Acoustic Correlates of Word Stress in American English

by

Anthony O. Okobi

B.S., Neurobiology and Behavior
Cornell University, 1999

S.M., Electrical Engineering and Computer Science
Massachusetts Institute of Technology, 2002

SUBMITTED TO THE
HARVARD-MIT DIVISION OF HEALTH SCIENCES AND TECHNOLOGY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY
IN SPEECH AND HEARING BIOSCIENCE AND TECHNOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2006

Author ______________________________
Harvard-MIT Division of Health Sciences and Technology
September 11, 2006

Certified by ______________________________
Kenneth N. Stevens
Clarence J. LeBel Professor of Electrical Engineering
and Professor of Health Sciences and Technology
Thesis Supervisor

Accepted by ______________________________
Martha L. Gray, Ph.D.
Edwin Hood Taplin Professor of Medical and Electrical Engineering
Co-Director, Harvard-MIT Division of Health Sciences and Technology

Acknowledgements

I would like first and foremost to dedicate this thesis to my father, Dr. Anthony N. Okobi, who passed away on October 11, 2005, and my maternal grandmother, Josephine M. Emodi, who passed away on August 7, 2005. Both these beautiful people were the most influential figures in my life. As a child, my grandmother fueled my curiosity and encouraged me to explore new things, knowing that she and safety were never far. She made me realize at an early age what a wonderful and exciting world we inhabit. My father made sure that I did not go outside and play until my homework and chores were completed. His tough love kept me out of trouble, disciplined me, and instilled in me a sense of determination. From both of them came the foundation from which I have achieved all that I have. I will forever be grateful to them and the rest of my family.

I would also like to thank members of my thesis committee. Special thanks to my thesis supervisor, Ken Stevens, who never gave up on me as I struggled to understand the biophysics of the human form of communication, known as speech. I would also like to thank Stefanie Shattuck-Hufnagel, a member of my thesis committee, for her guidance and emotional support as I attempted to bridge the knowledge gap between the physics, linguistics, and psychology of speech communication. Where most saw obstacles, she saw opportunities. Another member of my thesis committee that I would like to thank is Helen Hanson for writing her thesis and other journal articles from which I derived many of my measurement techniques, as well as for her help and suggestions on how to refine my thesis. Special thanks also to Adam Albright, also a member of my thesis committee, for his brilliant insights and pages of critique that helped structure my thesis.

Another person who was instrumental in shaping my thesis was Harold Goodglass, who first brought me on board on his project to quantify the prosodic deficits of aphasic speakers. Arthur Wingfield and Hiram Brownell were also involved in this project and a special thanks for their early guidance. I would also like to thank Alfonso Caramazza and Kevin Kearns, who were on my proposal advisory committee and oral qualifying exam committee, respectively. This thesis could not have been completed without the assistance of Arlene Wint. My gratitude also goes to Mark Tiede, Seth Hall, Majid Zandipour, and Satra Ghosh for their technical assistance. I would also like to thank Joe Perkell, Janet Slifka, Nanette Veilleux, David Gow, and Alejna Brugos for their idea and support.

I would like to give special thanks to my officemates Xuemin Chi, Elisabeth Hon, and ex-officemate Virgilio Villacorta, for putting up with my bad jokes and philosophical ranting. I am also grateful to the other graduate and past graduate members of the Research Laboratory of Electronics' Speech Communication Group, especially Neira Hajro, Steven Lulich, Laura Dilley, Annika Imbrie, Xiaomin Mou, Julie Yoo, Lan Chen, Sherry Zhao, Yoko Saikachi, Chi-youn Park, Nancy Chen, and Tamas Bohm. I would also like to thank members of my ACME study group for their support and encouragement.

This research was supported in part by an NIH training grant T32-DC00038 and the MIT Graduate Students Office.

This page is intentionally left blank.

Acoustic Correlates of Word Stress in American English

by

Anthony O. Okobi

Submitted to the Harvard-MIT Division of Health Sciences and Technology on September 11, 2006 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Speech and Hearing Bioscience and Technology

ABSTRACT

Acoustic parameters that differentiate between primary stress and non-primary full vowels were determined using two-syllable real and novel words and specially constructed novel words with identical syllable compositions. The location of the high focal pitch accent within a declarative carrier phrase was varied using an innovative object naming task that allowed for a natural and spontaneous manipulation of phrase-level accentuation. Results from male native speakers of American English show that when the high focal pitch accent was on the novel word, vowel differences in pitch, intensity prominence, and amplitude of the first harmonic, H1* (corrected for the effect of the vocal tract filter), accurately distinguished full vowel syllables carrying primary stress vs. non-primary stress. Acoustic parameters that correlated to word stress under all conditions tested were syllable duration, H1*-A3*, as a measurement of spectral tilt, and noise at high frequencies, determined by band-pass filtering the F3 region of the spectrum. Furthermore, the results indicate that word stress cues are augmented when the high focal pitch accent is on the target word. This became apparent after a formula was devised to correct for the masking effect of phrase-level accentuation on the spectral tilt measurement, H1*-A3*. Perceptual experiments also show that male native speakers of American English utilized differences in syllable duration and spectral tilt, as controlled by the KLSYN88 parameters DU and TL, to assign prominence status to the syllables of a novel word embedded in a carrier phrase. Results from this study suggest that some correlates to word stress are produced in the laryngeal region and are due to vocal fold configuration. The model of word stress that emerges from this study has aspects that differ from other widely accepted models of prosody at the word level. The model can also be applied to improve the prosody of synthesized speech, as well as to improve machine recognition of speech.

Thesis Supervisor: Kenneth N. Stevens
Title: Clarence J. LeBel Professor of Electrical Engineering
and Professor of Health Sciences and Technology

This page is intentionally left blank.

Table of Contents

Chapter **Page**

Table of Figures

Table of Tables

Table of Equations

1. Introduction

1.1 Significance of Word Stress

Word stress is prosodic prominence within a word. Prosody can be defined as a time series of speech-related information that is not predictable from the simple sequence of phonemes. According to Terken (1991), prosodic prominence is defined as the property by which linguistic units are perceived as standing out from their environment. Thus word stress is prosodic prominence that characterizes the relationship between the syllables of a word, such that one of these syllables is considered more prominent than the others. For most languages, prosody can be used to convey meaning at various levels of conversation (e.g., discourse level, phrase level, and word level). Prosodic composition of an utterance is often thought of as a means of organizing and delivering content and meaning (Beckman and Edwards, 1994).

English is a stress language that specifies one syllable in a content word to have primary word stress. In general it is the primary stressed syllable that is pitch accented when the word of interest is the focus of a phrase (i.e., high focal pitch accent). Prosodic information is part of the lexical entry of each English content word, although it is usually not a contrastive property (Kager, 1995; Wingfield *et al.*, 2000). Exceptions to this non-contrastive rule are noun-verb minimal stress pairs, which are pairs of words with the same spelling and similar pronunciations, but different meanings, such as the noun *'abstract*, meaning a summary of a text or scientific article, and the verb *ab'stract*, which means to take away or remove. Primary word stress is on the first syllable for the noun and on the second syllable for the verb. Such word pairs can in general be distinguished only by their different stress patterns, although vowel quality differences may also exist. Figure 1 shows spectrograms of (a) the minimal stress verb *di'gest* when it is the focus of the utterance and high focal pitch accented and (b) when it is not the focus of the utterance.

Within an utterance, prominent syllables can serve as signs indicating what possible words one might encounter along the speech-path. Studies have shown that stressed syllables are informative when inferring words, such that knowing the stress pattern of a word can greatly reduce the number of competing word candidates (Mattys and Samuel, 2000; Wang and Seneff, 2001). There are suggestions that prosodic information about a word may be independently retrieved in word production, as in the case when a speaker in a tip-of-the-tongue state can give the correct number of syllables and the stress pattern of the word, but cannot produce the phonemic segments of the word (Wingfield *et al.*, 2000). In the field of speech therapy, information conveyed by prosodic characteristics of words has served as the basis for the development of therapies to help patients with dysarthria, because such traumatic brain injury disorders are often accompanied by prosodic deficits (Wang *et al.*, 2005).

According to Beckman and Edwards (1994) stressed syllables are anchor points for the pitch accent within an utterance. A study conducted by Fry (1958) showed that the salience of the F0 contour was involved in the cueing of stress in minimal noun-verb stress pairs, such as *'permit* versus *per'mit*. Unfortunately, this study gave rise to a common misunderstanding in experimental literature that fundamental frequency (F0) prominence is a direct acoustic correlate of word stress. This is a misunderstanding that has been incorporated into standard textbooks (as pointed out by Beckman and Edwards, 1994). In contrast, Bolinger (1958) suggested that vowels with primary versus non-primary word stress do not differ in their acoustic properties or in the nature of their articulation. Instead such word stress distinctions were suggested to be rule based. However studies by Fry (1955 and 1958), Lieberman (1960) and Harrington *et al.* (1998) indicate that physical correlates that distinguish between primary stress and non-primary full vowels do exist, at least when the word of interest is pitch accented. These word stress distinctions are fundamentally different from the segmental or phonemic specifications of a word. While segmental specifications give information about the make-up of a word, word stress prosodic specifications indicate the relationship between these segments, as to which is the most prominent.

