

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

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

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

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI[®]

H

**TEMPORAL CHARACTERISTICS OF SPANISH-ACCENTED ENGLISH:
ACOUSTIC MEASURES AND THEIR CORRELATION WITH
ACCENTEDNESS RATINGS**

by

AMEE SHAH

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

2002

UMI Number: 3063882

Copyright 2002 by
Shah, Ameer P.

All rights reserved.

UMI[®]

UMI Microform 3063882

Copyright 2002 by ProQuest Information and Learning Company.
All rights reserved. This microform edition is protected against
unauthorized copying under Title 17, United States Code.

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

© 2002

AMEE SHAH

All Rights Reserved

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

9/18/02
Date

Winifred Strange
Winifred Strange, Ph. D.

Chair of Examining Committee

9/18/02
Date

Robert Goldfarb
Robert Goldfarb, Ph.D.

Executive Officer

Supervisory Committee:

Lawrence Raphael, Ph.D.

The City University of New York

Loraine Obler, Ph.D.

The City University of New York

Catherine Best, Ph.D.

Wesleyan University

Haskins Laboratories

THE CITY UNIVERSITY OF NEW YORK

Abstract**TEMPORAL CHARACTERISTICS OF SPANISH-ACCENTED ENGLISH:
ACOUSTIC MEASURES AND THEIR CORRELATION WITH
ACCENTEDNESS RATINGS**

by

Ameesh Shah

Co-Chairs: Professors Winifred Strange and Lawrence Raphael

This study was designed to identify acoustic parameters of Spanish-accented English that give it the perception of being “accented.” Recordings of eight multisyllabic (3, 4, and 5 syllable) target words spoken in sentences by 22 Spanish speakers of English and five native speakers of American English (AE) were analyzed for temporal acoustic differences. Temporal characteristics of Spanish productions, including overall word duration, unstressed vowel duration, stressed-unstressed (S/U) vowel duration ratios, Voice Onset Time (VOT) and closure duration in intervocalic flaps/stops, differed systematically from native productions. Recordings of the sentences and excised target

words were presented to native (AE) listeners (N=10) who judged the degree of accentedness on a 9-point scale. Degree of accentedness of target words correlated strongly ($\rho = +0.82$) with accentedness ratings for the eight sentences (used as a measure of "global" accentedness.) Spearman rank-order correlations of overall word duration and native listener ratings of accentedness of target words varied from +0.04 to +0.56. Correlations of S/U vowel duration ratios and ratings of accentedness of target words varied from -0.01 to +0.35. VOT duration of initial voiceless /k/ (but not /p/) stops correlated positively with accentedness ratings (+0.26 to +0.36). Closure duration of intervocalic /t/ yielded positive correlations with perceived accentedness (+0.29 to +0.59). Overall, results suggest that Spanish-accented English is characterized by systematic temporal differences from native American English, and that these temporal differences contribute to the perception of accentedness as judged by native American English listeners.

Acknowledgments

I dedicate this dissertation to my ever-supportive parents who have always believed in me, and stuck with me through all of my unconventional choices. From indulging my precocious questions through childhood to analytical observations of adulthood, they have helped mold the scientific researcher in me.

I thank my advisor, Dr. Winfred Strange, for being my professional “parent”. Her constant guidance, contagious enthusiasm, and high scientific standard have set a fine example for me to emulate. Without her meticulous, painstaking direction, this dissertation would not have been possible. Thanks to Dr. James Jenkins not only for lending his research and statistical expertise, but also for his patient indulgence of scientific curiosity, and his ever present help.

Thanks to Dr. Lawrence Raphael for getting me “addicted” to speech science from my very first class in the doctoral program and all along the dissertation process—this certainly has been an educational trip through “thoracic park”! Thanks, Larry, for all your support beyond the call of duty.

Thanks goes out to Dr. Loraine Obler for being the best mentor a student could have. A ready shoulder to lean on, steady encouragement, utmost ethical standard, Loraine, you have helped me with so much I cannot even begin making a list!

Thanks to ex-Executive Officer, Dr. Irving Hochberg, for being the nurturing, caring “gel” that held the department together for so long. My pride is having known you, Irv, and my regret is not being able to share with you the joy of graduating.

I am indeed proud to be a “product” of this program, with such par-excellent role models. With skills and knowledge gained by interacting with Dr. Katherine Harris, Dr.

Harry Levitt, Dr. Arthur Boothroyd, Dr. Arlene Neuman and Dr. Glenis Long, I am confident to embark on a research career that can hopefully pay back to the scientific community.

Thanks to the Neurolinguistics Laboratory and the Speech Acoustics and Perception Laboratory for providing excellent brainstorming sessions, and a platform for scientific curiosity, creativity and camaraderie. Thanks in particular to my fellow student, Kanae Nishi, for ready help with tools and techniques of data management and statistics. Thanks also to my fellow laboratory-member, Paul Argodale, for help with recruiting subjects.

An acknowledgment list would not be complete without personal thanks to my dear friends. Thanks, Jackie, for always assuring me through last-minute panic attacks with a confident, “You’ll make it—you always do!” Thanks to my dear sister, Arpi, and supportive friends, Zahara, Julie, and Irina for being the supportive sounding boards I needed through the angst and tribulations of the trying dissertation-days. Thanks to Carmen, Assistant Registrar at CUNY, for being an invaluable guiding light to help me navigate through the murky waters of academic bureaucracy. Thanks also to my roommate, Dorothy Ross, for late-night philosophical discussions, proofreading/editing of drafts, and most of all, for believing in, and encouraging, my professional dreams.

I am thankful for the various sources of funding that supported my doctoral study and research. Academic tuition was funded by University Scholarship, Speech and Hearing University Fellowship, and PSC-CUNY Tuition Award. This dissertation was supported by the “Doctoral Dissertation Research Grant” (BCS 0001862) from the

National Science Foundation to Ameer Shah. Support was also received from NIDCD grant (DC 00323) to Winifred Strange.

TABLE OF CONTENTS

ABSTRACT.....	iv
ACKNOWLEDGMENTS.....	vi
TABLE OF CONTENTS.....	ix
LIST OF TABLES.....	xii
LIST OF FIGURES.....	xiii

Chapters

I. STATEMENT OF THE PROBLEM.....	1
II. REVIEW OF THE LITERATURE.....	3
Methods of Studying Accented English.....	3
Contrastive Phonological Analysis.....	3
Acoustic Analysis.....	4
Imitation of Foreign Accent.....	5
Perceptual Analysis.....	7
Spanish-Accented English.....	10
Phonological Characteristics.....	10
Perceptual Characteristics.....	17
Acoustic Characteristics.....	19
Prosodic Cues.....	19
Voicing Cues.....	20
Other Temporal Characteristics of Accented English.....	22
Temporal Differences between Native and Nonnative Productions.....	22
Temporal Differences serve as Perceptual Cues in English.....	26
Temporal Deviations in Nonnative Speakers' Utterances are Relevant Perception by Native Listeners.....	27
Research Questions and Hypotheses.....	28
III. STUDY 1: RATING OF ACCENTEDNESS (PERCEPTION STUDY)	
Methodology	
Subjects.....	31
Nonnative Speakers.....	31
Native Speakers.....	32
Listeners.....	32
Stimuli.....	33
Stimuli for Global Accent-Judgment (Sentence-Rating) Task.....	33
Stimuli for Word-Rating Task.....	38

Procedure for Rating Tasks	38
Instructions for Rating Tasks.....	40
Results	41
Reliability of Accentedness Judgments.....	41
Intrajudge Reliability.....	41
Interjudge Reliability.....	41
Rationale for Pooling Listeners.....	43
Correlation of Accentedness on Words vs. Sentences.....	48
Discussion.....	51

IV. STUDY 2: ACOUSTIC ANALYSIS (PRODUCTION STUDY) AND CORRELATION OF ACOUSTIC PARAMETERS WITH ACCENTEDNESS RATINGS

Segmentation Methodology.....	54
Segmentation Problems.....	56
Results.....	58
Comparison of Acoustic Data of American English vs. Nonnative Speakers and Correlation with Accentedness Ratings.....	59
Overall word duration.....	59
Vocalic segmental duration.....	62
Stressed/Unstressed (S/U) vowel ratio.....	64
Voice Onset Time.....	68
Closure duration for Flaps/ Medial Stops..	70
Summary and Discussion.....	73

V. GENERAL DISCUSSION.....76

VI APPENDICES

Appendix A: List of Original 98 Target Words in Sentences.....	83
Appendix B: Language Background Questionnaire.....	88
Appendix C: Tables of Listeners' Medians and Ranges.....	89
Appendix D: Demographic Information of Subjects....	94
Appendix E: Scatterplots of correlations of accentedness ratings on sentences to accentedness ratings on words..	95
Appendix F: Scatterplots of correlations of overall word duration to accentedness ratings (only NN).....	96
Appendix G: Scatterplots of correlations of S/U ratio to accentedness ratings (only NN).....	97
Appendix H: Scatterplots of correlations of VOT to accentedness ratings (only NN).....	98
Appendix I: Scatterplots of correlations of Flap/closure durations to accentedness ratings (only NN).....	99

VII. BIBLIOGRAPHY100

LIST OF TABLES

Table

2.1	Production errors of Spanish speakers of English (MacDonald, 1989).....	12
2.2	Phonological processes with potential to affect intelligibility and their frequency of occurrence in native Spanish speakers' English (Ortega-Llebaria, 1997).....	14
3.1	List of target words within the revised sentence contexts.....	34
3.2	List of the 8 words, transcription of acoustic segments, and morphemic and stress status of target words.....	36
3.3	Interjudge reliability: median and range of accentedness ratings collapsed across all speakers (AE + NN) and all 8 words.....	42
3.4	Spearman rank order correlations for each word/sentence pooled over Listeners.....	50
4.1	List of 8 words and individual acoustic segments within the words.....	57
4.2	Differences in overall word duration (in ms) for native (AE) and nonnative (NN) groups.....	60
4.3	Vocalic segmental durations (%) of total word duration for native (AE) and nonnative (NN) groups.....	63
4.4	Vowels used in estimating stressed to unstressed (S/U) vowel ratio.....	66
4.5	Stressed to unstressed vowel ratios (S/U) for native (AE) and Nonnative (NN) groups.....	66
4.6	Coefficients and significance levels for correlation of S/U ratio and accentedness ratings.....	67
4.7	Word-initial stop VOT (ms) for native (AE) and nonnative (NN) groups.....	69
4.8	Coefficients and significance levels of correlation of word-initial stop VOT and accentedness ratings.....	70
4.9	Closure duration of intervocalic /t/ (ms) for native (AE) and nonnative (NN) groups.....	71
4.10	Coefficients and significance levels of correlation of flap/ closure duration and accentedness ratings.....	72
4.11	Summary table of all correlations of acoustic parameters and accentedness Ratings.....	74

LIST OF FIGURES

Figure

3.1a	Frequency distribution of listeners' accentedness ratings: AE sentence condition.....	44
3.1b	Frequency distribution of listeners' accentedness ratings: AE word condition.....	45
3.1c	Frequency distribution of listeners' accentedness ratings: NN sentence condition.....	46
3.1d	Frequency distribution of listeners' accentedness ratings: NN word condition.....	47
3.2	Correlation of overall median accentedness ratings of words and sentences (AE +NN).....	49
3.3	Overall median/median of NN speakers : sentence/word correlations.....	49

CHAPTER I

Statement of the Problem

The purpose of this dissertation was to address the distinct speech characteristics of Spanish-accented speakers of English that may be related to native listeners' perception of accentedness. Previous research (MacDonald, 1989; Ortega-Llebaria, 1997) involved contrastive phonological analyses of speech characteristics as they related to perceived accentedness. Only a few studies have employed acoustic measures (e.g. Magen, 1998; Backman, 1978; Flege and Eefting, 1987; Flege, Munro and Skelton, 1992), and those that did studied only a few parameters such as differences in nonnative voice onset time, vowel formant structure, vowel duration and rate of speech.

Previous research suggests that nonnative temporal patterns account, at least in part, for native listeners' perception of Spanish L2 English learners' speech as "accented." The previous phonological and acoustic research led to the hypotheses motivating the current project that a) Spanish-accented speakers' segmental durations will differ from those of native English speakers in multisyllabic words, and b) these duration differences will be related to perceived accentedness.

Two studies were designed to test these hypotheses. In the perception study, native English-speaking listeners rated nonnative productions of short sentences as a measure of global accentedness, and isolated multisyllabic words excised from the sentences as a predictor of global accentedness. The rating task involved making judgments of accentedness on a nine-point rating scale from "least-foreign accented" to "most-foreign accented." In the production study, vocalic segmental duration, initial voiceless stop VOT, and flap duration/closure duration for intervocalic /t/ were measured. Overall word duration was also determined and a measure of lexical stress (the ratio of

stressed to unstressed vowels within a word) was derived. These production measures were compared for the native and nonnative speakers. Based on phonological descriptions of Spanish-accented English and Spanish phonological rules, predictions were made for differences in durations between the native and nonnative speaker groups for each of the acoustic segments measured. The nonnative group was expected to produce longer overall word durations due to an expected slower rate of speech. VOT values for voiceless stops were expected to be smaller in the nonnative group since voiceless stops in Spanish are unaspirated. Flap/closure duration of intervocalic /t/ was expected to be greater in the nonnative group, approximating durations for stops. Vocalic segmental durations were expected to be approximately equal for stressed and unstressed syllables (S/U ratio=1.0) in the nonnative productions, as would be expected if the syllable-timed pattern from Spanish phonology were used in English pronunciation. This would lead to perception of errors in lexical stress patterns.

The non-native speakers' temporal deviations from native production patterns were then correlated with accentedness ratings of the multisyllabic words in order to test the hypothesis that nonnative differences in segmental duration are related to perceived accentedness of Spanish speakers' productions of English.

CHAPTER II

Review of the Literature

Foreign-accented speech is defined as “non-pathological speech produced by second language learners that differs in partially systematic ways from the speech characteristics of native speakers of a given dialect” (Munro, 1998). It is a phenomenon that is ultimately dependent on the perception of native listeners. A substantial amount of research has shown that non-native speakers of a language may not be able to produce utterances in a native-like manner, and that their speech is clearly “foreign,” that is, different from the speech of native speakers, whatever regional or social dialect the native speakers may represent. These pronunciation differences can be of various types, from difficulty in producing individual phonetic segments to prosodic differences in intonation, lexical stress or sentence focus.

Methods of Studying Accented English

Contrastive Phonological Analysis

Various methodologies have been used to address the issue of what contributes to the perception of foreign accent. A popular method has been contrastive phonological analysis, that is, description using impressionistic transcription of phonological patterns in accented speech. This approach explores the role of, and transfer from, native phonology realization rules to the production of non-native phonetic segments and sequences. The method involves comparing the entire phonological inventory of the native/first language (L1) to the nonnative/second language (L2) inventory. Aspects of the L1 phonological system that transfer to the acquisition and maintenance of the L2 system are identified and inferences are made about their probable contribution to the perceived foreign accent. Some examples of L1 phonological transfer leading to accented

pronunciation are: phoneme substitutions, e.g. /t/ for /θ/, as for speakers of French learning English; /a/ for /æ/ as in /apəl/ for “apple”, and vowel epenthesis, e.g. /əskul/ for “school” as for speakers of Spanish learning English. Stockwell and Bowen (1965) and Dalbor (1980), for example, compared the Spanish phonological system to that of English. The method of contrastive phonological analysis has also been used across dialects within a single language (Canfield, 1962; Resnick, 1975; and del Rosario, 1970).

Acoustic Analysis

Another method of studying foreign accent has been acoustic analysis of accented productions as compared to native speakers’ productions. This approach uses measurement of segmental and suprasegmental features. Segmental features measured include durations and formant frequency patterns of consonants and vowels. Suprasegmental measurements include patterns of fundamental frequency contours as related to perceived intonation, phrase and sentence duration as related to the perceived rate of utterance, duration, intensity and fundamental frequency changes as related to perceived lexical and sentence stress. Several studies have employed the acoustic approach. Flege and Port (1981) compared the Voice Onset Times (VOT) of English stop consonants produced by monolingual English and bilingual Arabic-English speakers. Flege and Eefting (1987) compared VOT differences of English stop consonants produced by monolingual English speakers and Spanish speakers learning English. Munro (1993) compared vowel durations and formant frequencies of ten English vowels in /bVt/ and /bVd/ contexts produced by monolingual English speakers and Arabic speakers learning English.

Magen (1998) went beyond acoustical analysis by adding an experimental manipulation of accented speech materials to investigate which kinds of deviations contributed to perceived accentedness. In her study, native Spanish speakers with English as L2 and native American English (AE) speakers were recorded as they read 96 English sentences each containing a target phrase containing sounds or sequences of sounds whose production had been found to be characteristically difficult for them. The phrases of the nonnative speakers were then acoustically edited to match the productions of the native speaker, e.g. by deleting epenthetic schwa, by reducing vowel duration and by changing non-English F_0 contour characteristics. The edited and unedited phrases were played to native listeners who rated them for foreign accent on a scale of 1 to 7, which ranged from “closer to native English” to “less close to native English” respectively. The results indicated that the edited forms were perceived to be less accented than the unedited forms. (This study is described later in detail.)

Imitation of Foreign Accent

A third approach to the study of accented speech is the use of simulated foreign-accented speech. In a study that sought to reveal the phonetic characteristics that are associated with perception of foreign accentedness, Cunningham-Anderson and Engstrand (1989) used simulated productions of Finnish and British accents of Swedish as well as features of regional dialects of Swedish. This approach involved training a phonetician to produce characteristics of Finnish- and British-accented Swedish. Some examples of this approach are, the replacement of the Swedish “grave word accent” by the “acute word accent” that occurs in Finnish and British forms, the replacement of the Swedish apico-post-alveolar consonants by the corresponding dental consonants as in the

Finnish forms, and velarizing of /l/ as in Finnish forms. The researchers then recorded a large number of readings of a short passage in Swedish, incorporating or “artificially introducing” the deviant pronunciation features in various combinations as a simulation of Finnish-accented Swedish or British-accented Swedish. These readings were played to 35 monolingual listeners in a task of judgment of foreign accent. The listeners were required to indicate if they perceived each of the readings to be either foreign-accented, or a possible regional accent of Swedish, or if they merely sounded strange. They were also required to rate the degree of deviation on a scale of 0=no deviation to 4=maximum deviation. Results indicated that the listeners were able to reliably separate the three categories, namely, regional Swedish accent, foreign accent, or peculiar sounding. Results also established that it is possible to imitate phonetic characteristics so that the listeners perceive them as foreign or regionally accented. Moreover, the greater the number of deviant features introduced, the more a reading was perceived to be foreign-sounding. Qualitatively, some deviant features were found to be perceived as more accented than others.

Similarly, Wingstedt and Schulman (1987) used a form of simulated accented stimuli. They attempted to discover how listeners comprehend accented speech. Specifically, they asked what strategies listeners employed to improve their ability to interpret a foreign accent after increased exposure to that accent. They created a “cryptic accent” using a set of deviant phonological rules previously found in the errors of nonnative speakers of Swedish. These phonological rules were implemented in various combinations in the creation of a new text that was transcribed and recorded by one of the authors. The recordings of sentences as well as excised words were played to native

Swedish listeners who were required to repeat these sentences as they heard them. A control group, which listened only to the word list, was also included to counter the claim that listeners would be able to understand the words without being exposed to the sentences. Results showed that the Experimental group (the group receiving exposure to sentences) interpreted the cryptic words better than the Control Group. Different combinations of the cryptic rules resulted in different degrees of difficulty in comprehension. Additionally, sentences where only the segmental rules were applied were the easiest to comprehend. The addition of vowel insertion errors or lexical stress misplacement further increased difficulty for the listeners. Interestingly, the situation was reversed in comprehension of isolated words, where segmental errors were found to be most difficult for comprehension. Overall, the authors concluded that prior exposure to a particular foreign accent, even if it was a heavy accent, resulted in improved comprehension of the “cryptic accent” created by the experimenters.

Perceptual Analysis

Yet another methodology employed to identify characteristics of foreign-accented speech has been the direct perceptual approach that uses native-listener judgments of the degree of perceived global accentedness. Native listeners use rating-scales of one or more dimensions and/or accent questionnaires for judgment of accent or intelligibility. These scales have been either continuous (Flege, 1988; Munro, 1993) or equal-appearing interval fixed-point scales, e.g., 5-, 7-, or 9-points (Calla McDermott, 1986; Southwood and Flege, 1999). Direct Magnitude Estimation (DME) has also been used occasionally to scale foreign accent (Brenan, Ryan & Dawson, 1975). A comparison of continuous, interval versus DME scale (Southwood and Flege, 1999) has shown that a 9 or 11-point

scale reliably measures the degree of perceived foreign accent for speakers with English as L2.

Research using rating tasks has shown that native listeners are able to accurately discriminate foreign accent from regional dialectal variations within the native language. Foreign-accented speech segments can be identified from stimuli ranging from 30 ms-excerpts of word-initial consonants to isolated syllables and larger segments such as phrases (Flege, 1984). Further, the identity of the native languages of non-native speakers can be determined based on their accent (Van Els and De Bot, 1987; Anderson-Hsieh, Johnson, Koehler, 1992), and degrees of foreign accent can be differentiated (Munro, 1993).

