

7-17-2007

# Effects of Clear Speech and Linguistic Experience on Acoustic Characteristics of Vowel Production

Michelle Bianchi

*University of South Florida*

Follow this and additional works at: <https://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

---

## Scholar Commons Citation

Bianchi, Michelle, "Effects of Clear Speech and Linguistic Experience on Acoustic Characteristics of Vowel Production" (2007).  
*Graduate Theses and Dissertations*.  
<https://scholarcommons.usf.edu/etd/631>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Effects of Clear Speech and Linguistic Experience  
on Acoustic Characteristics of Vowel Production

by

Michelle Bianchi

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science  
Department of Communication Sciences and Disorders  
College of Arts and Sciences  
University of South Florida

Major Professor: Catherine L. Rogers, Ph.D.  
Jean C. Krause, Ph.D.  
Stefan A. Frisch, Ph.D.

Date of Approval:  
July 17, 2007

Keywords: Spanish vowels, acoustic analysis, spectral change, second  
language, bilingualism

© Copyright 2007, Michelle Bianchi

## Dedication

It is only by the grace of God that I have been blessed with the gifts necessary for this accomplishment. If not for His provisions of time, talent, patience, and determination, this work would not have been possible. Therefore, I dedicate this thesis to the One who called me to complete it.

I thank my amazing husband, Dominick, for his love, support, encouragement, and understanding over the last three years of this journey. I remember standing in the kitchen when you told me to go for it. We had no idea how hard it would be, but we both know that *good* things don't always come easy.

I thank my loving children, Trey and Tanner, for always praying for me to get A's and for all of the times you occupied yourselves while I wrote. You won't have to go to Dad's shop anymore!

I thank my wonderful advisor, Dr. Rogers. You have been in my corner from day one when I nervously called you to ask if I could take your class as a non-degree seeking student. You took me under your wing and spoke on my behalf during the admissions process. You always understood the needs of my family and helped me to balance them with school. You taught me how to think and shared your gifts and your life stories with me. You've laughed with me and cried with me. You have inspired me with your brilliant mind, and I will never forget you. Thank you for your kindness. You are an angel from heaven.

## Acknowledgments

The research in this thesis was supported by NIH-NIDCD grant #5R03 DC005561 to Catherine L. Rogers (major professor).

I would like to thank Teresa DeMasi for the countless hours she spent finding and running subjects, teaching me, watching my boys, and especially for EDITING! We thought we'd never finish girl! You have been a wonderful friend, and I will miss you terribly.

Thank you, Dr. Krause and Dr. Frisch, for taking the time to read this thesis and make comments. Your comments were very helpful and much appreciated.

Thank you Merete Møller Glasbrenner for the foundation you laid for my work, for making me feel welcome in the lab, and for making me laugh.

Thank you, Dr. Diehl, for helping me with the overall organization of my thesis when I began writing. The time you invested in me was much appreciated.

Thank you, Mary, Jen, and Barb, for taking care of my boys. I never worried when they were with you. You gave me peace of mind so I could focus. I love you all.

## Table of Contents

List of Tables	iii
List of Figures	iv
Abstract	v
Chapter One Introduction	1
Overview and Statement of the Problem	1
Linguistic Experience	6
The Speech Learning Model	6
Vowel Inventories of Spanish and English and Predictions of the SLM	8
Acoustic Properties of Vowels and Studies of Vowels Produced by L2 Learners	9
Acoustic Properties of Vowels	9
Vowels Produced by L2 Learners	12
Clear versus Conversational Speech	17
Purpose of the Present Study	22
Chapter Two Method	24
Inclusion and Exclusion Criteria	24
Participants	25
Materials	27
Recording Procedure	28
Editing Procedure	31
Settings for Acoustic Analysis	32
Vowel Duration Measurement	33
Frequency Measurements	35
Chapter Three Results	39
Vowel Duration	40
F1 at 50% of Vowel Duration	43
F2 at 50% of Vowel Duration	49
Length of Vector from 20% to 80% of Vowel Duration	52
Chapter Four Discussion	57
Summary of Results	57
Comparisons to Previous Studies	59
Limitations and Implications for Future Research	61

Conclusion	61
References	63

## List of Tables

Table 1	Demographic data for early bilingual talkers	26
Table 2	Demographic data for late bilingual talkers	27
Table 3	Statistical results for vowel duration	40
Table 4	Statistical results for F1 at 50% of vowel duration	44
Table 5	Statistical results for F2 at 50% of vowel duration	50
Table 6	Statistical results for two-point vector length	53

## List of Figures

Figure 1.	Average durations (in ms) of target vowels for words produced in conversational and clear speech styles.	41
Figure 2.	Average steady-state (50% of vowel duration) F1 and F2 frequencies (in Barks) for vowels in conversational and clear speech (MO and EB talkers).	45
Figure 3.	Average steady-state (50% of vowel duration) F1 and F2 frequencies (in Barks) for vowels in conversational and clear speech (MO and LB talkers).	46
Figure 4.	Average F1 and F2 frequencies (in Barks) at 20% and 80% of vowel duration for vowels in conversational (black arrows) and clear (gray arrows) speech (MO and EB talkers).	54
Figure 5.	Average F1 and F2 frequencies (in Barks) at 20% and 80% of vowel duration for vowels in conversational (black arrows) and clear (gray arrows) speech (MO and LB talkers).	55

Effects of Clear Speech and Linguistic Experience  
on Acoustic Characteristics of Vowel Production

Michelle Bianchi

ABSTRACT

The present study investigated the hypothesis that later and/or early learners of English as a second language may exhibit an exaggerated or restricted degree of change in their production performance between clear and conversational speech styles for certain acoustic cues. Monolingual English talkers (MO), early Spanish-English bilinguals (EB) and late Spanish-English bilinguals (LB) were recorded using both clear and conversational speaking styles. The stimuli consisted of six target vowels /i, ɪ, e, ε, æ/ and /a/, embedded in /bVd/ context. All recorded target-word stimuli were isolated into words. Vowel duration was computed, and fundamental frequency (F0), and formant frequency values (F1-F4) were measured at 20%, 50%, and 80% of the vowel duration.

Data from the MO and EB talkers indicates that these two groups are very similar in that they emphasize duration differences in clear speech, have similar spacing of vowels (static & dynamic properties), and have similar frequency changes in clear speech. Data from the LB talkers indicates that this group failed to emphasize differences in clear speech, particularly duration differences. In addition, the high-mid front vowels (/i, ɪ, e/ and /ε/) were found to be very poorly

separated in the F1-F2 space for the LB talkers. In support of the hypothesis, the data showed that LB talkers exhibited a restricted degree of change in their production performance between clear and conversational speech styles for duration, as compared to monolingual talkers. Data analyzed for the EB talkers do not reveal systematic reductions in the degree of change in their production performance between clear and conversational speech styles, as compared to monolingual talkers.

## Chapter One

### Introduction

#### *Overview and Statement of the Problem*

Non-native speakers of English living in the United States must learn to adapt to many environmental challenges in speaking conditions if they are to be as well understood as native talkers in their daily lives. Some of the environmental challenges that occur quite frequently are background noise, reverberation and the filtering that occurs in telephone communication. All of these factors have been shown to affect intelligibility of native talkers (Payton, Uchanski, & Braida, 1994; Bradlow & Alexander, 2007; Bradlow & Bent, 2002) and native talkers have been shown to develop speaking strategies that can partially overcome these challenges (Bradlow & Alexander, 2007; Ferguson, 2004).

Yet the development of speech sound production abilities across different speaking conditions by adult Spanish speakers of English has received relatively little investigation. Each phoneme is identified by listeners by a range of speech cues and differences on any of these can result in a detectable foreign accent and may impede communication in difficult environments.

The rapid growth of Spanish speaking bilinguals in the United States (approximately 28 million persons at the 2000 Census; United States Census

Bureau, 2000) has given rise to the need for research in the area of speech production by this population. As evidenced by the recent growth of accent modification therapy by speech-language pathologists, second-language (L2) learners are eager to learn native-like pronunciation. Non-native speakers of English often have difficulty being understood. Under difficult speaking conditions, intelligibility differences between native and non-native speakers can be increased. Rogers, Dalby, & Nishi (2004) found that even mildly accented non-native speakers of English, who were nearly as intelligible as native speakers in quiet, were substantially less intelligible than native English speakers in noise. Research is needed to understand the conditions in which non-native speakers may have particular difficulty being understood and for the development of effective treatment techniques for non-native speakers of English.

During second language acquisition, L2 learners strive for native-like pronunciation. Spanish learners of English must learn a range of speech cues and their relative importance to achieve native-like performance. For several decades, researchers have examined the overall degree of foreign accent in L2 (cf. Flege, 1995). Due to the many factors that may influence the degree of L2 foreign accent, numerous studies have been completed in an effort to identify the most important predictors of foreign accentedness. Linguistic experience, including age and duration of immersion in an environment where the L2 is spoken, is a variable that has emerged as a major area of research in the field

(Bohn & Flege, 1997; Flege, 1995; Flege, Bohn, & Jang, 1997; MacKay & Flege, 2004; Piske, MacKay, & Flege, 2001).

Accuracy of production of target phonemes and their acoustic correlates is another area that has been extensively investigated in terms of its relationship to degree of foreign accentedness. For vowels, the most frequently investigated acoustic variables in studies of L2 speech have been vowel duration and target formant frequencies, typically measured as formant frequencies at the vowel midpoint (Bohn & Flege, 1997; Flege, Bohn, & Jang, 1997). However, recent studies of vowels produced by native speakers of English have begun to focus on dynamic properties of vowels, defined as the degree and direction of change in formant frequencies during vowel production (Hillenbrand, Getty, Clark, & Wheeler, 1995; Hillenbrand & Neary, 1999). Hillenbrand et al. (1995) found that even the “monophthongal” vowels of American English showed characteristic differences in the direction and degree of change in formant frequencies measured from 20% to 80% of the vowel duration. In a follow-up study, Hillenbrand & Nearey (1999) found that vowels were about 14.7% more intelligible, on average, when this dynamic information was retained than when it was not.

Very few studies of L2 vowel production have examined the dynamic properties of vowels and how these properties may contribute to accentedness and intelligibility of second-language learners. In one of the few studies relating vowel formant dynamics to accentedness or intelligibility of L2 speech, however, Kewley-Port, Akahane-Yamada & Aikawa (1996) found that the appropriate use

of spectral change in vowel production greatly contributes to the intelligibility of vowels produced by Japanese-accented English speakers. Thus, further research examining the acquisition of dynamic properties of vowels is important to understanding the acquisition of native-like proficiency in vowel production by L2 learners.

Another area that has received relatively little attention in studies of second-language speech production is the degree to which non-native speakers can change speaking style to adapt to challenging speaking environments. There is, however, some literature on the ability of both native and non-native speakers to modify their speaking style in response to speaking environment and the effects of these modifications on intelligibility (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Ferguson, 2004; Ferguson & Kewley-Port, 2002; Johnson, Flemming, & Wright, 1993; Picheny, Durlach, & Braid, 1985a; Picheny, Durlach, & Braid, 1985b).

Clear speech is a speaking style that is often used to increase the effectiveness of communication. It is typically used when speaking with those who are hearing impaired or in other situations when a listener may have trouble understanding (Picheny, Durlach, & Braid, 1985a). Researchers have found that the use of clear speech by native speakers positively affects intelligibility for native listeners (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Ferguson, 2004; Ferguson & Kewley-Port, 2002; Johnson, Flemming, & Wright, 1993; Picheny, Durlach, & Braid, 1985a; Picheny, Durlach, & Braid, 1985b). For sentences presented in noise to normal-hearing native listeners, clear speech

has been shown to be about 10-17% more intelligible than normally produced or “conversational” speech (Bradlow & Alexander, 2007; Bradlow & Bent, 2002; Krause & Braida 2002). This increase in intelligibility is typically referred to as the “clear speech benefit.” For identification of vowels presented in noise to normal-hearing native listeners, a clear speech benefit of about 8% has been found (Ferguson, 2004).

Bradlow and colleagues (Bradlow & Bent, 2002; Bradlow & Alexander, 2007) have compared the intelligibility of clear speech produced by native English speakers for native English-speaking listeners to its intelligibility for listeners for whom English is a second language. They have found a significantly smaller clear speech benefit for the non-native listeners than for the native listeners. They attribute the decreased clear speech benefit for the non-native listeners to an incomplete linguistic knowledge of the cues enhanced in the clear speech context.