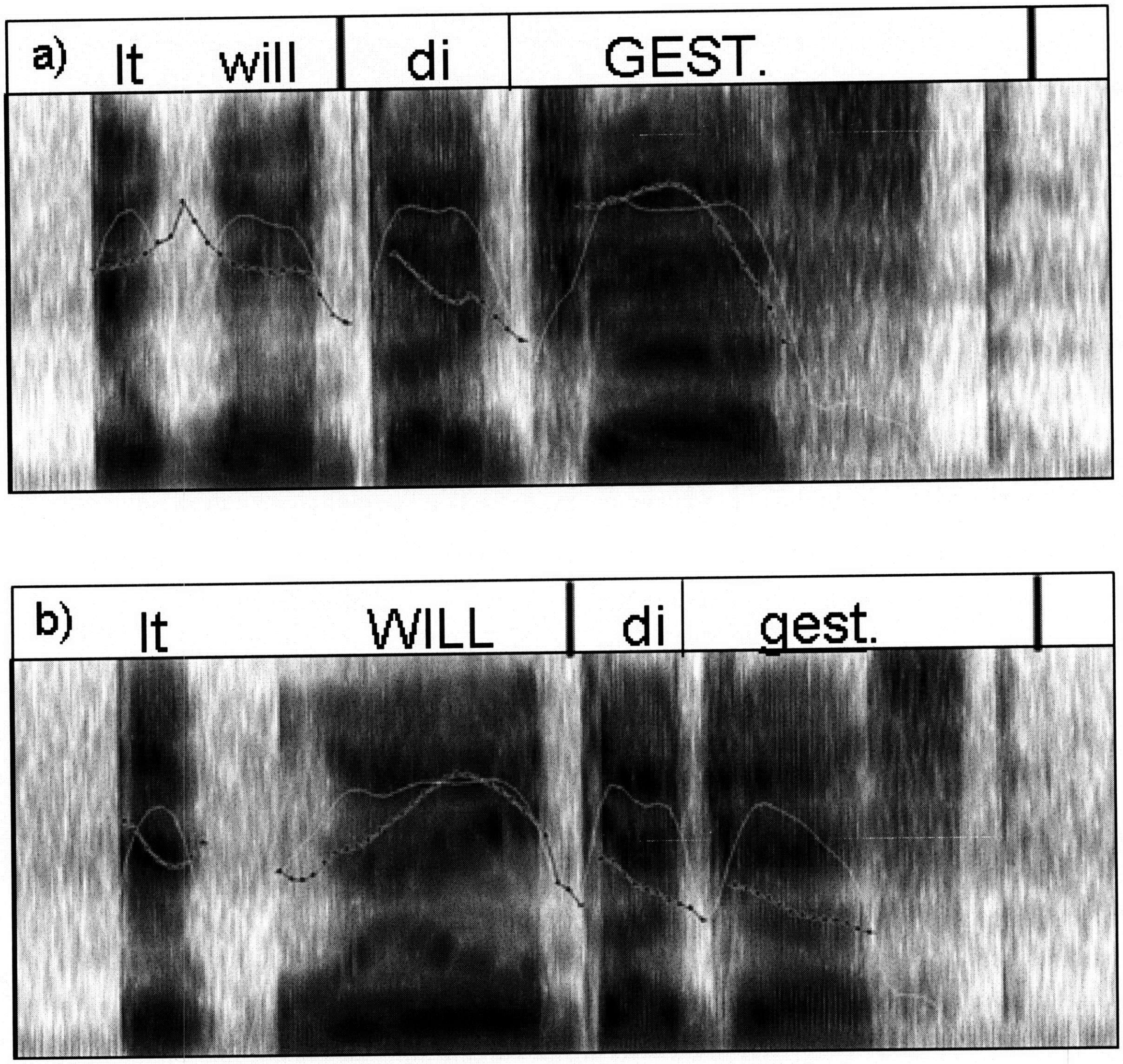


Figure 1. *Labeled spectrogram of the minimal stress verb di'gest when it is (a) the focus of the utterance and is high focal pitch accented and (b) when it is not the focus of the utterance. The solid yellow line is the intensity contour, while the blue-dotted line is the pitch (F0) contour. Words or syllables in all capital letters indicate focal pitch accent on that word or syllable of word.*

1.2 Previous and Related Studies on Word Stress

Recent studies by Beckman and Edwards (1994) demonstrate that unaccented stressed vowels can differ from reduced vowels by vowel quality, duration, and possibly amplitude, while pitch accented vowels are distinguished from unaccented full vowels by an F0 prominence marker. Sluijter and van Heuven (1996a-b) in their study using reiterant speech copies of noun-verb minimal stress pair words showed that, for native speakers of both American English and Dutch, stressed full-vowel syllables in reiterantly imitated words can be distinguished from non-primary full-vowel syllables, even in non-pitch accented contexts. They showed that primary stressed and full vowels can be differentiated based on the relative level of energy at their high frequencies (i.e., degree of spectral tilt), where the primary stressed vowels had more energy at their high frequencies. Stevens (1994) also gave evidence that the glottal excitation waveform differs for the vowels of syllables that are accented from vowels that are full, but unaccented, as well as from reduced vowels. These results support the claim that these three types of vowels can be distinguished based on their acoustic properties.

Assuming that the source of word stress prominence differences between these vowels is at the laryngeal level, how might this distinction arise during speech? During vowel production, the configuration of the vocal folds can be varied in several different ways. Four types of normal glottal configuration were considered by Hanson (1997a): (1) the arytenoids are approximated and the membranous part of the vocal folds close abruptly; (2) the arytenoids are approximated, but the membranous folds close sequentially from front to back along the length of the vocal folds; (3) there is a posterior glottal opening at the arytenoids that persists throughout the glottal cycle (a glottal chink), and the folds close abruptly; (4) a posterior glottal opening extends into the membranous portion of the folds throughout the glottal cycle, forcing the vocal folds to close from front to back in a non-abrupt manner. According to Hanson (1995) and Stevens (1998), the presence of a posterior glottal chink throughout a glottal cycle introduces modifications to the spectrum of a vowel. Formant bandwidth, in particular that of the first formant (F1), is increased due to additional energy loss at the glottis. Hanson (1997a) also determined that the amplitude of the first harmonic (H1) relative to that of the first formant (A1) can reflect the bandwidth of the first formant (B1). Thus, assuming a constant effect of the vocal tract on the first formant bandwidth, H1-A1 can be used to reflect changes in B1 caused by the presence of a posterior glottal chink.

Another acoustic consequence of the glottal chink is the production of additional tilt in the source spectrum. This additional tilt is due to the fact that the airflow through the glottal chink cannot undergo a discontinuous change because of the acoustic mass of the moving air through the glottal area (Stevens, 1994). Approximations of the spectral tilt can be made by measuring the amplitude of the first harmonic (H1) relative to that of the third formant spectral peak (A3), which is near 3kHz for most speakers. Measurements obtained using this method show that the mid- to high-frequency components are influenced by how abruptly the air flow returns to its minimum value, as well as by the presence of an opening in the posterior region of the glottis (Hanson and Chuang, 1999).

Stevens (1994) found that the average drop in amplitude of the first formant (A1) for the reduced vowels relative to the pitch accented vowels range from 7 to 13 dB for different speakers, with considerable variability for different vowels for the same speaker. Corrections for these spectral differences between vowels were applied by Hanson (1995 and 1997a-b) and further modified by Iseli and Alwan (2004). There are also differences between reduced vowels and pitch-accented vowels in the F1 bandwidth (B1), as determined from the waveform, with the bandwidth being wider for the reduced vowels, indicating a more abducted glottal configuration for those vowels (Stevens, 1994). Furthermore, the glottal source spectrum amplitude at higher frequencies is much weaker for reduced vowels (Stevens, 1994; Sluijter *et al.*, 1995; Sluijter and van Heuven, 1996a-b). This increased spectral tilt is also consistent with a more abducted glottal configuration, which leads to a less abrupt discontinuity in the waveform at the time of closure. Thus spectral analysis techniques used by Stevens (1994), Hanson (1995, 1997a), and Hanson and Chuang (1999) can be used to determine the acoustic variations between the vowels in the syllables within a word that best predict the word stress pattern of that word.

1.3 Unanswered Questions

The complication with the studies by Fry (1955 and 1958) and Lieberman (1960), as well as other earlier studies to determine the correlates of word stress, is that they did not control for the phrase level pitch accent. It seems that they assumed that the correlates of pitch accent were also correlates of word stress. However, studies by Beckman and Edwards (1994), Sluijter *et al.* (1995) and others show that high fundamental frequency (F0), greater intensity, and longer duration are correlates that distinguish accented primary stressed syllables from the neighboring non-primary syllables. Figure 1a shows that when the primary stressed second syllable of the minimal stress pair word, *di'gest*, is accented, it has a higher F0, more intensity, and longer in duration than the non-primary first syllable. However, as Figure 1b shows, if the word *di'gest* is not the focus of the utterance and not high focal pitch accented, the primary stressed second syllable no longer has the higher F0, greater intensity, and the durational difference between second and first syllables is now reduced. Is it possible to distinguish the primary stressed syllable from the non-primary full vowel syllables when the word of interest is not accentuated?

Studies done by Sluijter *et al.* (1995, 1996a-b, and 1997) attempted to answer this question using reiterant speech repetitions of noun-verb minimal stress pairs embedded in a carrier phrase. Although it is still uncertain as to what properties of language reiterant speech captures, Sluijter *et al.* (1995, 1996a-b, and 1997) found that when the reiterant speech version of the target word was not pitch accented they could still distinguish between the reiterant speech primary stressed syllable from the reiterant speech unstressed syllable. They found that duration, spectral tilt (measured as H1*-A3*, where "*" indicates correction for vocal tract shape), and first formant bandwidth (measured as H1*-A1) could be used to distinguish a primary stressed reiterant speech syllable from an unstressed reiterant speech syllable.

In their studies, Sluijter *et al.* (1995, 1996a-b, and 1997) manipulated the high focal pitch accent of a carrier phrase such that it was either on the reiterant speech version of the target word or not. They do not however indicate the location and proximity of the pitch accent to the reiterant target word. The importance of the location and proximity of the pitch accent will be discussed in the next chapter. Furthermore, Sluijter *et al.* (1995, 1996a-b, and 1997) did not mention that they controlled for vowel reduction. This is important because native speakers of American English often reduce the non-primary vowels of noun-verb minimal stress word pairs, like the ones they used in their studies. It is possible that the reiterant speech was capturing the difference between primary stressed syllables and reduced vowel syllables, not the difference between full vowels one of which has primary stress. Campbell and Beckman (1997) tried to replicate the studies done by Sluijter *et al.* (1995, 1996a-b, and 1997) and were unsuccessful. They concluded that contrary to the findings of Sluijter *et al.* (1995, 1996a-b, and 1997), there were no spectral correlates to word stress in English for real words with full vowels.

Thus unanswered questions remain with regards to the correlates of word stress. The first question about whether there exist acoustic properties of primary stressed syllables that can be used to distinguish them from non-primary syllables has been answered with regards to comparisons between accented full vowel syllables, unaccented full vowel syllables, and reduced vowel syllables (Beckman and Edwards, 1994; Stevens, 1994; Hanson, 1997b). However, the question has not been answered for unaccented primary stressed full vowel syllables versus non-primary full vowel syllables, for real English words with full vowels. This is the central question that will be addressed in this thesis. It can be broken down into three specific questions: Are there acoustic production correlates of word stress for non-reiterant speech words with full vowels, when they are not pitch accented? Are these acoustic correlates also perceptual cues for syllable prominence when the target word is not pitch accented? What is the range of syllable difference in these acoustic correlates that is considered natural by native speakers of American English?