Listener perception of degree of accentedness has been found to be related to several characteristics of the nonnative speakers' language history, namely age of L2 learning (AOL), and period of exposure to the target language. AOL, usually operationally defined as the age of arrival (AOA) in the L2 country, is correlated negatively with the presence and degree of the perceived accent (Asher and Garcia, 1969; Fathman, 1975; Seliger, Krashen, and Ladefoged, 1975; Oyama, 1976; Suter, 1976; Tahta et al., 1981; Thompson, 1991; Flege, 1988; Patkowski, 1990; Flege and Fletcher, 1992; Flege, Takagi and Mann, 1995; Flege, Yeni-Komshain and Liu, 1999). At AOLs above a cut-off threshold, called the "critical period of language learning," nonnative speakers are usually judged by native listeners as having a foreign accent. This critical period has been reported to range from as low as 3 years (Flege et al., 1995), to as high as 12 years (Scovel, 1988) or 15 years (Patkowski, 1990). It can be concluded that speakers

with AOL over 15 years will be most likely to speak English with a perceptible L1 accent (Munro, 1993).

The duration of exposure to L2, usually defined as the length of residence (LOR) in the L2 country, has also been found to correlate with the presence and degree of perceived foreign accent (Asher and Garcia, 1969; Flege and Fletcher, 1992; Flege et. al., 1995). Not every study has shown a significant effect of LOR on degree of L2 foreign accent. In those studies where an LOR effect was found, LOR was a less important predictor of degree of L2 accent than AOL. Flege and Fletcher (1992) reported that adult learners showed an LOR effect when diverse groups (LOR ranging from less than a year to several years) are compared. For highly experienced subjects, additional years of experience in the L2 appear to be unlikely to lead to a significant decrease in degree of perceived foreign accent. Conversely, in initial stages of learning, additional experience in the L2 may well lead to less foreign-accented L2 speech. Flege (1988b) summed up the LOR findings with the conclusion that after an initial period of rapid learning, LOR does not significantly affect the degree of foreign accent for adult learners of a second language.

Several studies have used the approach of making measurements of several acoustic characteristics typical of accented speech and then relating them to the perceptual ratings by native listeners (Magen, 1998; Munro, Flege and Mackay, 1996; Munro, 1993). This approach helps to identify those production characteristics that are potentially most significant in the perception of accent. (These studies are reviewed in more detail below.)

The majority of studies on accented speech has been conducted using subjects speaking English as a non-native language whose L1s include Arabic (Flege and Port 1981), French (Hazan and Boulakia 1993; Laeufer 1996; Neufeld 1979), Dutch (Elsendoorn 1985; Snow and Hoefnagel-Hohle 1977), German (Olson and Samuels 1973), Italian (Oyama 1976; Flege, Munro and Mackay 1995), Mandarin/Taiwanese (Flege 1988b), Korean (Flege, Yeni-Komshain and Liu 1999), Russian (Thompson 1991), and Polish (Neufeld 1980). Among the various immigrant languages spoken in the United States, a widely spoken language is Spanish. Recent US Census data [US Census, 2000]¹ show that the Hispanic population represents the largest number of individuals in the United States whose native language is not English. From these data we can infer that Spanish is the most common form of accented English in the United States.

Speakers of Spanish arrive in the United States primarily from regions of Central and South America, and thus they speak different dialects of New World Spanish and, in addition, they speak English with different regional accents depending on their location within the United States. Because of the high influx of Spanish-speaking people to the United States and their regional and occupational spread, it is of practical interest to study Spanish-accented English to determine what speech characteristics make Spanish speakers sound “foreign accented.”

Spanish-Accented English

Phonological Characteristics

Among the previously mentioned methods, contrastive phonological analysis has been widely used to describe characteristics of Spanish-accented English. MacDonald

¹ Lollock, Lisa (2000), *The Foreign Born Population in the United States: March 2000, Current Population Reports, P20-534*, US Census Bureau, Washington, D.C.

(1989) published a compilation of studies of Spanish phonology as well as Spanish-accented English phonology. She attempted to explain the phonological deviation of Spanish-accented English using this contrastive method. Her description of Spanish-accented English was based on groups of studies in three dialects of Spanish: Mexican-American, Puerto-Rican-American, and Cuban-American. Overlapping features of phonological variations found across all studies were used to draw conclusions about characteristics of Spanish-accented English. The frequencies of occurrence of common features across studies were taken into account to arrive at Table 2.1, which lists phonological characteristics of Spanish-accented English. Additional characteristics were also found, but those listed in the table occurred most frequently.

MacDonald (1989) did not report the nonnative speakers' hierarchy of difficulty in producing different phonemes that could have been derived from a proportional analysis of phonetic feature errors. Knowledge of this hierarchy would be useful to predict native listeners' perception of presence or degree of accentedness. A more recent study by Ortega-Llebaria (1997) reported a hierarchy of frequency of occurrence of a set of phonological characteristics for Spanish-accented speech. In this study, thirty Spanish speakers from all American dialect areas defined by Dalbor (1980) and peninsular dialects were administered the Templin Darley Test of Articulation (TDTA), a labeling test that examines all English phonemes in simple and complex onsets and codas. Fifteen native listeners made phonemic transcriptions of the Spanish productions of the TDTA words as a measure of intelligibility. A database was created of Interlanguage (IL) transcriptions across subjects computed in frequency of occurrence (in percent) of the IL transcriptions for each target sound produced by all 30 speakers.

Table 2.1 Production Errors of Spanish Speakers of English (MacDonald, 1989).

Target Phoneme	Substituted phoneme	Description
/ð/	[d]	The fricative is replaced by a noncontinuant counterpart i.e. interdental fricative → alveolar stop (voicing preserved)
/θ/	[t]~[θ]	Most frequently in word-initial position (Wolfram 1973) i.e. interdental fricative → alveolar stop (for voiceless stops)
/t/	[s]~[f]	Mostly in word-final position (Wolfram 1973) Stop → fricative
/tʃ/	[ʃ]	/ʃ/ substituted for /tʃ/ more than vice-versa. Varying degree of closure present; affricate → fricative
/v/	[b]~[β]	Labiodental fricative realized as bilabial stop or fricative
/z/ and other word-final obstruents	/s/	Devoicing of voiced fricative in word-final obstruents
Word final clusters		Deleted or simplified
/dʒ/	/j /	Neutralized to varying degrees of closure affricate → glide
/ʌ/	[a]~[ɑ]~[ɔ]	/ɔ/ most often substituted for [ʌ]
/æ/	[a]~[ɛ]	Vowel substitution
/iy/ or diphthongized / iə /	[i]	Vowel substitution
Offglides		Released as tenser, higher, shorter vowels

The non-target productions were examined to find Spanish L2 English traits that affected intelligibility. These speech-substitution patterns were called phonological processes or “tendencies,” and were originally 30 in number. Nine of these 30 tendencies obtained a frequency-of-occurrence score of less than 1%, and were therefore excluded from the analysis, leaving 21 tendencies that affected intelligibility. These 21 phonological processes are shown, with examples and corresponding frequency of occurrences, in Table 2.2. On the basis of these, Ortega-Llebaria inferred a hierarchy of production difficulties and their influence on intelligibility for native listeners.

Currently, the question of the relationship between these production difficulties and their influence on listeners’ perception of foreign-accentedness remains unanswered. Studies of intelligibility cannot be used to infer perceived accentedness, since the factors that affect intelligibility are not identical to those that affect perception of accentedness. Intelligibility and accentedness are related but partially independent. Although heavily accented speech tends to be less intelligible than unaccented speech, it is not always so. Munro and Derwing (1995a) found that Mandarin speakers’ read productions of English sentences that were rated as moderately or heavily accented were often perfectly intelligible. Parallel findings have been reported in studies of extemporaneous utterances from other L1 groups (Derwing and Munro, 1995) and of identification and ratings of L2 speakers’ vowel productions (Munro, Flege and Mackay in press) and consonant productions (Flege, Takagi and Mann, 1995).

In a comprehensive study, Munro and Derwing (1995b) tested the relation between perceived accentedness and intelligibility, and also the relation between perceived accentedness and production errors in extemporaneous speech by native

Table 2.2 Phonological Processes with Potential to Affect Intelligibility and their Frequency of Occurrence (%) in Native Spanish Speakers' English (Ortega-Llebaria, 1997)		
Phonological Process	Example	(%)
Target lax vowels becomes tense	[lɪft] → [lift] "lift"	44.95
Target /ɑ/ is produced as /ɔ/ and /ʌ/	[klɔk] or [klʌk] "clock"	27.82
Final position: Target voiced obstruents become voiceless	[tʌb] → [tʌp] "tub"	24.82
Final position: Deletion of a consonant of a target cluster	[lakt] → [lak] / [lat] "locked"	17.73
Target /ʌ/ is produced as /ɑ/ and /ɔ/	[trʌk] → [trɑk]/[trak] "truck"	15.33
Target /æ/ is produced as /ɑ/	[kæt] → [kɑt] "cat"	12.07
Addition of an epenthetic vowel in the word-initial position	[snou] → [əsnou] "snow"	8.88
Target /z/ becomes voiceless /s/	[zɪpə] → [sɪpə] "zipper"	6.43
Target medial voiced stops are perceived as voiced fricatives, and the affricate as a glide	[rʌbə] → [rʌvə] "rubber"	5.70
Target voiced stops deleted in final position	[tʌb] → [tʌ] "tub"	3.07
Target nasals neutralize to the point of articulation	[drʌm] → [drʌn] "drum"	2.96
Target initial voiceless stops are perceived as voiced	[p ^h aɪ] → [baɪ] "pie"	2.74
Final position: Target voiceless fricatives become voiced	[skɑrf] → [skɑrv] "scarf"	2.56
Target /tʃ/ becomes /ʃ/	[skrætʃ] → [skræʃ] "scratch"	2.08
Target /dʒ/ becomes continuant	[keɪdʒ] → [keɪʒ] "cage"	2.04
Target /ʃ/ becomes /tʃ/	[ʃɑrp] → [tʃɑrp] "sharp"	1.97
Target /y/ devoices	[gəɹɑʒ] → [gəɹɑʃ] "garage"	1.86
Target glides become noncontinuant segment	[jeləʊ] → [dʒeləʊ] "yellow"	1.68
Target /θ/ becomes /t/ or /f/	[θʌm] → [tʌm] "thumb"	1.57
Target /ð/ becomes /d/	[ðer] → [der] "there"	1.50
Target /v/ becomes /b/	[væləntaɪm] → [bæləntaɪm] "valentine"	1.35

Mandarin speakers of English as an L2². Intelligibility referred to the actual process of decoding the message, in that native English listeners were required to transcribe each utterance presented. Ten native Mandarin speakers and two native speakers of English were recorded as they produced an extemporaneous speech sample describing a story card. These samples were presented to 18 native English-speaking listeners who orthographically transcribed each utterance (intelligibility), and also rated degree of foreign accent on a scale of 1 to 9, where 1=no foreign accent, and 9=very strong foreign accent. Analyses involved a tally of the total number of phonemic errors (deletions, insertions or substitutions of a segment), phonetic errors (“involved the production of a segment in such a way that the intended category could be recognized but the segment sounded noticeably nonnative,” p. 80), and morphosyntactic errors and grammatical errors (including inappropriate use of preposition, errors of subject-verb agreement, errors of verb tense and verb form, as well as the use of inappropriate articles, incorrect number, omission of subject, object and relative pronouns). Intonation was also rated on a scale from 1=native-like to 9=not at all native-like. A second analysis involved coding the listener-transcribed responses. Means of the accentedness ratings were calculated. The listeners’ orthographic transcriptions were compared to those of the authors and intelligibility scores were assigned to each segment. Results showed that accentedness ratings were higher, i.e more accented, for nonnative speakers (mean ratings=2-8) compared to native (mean ratings=1-2). In spite of the relatively higher accented ratings, the nonnative speech samples were found to be highly intelligible. Phonemic, phonetic,

²Not relevant to this discussion of the relationship between intelligibility and accentedness, this study also addressed a measure of comprehensibility of utterances where the listeners assigned a perceived comprehensibility rating on a scale of 1 to 9, where 1=extremely easy to understand and 9=impossible to understand. Nonnative speakers were found to be less comprehensible than the native speakers.

intonation and grammar errors were found to be highly correlated with perceived accentedness, but less correlated with intelligibility. While all the nonnative stimuli were highly intelligible, there was a wide range of scores for accentedness, with a fairly even distribution of accent scores from 2 to 8 categories on the 9-category scale. Only 5 listeners (28%) showed a significant correlation between the accent scores and intelligibility scores, with values ranging from -0.37 to -0.48. Listeners rated utterances as moderately or heavily accented even when they were able to transcribe them perfectly. Therefore, we can conclude that the presence of a strong foreign accent does not necessarily result in reduced intelligibility.

Most studies of accented English have been concerned with its effects on intelligibility. Clinically, ESL training and accent reduction programs are typically geared towards addressing only those speech characteristics that affect intelligibility by native listeners. However, as shown above, L2 speech can sound foreign even when it is highly intelligible. Accented speech, even when free of errors affecting intelligibility, may still be stigmatized. Therefore, it may pose difficulties in L2 speakers' personal and professional interaction with native speakers. This may have implications for professional growth (e.g. difficulty in dealing with clients in the corporate world) and personal wellbeing (e.g. the negative response of native listeners to the accented speech may cause the nonnative speakers to feel alienated). Therefore, issues of accentedness should be addressed using a measure other than intelligibility scores. Global accent rating by native listeners is a more valid measure of accentedness according to several of the studies mentioned above.

In a study correlating production deviations of nonnative speakers and accentedness ratings by native speakers, the predictor variables must, of course, be quantifiable characteristics. Phonological deviations derived from impressionistic transcriptions do not lend themselves to quantitative analyses. Magen (1998) used temporal and spectral acoustic correlates of phonological characteristics because they allowed for quantifiable measures of production deviation. The acoustic measures also allowed for direct experimentation, to show that modifying the acoustic signal affected listeners' ratings of accentedness.

Perceptual Characteristics

The only study that directly addresses the causal relationship between production differences and listeners' perception of Spanish-accentedness is a recent study by Magen (1998). Magen edited Spanish-accented utterances to make them acoustically more similar to native English utterances and tested the influence of these modifications on listeners' perception. Two native Spanish speakers and two native American-English speakers recorded 96 sentences each containing a target phrase designed to contain sounds or sequences of sounds whose production had been found to be characteristically difficult (based on previous reports by Stockwell and Bowen, 1965, and by MacDonald, 1989) for Spanish speakers learning English. Each target phrase consisted of 5-11 syllables and presented the potential for evoking characteristic differences from American English, such as differences of syllable structure (including initial epenthetic schwa, and past-tense “-ed endings”), vowel quality (including vowel substitution and tense-lax errors), consonant errors (including final /s/ deletion, /tʃ-/ʃ/ confusions, /s-z/ confusions, and stop voicing) and lexical and phrasal stress. Digital waveforms were

edited (using the Bliss Waveform Editor) to make the resulting phrases temporally match those of the native American English speakers. Examples of modifications in each category are as follows. Modification of syllable structure included deleting epenthetic schwa and deleting final schwa in –ed endings which were inappropriate. Modification of vowel quality referred to reducing vowel duration or lengthening surrounding vowel duration and deleting/lengthening /i/ or /u/-like portions associated with tense-lax aspects of vowel quality. Modification of consonantal errors consisted of changes in amplitude associated with final /s/, deleting the burst or increasing the rise-time of the burst for /tʃ-/ʃ/ substitutions, or eliminating coarticulation in the case of replacement of /tʃ/ with /ʃ/, and addition of aspiration by amplitude adjustment for stop voicing in the case of /p/-/p^h). Suprasegmental features of lexical and phrasal stress were edited by resynthesizing the F₀ contour.

Ten native speakers of American-English rated the non-native utterances in their unedited and edited forms. The listeners assigned ratings using a seven-point scale ranging from “closer to native English” on the left to “less close to native English” on the right. The results indicated that the edited forms were perceived to be less accented than the unedited forms. Listeners’ judgments of foreign accentedness were affected significantly by modifications in epenthesis, final /s/ deletion, and consonant manner substitution. By contrast, vowel reduction, tenseness-laxness, and voicing errors did not significantly affect listeners’ judgements of foreign accentedness.

Although the Magen (1998) study is an important contribution of the relationship between acoustic deviations and listeners’ ratings of accentedness, it has several limitations that need to be addressed in a new study. Magen (1998) tested only a few

phonological/acoustic deviations and did not account for all deviations of Spanish-accented English. Thus far, a database of acoustic measures of the production deviations of Spanish-accented speakers, such as the database of phonological deviations by Otega-Llebaria (1997), is unavailable. Also, Magen (1998) attempted to correlate production and perception *after* acoustic editing. No attempt was made to correlate the production measures with the perception *before* doing any editing, as a means of exploring the baseline production-perception relationship. Therefore, there remains the need for a study of the relationship between production deviations and perceived accentedness.

Other studies investigating the acoustic characteristics of Spanish-accented English have examined a particular feature, such as voicing, to test its relevance to the perception of accentedness. Most of these studies have included only acoustic measurements without establishing listener ratings of accentedness. The following studies report the acoustic characteristics of Spanish-accented English.

Acoustic Characteristics of Spanish-Accented English

Prosodic Cues. Intonation has been studied as an acoustic characteristic of Spanish-accented English. Backman (1978) compared the intonation contours of Venezuelan-accented English speakers with those of native AE speakers. Trained linguists judged 368 interrogative (wh-questions and yes-no questions) and declarative sentences for appropriateness of intonation. Stimuli were sentences extracted from short dialogues read aloud by the subjects. The judges made appropriateness ratings for the intonation and identified categories of errors for pitch-related problems. Perceptual results for the nonnative speakers showed that for both interrogative sentences and declarative sentences, the overall pitch range was judged to be too narrow and pitch

prominence over the terminal contour was small or shifted earlier in the sentence. As part of the acoustic analyses, 550 wide band spectrograms with a scale magnification of the fundamental frequency were analyzed in order to compare quantitatively fundamental frequency (F_0) patterns of the native and nonnative speakers. Acoustic findings corresponded with the perceptual results in that the F_0 range for all sentences by the nonnative subjects was narrow, and also, the final rise in F_0 (terminal contour) was smaller and placed earlier in the sentence. Additionally, for declarative sentences, the unstressed words and syllables were found to be high in F_0 values, as well as perceived to be high in pitch relative to native utterances. The nonnative subjects' utterances showed a lack of pre-final word rise in F_0 , thus making the final fall not as distinct. This resulted in the perception of insufficient final pitch fall.

Voicing Cues. Voicing differences have been reported to contribute to production differences in Spanish-accented English. Flege and Eefting (1987) compared word-initial VOT in the productions of voiced and voiceless stops in English and Spanish words produced by Spanish-speaking adults and children learning English. Both the Spanish adults and children produced shorter VOTs in English initial voiceless stops than did age-matched English-speaking monolinguals and also longer VOTs in Spanish voiceless stops as compared to Spanish monolinguals. Therefore, although these native Spanish speakers of English developed distinct phonetic categories for Spanish and American English /p,t,k/, they did not realize either Spanish or American-English stops in the same way as the Spanish-speaking monolinguals did.

Similarly, Flege (1991) examined VOT for both early (mean Age of Arrival=2 years) and late (mean Age of Arrival=21 years) Spanish-speaking bilinguals and Spanish-

and English-speaking monolinguals. The results were similar to those of the previous study, i.e., the VOT values of late bilinguals were intermediate to the short-lag values of the Spanish-speaking monolinguals and the long-lag values of the English-speaking monolinguals. The early bilinguals' VOT values for English stops did not differ from those of the English monolinguals. In another study of Spanish-accented VOT, Schmidt and Flege (1996) studied the effect of speaking rate on the production of voiceless stop consonants for monolingual Spanish speakers, bilingual Spanish-accented English speakers and monolingual native American-English speakers. Voice Onset Time (VOT) was measured for word-initial tokens of /p/ and /t/ in sentence-initial, medial, and final position. The sentences were produced at either a normal, slow, or fast speaking rate. Results showed the following: 1) speaking rate exerted less effect on the VOT of Spanish stops than on the VOT of the English stops produced by native AE speakers, 2) AE speakers produced /p/ and /t/ with shorter VOT at a faster rate than at a normal rate, whereas Spanish-speaking monolinguals showed a trend in the opposite direction, i.e., VOT increased as rate became faster, 3) Early Spanish-English bilinguals performed like monolingual English speakers with respect to VOT, whereas late bilinguals showed a smaller speaking rate effect, and thus were more like the Spanish-speaking monolinguals.

Flege, Munro, and Skelton (1992) studied the production of word final /t/ and /d/ by Spanish L2 speakers of English and monolingual English speakers. Spanish does not contrast word-final /t/-/d/. Acoustic analyses showed the following for the monolingual English speakers: a) preceding vowels tended to be significantly longer before /d/ than /t/, b) closure voicing was of greater duration for /d/ than for /t/, and c) stop closure duration was longer for /t/ than for /d/. For the Spanish speakers of English, similar acoustic

differences were found, although the magnitudes of the differences were smaller. In this study and in the studies of VOT in initial stops (Flege 1991, 1996) native English-speaking listeners were not asked to rate the relative accentedness of the nonnative speakers' utterances. Therefore, it is not known whether the differences in acoustic parameters for voiced and voiceless stops in initial and final position correlated with the perception of accentedness by native American-English listeners.