If this hypothesis is true, the same incomplete linguistic knowledge may contribute to a reduction for non-native *speakers* in the acoustic enhancements that occur in clear speech, relative to native speakers. Thus, comparing the acoustic characteristics of phonemes produced in conversational and clear speech styles by native and non-native speakers may be a useful way of examining productive linguistic knowledge in these populations. Understanding these differences may then result in improved methods of accent reduction training for non-native speakers. No research, however, has been found comparing the acoustic properties of clear and conversational speech *produced*

**by** native English speakers to the properties of the clear and conversational speech produced by non-native speakers of English. Thus, the purpose of the present study is to compare the acoustic characteristics of vowels spoken by native and non-native (Spanish-English bilingual) speakers of American English in both conversational and clear speech styles.

To develop the methodology for the present study, a number of factors had to be considered in detail. The remainder of this chapter will therefore be used to review in more depth the following topics: theory and research on the role of linguistic experience in second-language speech production; acoustic characteristics of American English vowels and research on vowels produced by L2 learners; and previous research on acoustic and perceptual characteristics of clear speech.

### *Linguistic Experience*

*The speech learning model.* The speech learning model (SLM) developed by Flege (1995) attempts to explain the way age and the primary language (L1) phonological system affect one's ability to achieve native-like performance in pronunciation and perception of L2 phonemes. The model's premise is that when learning our L1, we perceive the phonetic differences between sounds and create separate phonetic categories for all of the sounds of our L1, including separate categories for at least some of the allophonic variants of phonemes (Flege, 1995). When learning the L2, however, the model asserts that learners may either fall short of perceiving the differences between pairs of speech sounds within the L2, or may fail to perceive differences between certain L2 and

L1 speech sounds (Flege, 1995). The model further hypothesizes that the L2 learner's failure to discriminate between certain L2 and L1 sounds may be due to assimilation of the L2 sounds to familiar L1 phonetic categories and that the L1 phonology may filter out features of L2 sounds that are not distinctive in the L1 (Flege, 1995). Another important feature of the SLM is the proposal that the L1 phonemes become stronger "attractors" of L2 phonemes as age of onset of learning a second language increases (Flege, Schirru & MacKay, 2003).

The SLM also makes predictions about changes in categorization of L2 sounds over time. During the early stages of L2 acquisition, the model asserts that some L2 sounds will be identified by the learner as being the same as an L1 phoneme or one of its allophones, while other L2 phonemes may fall into *uncommitted* space or may not be identifiable as any L1 phoneme. Over time, however, the model predicts that the L2 learner become more able to notice more of the differences between at least some of the L1 and L2 sounds. At this point, the learner may develop a new sound category, or as Flege terms it *phonetic category*, to represent differing L1/L2 sounds.

The SLM reflects the idea that if an L2 sound is perceptually linked to an L1 sound, production of the L1 and L2 versions may eventually merge (Flege, 1995). According to the model, the likelihood that L1 and L2 phonemes will merge is influenced by the age of onset of learning (AOL) of the L2, and the distance between L1 and L2 sounds as perceived by the learner. The likelihood that an L2 sound will be placed into a new phonetic category increases with an increase in perceived distance between the L1 and L2 sounds by the learner.

Similarly, the earlier the AOL, the smaller the distance between sounds needs to be in order for the learner to categorize the L2 sound as different from the L1 sound.

*Vowel inventories of Spanish and English and predictions of the SLM.*

According to most sources, English is assumed to have approximately 12 “monophthongal” vowels (/i, ɪ, æ, ɛ, e, o, ɔ, u, ʊ, ʌ, ɜ/) (Ladefoged, 1982), while Spanish has five (/i, e, a, o, u/) (Dalbor, 1969). Thus, Spanish learners of English must adapt their acoustic vowel space to include the new English vowels. Although some English vowels have a phonemic counterpart in Spanish, namely /i, e, a, o, u/, others do not. According to Bradlow (1995), the vowel spaces of English and Spanish differ in several ways. She states that although some vowel categories occupy similar positions in the acoustic space of English and Spanish, they are not precisely in the same position. So in addition to Spanish speakers needing to find a position in their articulatory vowel space for about seven new vowels in order to have native-like vowel production, they also must fine tune the production of similar vowels in English. Spanish vowels are also assumed to be produced with little or no spectral change as compared to English vowels, although this issue has not been extensively investigated for Spanish (Flege, 1991).

For a Spanish learner of English, a prediction of the SLM is that the difference in size of the vowel inventories of Spanish and English might result in a large number of English vowels being assimilated to Spanish vowel categories,

especially by later learners. Thus, a native Spanish (NS) speaker who began learning English at an early age should be more likely to differentiate between all English vowels than a native Spanish speaker who began learning English later.

Flege (1995) suggests an earlier learner's production of target L2 vowels should exhibit greater accuracy, but suggests that may be deflected from target positions for native talkers due to the need to maintain phonetic distance between similar L1 and L2 sounds. The present study will help to address these hypotheses by examining vowel productions of earlier and later Spanish learners of English and by comparing dynamic features of these vowels. Because formant dynamic properties have not been extensively investigated, they should offer a unique means of providing supporting (or disconfirming) evidence for the predictions of the SLM.

#### *Acoustic Properties of Vowels and Studies of Vowels Produced by L2 Learners*

*Acoustic properties of vowels.* In the classic study by Peterson & Barney (1952), the authors conducted an experiment that addressed target formant frequencies (F1- F3), vowel spaces and variation across vowels produced by men, women and children. Formant frequencies, formant amplitudes, and fundamental frequency (F0) were measured at a single time slice. The authors found that formant frequencies were highly variable for each speaker. In addition, there was a considerable degree of overlap between vowel formant frequencies for vowels of different categories. In particular, considerable overlapping existed between /ɜ/ and /ɛ/, /ɜ/ and /ʊ/, /u/ and /ʊ/, and /a/ and /ɔ/. The F3 values were

largely variable between all three groups of talkers. The men had the lowest, the women's were intermediate, and the children had the highest frequencies.

In Hillenbrand et al. (1995), the authors attempted to replicate and to address the limitations of the study of vowel acoustics by Peterson & Barney (1952) (PB). The limitations included: (1) measurements were taken at a single time slice; (2) duration measurements were not made; (3) measurements of spectral change over time were not made; (4) speaker and listener dialect was not considered; (5) data on age and gender of child talkers were not provided; (6) the child group was small; (7) identifiability of tokens could not be determined; (8) reliability of measurements was not reported; and (9) the database is no longer available and cannot be used to make F0 and formant frequency comparisons.

The authors extended the PB study to include measures of vowel duration and spectral change information by native speakers of English. To measure spectral change, vowel formant measurements were made at 20%, 50%, and 80% of the vowel duration as measured from onset of voicing for the vowel to onset of closure for the stop for the /hVd/ words recorded. The authors also attempted to replicate the "target" vowel measurements of PB by making formant measurements at the location within the vowel judged to have the least amount of change in the first and second formants (*F1* and *F2*, respectively).

Hillenbrand et al. (1995) also converted formant frequencies from Hz to mels for analysis of spectral change in order to present the data in a way that would be better correlated with listeners' perceptions. The vowels with the longest durations were /ɔ/, /æ/, and /e/, and the vowels with the shortest

durations were /ʌ/, /ɛ/, /ʊ/ and /ɪ/. Durations of vowels produced by male speakers were shorter than those for vowels produced by women and children. The vowels with the greatest degree of spectral change were /ʊ/, /æ/, /ɔ/ and /ʌ/, and the vowels with the smallest degree of spectral change were /u/, /i/, and /ɛ/. The vowels that are in close proximity to each other vary by the changes in F1 and F2 and durational differences. For example, although the vowels /ɔ/ and /a/ are located close together at 80% of vowel duration, both F1 and F2 are substantially higher for /a/ than for /ɔ/ at 20% of the vowel duration.

Average formant values for the three talker groups from Hillenbrand et al. (1995) reflected a general tendency toward crowding among adjacent vowel categories as compared to the PB data. The only vowels that did not occupy similar relative positions in Hillenbrand et al. (1995) and in the PB data were /ɛ/ and /æ/, with higher F2 values for /æ/ than /ɛ/ and lower F1 values for /æ/ than /ɛ/ in Hillenbrand et al. (1995) than in PB.

In a follow-up study, Hillenbrand and Nearey (1999) showed that the vowels' formant dynamic properties are used by listeners for vowel identification. Hillenbrand and Nearey (1999) created two sets of synthetic versions of /hVd/ words modeled on the properties of vowels produced by the talkers in Hillenbrand et al. (1995). In one set of synthetic stimuli (dynamic vowels), they preserved the direction and degree of formant change observed in the natural

vowels and in another set (static vowels) they maintained a single “target” formant frequency throughout the vowels. The synthetic vowels were played to listeners who had to decide which word they had heard. The dynamic vowels were about 14.7% more accurately identified by the listeners than the static vowels. The vowels that were most affected by the addition of the dynamic information were /e/, /æ/, /ʊ/, /ʌ/, and /o/. Conversely, /ɑ/, /ɔ/, and /ɜ/ were least affected by the addition of the dynamic information.

*Vowels produced by L2 learners.* Spectral change is an important cue for vowel identification by native listeners (Strange, Jenkins, & Johnson, 1983; Hillenbrand & Nearey, 1999). Since native listeners rely on spectral change, further research that specifically addresses the use of spectral change in vowel production by non-native speakers is needed.

Little research has addressed the use of formant dynamic cues by Spanish speakers of English in vowel production. Appropriate use of spectral change in vowel production has been shown to contribute to non-native speech intelligibility for Japanese-accented English speakers, however (Kewley-Port, Akahane-Yamada & Aikawa, 1996). In Kewley-Port et al. (1996), the aim of the authors was to gain knowledge of the perception and production of American English (AE) vowels by Japanese talkers. Three experiments were conducted including open-set identification, minimal-pair identification, and acoustic correlation of perception and production. The major finding in this experiment was that spectrally similar AE vowels produced by native speakers of Japanese were less intelligible to native English speakers than were dissimilar vowels. The

authors concluded that the Japanese talkers were unable to effectively communicate all of the spectral properties of the target AE vowels. The authors used regression analysis to study the influence of three acoustic properties (target frequency, dynamic formant movement and duration) of vowels produced by Japanese-accented English speakers on the intelligibility of /æ/ and /ɪ/ for native English-speaking listeners. They found that spectral change of Japanese English vowels relative to the AE targets was the most important property influencing intelligibility of these two vowels. Although duration was found to be significant for /æ/, it was not independently responsible for increased intelligibility.

Bohn & Flege (1997) found that adult experienced German learners of productions of a vowel category that is not present in German were perceived as native-like by native English-speaking listeners. The authors recorded the production of /æ/ by three groups: monolingual English speakers, experienced German learners of English, and inexperienced German learners of English.

The general distribution of the vowels in the Bark-difference space revealed that the inexperienced German subjects' German vowels did not occupy the same space as the English subjects' /æ/. The authors concluded that this is sufficient evidence to support the premise that the English /æ/ is a new vowel for their German subjects.

Next, the three groups each recorded productions of the words *bat*, and *bet* in the carrier phrase *I will say \_\_\_\_*. The fundamental and formant frequency

measurements (F1, F2, and F3) and duration of the vowels were examined for all three groups. With regard to formant frequency, the authors concluded that both monolinguals and experienced subjects produced fairly clear distinctions between the two vowels; however, inexperienced subjects' productions revealed an almost complete overlap of the target vowels /ɛ/ and /æ/. With regard to vowel duration, the authors found that both the monolingual and experienced bilingual groups had similar durational ratios for the two vowels. Conversely, the inexperienced group had smaller ratios for the two vowels. The authors concluded that the results support the hypothesis that experienced adult learners will accurately produce a new vowel, but inexperienced adult learners will not.

In an effort to determine whether the perception of /æ/ related to the aforementioned findings, the authors conducted another experiment. Synthetic speech was created that manipulated duration and formant frequency values to simulate the target /ɛ/ to the target /æ/. Intermediate formant values between those appropriate for American English /ɛ/ and /æ/ were used to create a continuum of synthetic vowel stimuli between the end vowels. Each of the eleven synthetic stimuli created was presented at durations of 150, 200, and 250 ms. The same subjects as were recorded for the acoustic analysis listened to the stimuli and identified them as either *bet* or *bat*.

From the results of the perception experiment, the authors concluded that the monolinguals relied most on spectral differences to identify *bet* versus *bat*, followed by the experienced group, with the inexperienced group relying least on

spectral differences. Conversely, the inexperienced group relied most on duration, followed by the experienced group, with the monolinguals relying least on duration.