1.4 Research Objectives

The objective of the thesis research was to determine the acoustic parameters that change in response to word level prosody. In particular, the goal is to determine the acoustic parameters that consistently distinguish the primary stressed full vowel syllable from the non-primary full vowel syllable of target words in different pitch accented conditions, as well as those parameters that make this distinction only when the word of interest is pitch accented (i.e., correlates to pitch accent). A long-term goal of this thesis work is to derive a word stress model of American English that can be used to automatically extract quantitative word stress information in order to greatly improve automated speech recognition systems.

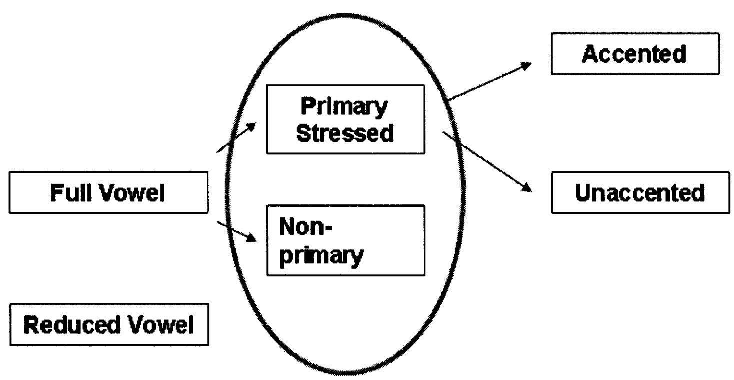


Figure 2. ***Main goal of research is to determine the correlates of word stress that can be used to differentiate between a primary stressed full vowel from a non-primary full vowel, even when the syllable containing vowel is not pitch accented.***

Information from this study can also be used to design a specialized diagnostic tool for probing patients with language or motor speech production deficits, in order to determine if the problem is of a prosodic nature. Furthermore, such a diagnostic tool could be used to determine if the prosodic deficit is on the phrase level or at the word level. The method used in this study to prompt speakers to accentuate and de-accentuate target words can also be used, with slight modification, to teach non-native speakers of American English how to produce native-like utterances with varying phrasal focus.

The specific aim of this thesis study is to determine the acoustic correlates of primary word stress and distinguish it from phrase level pitch accent correlates in order to derive a quantitative acoustic model of word prosody. On the assumption that the acoustic parameters associated with primary stressed and accented syllables are the result of articulatory mechanisms used in speech production, the acoustic characteristics of primary stressed syllables in American

English two-syllable nouns are analyzed and quantified in attempt to develop this model of articulatory-acoustic mapping.

1.5 Hypotheses

The general working hypothesis for this study is that native speakers of American English are expected to show differences between primary and non-primary stressed syllables in their production of both real and novel word utterances. This word stress distinction is expected to be indicated primarily by syllable duration, spectral tilt (H1*-A3*) and noise at high frequencies. It is also possible that word stress information might be carried by syllable vowel differences in first formant bandwidth approximated by H1*-A1*, as indicated by results from preliminary experiments on the acoustic differences between primary stressed and reduced vowels, and studies done by Klatt and Klatt (1990), Stevens (1994), and Hanson (1997b). Corrections were made to the spectral measurements to account for the effects of the vocal tract shape on the glottal source spectrum (Hanson, 1995; Iseli and Alwan, 2004). These corrected parameters are indicated by "*" in the text.

Evidence for syllable duration as a word stress cue comes from several studies (Oller, 1973; Klatt, 1976; Sluijter *et al.* 1995, 1996a-b, and 1997). Studies done by Sluijter *et al.* (1995 and 1996a) also indicated that for noun-verb minimal stress pair words, primary stressed vowels have less spectral tilt than unstressed vowels. Studies by Klatt (1976), Klatt and Klatt (1990) in a paradigmatic (i.e., across different words) comparison of primary stressed vowels to unstressed vowels showed that primary stressed vowels had less noise at high frequencies than unstressed vowels, which were not controlled for reduction.

Syllable differences in spectral tilt, noise at high frequencies, and duration are hypothesized to exist between primary and non-primary stressed full vowel syllables for cases when the phrase level prominence (i.e., high focal pitch accent) is on the target word and also when it is not on the target word. Based on previous findings by Klatt and Klatt (1990) and Sluijter *et al* (1996a-b and 1997), we expect that a non-primary full vowel would be shorter in duration, have greater spectral tilt, and be noisier, than the primary stressed vowel within the same word. However, it is possible that duration is also affected by phrase level accentuation, since syllable duration is known to be affected by location relative to phrase boundaries and discourse (Oller, 1973; Klatt, 1976; Beckman and Edwards, 1994; Turk and White, 1999).

Changes in the value obtained for syllable difference in F0 prominence, intensity and the spectral approximations of amplitude of voicing and open quotient, H1* and H1*-H2* respectively, are expected to correlate with the primary stressed syllable only when it is also accented (i.e., pitch accent correlates), but not when it is de-accented. This is based on the results from studies by Beckman and Edwards (1994) and Sluijter *et al.* (1995, 1996a-b, and 1997) discussed in Section 1.2. Primary stressed syllables of target words are expected to be identifiable by their higher F0 prominence and greater intensity only when they are pitch accented, as demonstrated in Figure 1. Increases in H1* and H1*-H2* give rise to increases in the overall amplitude and intensity and are therefore expected to line-up with intensity as a pitch accent correlate (Klatt and Klatt, 1990). Thus these parameters are hypothesized to be correlates for phrase level prominence, not word stress, in American English, as shown by Beckman and Edwards (1994) and Sluijter *et al.* (1995 and 1996a-b).

1.6 Approaches to Study

The hypotheses discussed in Section 1.5 can be organized into three general areas of interest (distinction, production and perception) which have to be addressed in order to meet the objectives of this thesis. The first area of interest is distinction. According to the hypotheses of Section 1.5, the primary stressed syllable of a two-syllable word should be acoustically different from the non-primary syllable in a non-accented situation, even if both syllables contain full vowels. In order to address this area, an object naming paradigm was developed that allowed the author to prompt native speakers of American English to put high focal pitch accent on the target words embedded in a carrier phrase, as well as to de-accent them. It is important that speakers be able to pitch accent the correct syllable (i.e., primary stressed syllable) of a target word because this shows that speakers know the relationship between the two syllables of the target word and can accurately distinguish them in a pitch accented condition. It is the objective of this thesis to determine if the same speakers continue to distinguish the primary stressed syllable from the non-primary syllable in non-pitch accented situations.

The second area of interest is production. Production differences between primary stressed and non-primary full vowel syllables should be consistent across vowels (e.g. /a/, /i/,/o/, and /u/). That is since vowel differences in vocal tract shape are corrected, the primary stress versus non-primary full vowel distinction should be present regardless of the formant characteristics of the vowel. This is because the events giving rise to this distinction are hypothesized to be occurring at the region of the glottis, which by first approximation is assumed not to be influenced by the changes in the vocal tract that give rise to the different vowels. In order to test this hypothesis two-syllable novel words with full vowels, discussed further in Chapter 2, were used in a production study to control for the phonological differences between syllables that might affect accurate measurements of the acoustic parameters of interest. Non-minimal stress pair real words with full vowels, but contrasting in the primary stress syllable location, were also used in the production study to determine the acoustic correlates to word stress and pitch accent. The object naming paradigm was used in the production study to accentuate and de-accentuate target novel and real words.

Perception is the third area of interest and is directly related to the results obtained from the production study. It addresses the issue of whether the acoustic correlates found in the production study are perceived as carrying word stress information to listeners. That is, production word stress acoustic correlates should be used perceptually as syllable prominence cues. In order to determine the perceptual cues of word stress, two-syllable novel words were synthesized and embedded in the same phonological environment used in the carrier phrase for the production study. The syllable difference in the correlates of word stress that were found in the production study were manipulated in order to change the prosodic relationship between the two full vowel syllables of the synthesized words and determine how changes in syllable differences in these correlates influence syllable prominence judgment.

2. Production Study: Novel and Real Words

2.1 Speakers

Five male native speakers of American English, between 18 and 50 years of age, participated in this study. None of the participants had a history of hearing or speech production difficulties. Participants were compensated for the amount of time they devoted to this study. They were individually recorded in a sound insulated booth using a directional condenser microphone, approximately 12 inches from the mouth. Utterances were digitally recorded at 10kHz sampling rate and low-pass filtered at 5kHz for speech analysis.

Although both male and female speakers were used in the preliminary experiments leading to this study, only male speakers were used in this thesis study. Preliminary experiments revealed that the object naming paradigm, used to prompt speakers to accent or de-accent the target word, was more affective with male speakers, who in general produced only one pitch accent corresponding to the high focal pitch accent in their utterance of the carrier phrase. Female speakers, tested in the preliminary experiment, often not only placed a high focal pitch accent in the right location, but also contrastively pitch accented the target word. This made it difficult to obtained non-accented target words to test our hypotheses stated in Chapter 1.

Furthermore, previous studies by Klatt and Klatt (1990) and Hanson and Chuang (1999) showed that there were gender differences with regard to some of the acoustic measurements that will be used in this study, such as the approximation for glottal spectral tilt, H1*-A3*, and noise at high frequencies. According to Hanson and Chuang (1999), it is possible that spectral tilt is an important cue for distinguishing male and female voices, while Klatt and Klatt (1990) found that female speakers tended to have more noise at high frequencies. Male speakers tended to have greater harmonic energy at high frequencies and less noise. Since we wanted to avoid incorrect or ambiguous results that might be interpreted as being due to gender differences, as well as narrowly focus on correlates of word stress between primary stress and non-primary full vowels, only male speakers were used in this study.

2.2 Stimuli

Speakers were required to name objects represented by digital pictures displayed on a 19 inch computer monitor. These pictured objects were visualizable nouns. Object names were said using the carrier phrase discussed in Section 2.3.

2.2.1 Novel Words

The difficulty with finding large numbers of two-syllable English names of objects with variable stress patterns and then controlling these words for vowel-consonant compositions and vowel quality, led to the use of reiterant speech-like novel words for this production study. The novel words were *'dada*, *'dodo*, and *'didi*, with first syllable primary stress, and their second syllable primary stress counterparts *da'da*, *do'do*, and *di'di*. The first syllable $[CV]_1$ and the second syllable $[CV]_2$ of the novel words contained the same consonant and vowel in order to control for the phonological composition of the syllables.