Other Temporal Characteristics of Accented English

Temporal Differences between Native and Nonnative Productions

Production deviations of nonnative speakers (that may affect perception of accentedness) are of two types: temporal and spectral. Temporal parameters have been found to influence the perception of accentedness. Temporal properties vary across different languages (Flege, 1988; Mack, 1982) and similar phonemes across different languages are reported to have quite different patterns of temporal implementation (Lehiste, 1970; Kohler, 1979; Port, Al-ani and Maeda, 1980). L1 temporal properties are reported to carry over to L2, affecting listeners' perceptions and making the L2 speech sound accented (Flege, 1984). Additional evidence (Flege and Port, 1981) reveals that durational differences in vowels preceding obstruents are smaller for Arabic speakers of English than those produced by native speakers. Mack (1982) showed a similar effect for French speakers of English. Mack showed that vowel duration in English varied depending on the voicing class of the following stop, but this difference was found to be significantly smaller for nonnative (French) speakers of English. The nonnative speakers showed vowel duration ratios that mimicked those of French monolingual speakers, reflecting a carryover from French. Duration plays a role as a phonetic rule in English but

not in French. This finding lends support to the claim that nonnative speakers have different temporal patterns of production compared to native English speakers, and that measuring duration for nonnative speakers may tap into production differences that affect perception of accentedness.

Elsendoorn (1985) examined vowel duration in speakers of Dutch as they spoke Dutch as well as English. Three groups of subjects (native Dutch speakers speaking Dutch, Dutch speakers of English, and native English speakers) produced vowels preceding obstruents in Dutch and English words. Acoustic analysis showed Dutch speakers of Dutch produced the shortest vowel durations. Dutch speakers of English produced vowels of intermediate duration in English words, i.e. they were longer than typical Dutch vowels, but shorter than vowels produced by native English speakers. Native speakers of English produced the longest vowel durations in English. In summary, Dutch speakers of English, on average, produced shorter vowel durations preceding obstruents compared to native English speakers. Furthermore, they did not make durational contrasts for vowels preceding voiced versus voiceless obstruents, whereas the native English speakers showed systematic reduction in duration in vowels preceding voiceless obstruents.

This study also included a perception experiment to test if the reduced vowel durations of Dutch speakers of English resulted in inaccurate identification of words by native listeners. That is, did they perceive the consonants as voiceless stops instead of voiced if the preceding vowel was shorter than expected or as voiced if the preceding vowel was too long? Stimulus words were nine CVC words with one of the three vowels /æ/, /u/, or /e/, and a final voiced/voiceless obstruent or nasal /n/. These recorded stimuli

words were acoustically edited to adjust the vowel duration to equal the mean duration of native English speakers or native Dutch speakers. Four conditions of edited vowel duration were presented to listeners: English speakers speaking English, Dutch speakers speaking English, English speakers speaking Dutch, and Dutch speakers speaking Dutch. Listeners (native speakers of English) were asked to write down the word they heard and also required to score the acceptability of the pronunciation on a 5-point acceptability scale, ranging from acceptable to unacceptable. Results showed the listeners were able to discriminate native speakers of English from the Dutch speakers of English. There was a significant “final-obstruent effect”, as the vowel contrasts for the native English speakers were highly intelligible and discriminable (1.4% errors) in both English and Dutch conditions. However, the errors significantly increased in the judgment of the vowels spoken by Dutch speakers of English (29.6% errors) for voicing contrasts in final obstruents. In light of this speaker effect, it was concluded that spectral composition clearly affected identification as well as acceptability judgments by native English listeners. Vowel duration differences were not found to influence listeners’ identification of the English words spoken by native English speakers. However, foreign vowel duration was indeed seen to influence the acceptability of the speech as perceived by native listeners.

In a study of native Arabic speakers of English, Port and Mitleb (1983) showed that English timing patterns were more difficult to acquire than spectral characteristics for Arabic speakers of English. Twelve Arabic speakers of English (a group of Arabic speakers in the US, and a group of Arabic speakers from Jordan) and 5 native AE speakers read English carrier sentences with a target word in each sentence. The target

words consisted of minimal pairs of stop contrasts /p/-/b/ and /t/-/d/ in both initial and final positions. Temporal measurements of productions were made for VOT for /p/-/b/ and /t/-/d/ in initial and final positions, vowel duration in /p/-/b/ and /t/-/d/ in final position, and final stop closure duration for final /p/-/b/ and /t/-/d/. Durations of whole sentence minus duration of the target word were also calculated as a measure of tempo. Results showed that compared to AE speakers; the Jordanian Arabic group had reduced overall rate reduced aspiration and VOT in initial stops, reduced differences in vowel duration between vowels before tense and lax stops in “sop-sob” and “sit-sid” contrasts. Final stops also differed between the native and non-native groups in that the native speakers flapped the final stops /d/ and /t/, as evidenced by the shorter closure duration and lack of a difference in production of /d/ versus /t/. The US Arabic group produced flaps whereas the Jordanian Arabic group did not produce flaps. Similarly, the closure duration differences contrasting the stops, /p/ and /b/, were found to approximate American pattern of implementation that the Arabic speakers in Jordan had not acquired. The authors hypothesized that “temporal implementation rules are located further down in the motor system such that adult speakers have far greater difficulty altering them” (p.228).

The modification of durational differences in voiced stops produced by Spanish L2 speakers has also been examined. Nathan (1987) designed an experiment to discover if native speakers of Spanish acquire the English VOT values for voiced stops. Spanish listeners listened to English words presented orally containing voiced and voiceless initial stops and were required to shadow the words. They were re-tested after 18 months to see if production patterns had changed. Results showed that after a period of 18 months of

exposure, Spanish ESL speakers changed their articulation patterns by reducing the duration of voicing lead and developing short-lag versions of voiced stops, but only for velar targets.

The pattern of findings of Spanish-accented English described earlier in the studies by MacDonald (1989) and Ortega-Llebaria (1997) also support the importance of temporal features to the perception of accentedness. They described several processes which are related to vowel identity and voicing of stops. Many of the phonological deviations they reported have temporal/duration correlates, e.g., stopping of fricatives, neutralization of aspirates, deletion of word-final consonants, simplification of word-final clusters, devoicing of obstruents, addition of epenthetic vowels, and spirantization of medial voiced stops. The vowel production differences also have a temporal component in addition to the more marked spectral component for vowel changes related to /æ,ʌ,ɔ/; and /i-ɪ/.

Temporal Differences Serve as Perceptual Cues in English

Klatt (1976) reviewed the acoustic and perception literature to support the claim that segmental duration in English systematically varies to convey information at several linguistic levels and is, concomitantly, a significant perceptual cue. Segmental duration carries a functional load at each of the following linguistic levels: At the extralinguistic level, the varying durations convey mood and speaking rate of the speaker by duration of pauses and differential shortening of vowels and consonants. At the discourse level, duration gives an indication of the position of the phoneme within a paragraph. Also, the last sentence of a read paragraph is reported to be longer in duration than if it were not in paragraph-final position. At the semantic level, an increase in duration has been found to

correlate with emphasis or contrastive stress. At the syntactic level, duration of syllables before pauses at phrase and clause boundaries increases by about 60-200 ms., mostly restricted to an increase in the duration of the vowel and any postvocalic sonorant or fricative. Duration also indicates word-boundaries: word-final syllables are longer in duration even in non-phrase-final position, consonants are longest in word-initial position, somewhat shorter in word-final position; and shortest in medial position. Of greatest relevance to the present research are the temporal characteristics associated with the perceptual differentiation of phonetic segments and lexical stress. The duration of vowels and consonants is relevant to phoneme identification. Four vowels (i, e, ʌ and u) are intrinsically shorter than other English vowels. Voiceless stops, fricatives and affricates are longer in duration than their voiced counterparts. Formant transition duration varies systematically with manner of articulation of consonants (stop vs. glides). Variation of vowel duration can cue the voicing classification of postvocalic consonants. Lexical stress also has a durational correlate, as stressed vowels are longer in duration than unstressed vowels. This difference is largest in a phrase-final syllable. Additionally, prestressed consonants are slightly longer in duration than poststressed consonants. All of these variations in duration are shown to have perceptual significance. Segmental timing allows listeners to perceive lexical stress and segment identity.

Temporal Deviations in Nonnative Speakers' Utterances are Relevant to Perception by Native Listeners

As described above, there is substantial evidence that listeners use temporal cues in the perception of words, phrases and sentences. Moreover, deviations in temporal patterns are related to judgments of accentedness by native listeners. Flege (1993)

showed that degree of perceived foreign accent, as judged by native listeners, correlated significantly with vowel duration differences preceding final voiced versus voiceless obstruents for Chinese-accented English speakers. Hutchinson (1973) reported that vowel duration differences between stressed and unstressed syllables spoken by Spanish L2 speakers of English resulted in lower accentedness ratings by native AE listeners. Similarly, Swedish listeners judged as “foreign” the native Swedish speakers’ utterances whose consonant and vowel durations were modified to match those of American English speakers (Jonasson and McCallister, 1972).

All of the above studies indicate that native listeners are sensitive to durational deviations in non-native utterances and that these temporal differences are significant contributors to perceived accentedness.

Research Questions and Hypotheses

The following is a summary of the findings based on the literature review described above:

- 1) A substantial amount of information is available about the phonological characteristics of native Spanish speakers’ production of English words, phrases, sentences, and extemporaneous speech.
- 2) A hierarchy is available of the degree to which these characteristics affect native listeners’ judgments of intelligibility.
- 3) A few studies provide information on acoustic characteristics of Spanish-accented English speakers.
- 4) Only indirect inferences can be made regarding which phonological and acoustic characteristics affect listeners’ perception of foreign accentedness.

- 5) **A sizeable amount of evidence points to the contribution of temporal/durational deviations in nonnative productions, in general, (and for Spanish speakers of English in particular) to perceived accentedness.**
- 6) **No study to date has provided a detailed description of temporal production characteristics of the nonnative Spanish speakers of English, as they relate to accentedness.**

Based on the above summary, the studies designed here addressed the following hypotheses:

- 1) **The temporal structure of Spanish-accented speakers' productions of multisyllabic words will differ from that of native American-English speakers, and**
- 2) **Temporal deviations in nonnative speakers' productions of multisyllabic words will be correlated with native AE listeners' perception of accentedness.**

In order to address these two hypotheses, two studies were designed. Study 1 (the perception study) established the relative accentedness of a set of 22 native Spanish speakers of English as compared with five native AE speakers. Two sets of materials were presented to a panel of ten native AE listeners: 1) eight short sentences each containing a multisyllabic target word, and 2) the target words excised from the sentences and presented in isolation. The purpose of this study was to establish the extent to which the perceived accentedness of the isolated target words was predictive of the listeners' judgments of speakers' "global accentedness," derived from their ratings of the sentences. If the ratings of the target words were highly predictive of global accentedness, then a detailed analysis of their temporal structure and its relationship to accentedness ratings would be justified.

Study 2 (the production study) included detailed acoustic analysis of the productions of the eight multisyllabic target words. Overall word durations (used as a measure of speaking rate), relative vowel duration, stressed to unstressed vowel ratios (a measure of lexical stress), VOT of initial voiceless stops, and intervocal /t/ flap/closure duration were the temporal measures examined. Each of these measures was predicted to deviate from native AE speaker patterns for the Spanish speakers' productions. Finally, each of the temporal measures of the nonnative speakers' production was correlated with native listeners' judgments of accentedness of the words to establish the extent to which each temporal deviation was predictive of accentedness.

CHAPTER III

Study 1: Rating of Accentedness (Perception Study)

This study had two objectives: first, to provide perceptual ratings of relative "global" accentedness of Spanish speakers uttering English sentences, as judged by native listeners of English; second, to determine the relation between judged accentedness of the multisyllabic words used in this study and the speakers' overall level of global accentedness, as judged by native listeners' assessment of the sentences. This comparison revealed the extent to which the accentedness of the target words was predictive of "global accentedness" of sentences.

Method

Subjects

Nonnative Speakers (NN). Twenty-two (7 males, 15 females) native Spanish speakers of English as a second language served as the nonnative speakers. They were fluent speakers of English with a substantial degree of Spanish accent (as judged by native listeners in pilot testing). The speakers were from the Spanish-speaking Dominican Republic and resided in the New York metropolitan area. They were between the ages of 18-47 years and had arrived in the United States after 12 years of age. Age of Arrival to the United States (AOA) was used to represent the age of second language learning. The Length of Residence (LOR) was at least two years. All nonnative speakers had a minimum of a high-school level of education (Demographic information of subjects given in Appendix D).

A language background questionnaire (Appendix B) was used in the interview with all participants to establish that they were bilingual with Dominican-Spanish as L1

and English as L2, and had no spoken knowledge of any other language. Also, by self-report (and by the observation of the experimenter, a trained speech-language pathologist), the absence of any history of speech and language therapy or speech/language or hearing problems was ascertained. Subjects were recruited by posting advertisements in graduate school programs, ESL programs, and student e-mail postings.

Native Speakers (AE). Five (2 males, 3 females) monolingual AE speakers served as control subjects. A language background questionnaire was used to select native speakers who had not learned a foreign language in the past, and who did not speak any foreign language beyond a few words. The discrepancy between the group size of the native and nonnative speakers followed other studies (Flege, 1993; Flege 1984). These speakers ranged in age from 26-50 years. They were all born and raised in the Greater New York metropolitan area. They all had a minimum of a high-school education. These speakers did not have any history of or current speech/language or hearing problems, by self-report and by the observation of the experimenter.

Listeners. Ten listeners were drawn from the same population as the native AE speakers. Criteria for inclusion were: 1) ages between 18-45 years, 2) a minimum of a high-school education, 3) born and raised in the Greater New York metropolitan area, and 4) absence of history of any speech, hearing or language problems as reported to, and observed by, the experimenter. Additionally, on the basis of self-report, none of the listeners had any formal training in Spanish or any extensive experience listening to Spanish-accented English (e.g. living or working every day with a native Spanish speaker). They were all required to pass a hearing screening performed in a sound-treated room, with thresholds not over 25 dB for 500Hz, 1000Hz, 2000Hz and 4000Hz.

Stimuli

Stimuli for global accent-judgment (sentence-rating) task. Eight sentences, each containing a high-frequency multisyllabic target word, were used as the stimuli in this study. These target words were selected from 98 multisyllabic, two to five syllable words appearing in the Francis and Kucera (1982) word lists, with frequency of occurrence between 53 and 729. These words were embedded in sentences with well-defined semantic contexts for the presence of the target words (Appendix A). In a pilot experiment the Spanish speakers recorded the sentences. The speakers had no knowledge of which words in the sentences were target words.

These 98 sentences were then presented to two trained phoneticians for analysis regarding production deviations and accentedness. On the basis of these judgments, 16 words were identified in which the lexical stress was shifted and/or vowel production was perceptibly accented, but in which such consonantal problems as cluster reduction or consonant omission were not minimized (See Table 3.1). These 16 words were considered representative of those that elicit temporal deviations from Spanish-accented speakers.

Acoustical measurement of tokens of these 16 words was completed to determine whether phone-sized segments could be determined. Some of the words did not lend themselves well to acoustic segmentation. For example, in the word “community” /kəmˈjʊnɪti/, the formant structure was continuous from the vowel /ə/ to the end of the word, making segmentation unreliable. Similar problems were present in the segmentation of the words “material,” “recognize,” “apparently,” “literature,”

Table 3.1 List of Target Words Within the Revised Sentence Contexts

Final Corpus

The Telephone Company closed down

They formed a committee to solve this problem

Her sister will move to California soon

He makes automobile spare parts

Mexico's economic situation is slowly improving

We have a new political candidate

The new student association seems active

The congress representative changed his vote

Sentences not selected for final corpus

I live right next to the community center

Read the material carefully

I cannot recognize any of these pictures

He has apparently stopped working

She enjoyed Bill's literature paper

I am presenting in a conference next week

The fresh atmosphere calms us

I want Governor Smith to win again

“conference,” “atmosphere,” and “governor,” where flap, contiguous glide or diphthong combinations rendered segmentation impossible. Therefore, the list of 16 words was shortened to 8 words that were acoustically segmentable.

These 8 words chosen varied in morphological status, as well as in the distribution of primary stress in the word (See Table 3.2). “Company” [k^hʌmpəni], “committee” [k^həmɪri] “representative” [rɛprəzɛnrɛrɪv], and “California” [k^hæləfɔrnjə] do not involve a shift in lexical stress from a morphologically related word (e.g. represent), as they are either not derived morphemes, or are derived from the root without a lexical stress shift. The remaining four words are derived from another morpheme (or are multimorphemic) with resulting shift in lexical stress as follows: “economic” /ɛkənómɪk/ from “economy” /əkónəmi/; “political” /p^həlɪtɪkəl/ from “politics” /pólətɪks/; “association” /əsosiéjən/ from “associate” /əsósiət/; “automobile” /ɔrəməbiəl/ from which the derived word “automotive” /ɔrəmórv/ involves a stress shift. These eight target words in their corresponding sentences are listed in Table 3.1 as the “Final Corpus.”

Recording Procedure. All speakers who responded to the advertisement were recorded if they met the inclusion criteria. Each speaker filled out a language background questionnaire (Appendix B). Instructions during recording were for the speakers to read each sentence aloud “at a normal speaking rate, as naturally as you speak in your everyday conversation and avoid using over-careful pronunciation.” They were instructed to read the sentence list, one sentence at a time, silently counting to four after each sentence (to provide a silent interval between sentences).

Table 3.2 List of the 8 Words, Transcription of Acoustic Segments, and Morphemic and Stress Status of Target Words

<u>Word</u>	<u>Phonetic transcription</u>	<u>No. of syllables</u>	<u>Syllable with primary stress</u>	<u>No. of morphemes</u>	<u>Stress Change</u>
company	/k ^h ʌmpəni/	3	1 st	1	<u>No change</u>
committee	/k ^h əmɪri/	3	2 nd	1	<u>No change</u>
California	/k ^h æləfɔrnjə/	4	3 rd	1	<u>No change</u>
automobile	/ɔrəməbiəl/	4	1 st	2	<u>Difference in stress from related morpheme “automotive”</u>
economic	/ekənɔmik/	4	3 rd	2	<u>Change of stress from “economy”</u>
political	/p ^h əlɪtɪkəl/	4	2 nd	2	<u>Change of stress from “politics”</u>
association	/əsosiəʃən/	5	4 th	2	<u>Change of stress from “associate”</u>
representative	/rɛprəzɛnrɛɪtɪv/	5	3 rd	3	<u>No change from “represent” or “présent”</u>

Prior to the recording, each speaker was familiarized with the sentences in order to elicit fluent samples. The experimenter read the sentences first while the speaker silently read the sentences at the same time in order to become familiar with them. After this, the speaker read the sentences aloud as the experimenter set the recording level of the tape recorder. The speaker continued to practice reading until he/she felt comfortable with the material and the task, following which the recording began. The speaker was allowed to self-correct productions during the recording. Each speaker read the entire list three times and the most fluent repetition was used in the studies. Fluent repetition was defined as productions that were free from filled or unfilled hesitations, repetitions, and prolongations according to traditional definitions of fluency.

The sentences were recorded on the left channel of a portable DAT recorder (TASCAM, model # 40056), at a sampling rate of 44.1K, using a dynamic microphone (Electrovoice Shure SM 48). Recordings were made in a quiet room, with the microphone placed on a stand, and positioned directly in front of the speaker at a distance of approximately 14 cms. Via headphones, the experimenter monitored the noise level in the room to ensure that there were no electronic or mechanical sounds due to equipment such as air-conditioners, refrigerators, computers, radiators, bells, fans, or lights. Recording levels were monitored by looking at the VU meter during recording to avoid peak-clipping of the input.

For each of the 27 (22 nonnative and 5 native) speakers, the eight sentences were digitized (DAT analog out to PC analog in) at 16 bit resolution and a 22.05 kHz sampling using the Sound Forge speech-editing program (version 4.5, Sonic Foundry, Inc., 1991-1998). Each sentence was saved as an individual sound file with a “.wav” extension. To

preserve anonymity, the subjects' sound files were coded by number NN1, NN2 and so on for nonnative speakers, and AE1, AE2, and so on for the American English speakers. These sentence-files were then played to the listeners for rating of global accentedness.

Stimuli for Word-Rating Task. The target words were excised from their sentence contexts and saved as separate files using the Sound Forge speech-editing program (version 4.5, Sonic Foundry, Inc., 1991-1998). The words were segmented from the boundaries after the end of the preceding segment up to the end of upper formants/periodicity of the target word. Each target word either a) began with a silent closure and was therefore easily separable from the upper formants of the preceding vocalic segment, or b) began with a vocalic segment that could be separated from the upper formants of the preceding fricative. Similarly, the final segment of the target word was segmented from its neighboring sentence context by separating final fricative or vocalic formants from the neighboring silent closure. The sentence context for splicing these target words was, thus, designed to make the splicing as efficient as possible. Nevertheless a few splicing problems did occur despite the precautions in selecting the preceding and following context within the sentences; these will be discussed below in the results' section of the next chapter. These target word files were then played to listeners for rating of accentedness.

Procedure for Rating Tasks: Global Accentedness (Sentences) and Target Words

Ten listeners rated eight blocks of sentences (as a measure of global accentedness) and eight blocks of words. Within each block, the sentence/word was the same utterance, produced by all 27 speakers. Thus, each block of sentences/words consisted of 27 (22 nonnative and five native versions) sentences or 27 words, respectively. The sequence of

presentation of these blocks was as follows: Set 1) 8 blocks of sentences; Set 2) 8 blocks of sentences randomized differently from those in Set 1, Set 3) 8 blocks of words; Set 4) 8 blocks of words randomized differently from those in Set 3. A rest period (5-10 minutes) was provided between each of the four sets.