Bohn & Flege (1997) theorize that, contrary to the predictions of the SLM, experience may influence production more than perception for the /ɛ/-/æ/ contrast for German learners of English because the experienced Germans' productions appeared to be more native-like than their perception. One explanation for this difference may be related to the feedback that immersed learners receive for this notoriously difficult vowel contrast. That is, immersed L2 learners gain more feedback on their production versus their perception in their second language. Conversely, L2 classroom learners would tend to have more feedback given to them on their perception of the new language.

Flege, Bohn & Jang (1997) studied vowel production by Spanish speakers of English; however, their study did not include spectral change information. The acoustic analysis in their study was limited to the midpoint of the vowel. Their aim was to explore the effect of L2 experience on non-native speakers' production of the English vowels /i, ɪ, æ, ɛ/ as judged by native English listeners.

The speakers included twenty each of German, Spanish, Mandarin, and Korean subjects, and 10 native speakers of English. Native speakers of English evaluated the intelligibility of the natives' and non-natives' productions of the English vowels /i, ɪ, æ, ɛ/ in bVt context, within the carrier phrase *I will say*. The native English speakers were given seven choices by which to identify each

production (“*beat, bit, bet, bat, bait, but*” and “*bottle*”). An intelligibility score of percent correct identification by the native listeners was obtained for each native and non-native talker.

Although the main effect of experience on intelligibility was not found to be significant, the interaction between experience and vowel was found to be significant. The Spanish talkers’ productions of /ɛ/ yielded a higher percentage of correct vowel identifications by native English listeners than did their productions of /i/ and /ɪ/. Spanish talkers’ intended /æ/ productions were often heard as /a/. The authors concluded that the Spanish talkers were producing a vowel for target /æ/ that was more posterior in vowel space than American English /æ/.

Conversely, the Spanish talkers’ productions of /ɛ/ were almost always correctly identified by the native English listeners. The authors concluded that this is due to an allophone of Spanish /e/ being directly transferred into English.

The authors found evidence that undermines the Contrastive Analysis Hypothesis (Lado, 1957, as cited by Flege et al., 1997). According to the authors, the theory by Lado suggests that the absence of a vowel from the L1 phonemic inventory may represent a source of learning difficulty. This theory is not supported by the authors’ finding for Spanish learners of English. They found that Spanish subjects’ intended productions of /ɛ/ (a phoneme not found in Spanish) were more often correctly identified than their intended productions of /i/ (a phoneme found in Spanish) and /ɪ/ (a phoneme not found in Spanish).

### *Clear versus Conversational Speech*

Clear speech is often used to increase the effectiveness of communication. It is typically used when speaking with those who are hearing impaired or in environments in which communication may be difficult (such as noise or reverberation). Researchers have found that clear speech positively affects intelligibility (Bradlow & Bent, 2002; Ferguson & Kewley-Port, 2002; Johnson, Flemming, & Wright, 1993; Picheny, Durlach, & Braida, 1985a,b). Many acoustic differences between phonemes are enhanced in clear speech produced by native talkers. The speech cues used by Spanish bilinguals during clear speech may give more understanding as to which cues these bilinguals think are important for distinguishing target phonemes.

Native speakers' clear speech is more intelligible than normal or "conversational" speech. Picheny, Durlach, & Braida (1985a) found that clear speech is 17% more intelligible than conversational speech for hard of hearing listeners. Fifty clear and conversational nonsense sentences were presented in quiet to five listeners with stable sensorineural hearing losses at three levels: most-comfortable-level, maximum listening level, and 10 dB below most-comfortable-level. In addition, each listener adjusted the listening level in four different frequency configurations to the highest level comfortable for long-term listening.

Johnson, Flemming, & Wright (1993) reported larger vowel spaces in hyperarticulated (clear) speech versus conversational speech of native speakers.

Therefore, it can be theorized that the cues used by native speakers in clear speech production are the same cues that are important for perception. Vowel spaces of Spanish-speaking bilinguals using clear versus conversational speech have not been studied thus far.

Phonetic knowledge of cues is needed in order to produce native-like speech. Ferguson & Kewley-Port (2002) examined formant frequency measures, degree of spectral change, and duration for target vowels produced in conversational and clear speech style by a single native speaker of American English. In order to assess acoustic differences between the two speaking styles, the authors used several metrics, including target formant values, vowel duration and a vector length measure of spectral change during vowel production. Formant frequency measures in Hertz were converted to the Bark scale (Trautmüller, 1990). The Bark scale was used because equal Bark differences are perceptually equal at different portions of the scale, while equal Hertz differences are not.

Clear speech tokens typically had higher F1 values, but values for F2 frequencies varied among vowels. In clear speech, F2 was higher for front vowels versus back vowels. In general, the vowel space occupied in clear speech was found to be larger than the vowel space occupied in conversational speech. In addition, due to the overall increase in F1 values, the space was shifted to occupy the higher values of F1.

When measuring duration, the authors found that the average duration of clear speech tokens was approximately twice that of conversational speech tokens. All ten vowels showed a significant positive effect for duration.

Dynamic formant movement was also studied for both speech styles. Vector length was used to measure the distances between F1 and F2 values at 20% and 80% of the vowel duration. The vector was computed by calculating the Euclidean distance (in Barks) between the F1 and F2 values at 20% and 80% of the vowel duration. Vector length in the more crowded areas of the talker's vowel space was found to be significantly greater in clear speech.

In the perception portion of their study, Ferguson and Kewley-Port (2002) found that young normal hearing (YNH) listeners derived a 15% benefit in intelligibility from the clear speech, compared to the conversational speech; however, elderly hearing impaired (EHI) listeners did not benefit from the clear speech in this study. Both YNH and EHI listeners were presented with a vowel identification task where each word was mixed with a segment of speaker babble. For the YNH listeners, words were presented at an overall level of 70 dB SPL with a speech-to-babble (*S/B*) ratio of -10 dB. The EHI listeners' *S/B* ratio was -3 dB. The listeners identified the vowel within each word by typing the vowel's corresponding number on a key board.

It was of interest, however, that although the EHI listeners did not benefit from clear speech for vowel identification, they did surpass the YHN listeners' percentage correct vowel identification for the conversational speech tokens.

This may have been due to the less difficult *S/B* ratio presented to the EHI listeners.

Bradlow & Bent (2002) studied the clear speech benefit derived by native versus non-native speakers of English. The subjects included 32 non-native listeners of English and 72 native listeners of English. Sixty-four simple English sentences containing three or four key words were recorded by two adult native English speakers, one male and one female. All sentences were produced in conversational and clear speaking styles.

The non-native listeners completed a perception and a production task. The sentences were presented in white noise (first  $-4$ , then  $-8$  dB signal-to-noise ratio) and in both speaking styles. Through headphones, the subjects heard either a male or female talker and were told to write down whatever they heard. On a separate day, a word-familiarity rating test was given. Keywords from the sentences were presented on a computer screen with other distractor words, and the subject rated his or her familiarity with that word. Each subject then read the same sentences from the perception task. The authors edited these sentences by adding noise at a  $+5$  dB signal-to-noise ratio (SNR), similar to the sentences used for the perception task.

Thirty-two native listeners participated in a sentence-in-noise perception task, and 40 additional subjects judged the non-natives' sentence production stimuli. The 32 listeners' perception task mirrored that of the non-natives. The 40 judges of the non-natives' production listened and transcribed what they heard. Intelligibility estimates were based on the perception of the judges.

One major finding of this study was that a smaller clear speech benefit was found for the non-native listener group than for the native listener group. In other words, the non-native listeners did not benefit as much from clear speech as did the native listeners. The average clear speech benefit for non-natives was about 5% versus the much larger average benefit of about 16% for the native listeners. The authors asserted that the finding for the native listeners was similar to those of previous studies that examined hearing impaired adults versus normal hearing (Schum, 1996; Picheny et al., 1985a; Helfer, 1997). In these three studies, the range of the clear speech effect for hearing impaired listeners and normal listeners with degraded signals is 16 to 20%. In a companion study to Bradlow and Bent (2002), Bradlow, Kraus, & Hayes (2003) found that the average clear speech benefit for learning impaired children and non-learning impaired children was the same (about 9% - somewhat lower than that found for adults).

Bradlow & Alexander (2007) found that the non-native listener average clear speech benefit was smaller than the average native listener clear speech benefit. In this study, both native and non-native listeners heard English sentences in plain (conversational) and clear speech that differed in the final word. The clear and conversational sentences were further subdivided into high and low context. The subjects were presented with sentences in noise and were to write the final word on an answer sheet. The authors hypothesized that non-native listener speech-in-noise perception would be improved by both semantic (high context) and acoustic-phonetic (clear speech) enhancements.

Bradlow & Alexander (2007) addressed the limitation of uncontrolled target word predictability in Bradlow & Bent (2002). By doing so, they isolated the effect of clear speech from higher-level semantic-contextual information. From the results, they conclude that non-native listeners do gain a significant benefit from clear speech independent from a decreased ability to use semantic-contextual information.

The authors further suggest that listeners with less exposure to their L1 (i.e., children and non-natives) will eventually develop a greater degree of the clear speech effect with increased exposure to the language in question. The authors maintain that native listeners utilize the language-specific, code enhancements of clear speech, but that non-natives utilize mainly the signal enhancements of clear speech. In other words, native listeners use the exaggerated acoustic distance between contrasting categories (less vowel reduction), increased duration, and the pronunciation norms typically heard in clear speech. Non-natives, they assert, use the overall acoustic improvement of the signal, such as a slower speaking rate, a wider dynamic pitch range and more precise stop consonant releases (Picheny et al., 1985b).

The authors' final remarks (Bradlow & Bent, 2002) include an admission of the need for a better understanding of how talker- and listener-related factors interact to influence overall speech intelligibility. This supports the need for further research in the area of acoustic analysis of bilingual clear speech production.

*Purpose of the Present Study*

The purpose of this study is to examine the effects of linguistic experience on the acoustic properties of six target vowels produced in clear and conversational speech styles. Three talker groups were recruited: monolingual native English speakers, early (relatively balanced or English dominant) Spanish-English bilinguals and late (primarily Spanish dominant) Spanish-English bilinguals. The acoustic variables analyzed include vowel duration, fundamental frequency and formant frequencies at vowel midpoint (50% of vowel duration), and extent of change in formant frequencies across the target vowel duration (from 20% to 80% of vowel duration).

The present study tests the hypothesis that later and/or early learners of English as a second language may exhibit an exaggerated or restricted degree of change in their production performance between clear and conversational speech styles for certain acoustic cues. On at least some features, the productions of early learners were expected to be similar to those of native speakers. The productions of late learners of English were expected to differ more from those of monolinguals, and certain target vowel pairs (e.g., /i/-/I/ were expected to overlap substantially in their production, especially for the late learners). The present study differs from previous studies of second language vowel production in that it examines the spectral change of L2 vowels versus vowels produced by native English speakers and examines non-native speakers' ability to modify acoustic properties of vowels when asked to change speaking style.

## Chapter Two

### Method

#### *Inclusion and Exclusion Criteria*

Monolinguals who participated included adults up to age 60 who were native speakers of English. They were required to have no history of speech or hearing impairment or a strong regional accent. Persons who rated themselves as fluent in a second language, or whose parents/caregivers used another language with them as a child were not included. It was preferred that talkers be born and raised in the Tampa Bay area, but other subjects not fitting this criterion were allowed.

Bilinguals who participated included adults up to age 60 who were native speakers of any New World variety of Spanish (Caribbean, South American, Central American, or Mexican). They were required to have no history of speech or hearing impairment, nor to speak any languages other than Spanish and English. The Spanish talkers were further divided into two groups consisting of ten late bilinguals and 15 early bilinguals, based on their age of onset of immersion in an English-speaking environment (AOI). The experienced *early* bilinguals' English AOI was age 12 or under. Furthermore, this group rated themselves as English dominant or balanced in at least two modalities (listening, speaking, reading and writing), one of which was required to be non-print (i.e.,

must be listening or speaking). The less experienced *late* bilinguals' English AOI was age 15 or later.

Participants were recruited through flyers placed around the university campus. All participants were prescreened over the phone for inclusion criteria. Each participant was paid \$20 upon completion of the one-hour recording session, which was preceded by a one-hour session of perceptual testing (associated with a related experiment) on a preceding day.

### *Participants*

The participants included in the results comprised three groups of talkers: 1) ten native English speakers (monolinguals - MO); 2) 15 early Spanish-English bilinguals (EB); and 3) ten late Spanish-English bilinguals (LB). Males and females were recruited equally, however, more females than males volunteered for all three groups, so that less than one fourth of any group was represented by males.