Precautions were also taken to control for the surrounding environment of the syllables. A single syllable name of a color ending in a vowel always preceded the novel word and a single syllable word beginning with the voiced stop-consonant /d/ always followed the novel target word in the carrier phrase used in this study. Thus both the first and the second syllable of the target word were preceded by a vowel and followed by the voiced stop-consonant /d/. The vowels in the target novel words were chosen because they are full vowels, capable of being primary stressed and are relatively far from each other in the vowel formant space. The consonant /d/ was chosen for easier identification of landmarks for the consonants and the vowels.

Three visually distinct novel objects were chosen and given the first syllable primary stressed names *'dada*, *'didi*, and *'dodo*. These same three objects were then slightly altered, so that they were recognizable but noticeably different. The second syllable primary stress names *da'da*, *di'di* and *do'do*, respectively, were given to the altered forms of three objects. Figure 3 shows the objects used to represent the novel words. Thus the first syllable primary stress novel word was a lexical item representing a different object and having a different meaning than the second syllable primary stress novel word, although they both shared the same CVCV composition (i.e., *'dada* and *da'da*).

2.2.2 Real Words

A total of four real words were used in this production study. Two of the object names had first syllable primary stress, *statue* and *sushi*, while the other two target words had second syllable primary stress, *tattoo* and *bouquet*. All the above target words were chosen because they contain a primary stressed syllable and a secondary/non-primary full vowel syllable. Pronlex, a component of the COMLEX lexical database, as well as The American Heritage College Dictionary, 3rd edition, were used to verify the word stress status of each of the syllables of the target words used in this study. Figure 4 shows the objects used to represent the real words.

The first syllable primary stressed word, *statue*, and the second syllable primary stressed word, *tattoo*, have identical vowels in their first and second syllables. This allows for direct comparison of the two vowels when they are primary stressed and when they are non-primary full vowels. Target words *sushi* and *bouquet* share the same vowel /u/ with *statue* and *tattoo*, but in the first syllable rather than the second. The different syllable location of the vowel /u/ allows for a six-way direct and syllable location comparison of the vowel /u/ between the four target words. None of the words contained liquids (i.e., [l] and [r]) and/or glides (i.e., [w] and [j]) because of the effect of these segments on the spectral composition of adjacent vowels.

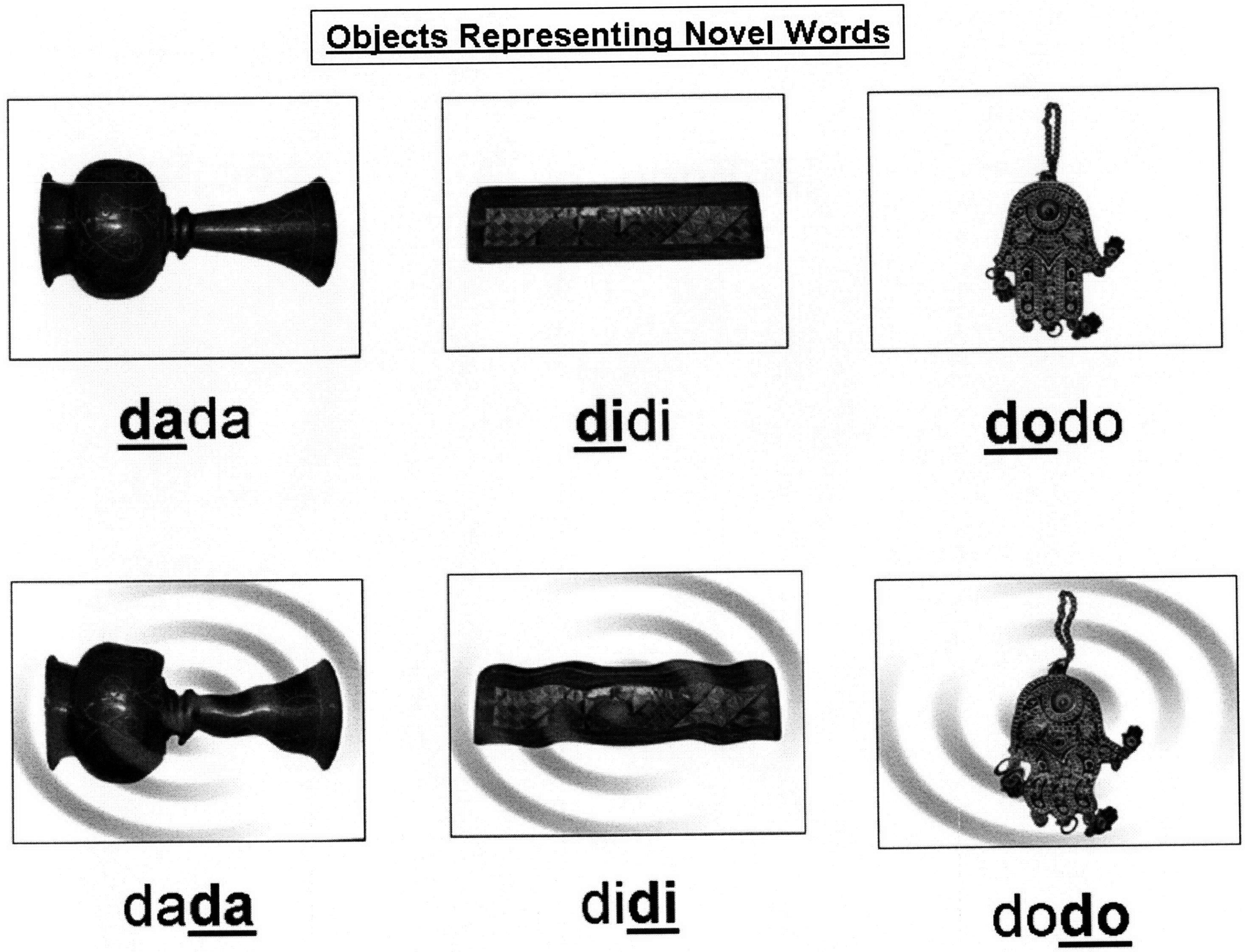


Figure 3. ***The objects used to represent the novel words.***

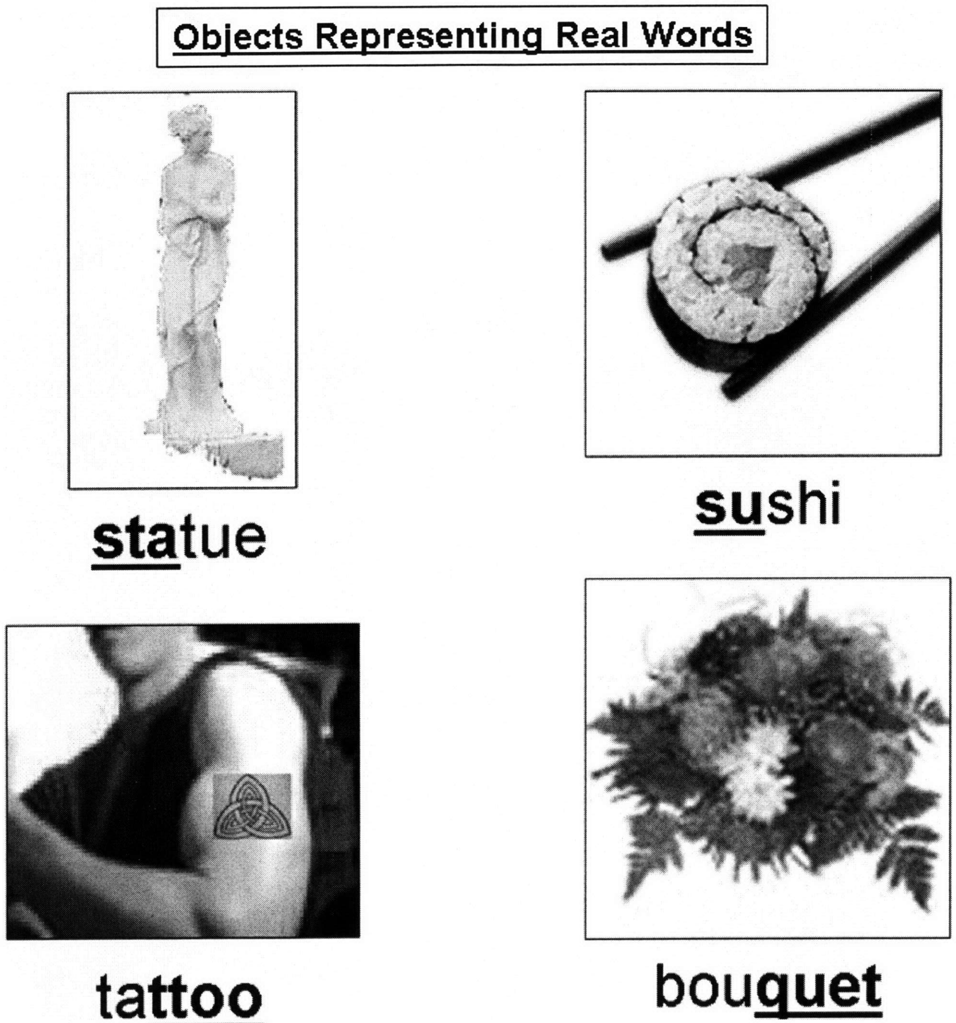


Figure 4. ***The objects used to represent the real words.***

2.3 Experiment Design

Before testing the participants on the target words, they were put through a preliminary training session. Two preliminary training objects were given the novel names 'gugu and gu'gu, respectively. The purpose of the preliminary training session was to introduce the speakers to the format of this production study. Following the preliminary training session speakers were presented the objects representing the target words, using the same format. Before the actual test, speakers were given a brief naming practice session, where they saw the orthographic spelling of each target word written underneath its corresponding object once and then practiced using the names of the objects (i.e., target words) in carrier phrases requiring them to verbally distinguish the minimal stress pairs of target words. In the practice session, two objects were presented together with the first object corresponding to a first syllable primary stressed target word and the second object corresponding to a target word with second syllable primary stress (i.e., *statue-tattoo*).

Digital pictures of the target words, referred to as objects, were presented to the participants within the object naming paradigm. The presentation of the objects was varied in three different conditions designed to produce systematic variations in phrase level accentuation. Results from these three conditions were used to determine the acoustic correlates of word stress that distinguished between the primary stressed syllable and the non-primary full vowel syllable of the two-syllable target words, as well as the correlates of pitch accent that indicate the presence of phrasal focus on the target word. The three conditions designed to separate phrase level focal pitch accent from word stress acoustic correlates are: The focal pitch accented condition (Fa); the post-nuclear pitch accented condition 1 (Fp1); and post-nuclear pitch accented condition 2 (Fp2).