Each of the 8 blocks of sentences/words was randomized (without replacement) differently for each listener. The order of Blocks 1-2 and Blocks 3-4 was also randomized differently across listeners; however, for all listeners, sentence blocks preceded word blocks so that the listeners would not know which words in the sentences were the target words. Ratings of foreign-accentedness were obtained for each sentence/target word using a 9-point rating scale presented to the listeners on a computer screen. Listeners were instructed to listen and rate the relative accentedness of the speakers using the entire range. The listeners were asked to rank each sentence from 1 to 9, where 1=Least foreign accented, 9=Most foreign accented³. The stimuli were presented binaurally through headphones. The presentation of the stimuli was self-paced, i.e., the listeners controlled the response-time, and each new sentence/word was played 750 ms after the listener had made a response. At the end of the presentation of the first sentence block, the listeners were reminded that they had just heard the entire set of speakers, and in subsequent presentations, the listeners should make their ratings based on this range of speakers. Therefore, only the second set of 8 blocks was used in the subsequent analyses in order to ensure familiarity with the range of accentedness of speakers and adequate calibration of

³The lower-end point "1" was based on previous studies using rating-scales (Munro, 1995; Munro, Flege, and Mackay, 1996; Munro and Derwing, 1995; Southwood and Flege, 1999; Magen, 1998, and Flege, 1988); the higher endpoint varied in all of these studies and was either "4", "5", "7", "9" on equal-appearing interval scales, or "256" on a continuous scale).

ratings. (This task, instructions, and labels of the end-point of the accent-rating scale were decided on to ensure that all of the listeners would experience a consistent and comparable task. The listeners could then presumably judge accents over a consistent range of speakers, as opposed to an arbitrary range of accent they may have intuitively come up with, if they were given open-ended instructions to judge for accents ranging from native to foreign). The productions of the native speakers of AE in the corpus were used as a baseline for comparison of the degree of relative foreign accent of nonnative speakers. The experimenter monitored the first few trials of the first block. Oral instructions, if needed, were given to remind the listeners to follow the instructions (given below) for the ratings.

Instructions for the Rating Task. The listeners were given the following instructions for the rating tasks (either sentences or words):

“You will hear (sentences/words) spoken by some American and some foreign speakers. Your task is to decide whether each (sentence/word) sounds foreign-accented compared to the native American accent. Please rate each (sentence/word) from 1 to 9, where 1=Least accented (among these speakers), and 9=Most accented (among these speakers). You can indicate your response by clicking on the rating-scale seen on the computer screen. Each (sentence/word) will be repeated several times, spoken by different speakers, so rate each one separately. LISTEN VERY CAREFULLY as you will not be able to replay the (sentence/word). Also, decide your response before clicking on the rating scale, as you will not be able to go back to change your response. Most importantly, be sure to use the entire rating scale to make your judgments, not limiting

yourself only to the extreme ends of the scale. If you have any questions, ask them now, or else click below to begin”.

Results

Reliability of Accentedness Judgements

Intrajudge reliability. A comparison of judgments of sentence and word conditions of each stimulus word for each of the 10 native listeners can be seen in Appendix C. Median ratings were calculated for each of the 8 words (pooled over 27 speakers each) as rated by each listener. Ranges of the ratings were also tabulated for each of the listener ratings of each of the 8 words. These medians and ranges were compared across the sentence and word conditions of each stimulus word for each listener as a measure of intrajudge reliability. All the listeners used nearly the full range on the 9-point rating scale. Each listener displayed internal reliability in his/her ratings across sentences and words, as demonstrated by the roughly comparable medians (medians of corresponding word and sentence condition for each stimulus differed no more than 2 points on the scale) and ranges for sentence and word conditions (See Appendix C).

Interjudge reliability. Interjudge reliability across the 10 listeners was fairly high. Table 3.3 shows the median and range of accentedness ratings collapsed across all speakers (including the native speakers) and all 8 words in both sentence and word conditions. Comparison of medians within each condition, across different listeners showed high consistency, with median ratings varying only 0.5 steps for the sentence condition, and 1.5 steps for the word condition. Correspondingly, comparison of ranges across all listeners for each sentence or word condition was completed. As the table

shows, with only four exceptions (shown in bold), listeners used the entire range (1-9) for both sentence and word conditions.

Each of the listener's ratings was also examined by separating their performance on only the native speakers (AE) from those of the nonnative speakers (NN), in the sentence and word conditions. Figure 3.1a shows the frequency distributions of ratings

Table 3.3 Interjudge Reliability: Median and Range of Accentedness Ratings Collapsed Across All Speakers(AE + NN) and All 8 Words				
Listener	Overall sentence stimuli		Overall word stimuli	
	median	range	median	range
1	6	1--8	5.5	1--9
2	6	1--9	5	1--8
3	6	1--9	5.5	1--9
4	6	1--9	5	1--9
5	6	1--9	6	1--9
6	6	1--9	7	1--9
7	6	1--8	7	1--8
8	6	1--9	7	1--9
9	5.5	1--9	7	1--9
10	6	1--9	7	1--9

for the sentences spoken by AE speakers for each listener. Each of the ten listeners gave a modal rating of 1 for all the AE sentences. The ranges of the ratings for AE sentences were also consistent across listeners, ranging from 1 to 3, with very few 2 or 3 ratings. Similarly, in Figure 3.1b, frequency distributions of ratings are displayed for the word-condition for AE speakers. Again, the modal ratings for all ten listeners were 1 and ranged from 1 to 3, except for one rating of 5 by Listener 2. Figures 3.1c and 3.1d show frequency distributions for the listeners' ratings of the nonnative (NN) speakers' sentence and word productions, respectively. (It is to be noted that in the previous two figures, Fig. 3.1a and 3.1b, the ordinate ranges from 0-6, whereas in Fig. 3.1c and the Fig. 3.1d, the range is from 0-20 because of the difference between group sizes.) The graphs for the NN data show more variability in range and shapes of distribution than the AE data. Figure 3.1c shows that the modes ranged from 6 to 9 across the listeners. The range in ratings of NN production varied across listeners from 9 steps (1 to 9) for Listener 4 to only 5 steps (5 to 9) for Listener 3. Some showed negatively skewed distributions while others were more symmetrical. Fig. 3.1d, showing the NN word ratings' distributions, displays the presence of multi-modes. The modes ranged from 4 to 9 across listeners, while individual listeners' distributions showed spread of 7 (2 to 9) to 9 (1 to 9) steps.

Rationale for pooling listeners. Figures 3.1a-d showed that across all listeners, the medians and ranges were more similar for ratings of the AE speakers' data, than for the NN speakers' data, as would be expected, while performance on NN speakers' productions varied somewhat across listeners. No listener's data were judged to be qualitatively atypical. All listeners separated the native from the nonnative speakers in both word and sentence conditions. Moreover, pooling all medians across all words and

Figure 3.1a: Frequency Distribution of Listeners' Accentedness Ratings: AE Sentence condition

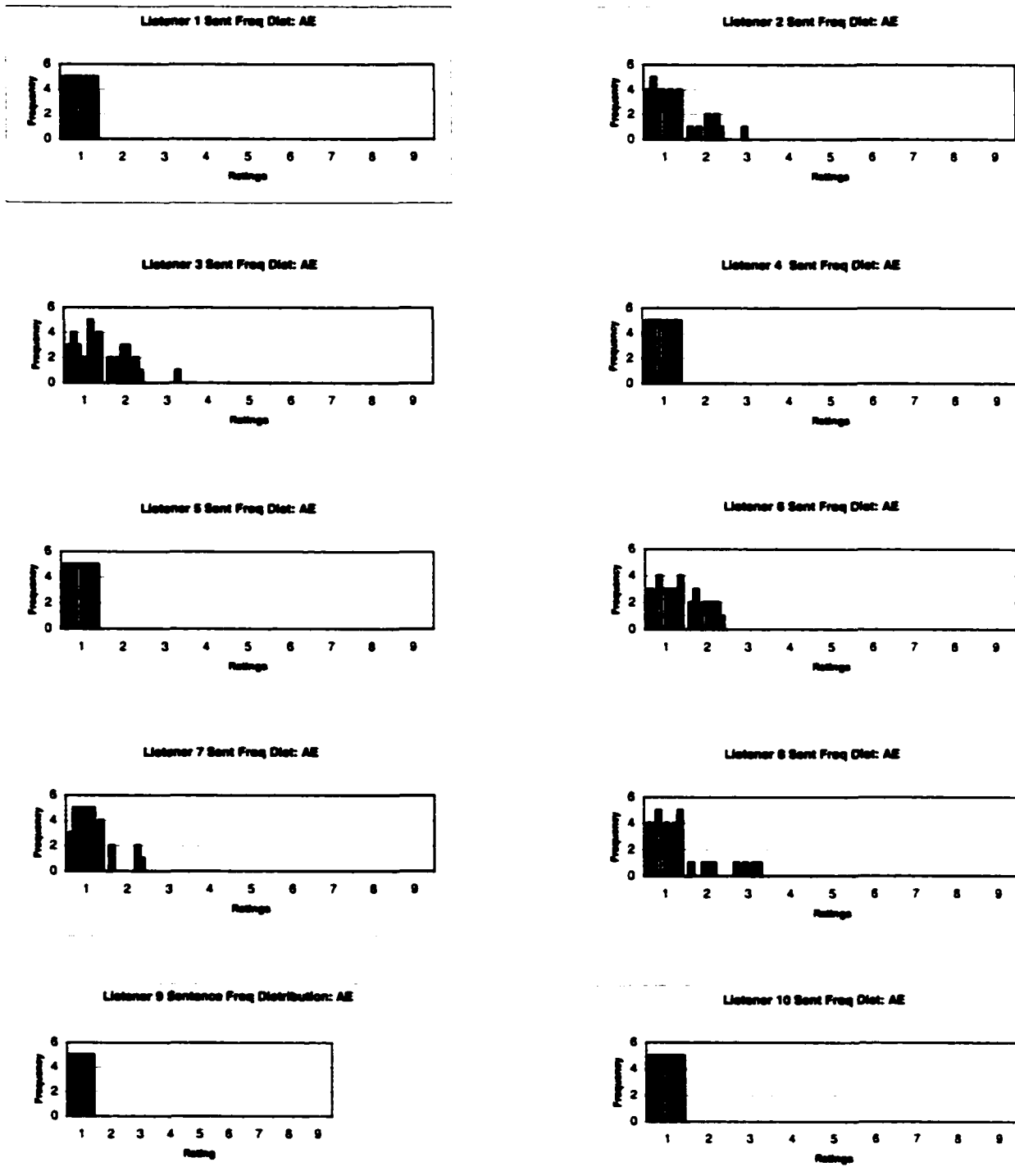


Figure 3.1b: Frequency Distribution of Listeners' Accentedness Ratings : AE Word condition

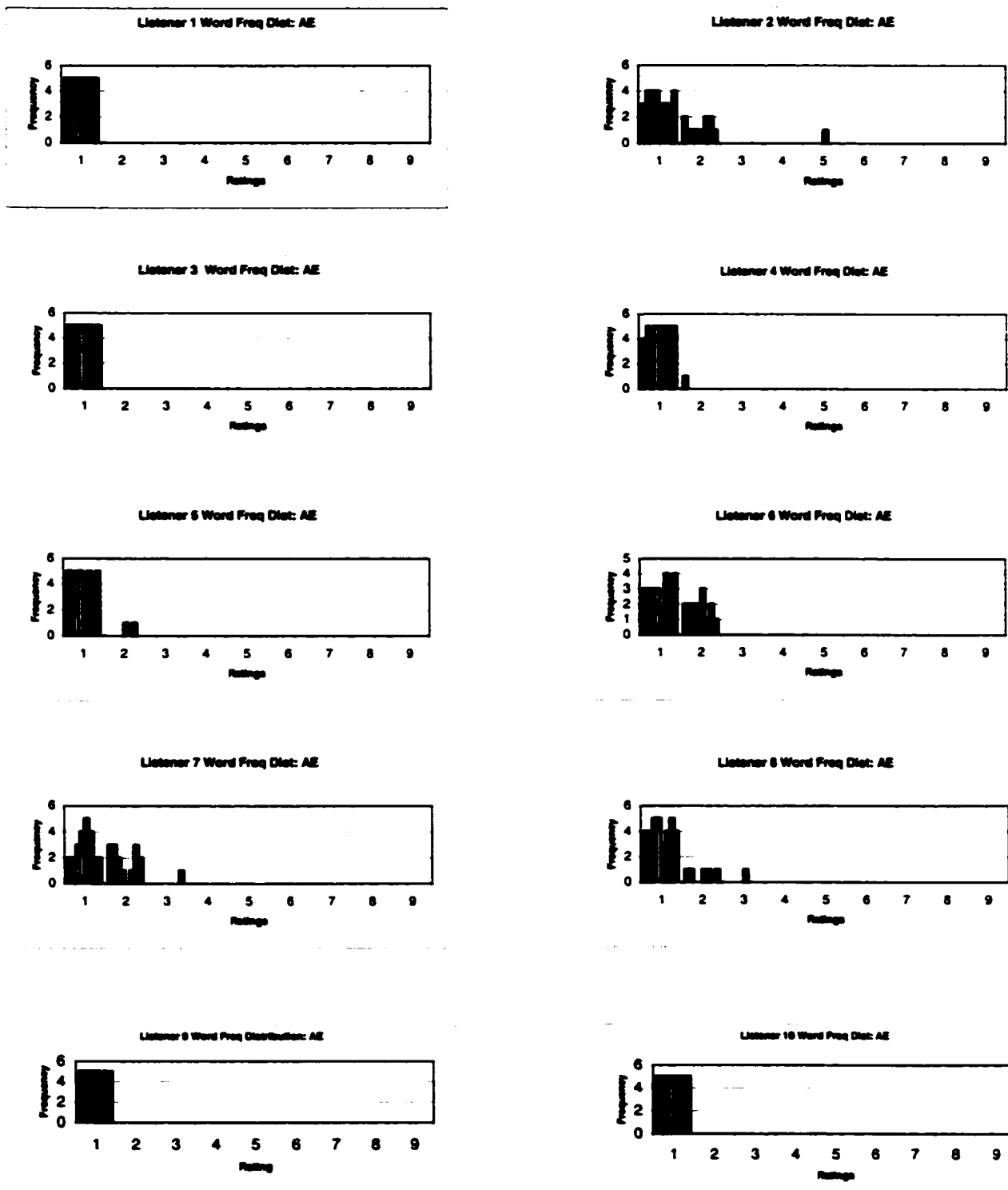


Figure 3.1c: Frequency Distribution of Listeners' Accentedness Ratings : NN Sentence condition

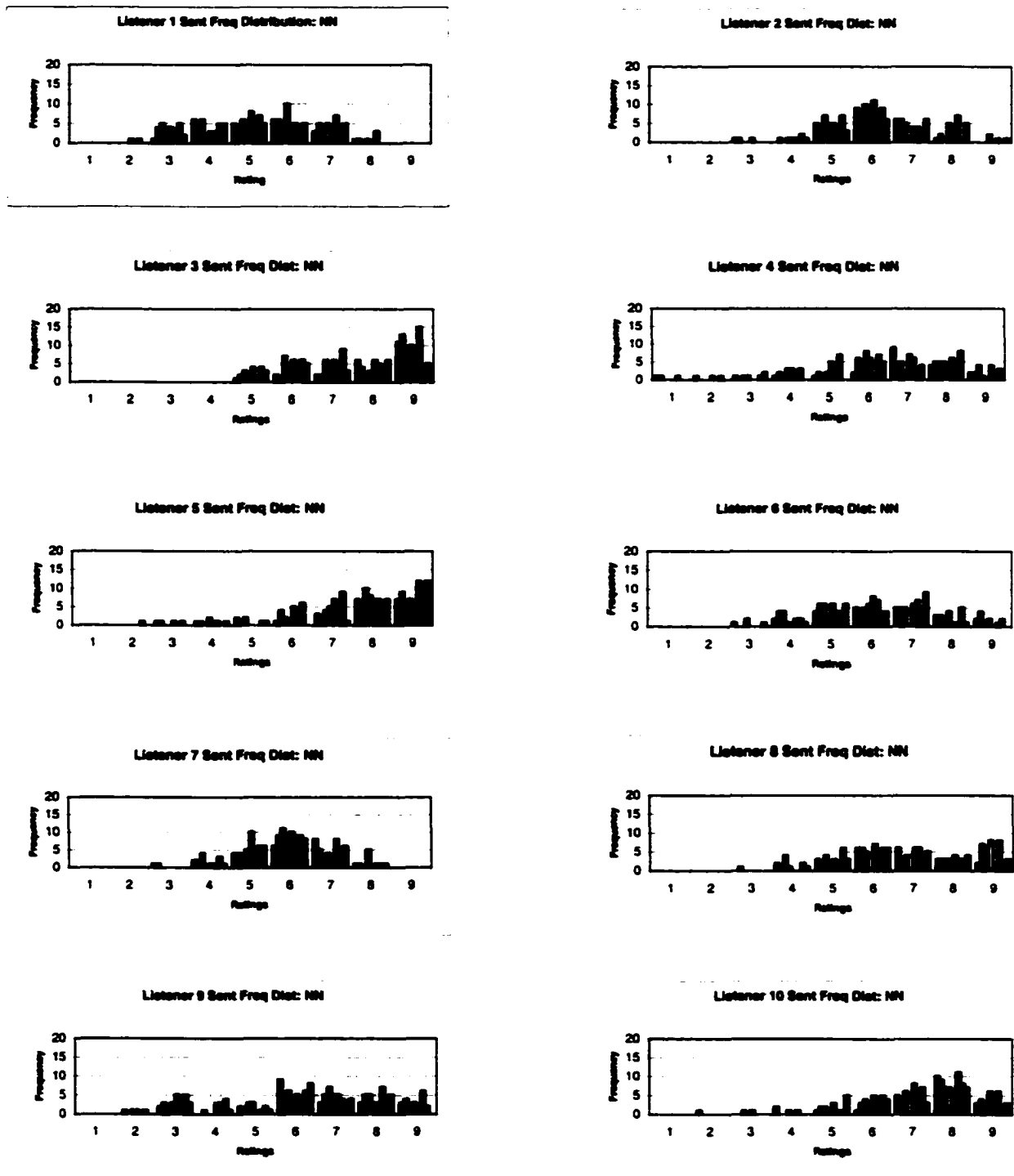
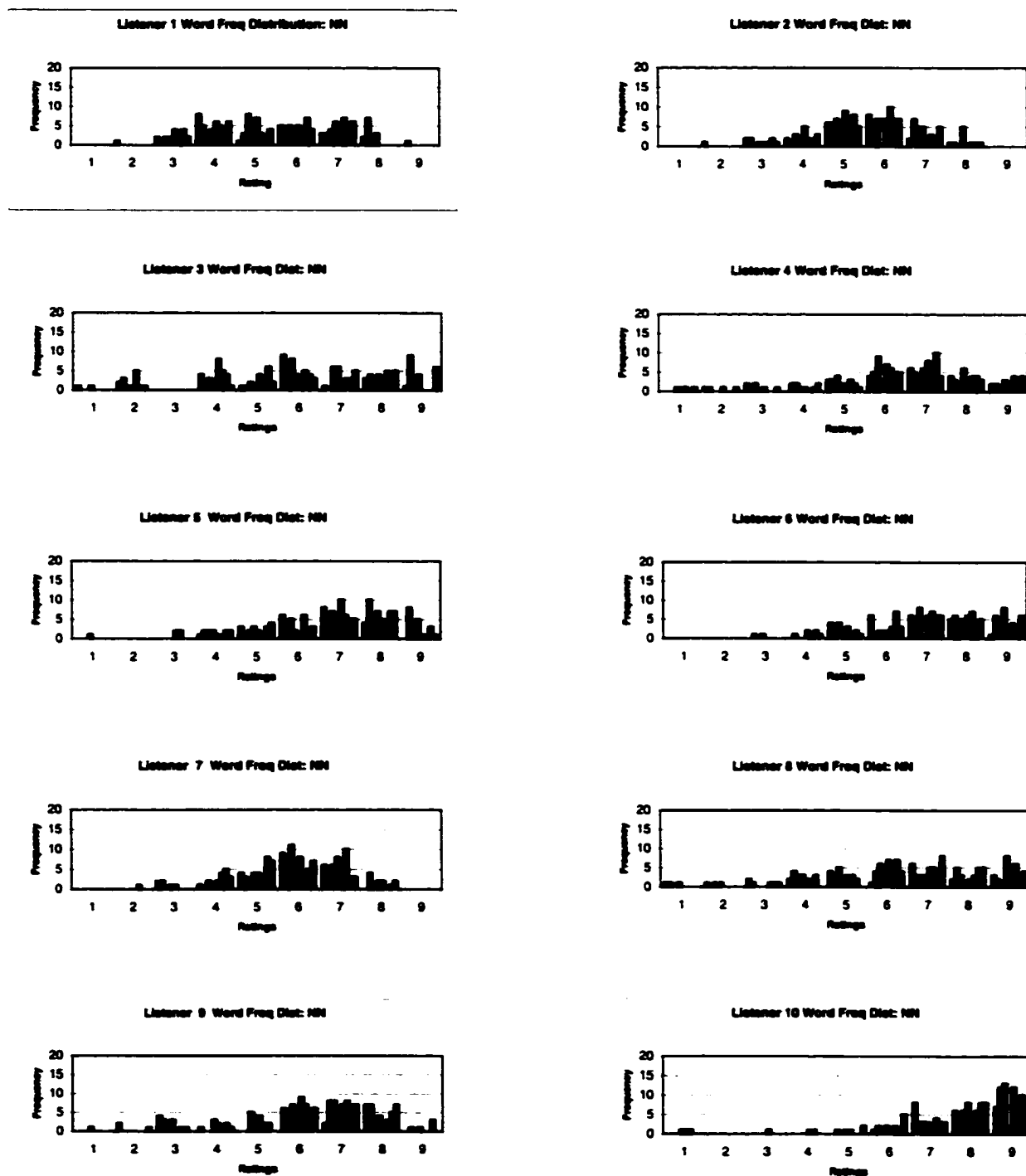


Figure 3.1d: Frequency Distribution of Listeners' Accentedness Ratings: NN Word condition



listeners, the median ratings assigned to each speaker showed that some speakers were identified as "least" accented, and some as "most" accented. Speakers NN15, NN19 and NN24 were all assigned low (less accented) medians (pooled over all words) by each listener, and speakers NN09, NN10, NN13, NN22 were all assigned high (more accented) medians by each listener. Given the above trends, inter-rater reliability was judged to be high enough to warrant pooling judgments across listeners. Therefore, ratings of the listeners are henceforth pooled and the typical (median) native listener judgments of accentedness of the speakers are used as the speakers' scores in correlational analyses.