The male participants were therefore dropped from the study due to their representation of a low proportion all three groups. A gender effect on degree of intelligibility difference between clear and conversational speech was found by Ferguson and Kewley-Port (2004). With the small proportion of males, gender effects could not easily be analyzed and their effects on the data would therefore be unknown. Other female participants who did not fit the criteria were allowed to participate, but were later dropped after detailed reading of their questionnaires. Of the total participants recruited, data for ten of 24 monolinguals, 15 of 33 early bilinguals, and 10 of 21 late bilinguals were included for analysis in the present

study. Some participants were dropped from acoustic analysis because their voice quality caused automatic formant tracking to be unreliable.

Table 1. *Demographic data for early bilingual talkers. Data are displayed for gender; age; country of origin (of listener or listener's parents if born in the U.S.); age of onset of immersion in an English-speaking environment (AOI); number of years spent living in the U.S.; and self-ratings of language dominance (E=English; S=Spanish; B=balanced) for the skills of listening, speaking, reading and writing.*

Code	Age	Born/ Raised in US?	Language background information					
			Country	AOI	Language most comfortable for:			
					Speak	Listen	Read	Write
EB05	19	Y	Cuba	4.5	E	E	E	E
EB06	19	N	Mexico	5	B	B	E	E
EB08	19	N	Nicaragua	8	E	E	E	E
EB10	19	Y	Nicaragua	6	B	B	B	B
EB11	20	Y	Cuba	6	E	E	E	E
EB12	24	N	Puerto Rico	10	E	E	E	E
EB16	19	Y	Mexico	6	S	E	E	E
EB17	19	Y	Cuba	4	E	E	E	E
EB19	18	Y	Cuba	4	E	E	E	E
EB24	26	Y	Colombia	5	E	E	E	E
EB25	21	N	Colombia	11	E	E	E	E
EB26	26	N	Venezuela	12	B	B	E	E
EB29	19	Y	Cuba	2	B	B	E	E
EB30	19	N	Venezuela	8	B	B	B	E
EB33	22	N	Colombia	6	S	E	E	S
Avg./	20.6	8 Y;	3 Colom.;	6.5	8 E;	10 E;	13 E;	13 E;
Sum.		7 N	2 Venez.;		5 B;	5 B;	2 B;	1 B;
			5 Cuba;		2 S	0 S	0 S	1 S
			2 Mexico;					
			3 Other					

Table 2. *Demographic data for late bilingual talkers. Data are displayed for gender; age; country of origin (of listener or listener's parents if born in the U.S.); age of onset of immersion in an English-speaking environment (AOI); number of years spent living in the U.S.; and self-ratings of language dominance (E=English; S=Spanish; B=balanced) for the skills of listening, speaking, reading and writing.*

Code	Age	Born/ Raised in US?	Language background information					
			Country	AOI	Language most comfortable for:			
Speak	Listen	Read			Write			
LB01	30	N	Panama	21	E	S	B	B
LB06	19	N	Colombia	16	S	S	S	S
LB07	50	N	Colombia	45	S	S	S	S
LB10	28	N	Colombia	28	S	S	S	S
LB11	22	N	Colombia	22	S	S	S	S
LB13	19	N	Puerto Rico	16	S	S	S	S
LB15	22	N	Colombia	18	S	S	S	S
LB16	49	N	Colombia	46	S	S	S	S
LB19	22	N	Cuba	19	S	S	E	E
LB21	21	N	Colombia	18	S	S	S	S
Avg./	29.6	10 N	11 Colom.;	24.9	13 S;	14 S	11 S;	12 S;
Sum			1 Cuba;		1 E		2 E;	1 E;
			3 Other				1 B	1 B

### Materials

Six target vowels /i, ɪ, e, ε, æ/ and /a/, embedded in /bVd/ context, were used as stimuli for the experiment. The target words were written as “bead, bid, bayed, bed, bad” and “bod” and were embedded in the carrier phrase “Say \_\_\_\_\_ again.”

Digitization and recording equipment included an Audio-Technica: AT4033 condenser microphone, an Applied Research and Technology microphone preamplifier with 48V phantom power supply, a Roland VS890 Digital Studio Workstation recorder, and Sennheiser HD265 headphones.

Editing software used included a signal editing software program (CoolEdit 2000, 2000) and Praat speech analysis software (Boersma & Weenink, 2006). The digitization/recording equipment was configured with the microphone connected to the input channel of the Applied Research and Technology microphone preamplifier. The preamplifier was connected to an analog input channel of the Roland VS890 Digital Studio Workstation. Recordings were digitized at 44.1 kHz with 24 bit resolution on AD conversion – 64 times oversampling. An antialiasing filter (20 kHz) was used and filtering automatically performed by the workstation; the effective response range was 20 Hz – 20 kHz.

The written stimuli were presented to talkers on a 15 inch flat screen monitor located inside the recording booth. The CPU of the computer was located outside of the recording booth.

Following recording, the experimenter transferred the files from the digital workstation to a PC. The files were transferred digitally using coaxial cable connected from the digital output of the workstation to the digital input of an M-Audio Audiophile 2496 sound card installed on the computer. Each recording session was transferred digitally, with separate files for conversational and clear speech stimuli.

#### *Recording Procedure*

Three experimenters conducted the recording of stimuli by the talkers. All were trained and judged by a trained linguist (the major professor) to be consistent in procedural manner.

An informed consent document, a race-ethnicity form and a language background questionnaire were filled out by every participant recruited. Each talker was recorded in a single-wall sound attenuating booth (IAC). Recording equipment (other than microphone) was located outside the booth. The microphone was positioned approximately six inches from talker's mouth and located at a 45 degree angle from the talker's mouth. Recording levels were monitored and adjusted as needed by the experimenters to avoid peak clipping and to maintain sufficiently high input amplitude.

There were two different speech styles (conversational and clear) produced by each talker. The experimenter showed the stimulus words to the talkers and read them aloud to the talker in order to avoid orthographic errors. Distractor words were included in the conversational style reading list to keep talkers from focusing too much on the /bVd/ frame of the target words. Distractor words were all single syllable /CVC/ (but not /bVd/) words (e.g., "cut, cape"). Target and distractor words were intermixed for the conversational condition. For the clear speech condition, only the target words were used.

Each word (embedded in carrier phase – e.g., "Say bad again") was presented using a Microsoft PowerPoint presentation file. A separate monitor and keyboard with dual control were located outside the recording booth. When the subject finished saying the sentence, the experimenter clicked on the screen (or pressed the right arrow key) to present the next sentence.

Twelve practice trials (one for each target and each distractor word) were conducted. On each practice trial, the subject heard the sentence to be read over

headphones and saw the text displayed on the screen. The subject was instructed to repeat the sentence in a normal speaking style. Audio of the 12 sentences to be repeated were produced by a single male talker (a monolingual native English speaker), recorded using the same procedures and equipment described above. These recorded stimuli were transferred to the computer in same way as described above. Each target phrase was saved to a separate file for presentation during the practice trials.

During the conversational style trials, the subject was instructed to remove the headphones used for the practice trials. The text of each target sentence was presented on the screen and the subject was instructed to read each sentence aloud in a normal speaking style. Each talker produced seven repetitions of each target and distractor word, for a total of 84 target sentences produced in the conversational style. Four lists of 21 sentences each were read by each talker with an opportunity for a short break given between each block of 21 sentences. The 84 target and distractor words were pseudorandomized so that no more than two /bVd/ words occurred in a row. Approximately half of the /bVd/ target words for each vowel were presented in the first two lists.

During the clear style trials, the talkers were instructed that some of the sentences they had produced needed to be spoken more clearly – as if speaking to someone who doesn't understand. The subjects were not given any particular instructions as to how to produce clear tokens. No distractor words were used for this condition. Each talker produced seven repetitions of each target word, for a total of 42 target sentences. Two lists of 21 sentences each were read by each

talker with an opportunity for a short break given between each block of 21 sentences. The 42 words were pseudorandomized so that no target word was occurred two times in a row. The entire recording session took approximately one hour for each talker, including completion of consent forms and questionnaires.

### *Editing Procedure*

Two trained experimenters edited all recorded target-word stimuli into isolated words. Each larger file (for session or style) was opened and subsequently edited in CoolEdit 2000. Each list of 21 sentences was isolated from the larger file and saved to a separate file. Each sentence containing a target word in the list of 21 sentences was then edited to isolate the target word only.

The target word was isolated by first locating and selecting the release of the initial /b/, plus 20 ms of the waveform preceding the /b/ release. The contents of the file preceding this 20 ms buffer were then deleted. The first 10 ms of the 20 ms buffer were then silenced. In cases where prevoicing of /b/ occurred, the next 3 ms were selected and linearly ramped from 0 to 100% of the original amplitude to prevent the perception of a click. Thus, the initial /b/ and up to 10 ms of prevoicing were preserved in the isolated word files. Next, the release of the word-final /d/ was located and selected on the waveform, plus 20 ms of the waveform following the /d/ release. The contents of the file following this 20 ms buffer were then deleted. The last 10 ms of the word-final 20 ms buffer were then silenced. Then the 3 ms of energy preceding the last 10 ms were linearly ramped from 100 to 0% of the original amplitude, again to prevent the perception of a

click. Thus the release of the word-final /d/ and 10 ms of the energy following were preserved in the isolated word files. Finally, the remaining waveform was saved to a new isolated word file.

Two of the seven tokens recorded from each talker for each of the target words were selected for analysis in the present study. The first and second tokens produced by each talker were used unless there was disfluency or poor voice quality or the talker clearly made an error in reading the word. If a token was not usable, the experimenter examined additional repetitions until an acceptable one was found.

Prior to acoustic analysis, all isolated word files were amplitude equalized for use in a separate experiment. For equalization, the average RMS of each file was set to -25 dB from the maximum amplitude. To accomplish this, the full duration of the isolated word file (including the silence of 10 ms of silence on the beginning and end) was selected and then the file's average RMS was computed using an automated procedure (CoolEdit 2000, 2000). The difference from -25 dB was computed and the amplitude adjustment procedure in CoolEdit was used to adjust amplitude up or down by the desired number of dB to get the average RMS of the file to equal -25. After amplitude adjustment, equalization was double checked by again obtaining the average RMS for the entire file and checking that it was equal to -25 dB.

#### *Settings for Acoustic Analysis*

All time and frequency measurements described below were made using Praat (Boersma & Weenink, 2006). The following settings were used, except in

cases where formant tracking did not provide a good match to observed formants on the wide-band spectrogram (see below): window length for spectrogram = 5 ms (wide-band spectrogram); spectrogram display range = 0-5500 Hz; spectrogram display dynamic range = 50 dB (Praat default); pre-emphasis for spectrogram display = 6 dB/octave (Praat default); method for automatic tracking of F0 = autocorrelation; range for F0 tracking = 75-500 Hz; method for formant tracking = Burg; pre-emphasis starting frequency for formant tracking = 50 Hz; number of formants to be tracked within 0-5500 Hz = 4, 5, or 6, depending on the experimenter's judgment based on visual inspection of the agreement between formant tracks and formants observed on the wide-band spectrogram; window length for formant tracks = 20 ms.

#### *Vowel Duration Measurement*

Measurement of vowel duration was performed by two trained experimenters (the author and a trained assistant). Agreement was checked and any additional measurement needed was performed by a trained linguist (the major professor). Criteria for determining vowel duration were specific. For the beginning of the vowel (vowel onset), experimenters located on the waveform the first large positive amplitude peak following the maximum negative of the first periodic cycle that had the same pattern as the rest of the vowel (i.e., not part of pre-voicing). The onset of F2 on the wide-band spectrogram was also used to confirm the location of the vowel onset. The first pulse where F2 was visible was a landmark for vowel onset. Typically, the waveform and spectrogram criteria for vowel onset agreed well; when they did not, the experimenters selected one of

the two criteria using their best judgment to determine the location of the vowel onset.

For the end of the vowel (vowel offset), the experimenters used the waveform display to locate the peak of the first negative pulse of the last cycle of voicing that had a similar shape as the rest of the vowel (last cycle prior to closure – not included in more sinusoidal cycles occurring during voicing during closure). The offset of F2 on the wide-band spectrogram was also used to confirm the location of the vowel offset. The last pulse where F2 was visible during the vowel was the spectrographic landmark for the vowel offset. Typically, the waveform and spectrogram criteria for vowel offset agreed well; when they did not, the experimenters selected one of the two criteria using their best judgment to determine the location of the vowel offset. Vowel onset and offset measures for each selected token were copied and saved to a spreadsheet. A spreadsheet formula automatically computed vowel duration and locations for 20%, 50% and 80% of vowel duration when onset and offset data were entered.