2.3.1 Focal Pitch Accented Condition (Fa)

In this object naming task, speakers were first shown a picture of the object representing the first syllable primary stressed word next to the picture of the object representing the minimal stress paired second syllable primary stressed word (i.e., *'dada-da'da*, *statue-tattoo*, etc.). Speakers were asked the question "Which object drove here?" and instructed to answer with the name of the circled object in the carrier phrase, "My grey (*target word*) drove here." This object naming task was designed to have the speaker place high focal pitch accent on the target word. In this high focal pitch accented condition (Fa), both objects were always the color grey and assigned the same owner, "my". Thus by varying the circled object, speakers were prompted to put the high focal pitch accent on the target word within the carrier phrase. The novel words were paired according to their CV composition, such that words with identical composition, but contrast in the syllable location of the primary stress vowel (i.e., minimal stress pairs like *'dada* and *da'da*). For the real words, *statue* and *tattoo* were paired, to allow for maximum contrast of word stress. *Sushi* and *bouquet* formed the second minimal stress pair of target real words, since they contrast in the syllable location of their primary stress. Speakers were presented a picture of the paired objects twelve times, with one of the paired objects circled. The first utterance of each target word was not used in analysis. Each utterance was checked for correct intonation before analysis. Figure 5 illustrates (a) the object presentation format and (b) an example utterance spectrogram from a speaker's response to the presentation.

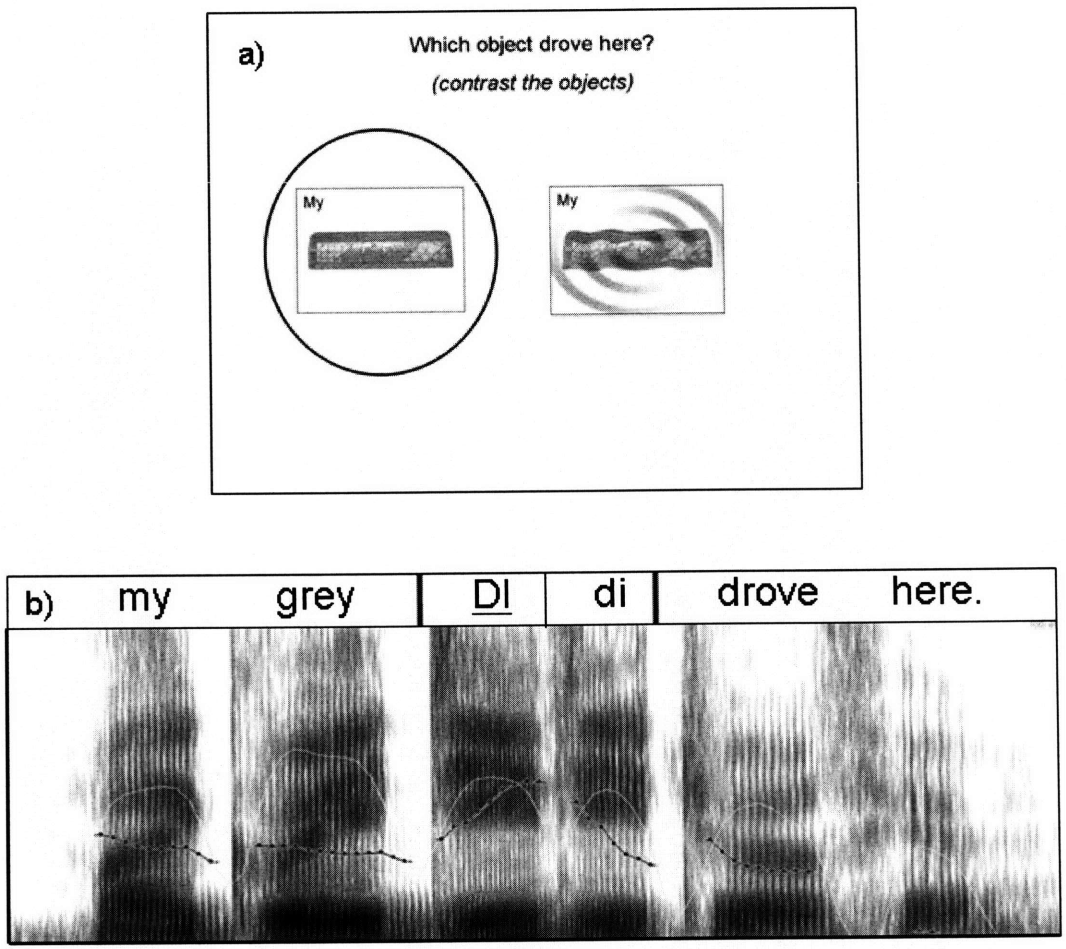


Figure 5. ***Focal Pitch Accented Condition (Fa). (a) object presentation and (b) labeled spectrogram. By varying the object, the speaker could be prompted to treat the name of the target object as new information and place focal pitch accent on it. The solid yellow line is the intensity contour, while the blue-dotted line is the pitch (F0) contour. Words or syllables in all capital letters indicate focal pitch accent on that word or syllable of word.***

2.3.2 Post-Nuclear Pitch Accented Condition 1 (Fp1)

Speakers were also tested in the non-focal pitch accented condition (Fp1), where the one-syllable word preceding the target word had the high focal pitch accent. The same pair of target words tested in the Fa condition was also tested in this Fp1 condition. Speakers were shown a grey version of the object representing one of the target words next to a blue version of the same object. They were then asked the question "Which object drove here?" The speakers were instructed to use the carrier phrase "Your (*color*) target word drove here." In this condition the object remained the same, as well as the owner, but the color of the circled object changed. Since the color of the object was the only thing different, speakers were prompted in this Fp1 condition to place the high focal pitch accent on the color in their utterance, instead of on the target word. Speakers were presented each object representing a target word six times in a row, with only the color of the circled object changing. As before, the first utterance of each target word was not used in analysis. Figure 6 illustrates (a) the Fp1 object presentation format and (b) an example utterance spectrogram from a speaker's response to the presentation.

2.3.3 Post-Nuclear Pitch Accented Condition 2 (Fp2)

An additional post-nuclear focal pitch accented condition was added to this production study in order to better understand the effect of location and presence of focal accent on both the primary stressed and non-primary full vowel syllables. This effect of high focal pitch accent on spectral measurements from the target words is discussed in detail in Section 2.6 of this Chapter. In this post-nuclear pitch accented condition (Fp2), objects of each target word were grouped into blocks of six presentations containing the exact same object all the same color. Each object was then assigned one of the possible two owners, "my" or "your", written on the object. All the target words, tested in both the Fa and Fp1 conditions, were also tested in this Fp2 condition.

They were then asked the question "Which object drove here?" The objects were presented in the same format as in the Fp1 condition, such that speakers were instructed to use the circled object's name in the carrier phrase "(*Owner*) blue target word drove here." Thus speakers were prompted to place the high focal pitch accent on the word two syllables in front of the target word. By only varying the owner of the pictured object, speakers were prompted to treat the owner of the object as the new information and place the high focal pitch accent on it. Figure 7 illustrates (a) the Fp2 object presentation format and (b) an example utterance spectrogram from a speaker's response to the presentation.

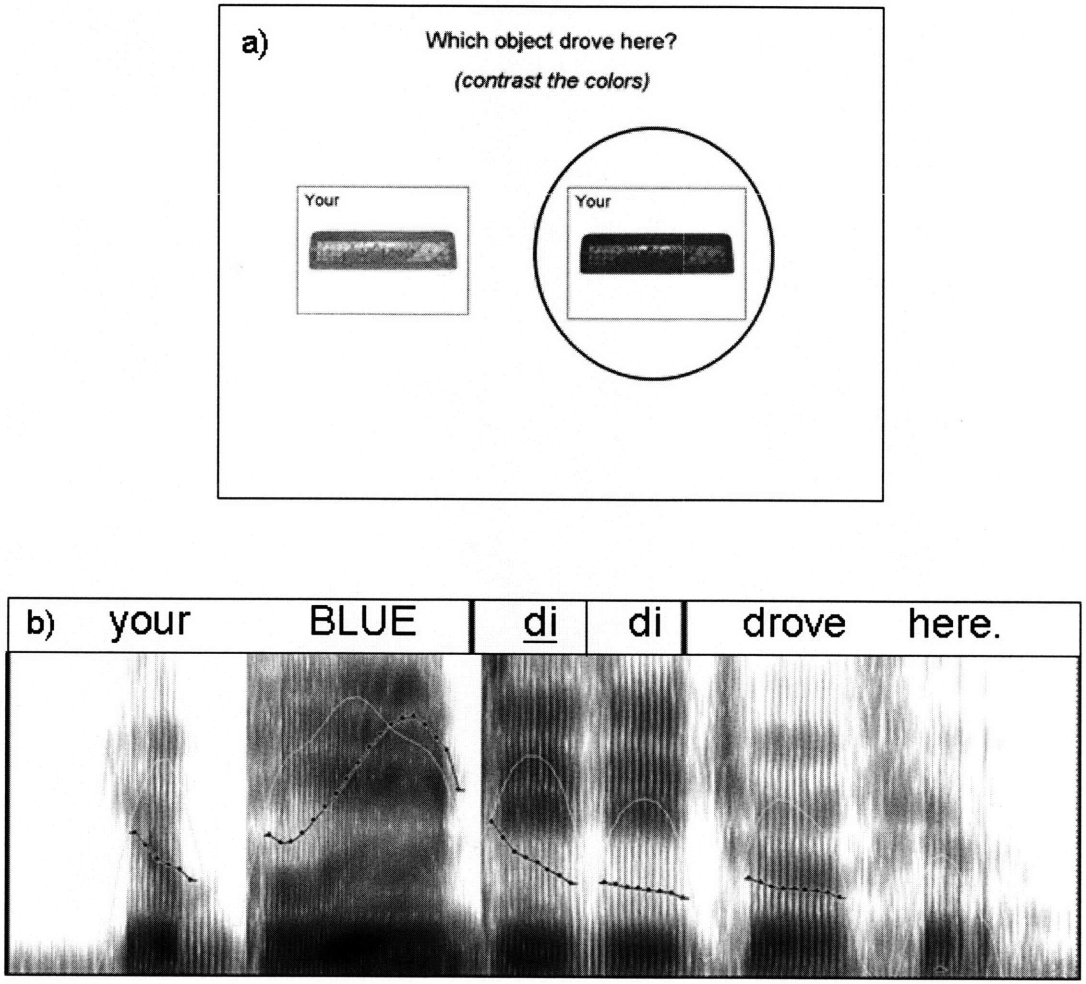
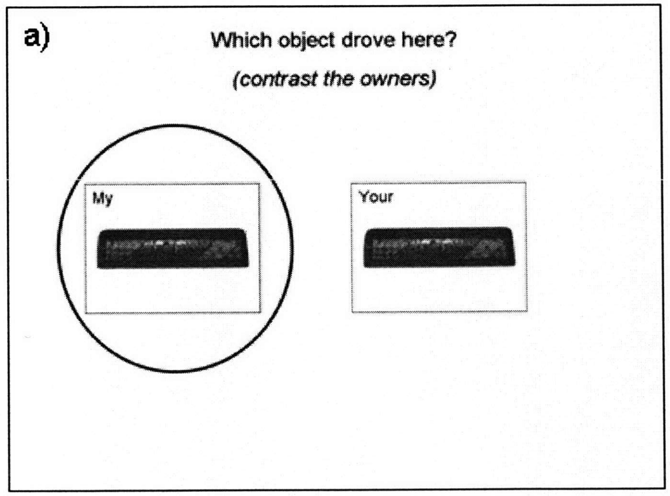