Correlation of Accentedness on Words vs. Sentences

Spearman rank order correlations of median accentedness ratings for sentence and word conditions were computed for each of the eight stimuli. In addition, an overall correlation was performed in which the overall median ratings on all 8 words/sentences for each speaker were calculated⁴. The overall sentence and word correlations were computed first including all 27 speakers (AE and NN) as seen in Figure 3.2. The ranked median word ratings are plotted on the abscissa and the ranked median sentence ratings are plotted on the ordinate. All 5 AE speakers were assigned (tied) ranks of 1. The correlation coefficient of +0.90 was thus inflated by including these AE speakers in the correlation. Therefore, the correlation was re-calculated with the AE speakers excluded, resulting in a correlation coefficient of +0.82, as seen in Figure 3.3. This positive coefficient is indicative of a high correlation between the word and sentence conditions across all stimuli. In general then, the excised multisyllabic target words were highly

⁴ Rank-ordered correlation was required in this study since an interval scale assumption for the 9-point accentedness rating scale could not be assumed.

Fig. 3.2: Correlation of overall median accentedness ratings of words and sentences: All subjects (AE + NN)

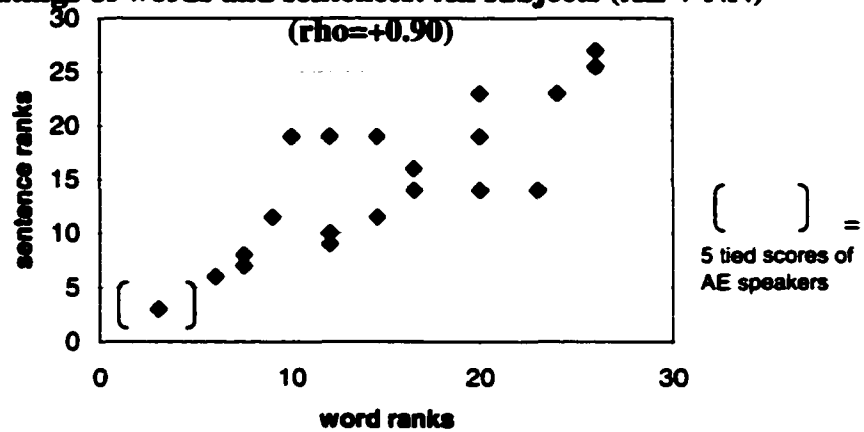
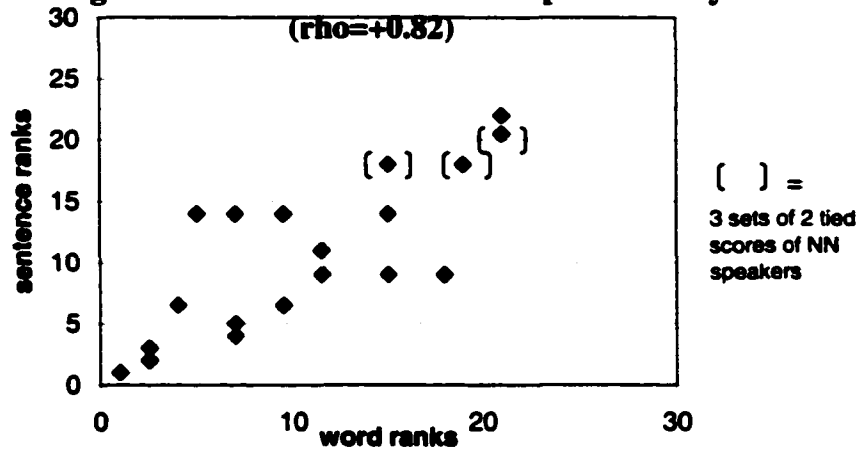


Fig. 3.3: Correlation of overall median accentedness ratings of words and sentences: NN speakers only



predictive of global (sentence) accentedness.

Each of the 8 isolated words was examined for differences in its predictability of global accentedness. Table 3.4 shows the results of rank order correlations for each stimulus word with its corresponding sentence. As seen in table 3.4, correlations across the 8 words ranged from +0.50 to +0.82. These correlations were all statistically significant with p values ranging from <0.02 to <0.001.

Table 3.4 Spearman Rank Order Correlations for each word/sentence, pooled over listeners		
Stimulus	Correlation (rho)	p-level
Association	+0.72	p<.001
Automobile	+0.80	p<.001
California	+0.61	p<.01
Committee	+0.82	p<.001
Company	+0.64	p<.01
Economic	+0.52	p=.01
Political	+0.50	p<.02
Representative	+0.69	p<.001

These results indicate that each target word was predictive of the speakers' judged accentedness from sentence productions (variance accounted for [ρ^2] ranged from 0.25% to 0.64%), albeit with some variability in the strength of the correlations (Scatterplots presented in Appendix E.)

Discussion

The objectives of this study were to a) establish the relative global accentedness of the NN speakers in the study, and b) to determine the correlation between ratings of accentedness in word and sentence conditions.

Results showed that intra- and inter-judge reliability was adequate to warrant pooling of listeners' responses. All listeners followed instructions to use the full scale. No listener deviated markedly in his/her use of the rating scale or in the ranges and distributions of ratings across speakers and conditions. Listeners were consistently able to differentiate the "least" and "most" accented speakers, showing reliable use of the rating scale.

Correlations of accentedness ratings on words and sentences showed that the isolated multisyllabic target words were good predictors of the overall accentedness of the sentences, although some words were better predictors than others.

The present results have a few methodological as well as theoretical implications. Consistent performance across the listeners indicates that native listeners are reliably able to detect the presence of, and rate the degree of, foreign accent. Moreover, listeners' reliability helps validate the use of the accent rating scale as a suitable measure of listeners' perceptions of range of foreign accentedness. This finding adds to evidence from prior research validating use of rating scales (c.f Southwood and Flege, 1999).

Multisyllabic words were found to be moderately predictive of overall accentedness. Hence, it appears that multisyllabic words have properties that contribute significantly to overall sentence accentedness. Inferences can, therefore, be made from productions of these multisyllabic words for the importance of lexical stress and segmental deviations in judgments of accentedness. The acoustic properties of these

productions of multisyllabic words need to be studied in order to narrow down those that underly the perception of accentedness by native listeners. Hence, Study 2 involved measuring the acoustic durations of segments of these multisyllabic words, and correlating selected temporal parameters with listeners' ratings of accentedness of the words.

CHAPTER IV

Study 2: Acoustic Analysis (Production Study) and Correlation of Acoustic Parameters with Accentedness Ratings

Durations of the acoustic segments in the target words produced by native and nonnative speakers were measured to determine deviations from native temporal patterns that may be associated with Spanish-accented English. This study was designed to yield several measures of temporal deviations of Spanish-accented productions that could be correlated with listeners' ratings to establish the relationship between production and perception of Spanish-accented English. Based on previous phonetic and phonological findings (described earlier in the review of the literature), the following temporal deviations were expected. Overall word duration for the target words was expected to be greater for NN speakers than AE speakers, since past studies have shown overall rate of speech is slower for nonnative speakers of English. Relative vowel durations were also expected to be different between the speaker groups, especially those that might be associated with the distinction between stressed and unstressed vowels within a word. Since Spanish is a syllable-timed language, it was hypothesized that the ratio of durations of stressed to unstressed vowels within each word would be close to 1.0 for the most accented speech. That is, all vowels would be approximately equal in duration. English, however, is a stress-timed language with reduction of vowels in unstressed syllables and therefore the ratio of stressed to unstressed vowel durations was expected to be greater than 1.0. Voice Onset Time (VOT) values for NN initial voiceless stops were expected to be smaller compared to those for AE stops because Spanish (L1) phonological rules generate voiceless stops that are unaspirated. The duration of flapped allophones of /t/

was expected to be greater in NN speakers, approximating the duration of stops, because Spanish phonology does not include a flapping rule for intervocalic alveolar stops.

The above-mentioned acoustic segments were measured using the Multispeech waveform editing software, Model 3700, version 2.3 (Kay Elemetrics Corp. 1999). Displays of waveforms and wideband spectrograms were used in combination to arrive at decisions regarding the acoustic segments. Spectrogram settings were: analysis size of 125 points, frequency range 0-4000 Hz, pre-emphasis of 0.8, Nyquist frequency OFF. The formant analysis settings were: Filter of 24 points, frame advance at 5 ms, bandwidth of 500 Hz, frequency range of 0-4000 Hz, Autocorrelation ON, and Nyquist frequency OFF.

Segmentation Methodology

Each of the 8 target word productions by the native (n=5) and nonnative speakers (n=22) in Study 1 was segmented using criteria listed below, and the duration of each segment was computed for each of the 216 words (27 speakers x 8 words). The following temporal segments were measured: 1) Closure duration for stops, 2) VOT of stops, 3) duration of flapped /t/, 4) Noise duration for fricatives, 5) Nasal murmur duration for /m,n/, 6) steady-state formant duration for liquids, and 7) vowel duration.

The following are the onset and offset criteria for each of the seven segments described above. Silence (Closure) duration for stops, as indicated by the duration of absence of energy in upper formants preceding stop release, was measured from the end of the upper formants (F2, F3) of the segment preceding the target phone up to the stop release (indicated by aperiodicity on waveform), not including the release. The duration of VOT of stops was measured from the onset of the burst to the onset of voicing, defined

as the first regularly-spaced vertical striations denoting periodicity that were evident above the baseline of the spectrogram. Flap duration was measured as the duration of the brief reduction of energy in upper formants accompanied by a decrease in the amplitude envelope. The duration of fricatives was measured from the onset to the termination of high frequency aperiodicity. The duration of nasal consonants was measured from the offset of F2 and/or a drop in amplitude envelope to the onset of F2 and a rise in amplitude envelope. Measurement procedures for the steady-state duration for liquids was adapted to specific target words. Since the two liquid sounds, /l/ and /r/ were preceded and followed by vowels in both the target words, “political” and “California” rendering segmentation impossible due to continuous periodicity, the duration of the entire syllable was measured. In “political”, the syllable /əl/ was measured from the onset of periodicity to the offset of the reduced amplitude associated with /l/. The syllable /ɪr/ was measured from the rise in amplitude to the end of the vocalic portion in the segment. In the word, “California”, the syllable /æɪ/ was measured from the onset of periodicity following the preceding burst stop and aspiration up to the point prior to the beginning of steady-state portion and/or reduction in energy. Correspondingly, the syllable /lə/ was measured from the beginning of the steady-state portion and/or reduction in energy, to the point prior to the beginning of high-frequency noise for the following fricative. The syllable /or/ was measured from the onset of periodicity in conjunction with the offset of frication noise of the preceding /f/, to the point just prior to the onset of nasal murmur for the following /n/. The syllable /jə/ was measured from the offset of nasal murmur to the offset of periodicity and formants. In the word “automobile”, the syllable /iəl/ included the final,

dark, /l/ which could not be dissociated from the preceding vowel; hence, the entire syllable was considered one segment.

Measurement of vowel duration was different depending on the phonetic context surrounding the target vowel. When preceded and followed by voiceless consonants, vowel duration was measured as the duration from onset to the offset of periodicity and upper formants. When preceded by a nasal consonant, the duration of the vowel was measured as the duration of the segment following the end of the drop in energy contour to the end of the periodicity. When followed by a nasal consonant, vocalic duration was measured as the duration from the onset of periodicity and formants to the offset of F2, in conjunction with initiation of rise in energy contour. As described above, in cases where the vowel was preceded or followed by glides and the vowel could not be segmented from continuant consonants with the help of F2 offset as the boundary indicator, the entire segment (vowel + consonants + vowel) was considered one acoustic segment. Specific onset and offset markers for each segment within each of the eight stimulus words can be noted with help of Table 4.1. This table includes the list of the eight words and their corresponding 7 types of acoustic segments.

Segmentation Problems

Certain vowels were difficult to segment from consonants, as indicated in Table 4.1. For “California”, /k^hæləfɔrnjə/, the underlined vowels could not be dissociated from the liquids /l/ and /r/ and hence the segments measured were /æ/, /ə/, and /or/ and /jə/. “Automobile”, /ɔ^hrəməbiəl/, was produced in the sentence “He makes automobile spare parts”. Occasionally, /ə/ could not be segmented from the preceding sound /s/, which some nonnative speakers realized as a voiced fricative, with periodicity

Table 4.1 List of 8 words and individual acoustic segments within the words

(Key: C.D= Closure Duration; VOT= Voice Onset Time)

Word		Type	of	Segme	nt					
Association <u>/əsoʃiəʃən/</u>	ə	S	o	s	ie	ʃ	ə	n		
Automobile <u>/ɔrəməbiəl/</u>	ɔ	r / C.D	ə	m	ə	C.D	b VOT	i	əl	
California <u>/kʰæləfɔrnjə/</u>	C. D	k ^h VOT	æ l	lə	f	o possibly	r one	n seg	j ə ment	
Committee <u>/kʰəmi:ri/</u>	C. D.	k ^h VOT	ə	m	ɪ	r / C.D	i			
						(for NN speakers if not flapped)				
Company <u>/kʰʌmpəni/</u>	C. D.	k ^h VOT	ʌ	m	C.D.	p VOT	ə	n	ɪ	
Economic <u>/ekənəmɪk/</u>	ɛ	C.D.	k VOT	ə	n	ə	m	ɪ	C.D.	k VOT
Political <u>/pʰəlɪtɪkəl/</u>	C. D.	p ^h VOT	əl ambi-s	l ɪ yllabic /l/	r / C.D	ɪ	C.D.	k VOT	ə l	
Representative <u>/reprəzɛntərɪv/</u>	r	ɛ	C.D.	p VOT	rə	z	e	n	r/C.D	(contd below)
	ə	r/C.D	ɪ	v						

that continued through the fricative to the onset of /ɔ/. Therefore the initial /ɔ/ was segmented from the preceding segment at the point of offset of high frequency striations of fricative noise. “Economic” /əkənómik/, was produced in the sentence, “Mexico’s economic situation slowly improved.” Again, periodicity continued from the phonetically voiced fricative in the word prior to the initial vowel in this target word. Therefore, the initial /ɛ/ was segmented from the preceding segment at the point of offset of high frequency striations of fricative noise, irrespective of the preceding periodicity. Moreover, the initial /ɛ/ was not used in comparisons of acoustic data across the groups. In cases when vowels were produced as diphthongs or as contiguous vowels, such as the ones underlined in the words “California” /jə/ and “association” /ie/, both the vowels were considered one unit. In segmentation of liquids, /l/ and /r/, the “steady-state” of preceding and final vocalic segments was not consistently steady, and the transitions were often not discernible; therefore the liquids were included with the vowels and measured as one acoustic segment, e.g., the [æɪ] in “Caliifornia”, the [iəɪ] in “automobile”, the [rɛ] and [rə] in representative”, and the [əl] in “political”. The final clusters were sometimes reduced and final syllables were omitted, so that in words such as “representative” and “association” many nonnative speakers partially produced the final syllable, (/reprəzentətɪ/), or omitted it, (/reprəzentə/); hence, the final syllable in “representative” was not used in comparisons of acoustic data between the speaker groups.

Results

Following the segmentation of words and the measurement of durations of each acoustic segment within the words, group differences in segmental durations were

compared and temporal deviations from AE productions of NN productions were tested for their relations to perceived accentedness. Only the nonnative speakers' scores were considered in the correlation analyses. Due to segmentation problems and/or missing segments in few NN speakers' data, not all correlations involved 22 NN speakers. The exact number of pairs considered in the correlation, where different from 22, is indicated in parentheses in the correlation analyses reported below.

Comparison of Acoustic Data of American English vs Nonnative Speakers and Correlation with Accentedness Ratings

Overall word duration. It was predicted that the overall duration of target words would be longer for NN speakers than for AE speakers, reflecting a slower production rate, which could have an effect on judgments of accentedness. As can be seen in Table 4.2, the NN speakers produced each of the seven words with longer durations, on average, than AE speakers⁵. Due to unequal group sizes (5 native speakers and 22 nonnative speakers respectively), and differences in variances, group differences were compared using *t*'-tests that corrected for unequal variances ("Welch *t*'-test"⁶ or "*t*'-test with separate variance estimate") for all temporal measures. The NN speakers' productions, averaged over all seven words, were significantly longer in duration than AE speakers' productions [$t(56)=-2.15, p<0.04$]. However, the differences between the group means for each word were significant in only 3 of the 7 comparisons (p-levels are shown in Table 4.2).

⁵ The word "representative" was not included in this analysis due to the aforementioned final-syllable segmentation problem.

⁶ Tomarken & Serlin (1986), Glass & Hopkins (1996)

The NN group showed greater variability across individuals in word duration than the AE group, as is indicated by the greater difference between longest and shortest duration tokens for all seven words (shown in parentheses in Table 4.2). However, the NN and AE ranges overlapped; that is, some NN speakers produced the words well within the duration ranges of AE speakers' productions. For all seven words, the upper end of the NN range exceeded the AE range, since some of the NN speakers spoke considerably slower than others. It was of interest to test whether this variability in word duration among the NN speakers correlated with accentedness ratings.

Table 4.2 Differences in overall word duration (in ms) for native (AE) and nonnative (NN) groups

Word Duration	AE Group			NN Group			p-values
	Average	Range across speakers	(Difference) across speakers	Average	Range across speakers	(Difference) across speakers	
Association	836	731-1012	(281)	861	591-1129	(538)	ns
Automobile	662	584-751	(167)	746	460-1019	(559)	ns
California	778	649-926	(277)	824	692-1239	(547)	ns
Committee	462	361-629	(268)	683	420-1760	(1340)	* (p<0.05)
Company	461	414-567	(153)	548	403-688	(285)	* (p<0.05)
Economic	534	471-709	(238)	596	379-939	(560)	ns
Political	535	483-607	(124)	660	509-943	(434)	* (p<0.01)
Representative		CND			CND		ns

CND = Could Not Determine

ns = not significant (p>.05)

To assess this, overall word durations were rank-ordered and correlated with ranked median ratings of accentedness on words, using a Spearman rank-order correlation. All words involved 22 NN speakers in the correlation, except for “automobile”(n = 20). Appendix F shows the scatterplots for each of the seven words, with word duration on the abscissa and accentedness ratings on the ordinate. Correlations of word duration and accentedness ratings ranged from very low to moderate ($\rho = +0.04$ to $+ 0.56$). No significant correlations of overall duration and accentedness were found ($p > 0.05$), except for the word “committee” ($\rho = +0.56$, $p < 0.01$). Thus, in most cases overall duration probably did not affect ratings of accentedness.

Since overall word duration was not found to correlate significantly with accentedness ratings, the individual acoustic segments within the word were examined for specific differences in temporal structure that were predictive of perceived accentedness. The acoustic segments examined for duration differences were: vocalic segment duration (as a percentage of the whole word), VOT duration for voiceless aspirated initial stops /p^h, k^h/, and flap/closure duration for intervocalic /t/. Since the target words were multisyllabic, it was expected that lexical stress would be deviant for NN speakers’ more accented productions of the words. Therefore relative duration of stressed and unstressed vowels, as a measure of lexical stress, was predicted to vary across the groups. To examine this, ratios of stressed and unstressed vowel durations (S/U ratio) were computed. All of the above-mentioned acoustic measures and the S/U ratios were each then correlated with perceived accentedness of the word.

Vocalic segment duration. For each word, the duration of the vocalic segments was measured (ms) and the averages of corresponding vocalic segments for the native and nonnative speaker groups were compared. While the absolute durations were expected to vary with overall speaking rate, the relative durations of segments were expected to be similar across speakers, at least within the AE group. Relative vocalic segment durations were computed as the percent of total word duration ($\{ \text{vocalic segment duration} / \text{whole word duration} \} \times 100$). The average relative durations for the AE speakers were then used as the baseline with which to compare the NN speakers' productions.

Table 4.3 shows the descriptive statistics for relative vocalic durations for the two speaker groups for each of seven⁷ of the target words. The vocalic segments of each target word have been categorized as stressed "S" (primary or secondary lexical stress) or unstressed "U" (reduced to schwa). Averages (rounded to nearest whole percentage), standard deviations, and ranges in relative durations for each individual vowel were compared across the two groups.

