
sop	134.4	1	1	0	0

Subject 2 Data					
Word	**Vowel (ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
cab	241.7	1	1	1	1
dug	207.8	1	1	0	1
bag	279.8	1	1	0	1
sob	196.7	1	1	1	1
pick	168.4	0	0	0	1
mob	195.4	1	1	0	1
bud	238.6	1	1	1	1
hid	188.8	1	1	1	0
hog	185.4	0	0	0	0
nap	189.2	0	1	1	1
duck	228.3	0	0	0	1
sad	277.6	1	1	1	1
dip	236.2	1	0	0	0
hock	205.8	0	0	0	1
sip	157.3	0	0	0	0
hat	163.9	0	0	0	0
fig	176.5	1	1	1	1
hag	258.4	0	1	0	0
mop	193	0	0	0	0
bug	236.5	1	1	1	1
dib	211.2	1	1	0	0
dock	228.7	0	0	0	1
kid	182.3	1	1	1	1
not	198.4	0	0	0	0
cub	220.8	0	1	0	0
hack	208.2	0	0	0	0
nod	219.4	0	0	0	0
cap	245	0	1	0	0
pub	224.5	1	0	0	0
cup	221.7	0	1	0	0
pig	166.8	1	1	1	1
sick	177.6	0	1	0	1
buck	239.3	0	0	0	1
hit	162.7	0	1	0	0
mutt	202.9	0	1	0	0
cod	193.8	0	0	0	0
pup	215.1	0	0	0	0
back	214.6	0	0	0	1
fib	179.6	1	1	1	1
sat	219.1	0	0	1	0

Subject 2 Data					
Word	**Vowel (ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
but	205.7	0	0	1	0
dog	210.9	0	0	0	1
kit	188.9	0	0	0	0
had	233.1	1	1	1	0
mud	216.8	0	0	0	0
cot	202.5	0	0	0	0
nab	216.1	0	0	0	0
sop	198	0	0	0	0
cab	215.5	1	0	0	0
dug	219.1	0	0	0	1
bag	248.7	1	1	0	1
sob	171.7	1	1	0	0
pick	173.8	0	0	1	1
mob	195.3	0	0	0	0
bud	253.3	1	1	1	1
hid	152.5	1	0	0	0
hog	154.8	0	0	0	0
nap	216	0	0	0	0
duck	204	0	0	0	1
sad	244.6	1	1	1	0
dip	165.5	0	0	0	0
hock	168.9	0	0	0	0
sip	144.4	0	0	0	0
hat	234.1	0	0	0	0
fig	208.3	1	0	1	1
hag	198.1	0	0	0	0
mop	169.5	0	0	0	0
bug	240.1	0	0	0	1
dib	198.2	1	0	0	0
dock	209.7	0	0	0	0
kid	203.7	1	1	0	0
not	229.8	0	0	0	0
cub	187.5	0	0	0	0
hack	185.4	0	0	0	0
nod	238.4	0	1	1	0
cap	202.7	0	0	0	0
pub	174.7	1	0	0	0
cup	208.4	0	0	0	0
pig	193.6	1	1	1	1
sick	184.1	0	0	0	0
buck	229.7	0	0	0	1
hit	157.8	0	1	0	0

Subject 2 Data					
Word	**Vowel (ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
mutt	190.2	1	0	0	0
cod	201.3	1	1	0	0
pup	197.9	0	0	0	0
back	258.4	0	0	0	1
fib	151.4	1	1	1	1
sat	227.4	1	1	1	0
but	190.5	0	0	0	0
dog	228.5	0	1	0	1
kit	195	1	1	0	0
had	197.4	0	1	0	0
mud	191.3	0	1	0	0
cot	192.3	0	0	0	0
nab	222.9	0	0	0	0
sop	192.5	0	0	0	1

Subject 3 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
cab	145.5	0	0	0	1
dug	218.2	1	1	1	1
bag	183.5	1	1	1	1
sob	187	1	0	1	1
pick	127.1	0	0	1	1
mob	153.3	1	0	0	1
bud	165.8	1	0	1	1
hid	147.8	0	0	0	0
hog	172.7	1	1	1	1
nap	148	0	0	0	1
duck	175.7	1	0	0	1
sad	180.5	1	0	1	0
dip	113.2	0	0	0	0
hock	127.8	0	0	1	1
sip	111.6	0	0	0	0
hat	178.1	0	0	1	0
fig	175.3	1	1	1	1
hag	152.8	1	1	1	1
mop	132.3	1	0	0	1
bug	186.9	1	1	1	1
dib	137.4	0	0	0	0
dock	122.8	0	0	0	1
kid	129.1	1	0	0	0
not	116.7	0	0	0	0
cub	123.1	1	0	0	1
hack	105.8	0	0	0	1
nod	166.4	0	0	0	0
cap	129	0	0	0	0
pub	142.2	1	0	0	0
cup	113.6	0	0	0	1
pig	170.3	1	1	1	1
sick	123.2	0	0	0	0
buck	163.6	0	0	0	1
hit	132.1	0	0	0	0
mutt	142.4	1	0	0	0
cod	170.6	0	0	0	1
pup	146.7	1	0	0	0
back	167.4	1	0	0	1
fib	131.7	1	0	1	1
sat	99.8	0	0	1	0
but	123.2	1	0	1	1
dog	173.9	1	1	1	1