Figure 6. ***Post-Nuclear Pitch Accent Condition 1 (Fp1). (a) object presentation and (b) labeled spectrogram. By varying the color of the object, speakers could be prompted to treat the color of the circled object as new information and place focal pitch accent on it. The solid yellow line is the intensity contour, while the blue-dotted line is the pitch (F0) contour. Words or syllables in all capital letters indicate focal pitch accent on that word or syllable of word.***



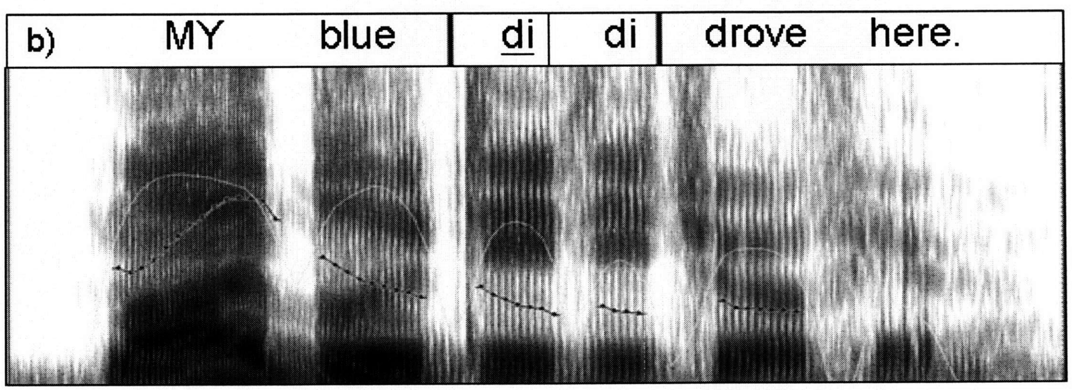


Figure 7. ***Post-Nuclear Pitch Accent Condition 2 (Fp2). (a) object presentation and (b) labeled spectrogram. By varying the owner of the object, speakers could be prompted to treat the owner of the circled object as new information and place focal pitch accent on it. The solid yellow line is the intensity contour, while the blue-dotted line is the pitch (F0) contour. Words or syllables in all capital letters indicate focal pitch accent on that word or syllable of word.***

2.4 Measurements

For each vowel of all the target words, the peak fundamental frequency (F0), maximum intensity, and the duration of each syllable of the target word were determined using the speech analysis application, Praat version 4.3.04 by Boersma and Weenink (2005). In this study, measurements were made of glottal source spectral parameters, using 512 DFT spectra of each target word vowel, at three different locations in the middle of the vowel that were at least 20ms apart. The spectra were constructed using a variable window size, depending on the average fundamental frequency of each speaker.

Spectral measurements of the first and second harmonics (H1 and H2, respectively), the first and second formant amplitudes (A1 and A3, respectively), as well as the frequencies of the first, second and third formants (F1, F2, and F3, respectively) were made for each vowel. Values obtained for H1 and H2 were corrected using a modified version of the correction formula proposed by Iseli and Alwan (2004) for the effect of F1 on H1 and H2 (Appendix A for more detail). The amplitude of the third formant (A3) was also corrected for the effect of F1 and F2, caused by vocal tract shape differences between vowels (Figure 8). The F3 of each vowel of a target word was 600Hz band-pass filtered and rated by the author for noise using a 7-point rating system, where a rating of 1 indicated evidence of no noise and a rating of 7 indicated completely noisy. Figure 9 shows the 7-point noise rating system which was adapted from the 4-point noise rating system used by Klatt and Klatt (1990), Hanson (1995 and 1997a), and Hanson and Chuang (1999). Utterances were pre-screened for the correct intonation. Only target words with vowels longer than 55ms in duration (both primary stressed and non-primary) were analyzed and used in the results reported in Section 2.5.

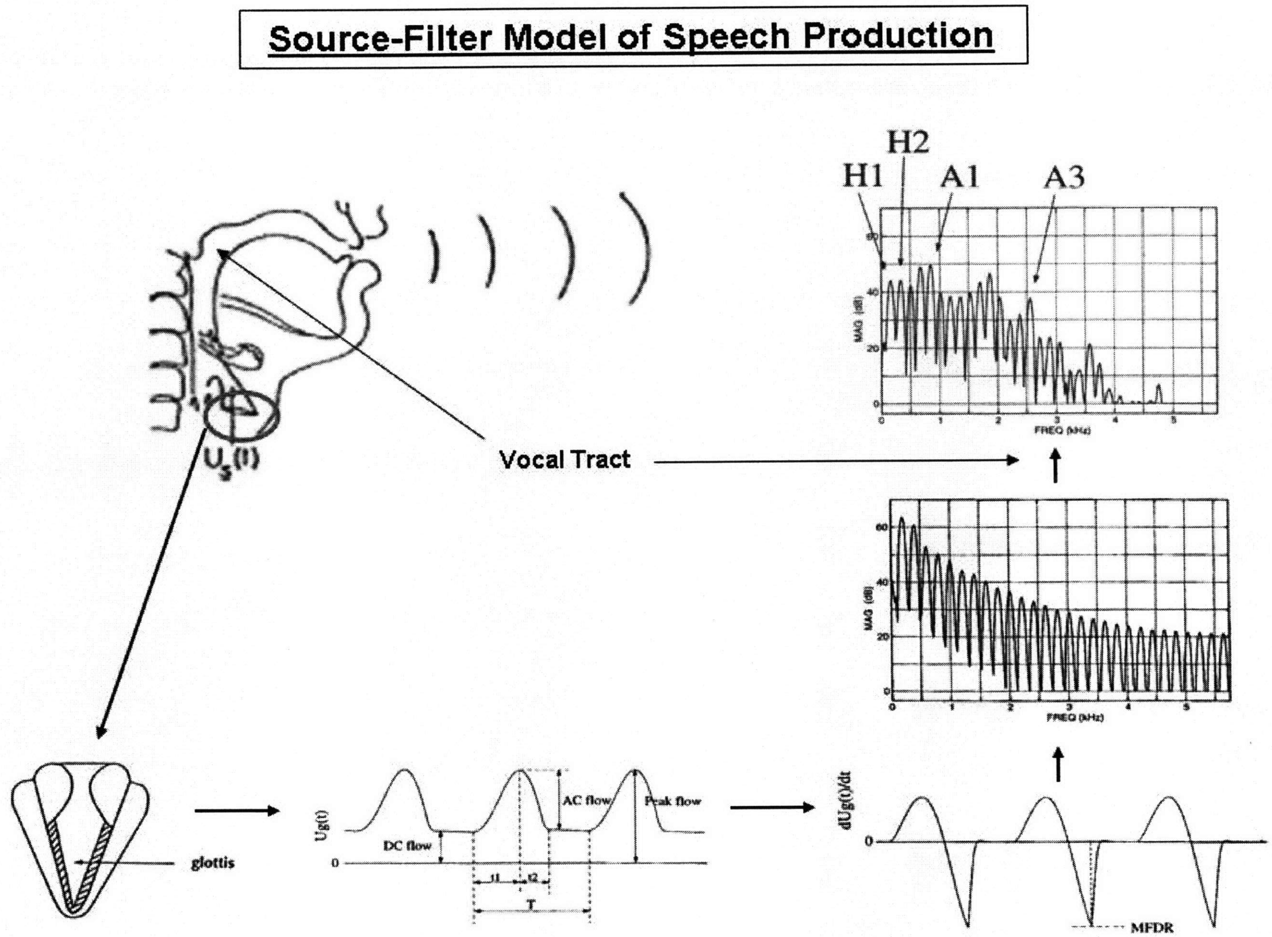


Figure 8. *Effect of vocal tract shape on glottal source spectrum (figures from Hanson, 1995; Stevens, 1998; and Hanson and Chuang, 1999).*

Filtered F3 Waveform Noise Rating

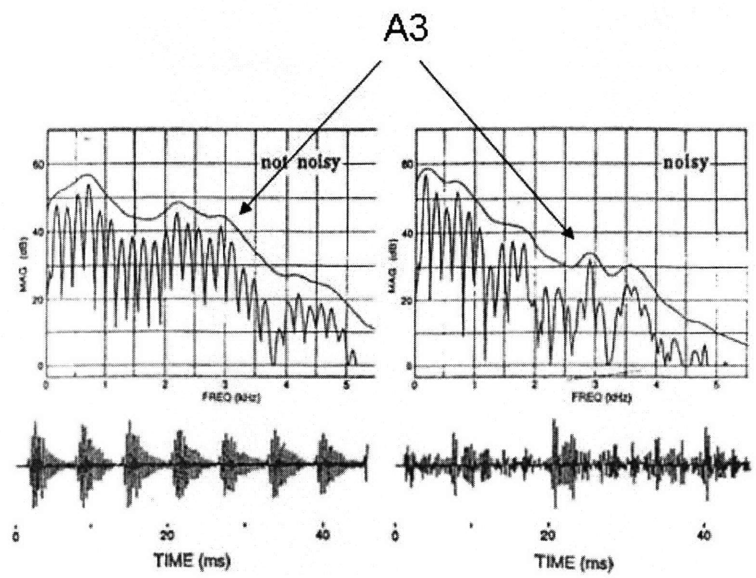


Figure from Stevens, 1998.