As can be noted from Table 4.3, 10 of the 12 unstressed vowels were, on average, proportionally longer for the NN group than the native group. The t'-tests revealed a significant effect of speaker group for 8 of the 12 unstressed vowels (p-values are shown in Table 4.3). That is, NN speakers, as a group, failed to produce reduced schwa vowels in syllables that AE speakers produced as unstressed. (Conversely, only 3 of the 11 stressed vowels were significantly shorter for the NN group than the corresponding AE vowels). However, the ranges in duration of unstressed vowels overlapped across the groups, with greater variability within the NN group (standard deviation ranged from 1.87

⁷ "Political" was not included in this analysis since the respective vocalic segments could not be reliably isolated from the surrounding consonants.

Table 4.3 Vocalic Segmental Durations (%) of total word duration for Native (AE) and Nonnative (NN) groups

Word	Vowel segment	AE Group			NN Group			p-value (if significant)
		Average	SD	Range	Average	SD	Range	
Association (v4 not computed)	v1(U)	5	1.3	3-6	10	5.1	3-20	*(p<0.01)
	v2(S)	14	2.1	12-16	6	2.1	3-10	*(p<0.01)
	v3(S)	24	4.8	21-31	21	4.7	15-30	ns
Automobile (v4 not computed)	v1(S)	19	3.3	15-23	19	8.0	7-32	ns
	v2(U)	6	1.9	4-9	9	7.3	5-32	*(p<0.05)
	v3(U)	13	2.4	11-16	17	5.0	8-25	*(p<0.05)
California	v1(S)	15	3.2	13-19	11	2.5	8-16	*(p<0.05)
	v2(U)	11	2.1	8-13	13	1.9	9-16	ns
	v3(S)	19	1.5	17-21	18	2.6	13-24	ns
	v4(U)	14	2.7	12-18	19	3.7	10-25	*(p<0.05)
Committee (v3 not computed)	v1(U)	6	1.5	5-8	11	4.8	3-23	*(p<0.01)
	v2(S)	19	2.8	16-22	18	8.0	8-31	ns
Company	v1(S)	19	3.1	15-23	19	3.8	13-28	ns
	v2(U)	7	2.8	3-11	13	3.1	6-18	*(p<0.01)
	v3(U)	23	4.9	19-31	19	5.4	8-31	ns
Economic	v1(S)	14	1.9	13-17	11	3.0	0-19	ns
	v2(U)	8	1.8	4-8	14	4.3	7-13	*(p<0.01)
	v3(S)	27	2.4	24-30	14	6.9	6-30	*(p<0.01)
	v4(U)	8	2.1	6-10	13	4.0	2-32	*(p<0.05)
Representative (v5 not computed)	v1(S)	12	5.4	6-20	14	4.6	5-22	ns
	v2(U)	7	4.0	1-11	7	2.3	4-14	ns
	v3(S)	11	4.1	4-15	10	3.5	5-20	ns
	v4(U)	7	2.1	3-9	8	4.4	3-21	ns

ns = not significant (p>.05) (Political was not included, see footnote no.6)

to 8.00 for NN group, and 1.26 to 5.42 for AE group). In light of consistent group differences for unstressed vowels and the variability in the NN speakers' productions, it was of interest to study vowel differences with respect to lexical stress. Hence group differences in the ratio of stressed and unstressed vowels (S/U ratio), and the relation of S/U ratio to perceived accentedness were examined.

Stressed/Unstressed (S/U) vowel ratio. The ratio of the stressed to the unstressed vowels (S/U ratio) in each of the seven measurable words was computed as a measure of lexical stress. The vowels used in computing the S/U ratio are shown in Table 4.4. In all S/U ratios, the numerator is either a stressed vowel, or an average of stressed vowels, and the denominator is either an unstressed vowel or an average of unstressed vowels. For "company," /k^hámpəni/, only the first two vowels were analyzed because the status of the final /i/ in the open syllable is controversial. Although /i/ is not considered to have either primary or secondary lexical stress, it is not reduced to schwa in AE pronunciation. For "California," /k^hæləfórnjə/, all four vocalic segments were included to compute the S/U ratio. (Recall that, the glides /ɪ/ and /ɪ/ were also included in the calculation of the vocalic segments, since they could not be reliably segmented from the vowels). In "automobile," /ɔ́rəməbiə̀l/, the final unstressed vowel /i/ was not considered in estimating the S/U ratio because it was diphthongized, and also because it could not be reliably segmented from the following liquid /l/. In "association," /əsòsiéʃən/, the vowel sequence, /ie/ was treated as a single acoustic segment in estimating the S/U ratio. Also, the final /ə/ was not included in computing the S/U ratio as it could not be segmented from the following nasal /n/. In "committee," /k^həmítri/, the final /i/ was not considered in the S/U ratio for the same reason as explained above for "company." In "representative," /rèprəzénrətiv/,

the final vowel /ɪ/ was not included since several NN speakers deleted that syllable entirely. In "economic," /ɛˈkənəmɪk/, the first vowel /ɛ/ was not considered in the ratio since it presented problems of segmentation from the preceding word (as described earlier).

Table 4.5 shows averages, standard deviations, and ranges of the S/U ratios for the two groups. For six of the seven words included in the analyses, the AE group produced larger S/U ratios than the NN group. As predicted, the native speakers made a temporal distinction between stressed and unstressed vowels, with ratios greater than 2 except for "California" (S/U ratio = 1.53) and "representative" (S/U ratio = 1.55). The AE groups also showed relatively low variability (standard deviations ranging from 0.25 to 1.97). In contrast, for the NN group, average ratios varied from less than 1.00 to 2.05. As the ranges in ratios for NN production show, the temporal distinction between stressed and unstressed vowels was not produced by some of the NN speakers, or was indeed reversed in some cases (ratio < 1.0). The *t*'-tests of group differences showed that AE and NN groups differed significantly in S/U ratio for five of the seven words (*p*-values are shown in Table 4.5).

Variability across speakers in the NN group was about the same as for the AE group, with standard deviations ranging from 0.20 to 1.53. The distributions of S/U ratios for the two groups showed overlap of ratios for some words ("California" and "automobile"), while distributions were more distinct for other words ("economic", "association") with only a few NN speakers producing ratios in the native range. With this variability in the NN group, and significant differences from the AE group, it was

Table 4.4 Vowels used in estimating Stressed to Unstressed (S/U) Vowel Ratio

Word	S/U ratio	Specific vowels
Association /əsoʊsɪˈʃən/	$(v_2+v_3)/2 : v_1$	$(o+ie)/2 : ə$
Automobile /ˈɔːrəməbiəl/	$v_1 : (v_3+v_2)/2$	$ɔ : (ə+ə)/2$
California /kəˈlæfɔːrnjə/	$(v_1+v_3)/2 : (v_2+v_4)/2$	$(æ+or)/2 : (lə+jə)/2$
Committee /kəˈmɪti/	$v_2 : v_1$	$ɪ : ə$
Company /kəmˈpəni/	$v_1 : v_2$	$ʌ : ə$
Economic /ɪˈkɒnəmɪk/	$v_3 : (v_2+v_4)/2$	$ɑ : (ə+i)/2$
Political /pəˈlɪtɪkəl/	$v_2 : (v_1+v_3)/2$	$ɪ : (ə+i)/2$
Representative /ˌreprəzənˈteɪv/	$(v_1+v_3)/2 : (v_2+v_4)/2$	$(re+ze) /2 : (rə+rə)/2$

Table 4.5 Stressed to Unstressed Vowel Ratios (S/U) for Native (AE) and Nonnative (NN) Groups

Word	AE Group			NN Group			p-value
	Average	SD	Range	Average	SD	Range	
Association	4.65	1.97	2.75-7.99	1.79	1.17	0.64-4.85	*p<0.05
Automobile	2.00	0.48	1.46-2.75	1.54	0.67	0.48-2.47	ns
California	1.53	0.32	1.08-1.88	0.93	0.20	0.61-1.40	*p<0.05
Committee	3.15	0.87	2.15-4.44	2.05	1.53	0.41-6.34	*p<0.05
Company	3.16	1.13	2.15-5.10	1.61	0.68	0.85-3.69	*p<0.05
Economic	3.64	0.80	3.02-4.82	1.36	0.92	0.21-2.58	*p<0.01
Political		CND			CND		
Representative	1.55	0.25	1.14-1.77	1.79	0.72	0.73-3.39	ns

CND= Could Not Determine

ns = not significant (p>.05)

appropriate to study whether differences in S/U ratio correlated with perceived accentedness. Spearman rank order correlations were performed between ranked S/U ratios and ranked median ratings of accentedness for each of the seven words analyzed. Each word involved 22 pairs in the correlation analysis, except “automobile” (n=20), and “representative” (n=21); pairs were missing in these because of measurement problems. The S/U ratio was ranked from high to low (e.g., highest ratio = lowest rank, 1), and median accentedness ratings were ranked from low to high (e.g., least accented = lowest rank). Thus a positive correlation would indicate that more deviant (lower) S/U ratios were associated with the perception of greater accentedness.

Table 4.6 Coefficients and significance levels for correlation of S/U ratio and accentedness ratings

Word	Correlation Coefficient (ρ)	p-level (if significant)
Association	+0.35	ns
Automobile	-0.01	ns
California	+0.15	ns
Committee	+0.09	ns
Company	+0.02	ns
Economic	+0.80	* $p < 0.001$
Political	CND	
Representative	+0.11	ns

CND = Could Not Determine
 ns = not significant ($p > .05$)

As can be seen in Table 4.6, correlations of S/U ratios with accentedness ratings yielded very low coefficients ($\rho = -0.01$ to $+0.35$) for six of the seven words; while the relation was highly significant for “economic” ($\rho = +0.80$; $p < 0.001$)⁸. The scatterplots of the correlations of S/U ratio with accentedness ratings can be seen in Appendix G. It was concluded that lexical stress, as measured by the S/U vowel ratio, varied systematically across groups, with many NN speakers failing to reduce vowels in syllables that AE speakers produced as unstressed. However, the nonnative group showed overlap with the native group, i.e. some NN speakers produced native like lexical stress patterns. Nonnative stress patterns did not correlate strongly with accentedness ratings of the whole word for most of the target words. Since S/U ratio correlated only weakly with perceived accentedness, deviations in the temporal pattern of consonantal productions (VOT and flap durations) were examined for group differences and correlations with perceived accentedness.

Voice Onset Time (VOT). The word-initial, voiceless, aspirated stops [k^h , p^h] in the words “company”, “committee”, “California”, and “political” were used to compare AE and NN productions of VOT. Table 4.7 shows the summary data for the two groups. As was predicted from previous studies, the averages show consistent group differences, with NN speakers producing shorter VOT (less aspiration) for all four stops. The t'-tests showed that these group differences were reliable for three out of four words (all containing [k^h] stops); p-levels are shown in Table 4.7.

The NN group showed larger standard deviations than the AE group for three of the four words and the ranges indicate some overlap across the groups in all four cases

⁸ (Note that “economic” also yielded a more robust significant difference ($p < 0.01$) in means of S/U ratio compared to the other words ($p < 0.05$)).

Several of the NN speakers fell within the AE range (9 for "committee", 13 for "company", 1 for "California" and 9 for "political"). The extent to which this variability in VOT in the NN group was correlated with perceived accentedness ratings was therefore assessed.

Table 4.7 Word-Initial Stop VOT (ms) for Native (AE) and Nonnative (NN) groups

Word	AE Group			NN Group			p-value
	Average	SD	Range	Average	SD	Range	
California /k ^h /	67	4.5	61.6-71.3	32	11.6	16.1-66.1	*p< 0.01
Committee /k ^h /	58	6.2	50.9-67.1	47	19.2	22.1-96.2	*p< 0.05
Company /k ^h /	50	13.7	31.5-63.6	32	11.2	19.3-61.4	*p< 0.05
Political /p ^h /	51	16.8	26.6-68.6	34	26.0	11.7-105.5	ns

ns = not significant (p>.05)

Spearman rank order correlations were computed for ranked VOT and ranked accentedness ratings. VOT durations were ranked from high to low (e.g. longest duration = lowest rank, 1) and median accentedness ratings were ranked from low to high (least accented to most accented). Appendix H shows the scatterplots with VOT on the abscissa and accentedness ratings on the ordinate. As shown in Table 4.8, correlation coefficients ranged from -0.01 to +0.36; none was statistically significant (p >.05).

It can be concluded that VOT for word-initial [k^h] was, on average, significantly shorter for the NN group than the AE group. NN speakers showed higher variability than AE speakers, with several NN speakers producing VOT values in the AE range, while

others produced /k/ with little aspiration, more appropriate for Spanish [k̟]. Lack of significant correlations suggests that deviation from native-like production of voiceless aspirated stops by NN speakers, contrary to expectations, was not reliably associated with perceived accentedness of these multisyllabic words. Closure duration for flaps/medial stops was examined next for temporal differences and correlation with perceived accentedness.

Table 4.8 Coefficients and significance levels of correlation of word-initial stop VOT and accentedness ratings

Word	Correlation Coefficient (ρ)	p-level
California /k ^h /	+0.36	ns
Committee /k ^h /	+0.29	ns
Company /k ^h /	+0.26	ns
Political /p ^h /	-0.01	ns

ns = not significant ($p > .05$)

Closure duration for medial /t/. Closure durations of 5 intervocalic /t/s in positions where a flap is expected in AE were considered: "committee", "automobile", "political", and two in "representative" (target segment underlined). The two flaps in "representative" are henceforth differentiated as "t₁" and "t₂". Table 4.9 presents the summary data (averages, standard deviations, and ranges) for the two groups.

Flap/closure durations showed consistent average group differences, with NN speakers producing longer closure duration (i.e. stops rather than flaps) in all five cases.

The t' -tests showed that these differences were significant for four of the five realizations (p-levels are shown in Table 4.9). The closure durations of the NN group were more variable than those of the AE group in all five cases (NN standard deviations ranged from 12.82 to 38.16, whereas AE standard deviations ranged from 1.82 to 8.43). Moreover, the distributions of flap/closure durations overlapped for the groups. Several NN speakers produced flap/closure durations in the AE range in each of the 5 cases: 3 NN speakers in "political," 11 in "automobile," 7 in "committee," 10 in t_1 , and 7 for t_2 in "representative." This variability in the production of intervocalic /t/ in the NN group was tested for its correlation with perceived accentedness ratings of these words.

Table 4.9 Closure Duration of Intervocalic /t/ (ms) for Native (AE) and Nonnative (NN) groups

Word	AE Group			NN Group			p-values (if significant)
	Average	SD	Range	Average	SD	Range	
Automobile /t/	16	5.2	9.70-22.64	38	38.2	7.35-169.53	* $p < 0.05$
Committee /t/	18	5.6	10.46-24.00	39	23.2	11.70-93.03	* $p < 0.01$
Political /t/	12	1.8	9.46-13.85	32	15.8	10.75-67.03	* $p < 0.01$
Representative /t ₁ /	26	8.4	13.47-32.41	36	12.8	14.54-60.33	ns
Representative /t ₂ /	28	8.1	19.55-39.12	68	34.2	19.48-141.78	* $p < 0.01$

ns = not significant ($p > .05$)

Spearman rank-order correlations were computed for ranked flap/closure durations (in ms) with ranked medians of accentedness. The number of speakers in the

correlation for each word varied: “automobile” (n=22), “committee” (n=21), “political” (n=20), “representative /t₁” (n=20), and “representative /t₂” (n=21). The flap/closure durations were ranked from low to high (e.g. shortest duration = lowest rank, 1), and median accentedness ratings were ranked from low to high (least to most accented). Appendix I shows the scatterplots of the correlations with flap/closure duration on the abscissa and accentedness on the ordinate. As can be seen in Table 4.10, correlation coefficients ranged from +0.29 to +0.59. Three of the five correlations (“political,” “automobile” and t₁ in “representative”) were statistically significant (p-levels are shown in Table 4.10).

Table 4.10 Coefficients and significance levels of correlation of Flap/Closure Duration and accentedness ratings

Word	Correlation Coefficient (ρ)	p-level
Automobile	+0.46	* p<0.05
Committee	+0.35	ns
Political	+0.59	* p<0.01
Representative (t ₁)	+0.58	* p<0.01
Representative (t ₂)	+0.29	ns

ns = not significant (p>.05)

It can be concluded that flap/closure duration was, on average, longer for the NN group than for the AE group and that NN speakers showed greater variability than the AE speakers. Several NN speakers produced flap durations within the AE range, while others

produced stops rather than flaps. Significant correlations for 3 of the 5 flap/closure durations suggest that differences in flap/closure durations in the NN speakers may have affected perceived accentedness of these words.

Summary and Discussion

Group differences were found in overall word duration, with NN speakers' being slower for all eight words; average differences were significant for three of these eight words. Relative vowel duration also varied between the two groups, as NN speakers produced unstressed vowels with longer durations, on average, with significant groups differences for eight of the twelve unstressed vowels. S/U ratios were smaller in the NN group, on average, for six of the seven words compared with significant group differences for five of these seven words. NN speakers, on average, produced smaller VOT values compared to the AE speakers, with significant group differences for initial [k^h] but not initial [p^h]. Closure duration for intervocalic /t/ was, on average, longer for the NN group than the AE group in all cases, with four of these five segments resulting in significant differences.

Considerable variability was found in temporal parameters of multisyllabic words produced by NN speakers, with some speakers falling within the native AE speakers' baseline ranges while others showed significant deviations from native production. Correlations with perceived accentedness of the target words are summarized in Table 4.11. Overall word duration showed low to moderate correlations with perceived accentedness, although only one word ("committee") yielded a statistically reliable relation. Stressed to unstressed vowel ratio, which was assumed to be a correlate of lexical stress, also correlated positively with perceived accentedness for all but one word,

although, again, only the correlation for one word ("economic") was significant.

Deviations in temporal parameters of stops (voiceless aspirated /p^h, k^h/ and flapped /t/) also correlated positively with accentedness (with the exception of a very low negative correlation of VOT and accentedness in "political"). While none of the VOT-accentedness correlations was statistically reliable, three of the five correlations of flapped /t/ and accentedness were highly significant.

Table 4.11 Summary table of all correlations of acoustic parameters and accentedness ratings

	Whole word duration	S/U ratio	VOT	Flap/ Closure duration
word	(ρ)	(ρ)	(ρ)	(ρ)
Association	0.08	-0.01		
Automobile	0.04	0.09		0.46*
California	0.36	0.15	0.36	
Committee	0.56**	0.35	0.29	0.35
Company	0.23	0.02	0.26	
Economic	0.23	0.80**		
Political	0.39		-0.01	0.59**
Representative	CND	0.11		t ₁ = 0.58**
				t ₂ = 0.29

CND = Could Not Determine

* p <.05

** p <.01

Findings from this study indicate that segmental durations in multisyllabic words differed for nonnative speakers' productions when compared to those of native speakers. However, overall group differences were small, and correlations with perceived accentedness were weak for each of the measured variables. It is hypothesized that, while each of the measures alone does not strongly predict accentedness, some combination of these temporal deviations may account, at least in part, for native listeners' judgments of perceived accentedness.

CHAPTER V

General Discussion

The purpose of this study was to examine temporal speech characteristics of native Spanish speakers of English as an L2 that relate to native English listeners' perception of accentedness. Previous studies have typically described nonnative speech in terms of phonological characteristics of nonnative productions, with only a few studies addressing the acoustic characteristics of those productions. These acoustic studies were mainly restricted to studying single parameters of accented English, such as vowel duration differences preceding final voiced/voiceless stops, acoustic vowel spaces, VOT of voiceless stops, change in f_0 as related to intonation, or some measure of overall rate of speech. Multiple acoustic parameters as they relate to perceived accentedness have not yet been studied in a correlational design.

The present study analyzed several temporal characteristics of multisyllabic words as they related to a measure of perceived accentedness. Temporal measures were selected based on predictions made from previous phonological descriptions of Spanish-accented speech. Duration differences were expected to underlie several features of interlanguage phonology of Spanish-accented speakers (Macdonald, 1989; Ortega-Llebaria, 1997), e.g., lack of differentiation of long vs. short vowels (/i/ vs. /ɪ/) and initial voiceless stops perceived as voiced. Lexical stress was also expected to be deviant in the production of Spanish-accented multisyllabic words (Mairs, 1989). Predictions of duration differences in NN vs AE productions included 1) overall word duration of multisyllabic words; 2) stressed to unstressed vowel ratio; 3) voiceless stop VOT; and 4) flap/closure duration of intervocalic /t/ (realized as a flap in AE productions.) Deviations from native temporal patterns in the productions of nonnative speakers were expected to be related to native

English listeners' perception of foreign accentedness. Specifically, the two hypotheses tested in this dissertation were: 1) The temporal structure of Spanish-accented speakers' productions of multisyllabic words will differ from that of native American-English speakers, and 2) Temporal deviations in nonnative speakers' productions will be correlated with native AE listeners' perception of accentedness of the multisyllabic words.

Two studies were designed in order to test the two hypotheses. Study 1 involved establishing global accentedness ratings of a group of NN speakers' productions of short sentences containing multisyllabic target words, using a task in which native AE listeners judged NN sentences on a 9-point rating scale. Then the same judges rated the multisyllabic words, excised from their sentence contexts. Results showed that perceived accentedness ratings of the multisyllabic words were, in general, good predictors of global accentedness, as overall correlation of words and sentences (pooled over listeners and stimulus words) yielded a strong relationship ($\rho=+0.82$). Therefore, it was concluded that multisyllabic words had properties that were representative of overall sentence accentedness.