When all vowel onset and offset measurements were completed independently by the two student experimenters, the trained linguist used a spreadsheet formula to determine agreement for the vowel onset and offset taken by the two student experimenters. The agreement criterion was set to 5 ms, which is approximately one pitch period for the average female, rounded to the nearest ms. That is, the average fundamental frequency (F0) for females is 219 Hz according to Hillenbrand et al. (1995), which converts to 4.57 ms per pitch period. The criterion for one pitch period for agreement was adapted from

Strange, Yamada, Kubo, Trent, Nishi & Jenkins (1998). For consistency's sake, the time measurements of a single student experimenter (the author) were used as the landmarks for frequency measurements for all instances in which the two students agreed.

The times of vowel onset and/or vowel offset were remeasured by the trained linguist for all tokens for which the measures of vowel onset or vowel offset of the two student experimenters disagreed by more than 5 ms. In nearly every case, the measurement of the trained linguist agreed with that of one of the student experimenters. In the few cases where the measurement of the trained linguist did not agree with that of either of the students, the trained linguist re-checked the measurement and recorded her own measurements in the spreadsheets of both raters.

### *Frequency Measurements*

Following time agreement measurement, fundamental frequency (F0) and the frequencies of the first four formants (F1-F4) were measured at the time points of 20, 50 and 80% of the vowel duration. Only measurements for duration, F1 and F2 will be used for the present thesis. As stated above, the time points of a single rater (identity dependent on agreement) were used to determine points from which to make formant measurements.

Frequency measurements were performed by three trained experimenters and the trained linguist. Frequency measurements were made by two of these four persons for each token and recorded to separate spreadsheets; agreement between the data on the two spreadsheets for each token was then computed by

the trained linguist. Agreement criteria for F1, F2 and F3 were +/- 50, 150 and 250 Hz respectively, following Strange et al. (1998). The agreement criterion for F4 was the same as for F3 (+/- 250 Hz).

In cases of agreement between the two spreadsheets, the measurements from a single spreadsheet (that of the author) were used. In cases where agreement within the specified criteria was not found, frequency measurements were made by a third experimenter and values for which at least two raters agreed were subsequently used; in the rare cases where all three raters disagreed, the measurements of the trained linguist were used.

For measurement of F0, automatic measurements were used almost exclusively. In the rare instances where the pitch tracking appeared to be in error, measurements were made by hand from the waveform by measuring the duration of the target pitch period and converting to Hz.

Two measurement techniques were used for measurement of formant frequencies. Automatic formant tracking was used in most cases, but analysis by hand was used in some cases. For automatic analysis, the automatic formant tracking feature (Formant → Show Formants) was used to overlay formant tracks on the wide band spectrogram display. The Praat (Boersma & Weenink, 2006) query feature was then used to automatically obtain the locations of F1-F4 and this information was then pasted into the spreadsheet for each token. The number of formants chosen as a setting in the automatic formant tracker was modified based on experimenter estimation of the best match between the formant tracker setting and the formants observed on the wide-band

spectrogram. Any extra formant tracks seen on the display (between formants observed on the wide-band spectrogram) were skipped for the purpose of measurement. The number of formants used for tracking was four, five or six for each token; this information was also recorded in the spreadsheet for each token.

By hand analysis from a narrow-band spectral slice was used for tokens that did not yield reliable formant tracks using the automatic formant tracking feature. This method was adapted from Monsen & Engebretson (1983). For this procedure, the spectrogram display was converted to a narrow band spectrogram by specifying a 29 ms analysis window. Then a spectral slice (frequency by amplitude display) was generated for the desired time point using an automatic feature of Praat (Boersma & Weenink, 2006). The frequency range 0-5500 Hz was selected for display. The location of the first four formants (or the desired formant or formants) was determined by clicking on the estimated location of the formant, causing a cursor to appear at that point. The frequency value at the cursor was automatically obtained by the Praat (Boersma & Weenink, 2006) query procedure and pasted into the spreadsheet. The formant locations were determined by visually estimating the location of the peaks in the spectrum according to the method described in Monsen & Engebretson (1983), in which a hypothetical triangle is created and superimposed over prominent harmonics and the peak of the triangle is adjusted to the left or right to a position that would result in the harmonic amplitude relationship observed. All formant frequency measurements determined by hand were noted as such in

spreadsheet by each experimenter. Formant frequency measurements were converted to the Bark scale for statistical analysis (Traunmüller, 1990).

## Chapter Three

### Results

Four separate three-way mixed-design analyses of variance (ANOVAs) were performed on four dependent variables (see below). In each case, the between-subjects independent variable was talker group (three levels: MO, EB and LB) and the within-subjects independent variables were speaking style (two levels: conversational and clear) and target vowel (six levels: /i, ɪ, e, ε, æ, a/). In each case, simple main effects post-hoc comparisons were used to explore significant effects and interactions.

The following dependent variables were derived directly from the vowel measurements described above: vowel duration (measured in ms), F1 (in Barks) at 50% of vowel duration and F2 (in Barks) at 50% of vowel. In addition, the two-point vector length for F1-F2 frequencies from 20% of the vowel duration to 80% of the vowel duration was computed by finding the Euclidean distance (in Barks) between the F1-F2 frequencies at these two time points (cf. Ferguson & Kewley-Port, 2002). These values were then used as the dependent variable in a fourth three-way mixed-design ANOVA. Note that the F0, F3 and F4 values for all target vowels, talker groups, and speaking styles are awaiting further analysis.

### Vowel Duration

Table 3. *Statistical results for vowel duration. Data on F values, degrees of freedom (df) and levels of significance (p values) for all main effects and interactions in the three-way ANOVA of the effects of talker group, speaking style and target vowel on duration of target vowels. Significant effects are indicated by an asterisk.*

Effect	F (df)	p value
Main effects		
Talker group	.215 (2,32)	.808
Speaking style *	88.79 (1,32)	<.001
Target vowel *	125.08 (5,160)	<.001
Two-way interactions		
Talker group by speaking style	.47 (2,32)	.631
Talker group by target vowel *	4.70 (10,160)	<.001
Speaking style by target vowel *	4.00 (5,160)	.002
Three-way interaction		
Talker group by speaking style by target vowel *	2.81 (10,160)	.003

Table 3 summarizes results for the three-way ANOVA on vowel duration. Significant main effects were found for speaking style and target vowel. The two-way interactions of talker group by target vowel and speaking style by target vowel were significant. The three-way interaction was also significant. F values

and p values for each significant effect are shown in Table 3. Only the three-way interaction will be discussed in detail because it alters the other effects.

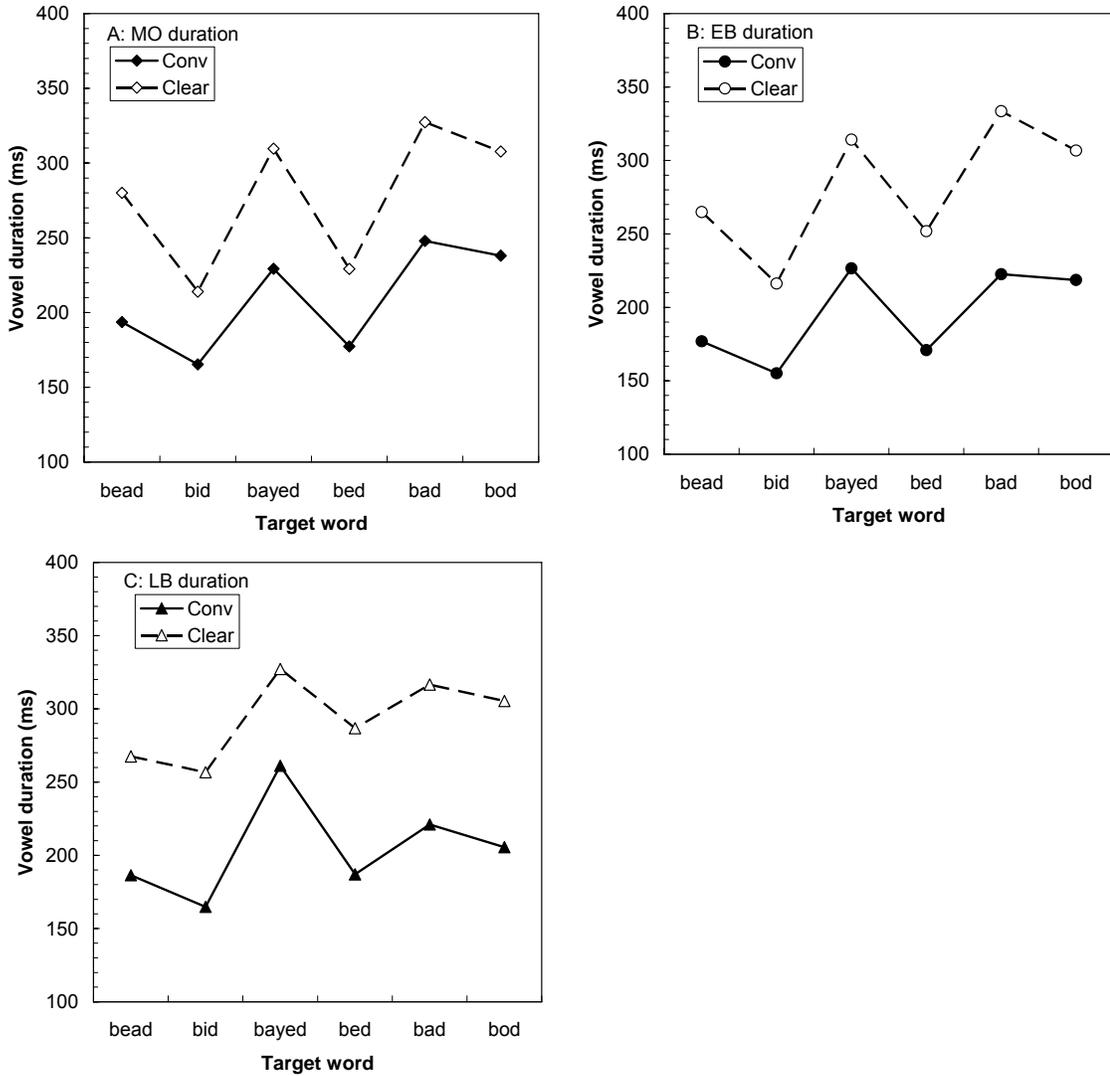


Figure 1. Average durations (in ms) of target vowels for words produced in conversational and clear speech styles. MO= monolingual talkers (panel A); EB= early bilingual talkers (panel B); LB=late bilingual talkers (panel C).

Figure 1 shows mean vowel durations (in ms) for each speaking style and target vowel, with a separate panel for each talker group. As can be seen from the figure, the “long vowels” (in particular the vowels in the words “bead, bayed” and “bad”) appear to be lengthened in clear speech more than their neighboring shorter vowels for the MO and EB talker groups (see Figures 1A and 1B). Thus, vowel durations are better distinguished for neighboring vowels in clear than in conversational speech for these two talker groups. For the LB talkers, on the other hand, the vowels in “bayed” and “bead” are lengthened less in clear speech less than their neighboring vowels, effectively reducing the degree of inherent vowel differences in clear speech (see Figure 1C).

Post-hoc tests comparing vowel durations within each level of group and style confirm these observations. For the MO talker group, the vowels /æ/ and /a/ did not differ significantly in duration in conversational speech (10 ms difference) but did in clear speech (20 ms difference). Although the duration difference between /i/ and /I/ was significant in both styles, it increased from about 28 ms in conversational speech to about 66 ms in clear speech.

For the EB talkers, the durations of the vowels /e,æ/ and /a/ were all within 8 ms of one another and did not differ significantly in conversational speech. In clear speech, /æ/ was significantly longer than both /e/ and /a/ (by 20 and 27 ms, respectively). Furthermore, the difference in duration between the vowels /i/ and /I/ increased from 22 ms to 48 ms from conversational to clear speech; the duration difference between /i/ and /I/ was significant for both styles.

For the LB talkers, on the other hand, the duration difference between the vowels /e/ and /ɛ/, while significant in both styles, decreased from 74 ms in conversational speech to 40 ms in clear speech. Similarly, the vowel /i/ is 22 ms longer than /I/ in conversational speech (a significant difference), but only 11 ms in clear speech (a non-significant difference). Together, the vowel duration results show the MO and EB talkers emphasizing vowel duration differences between neighboring vowels in clear speech. The LB talkers show less differentiation in duration between neighboring vowels in clear than in conversational speech.