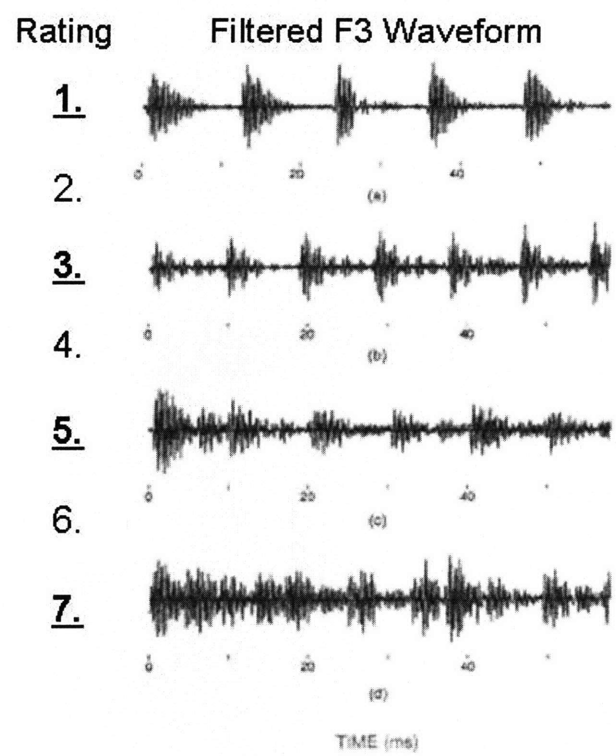


Figure from Hanson and Chuang, 1999.

Figure 9. ***The 7-point waveform noise rating system which was adapted from the 4-point noise rating system used by Klatt and Klatt (1990), Hanson (1995 and 1997), and Hanson and Chuang (1999).***

2.5 Results

Measurements made from the target words produced by the five speakers were organized into Tables shown in Appendix B according to the conditions in which they were produced (i.e., Fa, Fp1 and Fp2). The formant values obtained agreed with expected values for the vowels contained in the target words (Stevens, 1998). Although formant frequencies obtained for the vowels were within the expected value range for the novel words in all three conditions, the first formant (F1) of the primary stressed vowel in the novel words *'dada* and *da'da* was consistently greater than that of the non-primary full vowel (Tables 1-12 in Appendix B). However, this was not observed for the other novel word pairs. There were no consistent formant differences observed between primary stressed vowels and non-primary full vowels for the real words.

Syllable differences with regard to the remaining parameters were calculated from the values in these tables and graphed according to Figure 10. In this and later figures, what is graphed is the average speaker difference between the first syllable value and the second syllable value of the measured parameters (S1-S2). Thus if the value of the first syllable is greater, the difference is positive and if the second syllable has a larger value, the difference is negative. Equal values between the two syllables results in a difference of zero.

2.5.1 Correlates of Word Stress

Syllable difference values from the novel target words *'dada*, *'dodo*, and *'didi*, with first syllable primary stress, and their second syllable primary stress counterparts *da'da*, *do'do*, and *di'di*, revealed that consistent correlates of word stress do exist (Figure 11). The same correlates that distinguished primary stressed syllables from the non-primary full vowel syllables for novel words also correlated with word stress for the real words (Figure 12). These correlates of word stress are syllable differences in duration, spectral tilt (measured as H1*-A3*), and noise at high frequencies (indicated by the band-pass filtered F3 waveform ratings).

Results shown in Figures 11a-c and 12a-c illustrate how syllable differences in duration correlate to syllable prominence differences between the first and second syllables of the target words. When the first syllable has the primary stress it is greater than or equal to the duration of the second full vowel syllable. For the real words, the primary stressed first syllable was on average consistently longer in duration than the second full vowel syllable. This was not the case with the novel words, where in the non-pitch accented conditions Fp1 and Fp2, the primary stressed first syllable was often the same duration as the second syllable. The difference between the two types of words might be explained by noting that the primary stressed first syllable real word *statue* begins with a double consonant cluster, adding additional length to the first syllable. The first syllable of *sushi* contains the vowel /u/ which intrinsically has a longer duration than the vowel /i/. Thus it seems that it is the uncontrolled consonant-vowel composition of the real words that results in the observed primary stressed first syllable duration differences between novel and real words. However for both novel and real words, primary stressed second syllables were consistently longer than the preceding full vowel first syllable.

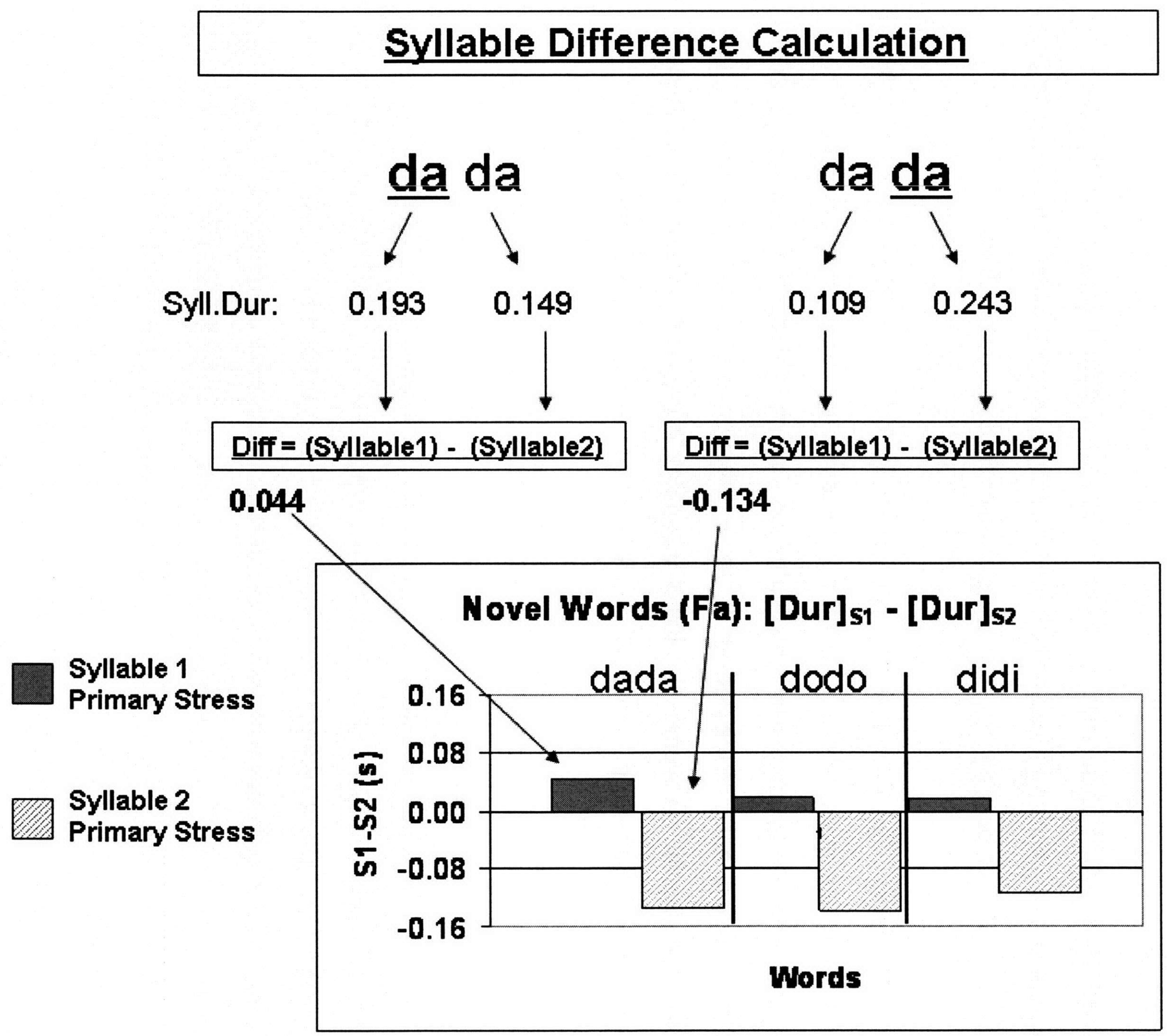


Figure 10. *Syllable difference calculation for production study measurements. What is graphed is the difference between the first syllable value and the second syllable value of the measured parameters (S1-S2).*

Syllable differences in the spectral tilt measurement H1*-A3* also distinguished the primary stressed syllable from the full vowel syllable, for both novel and real words, in all three pitch accent conditions (Figures 11d-f and 12d-f). In general the primary stressed syllable had less spectral tilt than the non-primary syllable. For both novel and real words, equal spectral tilt often corresponded to second syllable primary stress, with the exception being first syllable primary stressed *'didi* in the Fp2 condition. However, as will be demonstrated in Section 2.6, clearer measurements of spectral tilt can be obtained that more accurately depicts the spectral tilt syllable difference between primary stressed and non-primary full vowel syllables.

Figures 11g-i and 12g-i show that the average syllable difference in the band-pass filtered F3 waveform noise rating (Nw), which indicates relative amount of noise at high frequencies, accurately distinguishes the primary stressed syllable from the non-primary full vowel syllable. The syllable difference in noise rating goes in the same direction as that for H1*-A3*. That is, the primary stressed syllable on average has lower waveform noise ratings than the non-primary full vowel syllable for novel words, which have syllables with the same CV composition.

However, for real words two types of syllable differences seem to be captured by the Nw rating. The first is syllable differences in vowel composition. Note that for *statue* and *tattoo*, both having the vowel /u/ in the second syllable position, regardless of the syllable position of the primary stress, the second syllable had higher Nw ratings. For *sushi* and *bouquet*, both having the vowel /u/ in the first syllable position, it is the first syllable that consistently had higher Nw ratings. Thus syllables with /u/ in general have more noise at high frequencies. However, superimposed on this vowel distinction is the primary stress distinction. Notice that when the syllable with /u/ has primary stress, it has lower Nw ratings than the corresponding syllable with /u/ that is non-primary. Thus once syllable vowel differences are accounted for, primary stressed syllables can be distinguished from non-primary full vowel syllables using Nw ratings.