The strength of individual correlations of each of the 8 words with the corresponding sentences varied, with correlation coefficients ranging from $\rho = +0.50$ to $\rho = +0.82$. This variability across words suggested that some words were better predictors of global accentedness than others. These differences in perceived accentedness across words suggested a related difference in underlying acoustic properties of the words. The second study was designed to assess the temporal differences in NN productions of these

words, and to test whether the degree of temporal deviation from native patterns correlated with native English listeners' perception of accentedness of the words.

Study 2 compared differences across the native and nonnative groups for each of the following acoustic variables: overall word duration, relative vowel durations as well as ratios of relative durations of stressed and unstressed vowels, initial voiceless stop VOT, and intervocalic /t/ closure duration. Average differences in temporal patterns were found across AE and NN groups, as would be predicted from phonological analyses of Spanish-accented English. Overall word duration was, on average, longer for NN speakers than AE speakers, for all eight words, implying a reduced rate of speech in the NN group. Unstressed vowels were longer (relative to total word duration) for NN speakers, indicating that the Spanish speakers did not reduce unstressed vowels. The S/U vowel ratio of the AE speakers was considerably greater than 1.0 (1.53 to 4.65) for all eight words, indicating the temporal reduction of unstressed vowels in a stress-timed language. NN speakers, conversely, showed ratios much closer to 1.0 (0.93 to 2.05), indicating the minimal temporal distinction between stressed and unstressed vowels found in Spanish, a syllable-timed pattern of language. The difference was due to the failure of some NN speakers to reduce the duration of lexically unstressed vowels according to AE phonological rules.

As predicted, there were also differences in the temporal characteristics of consonants. The VOT of voiceless, initial stops was, on average, consistently shorter for the NN speakers compared to the AE speakers, a pattern expected from phonological transfer from Spanish, which has unaspirated stops. In addition, the durations of intervocalic /t/, realized as a flap in AE productions, was, on average, longer in NN than

AE productions, also conforming to previous phonological descriptions of stops being substituted for flaps in Spanish interlanguage phonology.

In summary, the analysis of segment durations indicated that NN speakers, as a group, produced the multisyllabic words with temporal deviations from AE speakers' productions that were predictable from phonological analyses. Both vowels and stop consonants showed temporal differences. Individual NN speakers, however, varied considerably in the extent to which segment durations differed from AE norms, with some speakers' productions having durations approximating native productions, and others being distinctly deviant. Therefore, the extent to which this variability in NN speakers' segmental durations influenced native listeners' perception of accentedness was tested. Correlations of each of these temporal variables with accentedness ratings on words from Study 1 were low to moderate. Differences in overall word duration yielded low to moderate correlations with perceived accentedness (ρ ranged from +0.04 to +0.56). S/U ratios correlated only weakly with native perception of accentedness ($\rho = -0.01$ to +0.35) except for the word, "economic" whose S/U ratio was highly correlated with accentedness ($\rho = +0.80$). VOT values for nonnative productions of initial /k/ were moderately correlated with accentedness ratings in three words ($\rho = +0.26$ to +0.36); initial /p/ VOT values were not related to accentedness in "political" ($\rho = -0.01$). Nonnative intervocalic /t/ closure durations showed moderate correlations with perceived accentedness in all five words ($\rho = +0.29$ to +0.58).

It can be concluded on the basis of the present data that temporal deviations in acoustic segments may indeed be related to perceived accentedness of NN speakers' productions of multisyllabic words. However, each of the temporal parameters alone did

not relate strongly to perceived accentedness of the whole word. This suggests that a combination of deviations in some, or all of the temporal parameters, would yield stronger correlations with native listener ratings of accentedness. This study, however, was not designed to address this issue directly, since the eight target words did not vary systematically in each of the variables under study. Furthermore, multiple regression analyses using nonparametric statistics are not available. Future research will need to address this prediction by way of a direct experimental study involving editing NN durations in the multisyllabic words, and making pre-and post-editing comparisons of native listeners' ratings of accentedness (e.g. Magen, 1998). Variables can be edited one at a time, and different combinations of segmental editing can be presented to listeners to test which combination affects listeners the most. A study of this nature can address questions about which combinations of durational segments determine listener ratings of accentedness.

This study was concerned only with how the temporal structure (segmental durations) deviated in NN speakers' productions of the target words. Based on previous phonological analyses of Spanish-accented English (Macdonald, 1989; Ortega-Llebaria, 1997), predictions can also be made about the role of spectral differences in nonnative acoustic segments on perceived accentedness. Specifically, differences in vowel quality and lexical stress in multisyllabic words are cued by spectral characteristics, in addition to the durational variables measured here. Therefore, future research also needs to address whether spectral differences are related to perceived accentedness. Additionally, other suprasegmental components, such as fundamental frequency contour and relative syllable intensity, may also influence perceived accentedness of multisyllabic words.

Therefore, these suprasegmental aspects also need to be analyzed in future correlation studies, in order to directly estimate their relationship to perceived accentedness.

The present research has made significant contributions to the existing literature on foreign-accented-English, in general, and Spanish-accented English, in particular. Multiple acoustic variables were studied as they affected perception of accentedness. Few studies have tested the direct influence of production variables on perception, as typically, most studies measured acoustic differences in NN speech, and made *inferences* about their contribution to native listener perception. The present study examined the correlation between production deviations and perceived accentedness. The findings from this project lend support to the existing phonological studies on Spanish-accented English. That is, measurements resulted in quantitative verification of the acoustic features which may underlie previously-identified phonological deviations of Spanish-accented speakers of English, such as flaps replaced by stops, voiceless stops perceived as voiced due to reduced VOT and aspiration, continuation of phonation in voiceless consonants such as stops and fricatives resulting in them being perceived as voiced.

The contributions of the present research have important implications for speech-language pathologists and English-as-Second-Language (ESL) teachers as well. Temporal differences in NN speakers were found to be correlated with perceived accentedness; hence, this information can be used in ESL and accent-modification to help correct misconceptions in such programs. As mentioned earlier, many accent-modification programs focus on speech factors that affect *intelligibility* and do not distinguish speech-related variables that may affect perception of *accentedness*. The temporal differences identified here are in direct relationship with perceived

accentedness, and can serve as the starting-point in the identification of variables to “modify” in accent-modification programs. Accent-modification programs can focus on teaching NN speakers to reduce the durational differences in target productions of voiceless stops, to reduce durations of medial stops in order to produce native-like productions of flaps, and to produce a durational distinction between stressed and unstressed vowels so that they may be perceived as more (AE) native-like in their speech.

In conclusion, temporal differences in acoustic segments of native Spanish speakers of English are correlated, though low to moderately, with native listeners’ perception of accentedness on multisyllabic words. Future research can add to the present findings of acoustic parameters as they relate to perceived accentedness.

APPENDICES

Appendix A: List of Original 98 Target Words in Sentences

**** Indicates the eight sentences in final selection**
***Indicates the eight sentences that were identified with stress and/or vowel problems but which had target words that were difficult to segment**

Word	Frequency of occurrence	Sentence
Between	729	I placed the TV between the table and the chair.
Number	658	My lucky number is seven.
Himself	601	He was talking to himself in the subway.
Against	625	I was leaning against the desk.
Program	529	The program of the conference was rather long.
Another	690	Tomorrow is another day.
Government	495	The present government is no good.
Something	454	I have something new for you.
** Company	453	The telephone company closed down.
President	424	Bill Clinton is the president of United States.
Business	412	I like doing business with him.
Development	377	There is a new development in the situation.
General	378	You can leave a message in the general mail box .
Experience	306	He has no experience in this kind of work.
Probably	263	You're probably right.
* Community	275	I live right next to the community hospital.
Department	272	The History department is on the first level.

Information	269	I'd like some more information in this matter.
Remember	250	I can never remember to carry an umbrella.
Available	246	Do you have any tickets available for tonight's show?.
** Economic	243	Mexico's economic condition is slowly improving.
Understand	240	I fail to understand how this works.
* Material	269	Read the material carefully.
Activity	231	The Student Activity building is right next door.
Difference	228	There is no difference between the two dresses.
Necessary	222	It is necessary to have a visa to go to Puerto Rico.
Washington	217	George Washington Bridge is better than Lincoln Tunnel.
Education	215	I think education is very important.
Secretary	210	Mr. Smith's secretary called yesterday.
Statement	209	My bank statement does not seem right.
Operation	197	The simple operation took a long time.
University	256	I go to the City University of New York.
** Political	258	Everyone is very political in this company.
Personal	196	Let me ask you a personal question.
Organization	192	That is the oldest organization on the east coast.
**Committee	188	We formed a committee to solve this problem.
Opportunity	172	This was the best opportunity I ever got.
Administration	169	The Clinton administration helped the country.
**Association	168	I did the project in association with Sears.

*Recognize	163	It is difficult to recognize any of these pictures.
Equipment	167	The film equipment is very expensive.
Including	166	I will be including you in my plans.
Individual	165	Each and every individual has a special role.
Temperature	161	I can't even adjust the temperature on this oven.
Performance	155	Tonight is the last performance of the season.
Production	155	Don't make a big production of this paper.
International	154	He does a lot of international travel.
Represent	163	Garcia will represent the Hispanic population.
Military	199	I served in the military last year.
Responsibility	143	No one will take responsibility for their mistake.
Industrial	143	My father works in the industrial complex.
Procedure	140	The doctor performed the procedure quickly.
Authority	135	She is an authority on this topic.
*Apparently	124	He had apparently broken up with his girlfriend.
Performance	155	The last performance is at 8 o' clock.
Possibility	130	There is a possibility you'll do well.
Technique	159	The new technique impressed them.
*Literature	134	She enjoyed the literature class.
Corporation	131	He joined the corporation at 25.
Relationship	126	He ended the relationship abruptly.
Discussion	126	This is not up for discussion anymore.

Immediately	116	These people were immediately contacted after the accident.
Particularly	106	I don't particularly care for high heels.
*Conference	122	I'm presenting at a conference next week.
Experiment	122	We did this experiment in the chem. lab.
Additional	120	I need two additional reference letters.
*Governor	118	The governor of NY is Pataki.
Institution	140	Peggy works for the institution in Boston.
Circumstance	99	It was a difficult circumstance for us.
Professional	99	That is not a professional attitude.
Requirement	110	There is only one requirement for the class.
Construction	98	The building construction began this year.
Communication	95	Keep the doors of communication open.
Considerable	96	It took considerable time to finish this job.
*Atmosphere	84	I enjoyed the atmosphere in the restaurant.
**California	88	Her sister will move to California in March.
Philosophy	88	His personal philosophy is "Live and let live".
Measurement	88	They took the measurement to build the shelves.
**Representative	85	The congressional representative gave his vote.
European	59	The European culture is very rich.
Experience	82	In my experience this will take a long time.
Appearance	71	He made his final appearance in the show.
Understanding	83	She has a good understanding of this issue.
Manufacturer	77	The auto manufacturer makes car parts.

Characteristic	76	Maria's most pleasant characteristic is her smile.
Neighborhood	75	I live in a nice neighborhood of Brooklyn.
Consideration	81	Please give enough consideration to this matter.
**Automobile	74	He makes automobile spare parts.
Interpretation	66	Her interpretation was different from mine.
Preparation	69	The preparation time for this dish is ten minutes.
Significance	66	The significance of the event occurred to him much later.
Personality	63	I did several personality tests.
Competition	63	There is no competition between us.
Phenomenon	62	Scientists have observed this phenomenon several times.
Intellectual	57	I like the company of intellectual people.
Restaurant	53	We had dinner in an Italian restaurant yesterday.
Probability	55	There is no probability of her ever dating him.

Appendix B: Language Background Questionnaire

1. **Name:**
2. **Age:**
3. **Gender:**
4. ***Profession:**
5. **Languages spoken and modality of each:**
 - Read only
 - Read, write, and speak:
6. ***First exposure to English (in USA or native country): (No native exposure preferred in own country)**
7. **Age of first learning English : (Preferred to be after 12 years or age)**
8. **Place of learning English (home, school, etc):**
9. **Self-rating of fluency of English (and *Spanish) :**
10. **Home country and region:**
11. **Dialect of Spanish:**
12. **Age of arrival to US: (12 or 18 years or more)**
13. **Length of residence in US: (explain if discrepant with #12):**
14. **Current use of English:(Exposure amount during the day, people who speak in English-preferable if English outside and Spanish at home)**
15. **Duration of stay in New York:**
16. **All other places resided in within US**
17. ***Time spent in any other English-dominant country: (preferable less than 3 months)**

listener for individual words

Table A: Listener 1

Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	5	1--7	5	1--7
Association	5	1--8	4	1--8
Committee	6	1--8	5	1--8
Company	5	1--7	4	1--6
Economic	6	1--8	5	1--7
Automobile	5	1--8	6	1--9
Political	4	1--6	5	1--7
California	4	1--7	5	1--7

Table B: Listener 2

Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--9	5	1--8
Association	6	1--8	5	1--8
Committee	6	1--9	6	1--8
Company	6	1--7	5	1--6
Economic	6	1--9	6	1--8
Automobile	6	1--8	5	1--8
Political	5	1--8	5	1--8
California	6	1--7	5	1--7

Table C: Listener 3

Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	7	1--9	7	1--9
Association	8	1--9	6	1--9
Committee	7	1--9	7	1--9
Company	6	1--8	4	1--7
Economic	9	1--9	5	1--8
Automobile	8	1--9	6	1--9
Political	6	1--9	5	1--8
California	6	1--9	6	1--8

Table D: Listener 4

Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	7	1--9	6	1--9
Association	7	1--9	6	1--9
Committee	6	1--9	7	1--9
Company	5	1--7	6	1--8
Economic	7	1--9	6	1--9
Automobile	6	1--9	6	1--9
Political	5	1--8	7	1--8
California	7	1--9	6	1--8

Table E: Listener 5				
Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	8	1--9	6	1--9
Association	8	1--9	6	1--8
Committee	8	1--9	7	1--9
Company	7	1--9	7	1--9
Economic	8	1--9	6	1--9
Automobile	8	1--9	8	1--9
Political	6	1--9	7	1--9
California	8	1--9	6	1--9

Table F: Listener 6				
Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--8	7	1--9
Association	6	1--9	6	1--9
Committee	6	1--9	7	1--9
Company	6	1--8	7	1--9
Economic	6	1--9	7	1--9
Automobile	5	1--9	7	1--9
Political	6	1--9	6	1--9
California	5	1--9	7	1--9

Table G: Listener 7

Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--8	5	1--8
Association	6	1--8	6	1--7
Committee	6	1--8	6	1--8
Company	5	1--7	6	1--8
Economic	6	1--8	6	1--7
Automobile	6	1--8	6	1--8
Political	5	1--8	5	1--8
California	6	1--7	6	1--8

Table H: Listener 8

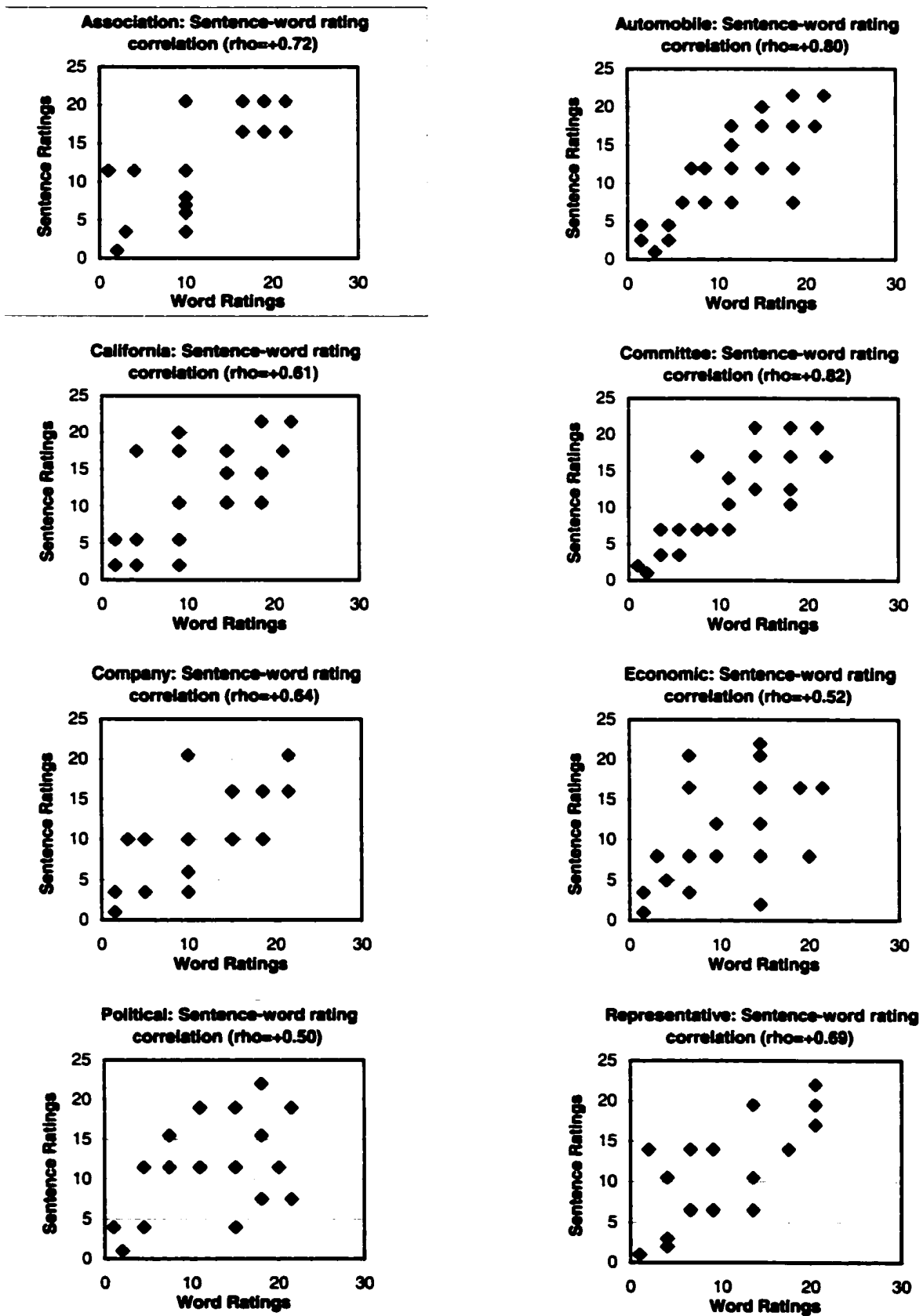
Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--9	7	1--9
Association	6	1--9	5	1--9
Committee	7	1--9	6	1--9
Company	6	1--9	6	1--9
Economic	7	1--9	7	1--9
Automobile	6	1--9	6	1--9
Political	6	1--9	6	1--9
California	6	1--9	5	1--9

Table I: Listener 9				
Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--8	7	1--8
Association	6	1--9	6	1--8
Committee	6	1--9	5	1--9
Company	4	1--8	6	1--7
Economic	7	1--9	6	1--8
Automobile	7	1--9	7	1--9
Political	4	1--9	6	1--9
California	6	1--8	6	1--8

Table J: Listener 10				
Stimulus	sentence stimuli		word stimuli	
	median	range	median	range
Representative	6	1--9	7	1--9
Association	7	1--9	7	1--9
Committee	7	1--9	8	1--9
Company	6	1--8	8	1--9
Economic	8	1--9	8	1--9
Automobile	7	1--9	8	1--9
Political	7	1--9	8	1--9
California	7	1--9	8	1--9

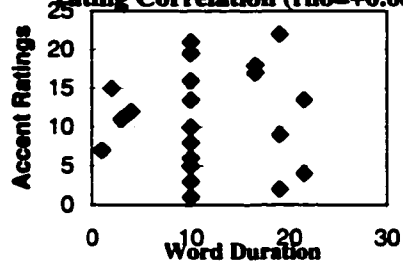
Appendix D: Demographic Information of Subjects					(CLIP=Cuny Language Immersion Program)		
code	age	gender	AOA (yrs)	LOR (yrs)	education	profession	Other
NN Speakers							
nn01	31	f	22		8 student:CLIP	Student	
nn02	21	f	14		7 student:CLIP	Student	
nn03	18	f	15		3 student:CLIP	Student	
nn04	18	f	12		6 student:CLIP	Student	
nn05	20	m	14		6 student:CLIP	Student	
nn06	22	m	17		5 student:CLIP	Student	
nn07	19	f	12		7 student:CLIP	Student	
nn08	20	f	18		2 student:CLIP	home health aide	
nn09	21	f	16		5 student:CLIP	supermarketcashier	
nn10	28	m	24		4 student:CLIP	truck driver	
nn11	18	f	17.6		0.6 high school	(not known)	
nn12	21	f	15		6 CLIP	cashier:wendy's	
nn13	24	f	18		6 CLIP	cashier:payless	
nn15	38	m	12		12 Bachelors	Accounting/GC	Interim in Puerto Rico (11 yrs from age 14 to 26)
nn16	34	f	30		4 High school	cashier:café	
nn17	41	m	14		20 High school	tech asst: NYPL	
nn18	44	m	18		26 High school	tech asst: NYPL	
nn19	47	f	15		32 Bachelors	Asst Registrar:GC	speaks Haitian
nn21	35+	f	17/25		9 High school	Teacher's Aide in Preschool	
nn22	44	f	22		22 High school	Teacher's Aide in Preschool	
nn24	21	f	12		8 High school	(not known)	
nn25	44	m	14		27 High school	mailman	
AE Speakers							
ae01	50	m	n/a	n/a		systems Adm. GC	
ae02	29	f	n/a	n/a	MA	Speech Pathologist	
ae03	30	f	n/a	n/a	BS	Physical Therapist	
ae04	26	f	n/a	n/a	BS	Occupational Therapist	
ae05	42	m	n/a	n/a	MBA?	Registrar:CUNY	