Subject 3 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
kit	101.6	0	0	0	0
had	177.8	1	0	1	1
mud	239.2	0	0	1	0
cot	145	0	0	0	0
nab	143.6	0	0	0	0
sop	155.5	1	0	1	1
cab	157.7	1	0	0	1
dug	210.3	1	1	1	1
bag	180.9	1	1	1	1
sob	183.2	0	0	1	1
pick	175.4	1	0	1	1
mob	170.6	0	0	0	1
bud	185.6	1	0	1	0
hid	173.7	1	0	0	0
hog	168	1	1	1	1
nap	167.4	0	0	0	1
duck	183.4	1	0	1	1
sad	232.7	1	1	1	1
dip	119	0	0	0	0
hock	138.8	0	0	0	1
sip	116.4	0	0	0	0
hat	211.4	1	0	1	1
fig	191.8	1	0	1	1
hag	203.3	1	1	1	1
mop	145.6	0	0	0	0
bug	248.7	1	1	1	1
dib	174	1	0	0	0
dock	163.2	0	0	0	1
kid	134.1	1	0	0	0
not	165.7	0	0	0	0
cub	139.9	0	0	0	1
hack	165	1	0	1	1
nod	186.1	0	0	0	0
cap	166.5	0	0	0	1
pub	157.3	1	0	0	0
cup	132.9	0	0	0	0
pig	191.8	1	1	1	1
sick	143.3	0	0	0	0
buck	174.2	0	0	1	1
hit	129.2	0	0	0	0
mutt	168.3	0	1	0	1
cod	161.3	0	0	0	1

Subject 3 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
pup	164	1	0	0	0
back	151.9	0	0	0	1
fib	160.2	1	0	0	0
sat	142.1	0	0	1	1
but	170.9	1	0	1	0
dog	234.3	1	1	1	1
kit	120.7	1	0	0	0
had	212.2	1	0	1	1
mud	169.5	1	0	0	0
cot	139.5	0	0	0	0
nab	143.8	0	0	0	0
sop	113.2	0	0	0	1

Subject 4 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
cab	123.1	1	1	1	1
dug	181.1	1	1	1	0
bag	173.9	1	1	1	1
sob	121.1	1	1	1	1
pick	77.1	0	0	0	0
mob	153	1	1	1	1
bud	214.2	1	1	1	1
hid	114.1	0	0	0	0
hog	139	1	1	1	0
nap	145.9	0	0	0	0
duck	160.7	0	0	0	1
sad	164	1	1	1	1
dip	100.2	0	0	0	0
hock	98	0	0	0	0
sip	93.4	0	0	0	0
hat	128.1	0	0	0	0
fig	139.5	1	1	1	1
hag	129.3	1	1	1	1
mop	134.9	0	0	0	0
bug	185.4	1	1	1	1
dib	108.3	1	1	1	0
dock	151.7	0	0	0	0
kid	97.5	0	0	0	0
not	162.7	0	0	0	0
cub	160.1	1	1	1	0
hack	128.4	0	0	0	0
nod	234.8	1	1	1	1
cap	138.2	0	0	0	0
pub	160.5	1	1	1	1
cup	139.1	0	0	0	0
pig	105.8	1	1	1	1
sick	74.4	0	0	0	0
buck	145.8	1	1	1	1
hit	95.3	0	0	0	0
mutt	157.6	0	0	0	0
cod	156.7	0	0	0	0
pup	167.3	1	1	0	0
back	173.7	1	1	1	1
fib	132.3	1	1	1	1
sat	154.4	1	0	1	0
but	174.5	0	0	1	0
dog	174.9	1	1	1	0

Subject 4 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
kit	89.7	0	0	0	0
had	149.7	1	1	1	0
mud	189.7	1	1	0	1
cot	151.3	0	0	0	0
nab	176.8	1	1	1	0
sop	112	0	0	0	0
cab	117.3	1	1	1	1
dug	199	1	1	1	1
bag	187.8	1	1	1	1
sob	126.5	1	1	1	1
pick	90	0	0	1	1
mob	148.3	1	0	0	1
bud	197.1	1	1	1	1
hid	126.9	1	1	0	0
hog	116	0	0	0	0
nap	149.9	0	0	0	0
duck	153.9	0	0	1	0
sad	142.7	1	1	1	1
dip	87.1	0	0	0	0
hock	104	1	0	0	0
sip	84.1	0	0	0	0
hat	148.3	0	0	0	0
fig	109.5	1	1	1	1
hag	171.2	1	1	0	1
mop	137.7	0	0	0	1
bug	191.6	1	1	1	1
dib	134.9	1	1	1	1
dock	155.4	0	0	0	1
kid	109.4	0	1	0	0
not	170.9	0	0	0	0
cub	157.4	1	1	1	0
hack	100.2	0	0	0	0
nod	190	1	1	1	1
cap	149.2	0	0	0	0
pub	159.5	1	1	1	1
cup	181.2	0	0	0	0
pig	138.7	1	1	1	1
sick	69.3	0	0	0	0
buck	160.1	1	0	1	1
hit	98.3	0	0	0	0
mutt	149.1	0	0	0	0
cod	155.9	1	1	0	0

Subject 4 Data					
Word	**Vowel (Ms)**	**Voicing T1**	**Voicing T2**	**Voicing T3**	**Voicing T4**
pup	182.7	1	1	1	1
back	156.3	1	0	0	1
fib	146.2	1	1	1	1
sat	158.1	1	0	1	0
but	196.6	0	0	0	0
dog	211.1	1	1	1	1
kit	92.2	0	0	0	0
had	138.2	1	0	1	1
mud	208.4	1	1	0	1
cot	135.9	0	0	0	0
nab	160.1	1	1	1	1
sop	92.4	0	0	0	0