#### *F1 at 50% of Vowel Duration*

Table 4 summarizes results for the three-way ANOVA on F1 at 50% of vowel duration. A significant main effect was found for target vowel only. The two-way interactions of speaking style by talker group, talker group by target vowel and speaking style by target vowel were significant. The three-way interaction was not significant. F values and p values for each significant effect are shown in Table 4.

Table 4. *Statistical results for F1 at 50% of vowel duration. Data on F values, degrees of freedom (df) and levels of significance (p values) for all main effects and interactions in the three-way ANOVA of the effects of talker group, speaking style and target vowel on the value of F1. Significant effects are indicated by an asterisk.*

Effect	F (df)	p value
Main effects		
Talker group	1.59 (2,32)	.219
Speaking style	.07 (1,32)	.790
Target vowel *	607.30 (5,160)	<.001
Two-way interactions		
Talker group by speaking style *	3.60 (2,32)	.039
Talker group by target vowel *	9.01 (10,160)	<.001
Speaking style by target vowel *	2.61 (5,160)	.027
Three-way interaction		
Talker group by speaking style by target vowel	.90 (10,160)	.534

Figures 2 and 3 show average Bark-frequency values of F1 (y-axis) and F2 (x-axis) at 50% of vowel duration for conversational (solid lines) and clear speech vowels (dashed lines). Each talker group is shown as a separate panel; data for the monolingual talker group are repeated in Figures 2 and 3 for easier comparison. Both axes are shown with values in reverse order, for better representation of jaw height and tongue position locations.

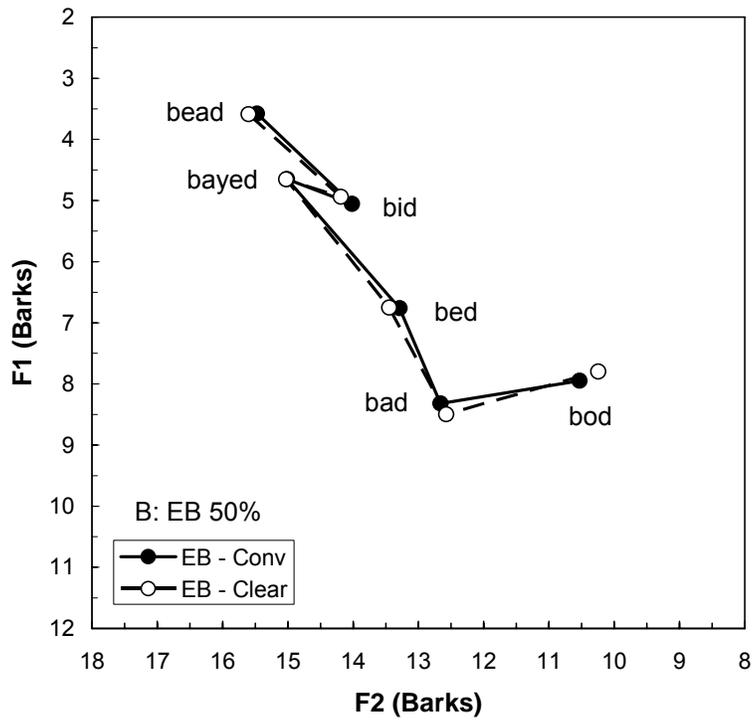
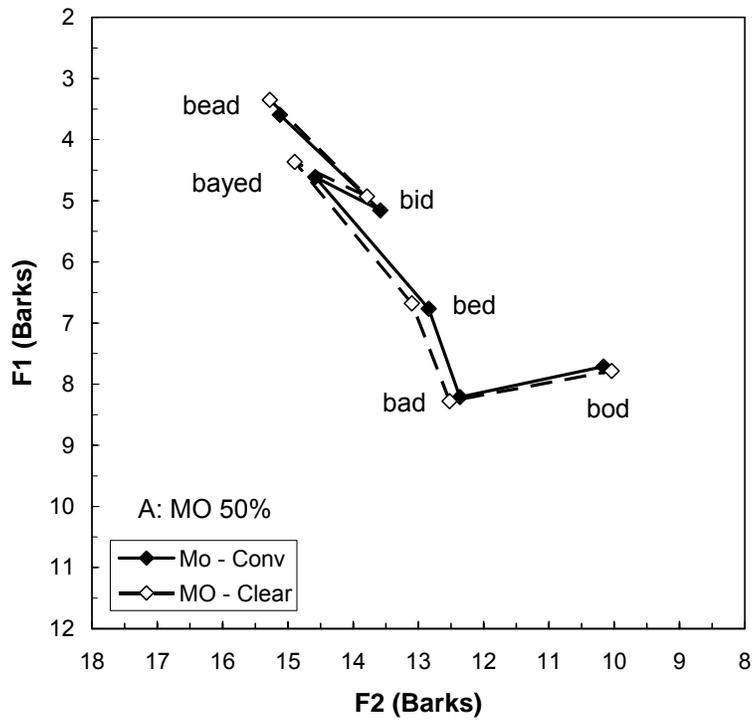


Figure 2. Average steady-state (50% of vowel duration) F1 and F2 frequencies (in Barks) for vowels in conversational and clear speech (MO and EB talkers). MO= monolingual talkers (panel A); EB= early bilingual talkers (panel B).

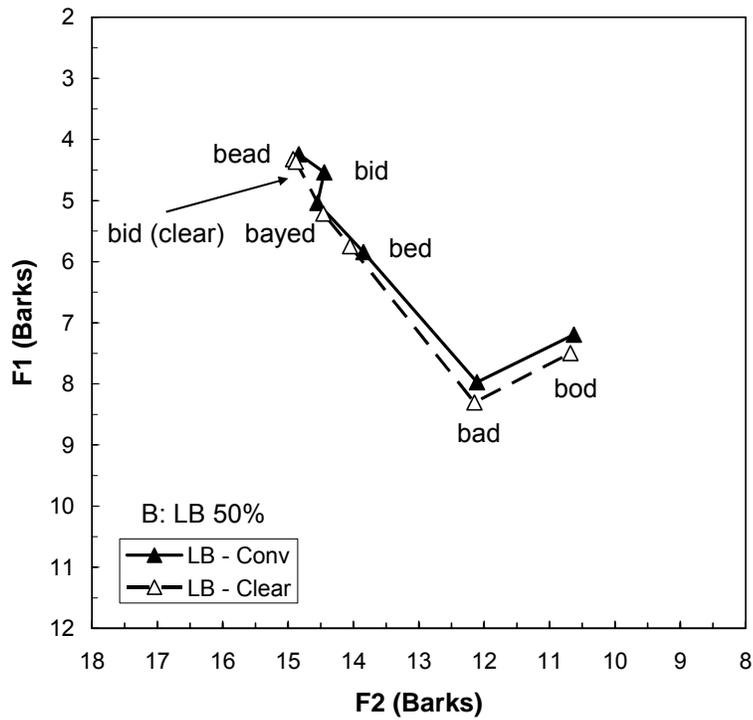
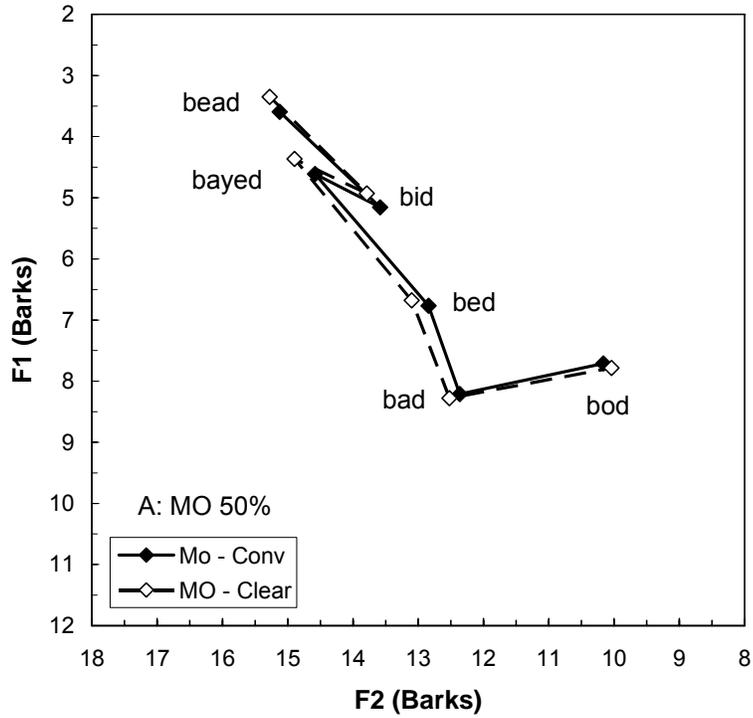


Figure 3. Average steady-state (50% of vowel duration) F1 and F2 frequencies (in Barks) for vowels in conversational and clear speech (MO and LB talkers). MO = monolingual talkers (panel A); LB = late bilingual talkers (panel B).

As can be seen from the figures, F1 values are slightly lower (indicating a higher tongue/jaw position in clear than in conversational speech for the vowels /i, I, e/ and /ɛ/ for the MO talkers (Figure 2A) and for the vowels /i, I, ɛ/ and /a/ for the EB talkers (Figure 2B). For the LB talkers (Figure 3B), only /ɛ/ and /I/ show decreases in F1 from conversational to clear speech. The values for /æ/ and /a/ on the other hand are higher in clear than in conversational speech for the LB talkers, as is that for /æ/ for the EB talkers, indicating a lowering of tongue/jaw position in clear speech.

A comparison of Figures 2A and 2B shows only minor differences in F1 values between the MO and EB talker groups. The relative positions and distances between the vowels on the F1 axis are nearly identical for the two groups. A comparison of Figures 3A and 3B, however, shows quite noticeable differences in vowel location between the MO and LB talkers. The vowels /i/ and /e/ are located lower in the vowel space (higher F1) for LB than for MO (and EB) talkers. The vowels /I, ɛ, æ/ and /a/, by contrast, are located higher in the vowel space (lower F1) for LB than for MO talkers. Thus, the maximum F1 distance between vowels appears to be reduced for the LB talkers, compared to the MO and EB talkers.

The post-hoc comparisons for the speaking style by talker group interaction revealed no significant speaking style effects for any of the three groups; however, the MO talkers' F1 values were nearly significantly lower in

clear than in conversational speech ( $p=.057$ ), partially confirming the observation of lower F1 values in clear speech for certain vowels. For the LB talkers, there is a nearly significant increase in F1 values from conversational to clear speech ( $p=.079$ ), partially confirming the higher F1 values observed in clear speech for /æ/ and /a/.

Post-hoc analyses of the group by vowel interaction showed significantly higher F1 values for LB than for MO and EB talkers for the vowels /i/ and /e/, confirming the observation of a lower position in the vowel space for these vowels. LB talkers had significantly lower F1 values than MO and EB talkers for the vowels /I/ and /ɛ/, confirming the observation of a higher position in the vowel space for these vowels. Finally, LB talkers had significantly lower F1 values than EB talkers for the vowel /a/, indicating a higher position in the vowel space. No significant differences in F1 values were found between MO and EB talkers.

Post-hoc comparisons of individual vowels' F1 values within each group showed all vowels to differ significantly from one another for both the MO and EB talker groups. The order of the F1 values was also the same for these two groups. For the LB talkers, no significant difference in F1 frequency was found between /i/ and /I/. Otherwise, all of the vowels differed significantly in F1 for the LB talkers, and the order of the F1 values was the same as for the other two groups.

Post-hoc analysis of the style by vowel interaction showed a significantly lower F1 value in clear than in conversational speech for the vowel /I/ (indicating

a higher position in the vowel space) and a significantly higher F1 value in clear than in conversational speech for the vowel /æ/ (indicating a lower position in the vowel space). No other vowels showed significant differences between clear and conversational speaking styles.

#### *F2 at 50% of Vowel Duration*

Table 5 summarizes results for the three-way ANOVA on F2 at 50% of vowel duration. Significant main effects were found for talker group, speaking style and target vowel. The two-way interactions of talker group by target vowel and speaking style by target vowel were significant. The three-way interaction was not significant. F values and p values for each significant effect are shown in Table 5. F2 values are shown along with F1 values for each talker group, target vowel and speaking style in Figures 2 and 3.