Appendix E: Scatterplots of correlations of accentedness ratings on sentences to ratings on words

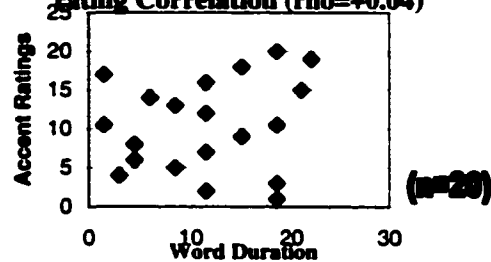


Appendix F: Scatterplots of correlations of overall word duration to accentedness ratings (only NN)

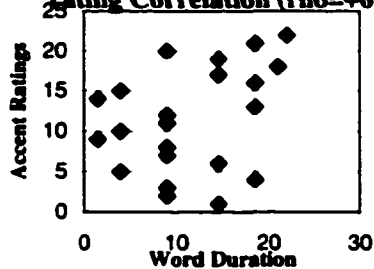
Association: Word duration-Accent rating Correlation ($\rho=+0.08$)



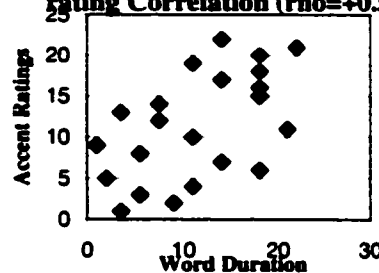
Automobile: Word duration-Accent rating Correlation ($\rho=+0.04$)



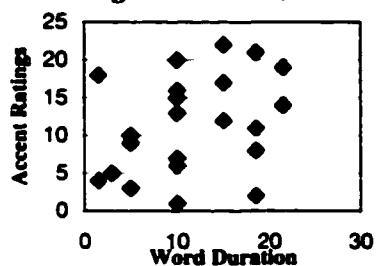
California: Word duration-Accent rating Correlation ($\rho=+0.36$)



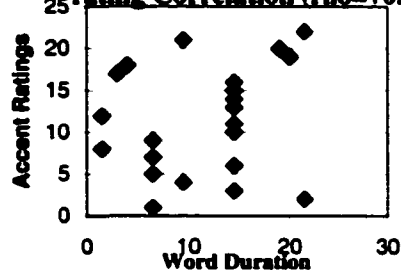
Committee: Word duration-Accent rating Correlation ($\rho=+0.56$)



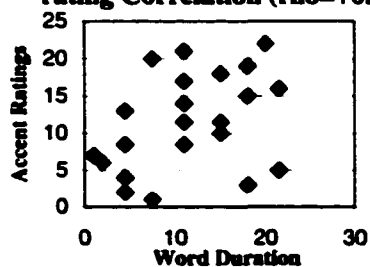
Company: Word duration-Accent rating Correlation ($\rho=+0.23$)



Economic: Word duration-Accent rating Correlation ($\rho=+0.23$)

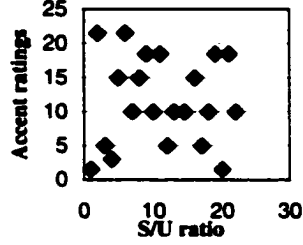


Political: Word duration-Accent rating Correlation ($\rho=+0.39$)

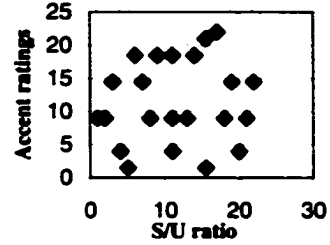


Appendix G: Scatterplots of correlations of S/U ratio to accentedness ratings (only NN)

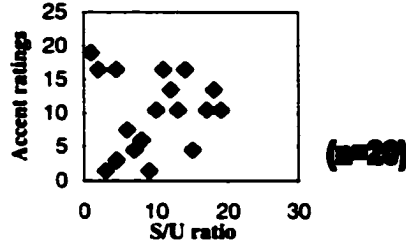
Company: S/U ratio-Accent rating
Correlation ($\rho=+0.02$)



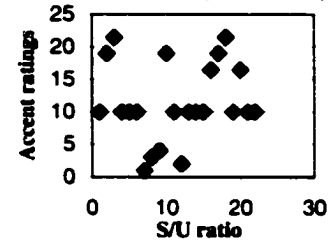
California: S/U ratio-Accent rating
Correlation ($\rho=+0.15$)



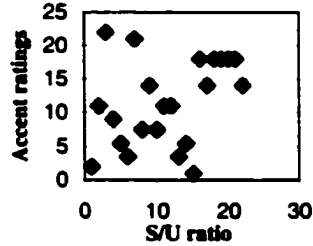
Automobile: S/U ratio-Accent rating
Correlation ($\rho=+0.09$)



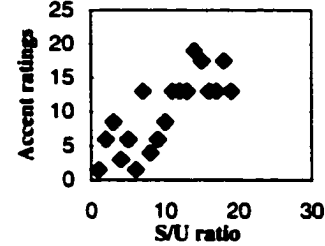
Association: S/U ratio-Accent rating
Correlation ($\rho=+0.17$)



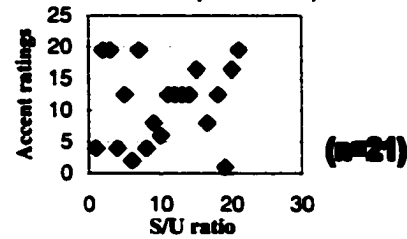
Committee: S/U ratio-Accent rating
Correlation ($\rho=+0.35$)



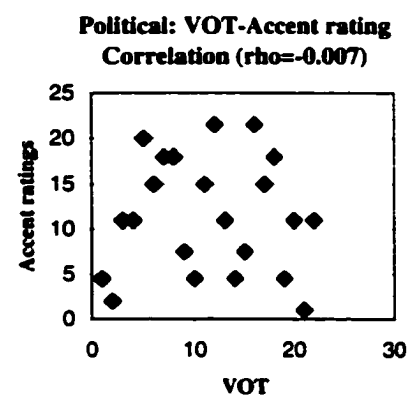
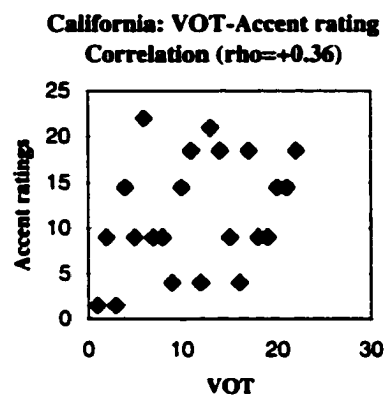
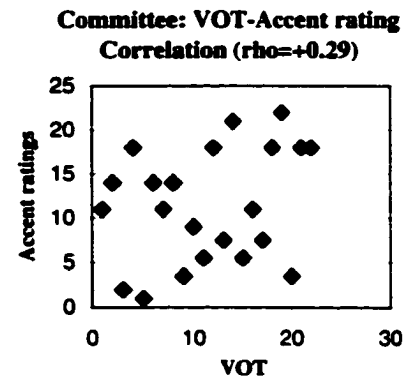
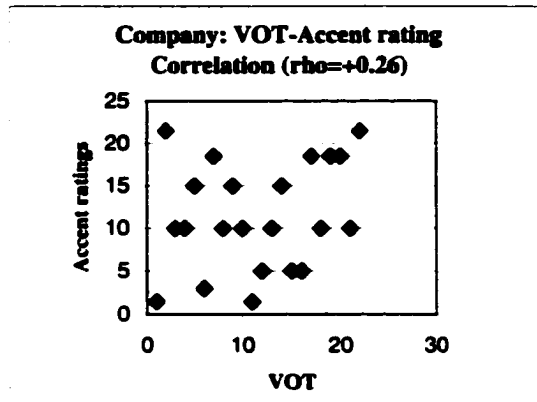
Economic: S/U ratio-Accent rating
Correlation ($\rho=+0.80$)



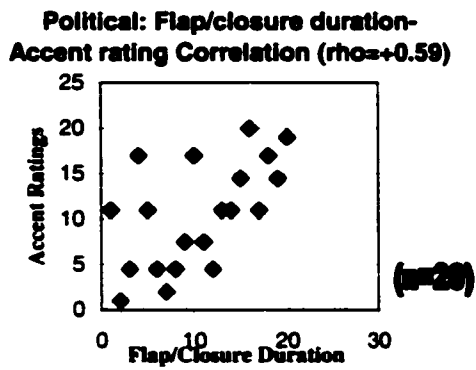
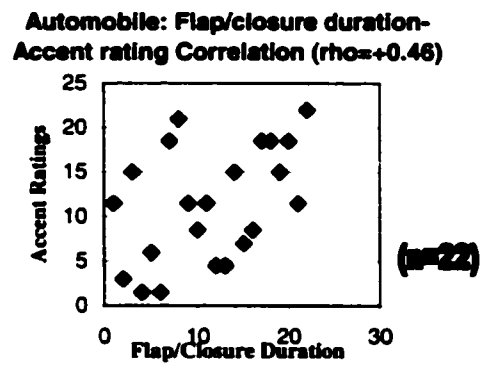
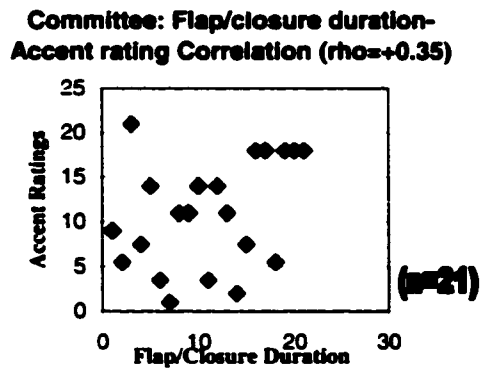
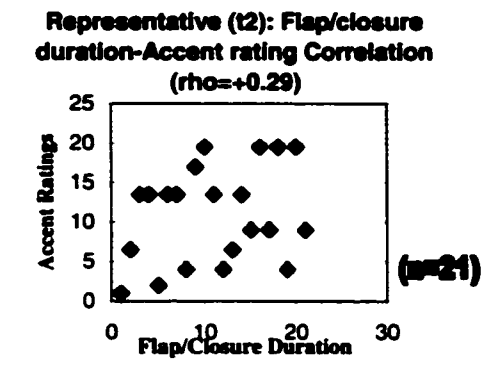
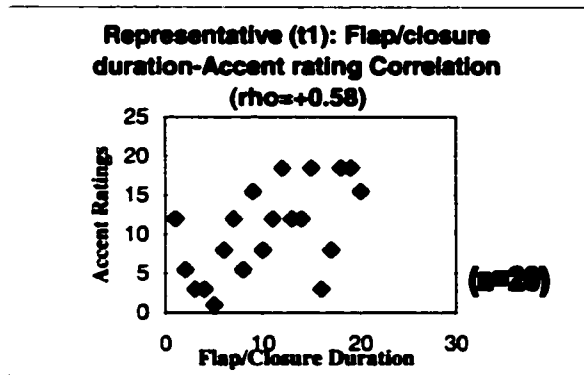
Representative: S/U ratio-Accent rating
Correlation ($\rho=+0.11$)



Appendix H: Scatterplots of correlations of VOT to accentedness ratings (only NN)



Appendix I: Scatterplots of correlations of Flap/closure durations to accentedness ratings (only NN)



Bibliography

- Anderson-Hsieh, Johnson, R., & Koehler, K. (1992). The relationship between native speaker judgments of non-native pronunciation and deviance in segmentals, prosody, and syllable structure. Language Learning, 42, 529-555.
- Asher, J. J. & Garcia, R. (1969). The optimal age to learn a second language, The Modern Language Journal, 53, 334-341.
- Backman, N. (1978). Intonation errors in second-language pronunciation of eight Spanish-speaking adults learning English. Paper presented at the Fifth Annual Congress of Applied Linguistics, Montreal.
- Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels, Journal of the Acoustical Society of America, 97 (3), 1916-1924.
- Brennan, E., Ryan, E., & Dawson, W. (1975). Scaling of apparent accentedness by Magnitude Estimation and Sensory Modality Matching. Journal of Psycholinguistic Research, 4 (1), 27-36.
- Calla McDermott, W. E. (1986). The scalability of degrees of foreign accent. Doctoral dissertation, Cornell University.
- Canfield, D. L. (1962). La pronunciacion del Espanol en America. Bogota: Caro y Cuervo.
- Cunningham-Andersson, U. & Engstrand, O. (1989). Perceived strength and identity of foreign accent in Swedish. Phonetica, 46, 138-154.
- Dalbor, J. B. (1980). Spanish Pronunciation: Theory and Practice. New York: Holt, Rinehart, and Winston.
- del Rosario, R. (1970). El Espanol de America. Sharon, CT: Troutman.

- Derwing, T. & Munro, M. (1995). Accent, intelligibility, and comprehensibility: Evidence from four L1s. Manuscript submitted for publication.
- Elsendoorn, B. A. G. (1985). Production and perception of Dutch foreign vowel duration in English monosyllabic words. Language and Speech, 28, (3), 231-254.
- Fathman, A. (1975). The relationship between age and second language productive ability. Language Learning, 25, 245-253.
- Flege, J. E. (1984). The detection of French accent by American listeners. Journal of the Acoustical Society of America, 76, 692-707.
- Flege, J. E. (1988a). The production and perception of foreign language speech sounds. In H. Winitz (Ed.), Human Communication and its Disorders (pp. 224-401). Norwood, NJ: Ablex
- Flege, J. E. (1988b). Factors affecting degree of perceived foreign accent in English sentences. Journal of the Acoustical Society of America, 84 (1), 70-79.
- Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. Journal of the Acoustical Society of America, 89 (1), 395-411.
- Flege, J. E. (1993). Production and perception of a novel, second-language phonetic contrast. Journal of Acoustical Society of America, 93, 1589-1608.
- Flege, J.E. & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English, Speech Communication, 6, 185-202.
- Flege, J.E. & Fletcher, K. (1992). Talker and listener effects on degree of perceived foreign accent. Journal of the Acoustical Society of America, 91 (1), 370-389.

- Flege, J. E., Munro, M. J. & Mackay, I.R.A. (1995). Factors affecting strength of perceived foreign accent in a second language. Journal of the Acoustical Society of America, 97, 3125-3134.**
- Flege, J.E., Munro, M.J., & Skelton, L.(1992). Productions of the word-final English /t-/d/ contrast by native speakers of English, Mandarin, and Spanish. Journal of Acoustical Society of America, 92 (1), 128-143.**
- Flege, J. E. & Port, R. (1981). Cross-language phonetic interference: Arabic to English. Language and Speech, 24, 125-146.**
- Flege, J., Takagi, N. & Mann, V. (1995). Japanese adults can learn to produce English /r/ and /l/ accurately. Language and Speech, 38, 25-56.**
- Flege, J.E., Yeni-Komshian, G. & Liu, H. (1999). Age constraints on second language acquisition. Journal of Memory and Language, 41, 78-104.**
- Francis, W. N. & Kuçera, H. (1982). Frequency Analysis of English Usage: Lexicon and Grammar. Boston: Houghton Mifflin Company.**
- Glass, G.V. & Hopkins, K.D. (1996). Statistical Methods in Education and Psychology. Boston: Allyn and Bacon**
- Hazan, V. L. & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. Language and Speech, 36, 17-38.**
- Hutchinson, S. P. (1973). An objective index of the English-Spanish pronunciation dimension. Unpublished master's thesis, University of Texas, Austin.**
- Jonasson, J. & McAllister, R. (1972). Foreign accent and timing: An instrumental study, PILUS, 14, 11-40.**
- Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and**

- perceptual evidence. Journal of the Acoustical Society of America, 59 (5), 1208-1221.
- Kohler, K. (1979). Dimensions in the perception of fortis and lenis plosives. Phonetica, 36, 332-343.
- Laeufer, C. (1996). The acquisition of a complex phonological contrast: Voice timing patterns of English initial stops by native French speakers. Phonetica, 53, 86-110.
- Lehiste, I. (1970). Suprasegmentals. Cambridge, MA
- Lollock, Lisa (2000), The Foreign Born Population in the United States: March 2000, Current Population Reports, P20-534, US Census Bureau, Washington, D.C.
- MacDonald, M.G. (1989). The influence of Spanish phonology on the English spoken by United States Hispanics. In Hammond, R., & Bjarkman, P. (Eds.), American-Spanish Pronunciation: Theoretical and Applied Perspectives (pp. 215-237). Washington D.C.: Georgetown University Press.
- Mack, M. (1982). Voicing-dependent vowel duration in English and French: Monolingual and bilingual productions. Journal of the Acoustical Society of America, 71, 173-178.
- Magen, H.S. (1998). The perception of foreign-accented speech. Journal of Phonetics, 26, 381-400.
- Mairs, J. L. (1989). Stress assignment in interlanguage phonology: An analysis of the stress system of Spanish speakers learning English. In Gass, S.M. & Schachter, J. (Eds.), Linguistic Perspectives on Second Language Acquisition, (pp.261-283). Cambridge: Cambridge University Press
- Munro, M. J. (1993). Productions of English vowels by native speakers of Arabic:

- Acoustic measurements and accentedness ratings. Language and Speech, 36 (1), 39-66.
- Munro, M. J. (1998). The effects of noise on the intelligibility of foreign-accented speech. Studies in Second Language Acquisition, 20, 139-154.
- Munro, M. J. & Derwing, T. M. (1995a). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. Language and Speech, 38 (3), 289-306
- Munro, M. J. & Derwing, T. M. (1995b). Foreign accent, comprehensibility and intelligibility in the speech of second language learners. Language Learning, 45 (1), 73-97.
- Munro, M. J., Flege, J. E., & Mackay, I. R. A. (1996). The effects of age of second language learning on the production of English vowels. Applied Psycholinguistics, 17, (3) 313-334.
- Nathan, G. S. (1987). On second-language acquisition of voiced stops. Journal of Phonetics, 15, 313-322.
- Neufeld, G. (1979), Towards a theory of language learning ability. Language Learning 29 (2): 227A1
- Neufeld, G. (1980). On the adult's ability to acquire phonology. TESOL Quarterly, 14, 285-298.
- Olson, L. L. & Samuels, S. J. (1973). The relationship between age and accuracy of foreign language pronunciation. Journal of Educational Research, 66, 263-267.
- Ortega-Llebaria, M. (1997). An explanatory intelligibility test for Spanish accented English. Doctoral dissertation, Department of Linguistics, Indiana University.

- Oyama, S. (1976). A sensitive period for the acquisition of a nonnative phonological system. Journal of Psycholinguistic Research, 66, 263-268.
- Patkowski, M. S. (1990). Age and accent in a second language: A reply to James Emil Flege. Applied Linguistics, 11 (1), 73-89
- Port, R.F., Al-ani, S. & Maeda, S. (1980). Temporal compensation and universal phonetics. Phonetica, 37, 235-252.
- Port, R. F. & Mitleb, F. M. (1983). Segmental features and implementation in acquisition of English by Arabic speakers. Journal of Phonetics, 11 (3), 219-229.
- Raphael, L. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. Journal of the Acoustical Society of America, 51, 1296-1303.
- Resnick, M.C. (1975). Phonological Variants and Dialect Identification in Latin American Spanish. The Hague: Mouton.
- Schmidt, A. M. & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. Phonetica, 53, 162-179.
- Scovel, T. (1988). A Time to Speak: A Psycholinguistic Inquiry into the Critical Period for Human Speech. New York: Newbury House / Harper and Row.
- Seliger, H. W., Krashen, S.D., & Ladefoged, P. (1975). Maturation constraints in the acquisition of second language accent. Language Sciences, 36, 20-22.
- Snow, C. & Hoefnagle-Hohle (1977). Age differences in the pronunciation of foreign

sounds. In Krashen, S.D., Scarcella, R.C., & Long, M.H. (Eds.), Child-Adult Differences in Second Language Acquisition (pp. 84-92), Rowley, MA: Newbury House.

Southwood, M.H. & Flege, J.E. (1999) Scaling foreign accent: Direct magnitude estimation versus interval scaling. Clinical Linguistics and Phonetics, 13, 335-449

Stockwell, R. P. & Bowen, J. D. (1965). The Sounds of English and Spanish. Chicago: The University of Chicago Press.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. Language Learning, 26 (2), 233-253.

Tahta, S., Wood, M. & Loewenthal, K. (1981). Foreign accents: Factors relating to transfer of accent from the first language to a second language. Language and Speech, 24 (3), 265-272.

Tajima, K., Port, R., & Dalby, J. (1997). Effects of temporal correction on intelligibility of foreign-accented English. Journal of Phonetics, 25, 1-24.

Thompson, I. (1991). Foreign accents revisited: The English pronunciation of Russian immigrants. Language Learning, 41 (2), 177-204.

Tomarken, A.J. & Serlin, R.C. (1986). Comparison of ANOVA alternatives under variance heterogeneity and specific noncentrality structures. Psychological Bulletin, 99, 90-99.

Van Els, T. & De Bot, K. (1987). The role of intonation in foreign accent. The Modern Language Journal, 71, (2), 147-155.

Wingstedt, M. & Schulman, R. (1987). Comprehension of foreign accents.

In W. Dressler (Ed.), Phonologica 1984: Proceedings of the Fifth International Phonology Meeting, Eisenstadt, 25-28, June 1984 (pp. 339-344). Cambridge: Cambridge University Press.