An examination of Figure 2A shows that all of the MO talkers' vowels except /a/ have slightly higher F2 values (are slightly more fronted) in clear than in conversational speech. A similar but smaller pattern is shown for the EB talkers (see Figure 2B). Figure 3B shows this pattern for the LB talkers only for the vowels /I/ and /ε/; however, the LB talkers' production of /I/ is sufficiently fronted (and raised) in the clear speech style that it nearly completely overlaps with target /i/. A comparison of Figures 2A and 2B also shows higher F2 values for the EB talkers than for the MO talkers for all six of the target vowels, suggesting that all vowels are slightly more fronted for the EB talkers than for the MO talkers, regardless of speaking style.

Table 5. *Statistical results for F2 at 50% of vowel duration. Data on F values, degrees of freedom (df) and levels of significance (p values) for all main effects and interactions in the three-way ANOVA of the effects of talker group, speaking style and target vowel on the value of F2. Significant effects are indicated by an asterisk.*

Effect	F (df)	p value
Main effects		
Talker group *	3.66 (2,32)	.037
Speaking style *	9.00 (1,32)	.005
Target vowel *	932.14 (5,160)	<.001
Two-way interactions		
Talker group by speaking style	1.83 (2,32)	.103
Talker group by target vowel *	11.37 (10,160)	<.001
Speaking style by target vowel *	5.12 (5,160)	<.001
Three-way interaction		
Talker group by speaking style by target vowel	1.35 (10,160)	.210

Post-hoc comparisons of the vowel by group interaction showed significant group differences for all of the target vowels, but the order of the groups' F2 values varied across target vowels. For /i, I/ and /ε/, all three groups differed significantly from one another in their F2 values. For /i/, F2 values were significantly higher for EB talkers than for MO and LB talkers, and values for MO

talkers were significantly higher than those for LB talkers. These differences indicate a more front position in the vowel space for the EB talkers than for the MO talkers and for the MO talkers than for the LB talkers.

For /I/ and /ε/, F2 values were significantly lower for the MO talkers than for the EB and LB talkers and lower for the EB talkers than for the LB talkers. These differences indicate a more back position in the vowel space for the MO talkers than for the EB talkers and for the EB talkers than for the LB talkers.

For the vowels /e/ and /æ/, F2 values were significantly higher for the EB talkers than for the LB talkers, but the F2 values for the MO talkers did not differ significantly from those for either of the other two groups. Similar to /i/, these differences indicate a more front position for the EB than for the LB talkers.

For the vowel /a/, F2 values were significantly lower for the MO group than for the LB group, but the F2 values for the EB talkers did not differ significantly from those for either of the other two groups. Similar to the results for /ε/, these differences indicate a more back tongue position for the MO than for the LB talkers. Overall, the group by vowel effect shows a smaller distance between the vowels /i/ and /a/ (most front vs. most back) for the LB talkers (/i/-/a/ distance = 4.2 Barks) than for the MO and EB talker groups (/i/-/a/ distance = 5.1 Barks).

Post-hoc comparisons of individual vowels' F2 values within each group showed all vowels to differ significantly from one another for both the MO and EB talker groups. The order of the F2 values was also the same for these two

groups. For the LB talkers, no significant difference in F2 frequency was found between /i/ and /I/ or between /I/ and /e/. The other three vowels differed significantly in F2 from one another (and from /i, I / and /e/) for the LB talkers, and the order of the F2 values for these vowels was the same as for the other two groups. The F2 difference between /æ/ and /a/ was about .8 Barks smaller for the LB than for the MO group; however, the F2 difference between /æ/ and /ɛ/ was about 1.3 Barks larger for the LB than for the MO group (due to the placement of /ɛ/ higher in the vowel space for the LB talkers).

Post-hoc analysis of the style by vowel interaction showed a significantly higher F2 value in clear than in conversational speech for the vowels /i, I/ and /ɛ/ (indicating a more front position in the vowel space) and a nearly significantly lower F2 value in clear than in conversational speech for the vowel /a/ (indicating a more back position in the vowel space). No other vowels showed significant differences between the clear and conversational speaking styles.

#### *Length of Vector from 20% to 80% of Vowel Duration*

Table 6 summarizes results for the three-way ANOVA on length of the vector in the F1-F2 space from 20% to 80% of the vowel duration. Significant main effects were found for speaking style and target vowel only. No interactions were statistically significant. F values and p values for each significant effect are shown in Table 6.

Table 6. *Statistical results for two-point vector length. Data on F values, degrees of freedom (df) and levels of significance (p values) for all main effects and interactions in the three-way ANOVA of the effects of talker group, speaking style and target vowel on the value of the Euclidean distance between F1-F2 frequencies at 20% and 80% of vowel duration. Significant effects are indicated by an asterisk.*

Effect	F (df)	p value
Main effects		
Talker group	1.06 (2,32)	.357
Speaking style *	13.08 (1,32)	.001
Target vowel *	59.24 (5,160)	<.001
Two-way interactions		
Talker group by speaking style	.72 (2,32)	.495
Talker group by target vowel	.95 (10,160)	.491
Speaking style by target vowel	.88 (5,160)	.494
Three-way interaction		
Talker group by speaking style by target vowel	.42 (10,160)	.934

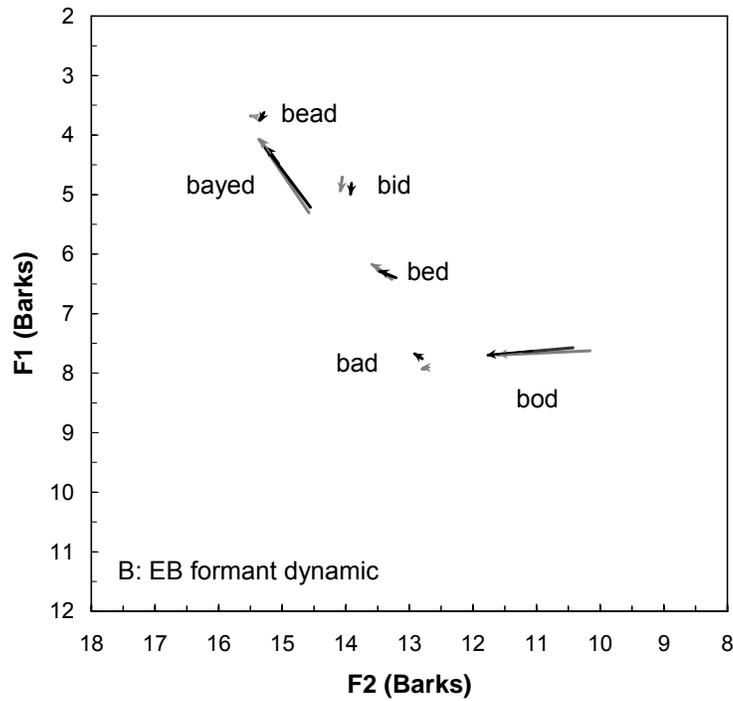
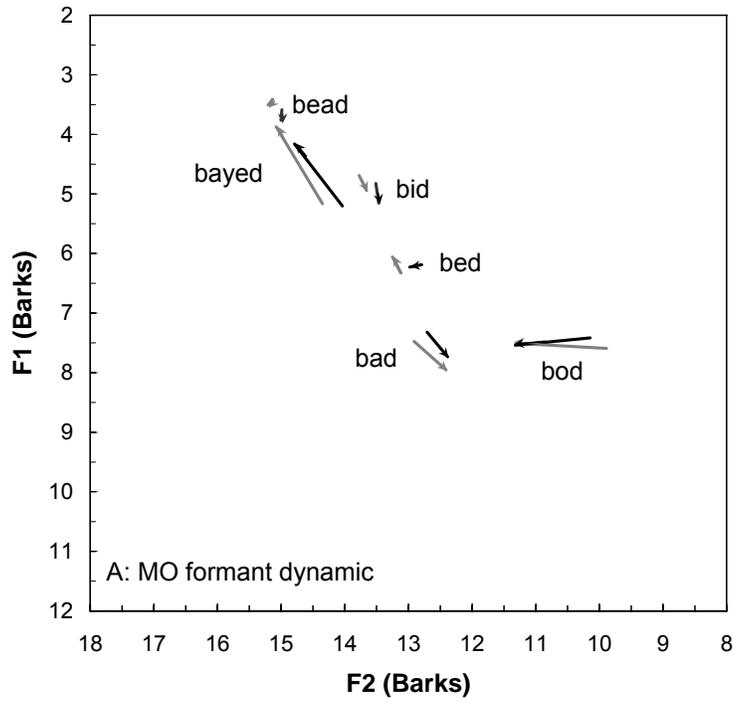


Figure 4. Average F1 and F2 frequencies (in Barks) at 20% and 80% of vowel duration for vowels in conversational (black arrows) and clear (gray arrows) speech (MO and EB talkers). The arrowhead indicates performance at 80% of vowel duration.

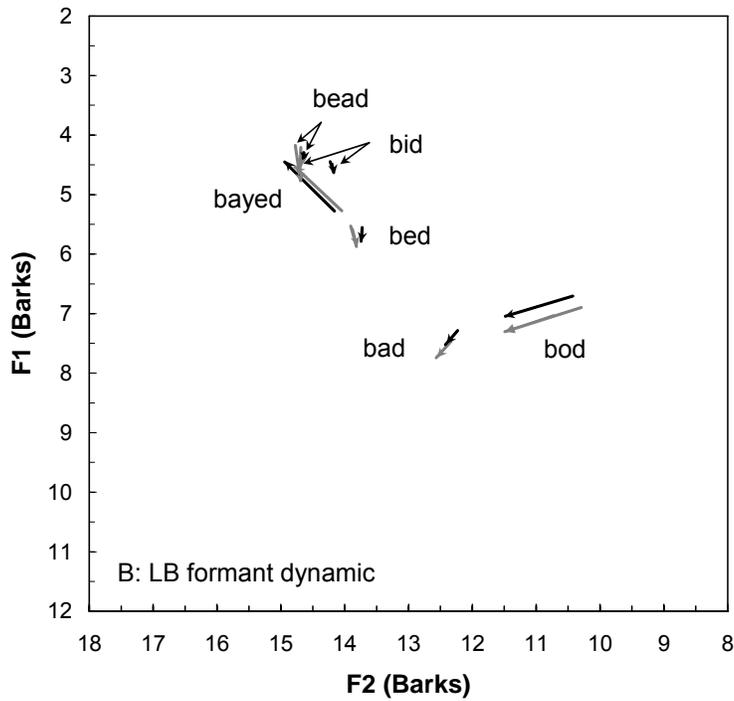
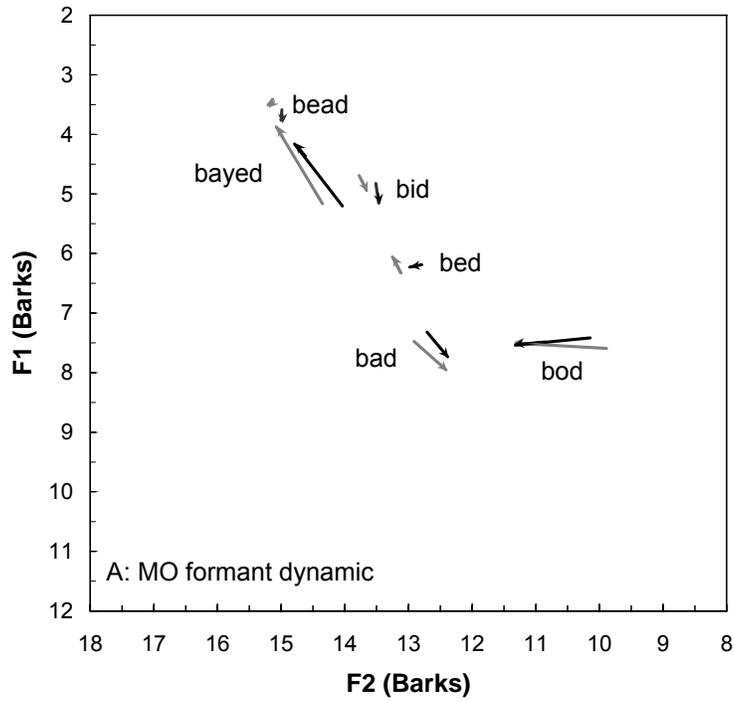


Figure 5. Average F1 and F2 frequencies (in Barks) at 20% and 80% of vowel duration for vowels in conversational (black arrows) and clear (gray arrows) speech (MO and LB talkers). The arrowhead indicates performance at 80% of vowel duration.

Figures 4 and 5 show the vectors in the F1-F2 space from 20% to 80% of the vowel duration for each target vowel in conversational (black lines) and clear (gray lines) speech. Figures 4A and 4B show the MO and EB talker's results; Figures 5A and 5B show the MO and LB talkers' results. The MO talkers' data are repeated in both figures for greater ease of comparison.

An examination of Figures 4 and 5 reveals no dramatic differences in vector length between clear and conversational speech tokens for any of the talker groups. Vector length appears slightly greater in clear than in conversational speech for /a, æ/ and /e/ for the monolingual talkers (see Figure 4A). For the EB talkers, vector length appears slightly greater in clear than in conversational speech for /e, ε / and /a/. For the LB talkers, vector length appears slightly longer in clear than in conversational speech for / æ, ε/ and /e/, but to a lesser degree than for the other two groups. Overall, the modestly greater vector lengths in clear than in conversational speech are reflected in the significant effect of speaking style on vector length.

Post-hoc comparisons of the main effect of vowel showed significant differences in vector length among all of the vowels except between /i/ and /I/ and between /I/ and /ε/. The order of vowels from greatest to smallest vector length was as follows: / a, e, æ, ε, I, i /.

## Chapter Four

### Discussion

#### *Summary of Results*

Both the MO and EB talkers were found to emphasize vowel duration differences between neighboring vowels in clear speech, as compared to conversational speech. To achieve this greater differentiation between neighboring vowels, the MO and EB talkers lengthened the “long vowels” (/e, æ, i/) in clear speech more than shorter vowels. The LB talkers, on the other hand, lengthened the vowels /e/ and /i/ less in clear speech than they lengthened the shorter vowels. Thus, the LB talkers were found to show less differentiation in duration between neighboring vowels in clear speech than in conversational speech.

At 50% of vowel duration, the relative positions and distances between vowels on the F1 axis are nearly identical for the MO and EB groups. Conversely, the maximum F1 distance between vowels appears reduced for the LB talkers as compared to the MO and EB talkers. In clear speech, the MO talkers decreased the F1 of the high vowels (/e, ε, I, i/) and increased the F1 of the lower vowels. In other words, the high vowels got higher and the low vowels got lower, so that the vowel space expanded slightly on the F1 axis in clear speech. The EB and LB talkers did not reflect an overall decrease in the F1 of all

high vowels and increase in the F1 of all low vowels. The LB talkers did, however, show a fairly sizeable lowering of F1 for low vowels in clear speech.

In clear speech, the MO talkers increased the F2 of the front vowels and decreased the F2 of the back vowel, again so that the vowel space expanded slightly on the F2 axis. The EB talkers did not increase F2 for all front vowels. This may be due to EB talkers being “more clear” to begin with, so little to no increase in F2 is seen in performance. Relative to the MO talkers, the EB talkers’ front vowels tended to be more fronted in conversational speech, so perhaps it would have been difficult for them to achieve additional fronting of these vowels. The LB talkers also increased F2 slightly in clear speech (/e/ was the exception).

Both the MO and EB talker groups appeared to increase the length of the vector in the F1-F2 space from 20% to 80% of the vowel duration in clear speech for several vowels. The LB group showed a similar pattern, but differences were smaller in extent. Vector lengths appeared to be largely comparable for the MO and EB talkers in both styles, except that the vector lengths for /æ/ appeared shorter for the EB than for the MO talkers in both styles. Vector lengths for /æ/ were appeared to be somewhat longer for the LB than for the EB talkers, but were shorter than those for the MO talkers. This cross-group difference in vector length for /æ/ was apparently not consistent or large enough to result in a statistically significant effect.

### *Comparisons to Previous Studies*

Hillenbrand et al. (1995) found that vowels showed characteristic differences in the direction and degree of change in formant frequencies measured from 20 to 80% of vowel duration. Of the vowels examined here, Hillenbrand et al. (1995) found that /æ, e/ and /a/ had the greatest degree of spectral change and /ɛ/ and /i/ had the smallest. The findings in this study showed that for monolingual talkers /e, a/ and /æ/ had the greatest degree of spectral change and /i/ and /ɛ/ had the smallest. For the EB and LB talkers, the main difference was that the vectors for /æ/ for these two groups were more comparable in length to those of /i/ and /ɛ/ (short vectors) than to those of /e/ and /a/. These between-group differences were apparently not large or consistent enough to yield statistically significant differences between the groups, however.

The steady state (50% point) frequency values appear to be in similar locations and spacing for the MO and EB talkers as for the adult female talkers in Hillenbrand et al. (1995), except that /æ/ is located lower in the vowel space in the present study than are the steady state values in Hillenbrand et al. (1995). The location of /æ/ in the present study appears to be a better match with the steady state values of Peterson & Barney's (1952) female talkers, as reproduced in Hillenbrand et al. (1995), except that /æ/ appears to be located lower in the

vowel space than /a/ for the talkers in the present study, whereas the two vowels are of approximately equal height in the Peterson & Barney (1952) data.

Ferguson & Kewley-Port (2002) examined formant frequency measures, degree of spectral change, and duration for ten target vowels produced in conversational and clear speech style by a single native speaker of American English. They found that in clear speech, F1 increased for all ten vowels. Conversely, the findings in the present study showed that F1 increased significantly for only /æ/ and /a/ (for the monolingual talkers). The findings of Ferguson & Kewley-Port (2002) were similar to those for the present study in that F2 increased in front vowels (/e, æ, ε, I, i/) and F2 decreased in the back vowel (/a/) in clear speech. In addition, in both studies the vowel space increased in clear speech for the monolingual talkers.

Ferguson & Kewley-Port (2002) found that vector length in the more crowded areas of the talker's vowel space was significantly greater in clear than in conversational speech. Of the vowels examined in the current study, /a, æ, ε/ and /e/ did show slightly greater vector lengths in clear speech than in conversational speech (with some variation across talker groups). An overall significant positive effect of speaking style on vector length was also seen in the present study.

When examining duration, Ferguson & Kewley-Port (2002) found that all vowels were significantly longer in clear speech. Similarly, the results of the

present study also showed a significant positive effect of clear speech on duration of vowels for all three talker groups.

#### *Limitations and Implications for Future Research*

A limitation in this study is that only six vowels were studied. Future research should include all monophthongal vowels. It should be noted, however, that everything measured in this study was not analyzed. Therefore, there is data that has been collected but not yet analyzed. Specifically, F0, F3, and F4 values for all target vowels, talker groups, and speaking styles are awaiting further analysis. In addition, data on spectral tilt may be gathered from this study.

Another limitation is that only ANOVAs were completed for this study. Ideally, a multivariate analysis of variance (MANOVA) should be completed. For example, the effects and interactions among the independent variables found in the present study might show different patterns when their effects on the relationships among the dependent variables are also explored.

Future research using these data should also include a correlational analysis between the acoustic variables and the intelligibility and degree of clear speech benefit shown for each talker. Individual differences across talkers in each group could be correlated with the acoustic measures from this study to determine which strategies used in clear speech result in the greatest intelligibility benefit.

#### *Conclusion*

One practical implication of this study for the speech-language pathologist (SLP) is the incorporation of these results for use in accent modification therapy

for Spanish-English bilinguals. The tendency of the LB talkers not to emphasize duration differences between neighboring vowels during clear speech suggests that they may be unaware of or unable to actively manipulate these differences. These differences might be drawn to the learner's attention during accent reduction therapy.

In addition, the location of the vowels /I/ and /i/ were located very closely to one another in the vowel space of the LB talkers. In the clear speech condition, the distinction between /I/ and /i/ for the LB group was essentially non-existent. In fact, the LB talkers tended to crowd all four of the high to mid front vowels. Training Spanish-English bilinguals to better differentiate high to mid front vowels in production could be highly beneficial in improving their intelligibility. Possible approaches to this training include the use of visual aids, indirect feedback in the form spectral displays of recorded vowels in the F1-F2 space, or direct articulatory feedback from ultrasound analysis of tongue position during vowel production.

## References

- Boersma, P., & Weenink, P. (2006). Praat: doing phonetics by computer (Version 4.4.24) [Computer program]. Retrieved June 19, 2006, from <http://www.praat.org/>.
- Bohn, O.-S., & Flege, J. (1997). Perception and production of a new vowel category by adult second language learners. In: A. James, & J. Leather (Eds.), *Second-language speech: Structure and process* (pp. 53-73). Berlin: Mouton de Gruyter.
- Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of the Acoustical Society of America*, 97, 1916-1924.
- Bradlow, A. R., & Alexander, J. A. (2007). Semantic and phonetic enhancements for speech-in-noise recognition by native and non-native listeners. *Journal of the Acoustical Society of America*, 121, 2339-2349.
- Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *Journal of the Acoustical Society of America*, 112, 272-284.
- Bradlow, A. R., Kraus, N., & Hayes, E. (2003). Speaking clearly for children with learning disabilities: Sentence perception in noise. *Journal of Speech, Language, and Hearing Research*, 46, 80-97.
- CoolEdit 2000 (version 1.1) [computer software] (2000). Syntrillium, Inc.: Phoenix, AZ.

- Dalbor, J. B. (1969). *Spanish pronunciation: Theory and practice*. New York, NY: Holt, Reinhart and Winston.
- Ferguson, S. H. (2004). Talker differences in clear and conversational speech : Vowel intelligibility for normal-hearing listeners. *Journal of the Acoustical Society of America*, 116, 2365-2373.
- Ferguson, S.H. & Kewley-Port, D. (2002). Vowel intelligibility in clear and conversational speech for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 112, 259-271.
- Flege, J. E. (1991). The interlingual identification of Spanish and English vowels: Orthographic evidence. *The Quarterly Journal of Experimental Psychology*, 43, 701-731.
- Flege, J.E. (1995). Second-language speech learning: Theory, findings, and problems. In: W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Baltimore, MD: York.
- Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 437-470.
- Flege, J. E., Schirru, C. & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, 40, 467-491.

- Helfer, K. S. (1997). Auditory and auditory-visual perception of clear and conversational speech. *Journal of Speech Language and Hearing Research, 40*, 432-443.
- Hillenbrand, J., Getty, L.A., Clark, M.J. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*, 3099-3111.
- Hillenbrand, J. & Nearey, T.M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America, 105*, 3509-3523.
- Johnson, K., Flemming, E. & Wright, R. (1993). The hyperspace effect: Phonetic targets are hyperarticulated. *Language, 69*, 505-528.
- Kewley-Port, D., Akahane-Yamada, R., & Aikawa, K. (1996). Intelligibility and acoustic correlates of Japanese accented English vowels. *ICSLP 96 (International Conference on Spoken Language Processing) Proceedings*.
- Krause, J.C. & Braida, L.D. (2002). Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility. *Journal of the Acoustical Society of America, 112*, 2165-2172.
- Ladefoged, P. (1982). *A course in phonetics*, 2nd ed. (Harcourt, Brace, Jovanovich: New York).
- MacKay, I.R.A. & Flege, J.E. (2004). Effects of the age of second language learning on the duration of first and second language sentences: The role of suppression. *Applied Psycholinguistics, 25*, 373-396.

- Monsen, R. B., & Engebretson, A. M. (1983). The accuracy of formant frequency measurements: A comparison of spectrographic analysis and linear prediction. *Journal of Speech and Hearing Research, 26*, 89-97.
- Payton, K. L., Uchanski, R. M. & Braida, L. D. (1994) Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing. *Journal of the Acoustical Society of America, 95*, 1581-1592.
- Peterson, G. & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America, 24*, 175-184.
- Picheny, M., Durlach, N. & Braida, L. (1985a). Speaking clearly for the hard of hearing I: Intelligibility differences between clear and conversational speech. *Journal of Speech and Hearing Research, 28*, 96-103.
- Picheny, M., Durlach, N., & Braida, L. (1985b). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research, 29*, 434-446.
- Piske, A., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics, 29*, 191-215.
- Rogers, C. L., Dalby, J. M., & Nishi, K. (2004). Effects of noise and proficiency level on intelligibility of Chinese-accented English. *Language and Speech, 47*, 139-154.
- Roland VS-890 manual. Appendices page 130. Strange, W., Jenkins, J. J., & Johnson, T. L. (1983). Dynamic specification of coarticulated vowels. *Journal of the Acoustical Society of America, 74*, 695-705.

- Schum, D. J. (1996). Intelligibility of clear and conversational speech of young and elderly talkers. *Journal of the Academy of Audiology*, 7, 212-218.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S.A., Nishi, K., & Jenkins, J.J. (1998). Perceptual assimilation of American English vowels by Japanese listeners. *Journal of Phonetics*, 26, 311-344.
- Traunmüller, H. (1990). Analytical expressions for the tonotopic sensory scale. *Journal of the Acoustical Society of America*, 88, 97-100.
- United States Census Bureau (2000). 2000 U.S. Census. Retrieved June 25, 2007, from <http://www.factfinder.census.gov>.