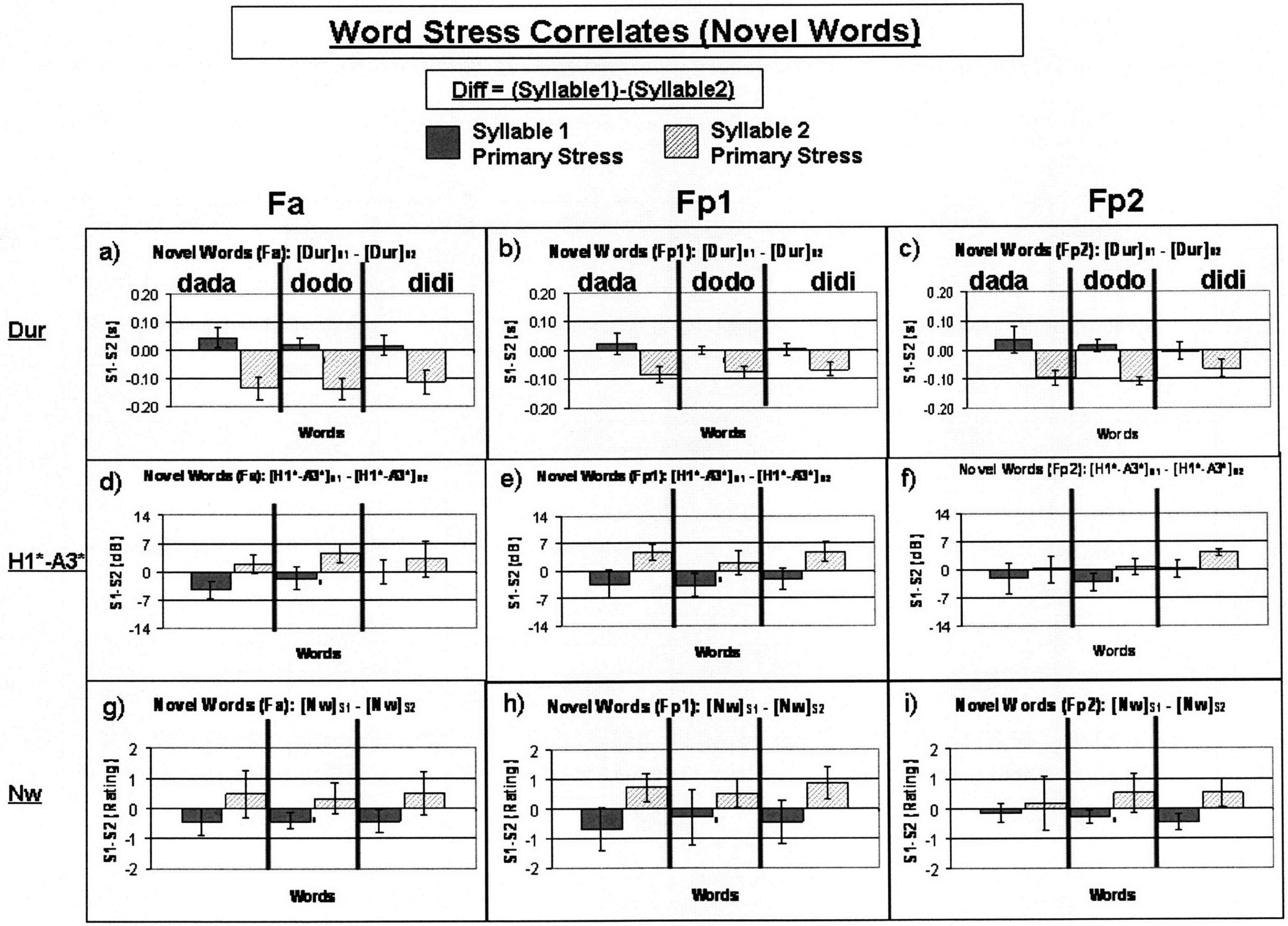


Figure 11. *Word stress correlates for novel words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

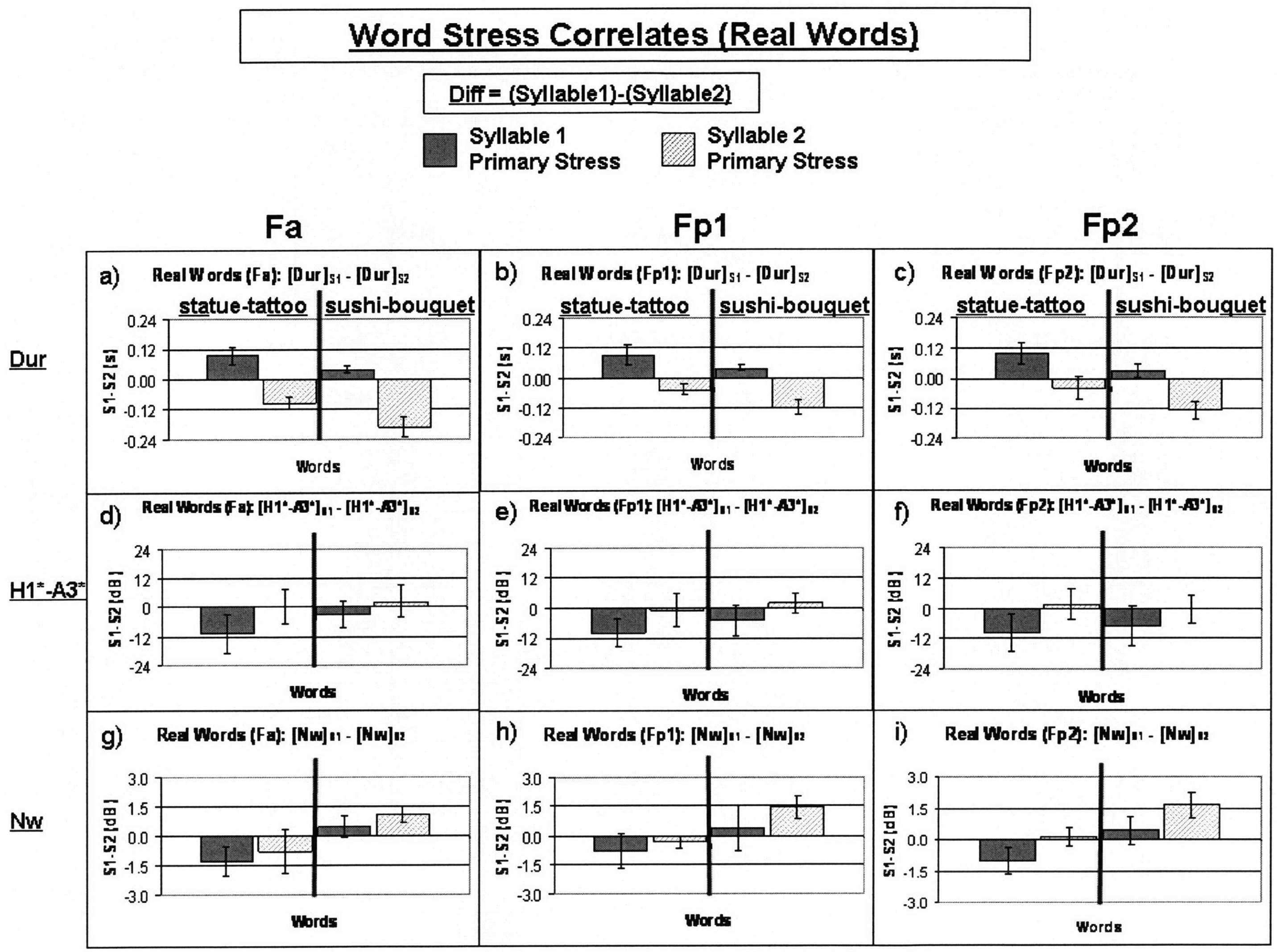


Figure 12. *Word stress correlates for real words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

2.5.2 Correlates of Pitch Accent

The same pitch accent correlates were found for both the target novel words and real words. These correlates only distinguished primary stressed syllables from non-primary full vowel syllables when the target word had high focal pitch accent. Syllable difference in peak fundamental frequency (F0), peak intensity, and amplitude of voicing, measured as H1*, all correlated to pitch accent. Figures 13 and 14 show that syllable differences in these parameters distinguished the more prominent syllable only in the Fa condition, when the target word had high focal pitch accent. This was true for both novel and real words.

Figures 13a-c and 14a-c show that syllable difference in F0 peak distinguished primary stressed from non-primary full vowels only in the Fa condition. In this pitch accented condition, the primary stressed syllable had the higher F0 peak. However, when the target word was not high focal pitch accented (i.e., Fp1 and Fp2 conditions), the first syllable had the higher F0 peak value, regardless of which syllable had primary word stress. This was true for both novel and real words. Furthermore, Figures 13b-c and 14b-c show that the further the high focal pitch accent is from the target word, the smaller the F0 peak difference is between the first and second syllables of the target word.

Syllable H1* differences also distinguished which of the syllables had the primary stress only in the Fa condition. Figures 13d-f and 14d-f show that like syllable difference in F0 peak, syllable difference in H1* was favored the primary stressed vowel only when the target word was high focal pitch accented in the Fa condition. However, when the target word was in the Fp1 and Fp2 conditions, the first syllable on average had the greater H1* value, regardless of which syllable had primary word stress. This was consistent for the novel, as well as the real words. As with the syllable difference in F0 peak, the further the high focal pitch accent is from the target word, the smaller the H1* difference is between the first and second syllables of the target word.

Another correlate of pitch accent was found to be syllable differences in peak intensity. Figures 13g-i and 14g-i show that only in the Fa condition does syllable difference in peak intensity accurately distinguish between primary stressed syllables and non-primary full vowel syllables. As with the other correlates of pitch accent, syllable difference in F0 peak and H1*, syllable intensity peak differences is positive in the Fp1 and Fp2 conditions, indicating that the first syllable had the greater intensity peak regardless of which syllable had the primary word stress. However, unlike the other correlates of pitch accent, the positive intensity peak difference between the syllables in the Fp1 and Fp2 conditions is smaller when the second syllable has the primary stress. Although this difference exists, it is also small, such that the syllable intensity peak difference when the first syllable has primary stress is often with 3dB of the syllable difference when the second syllable has the primary stress. At first glance this might seem like the same situation as with the correlate of word stress, Nw rating, however there are major differences.

One major difference between Nw rating and intensity peak is that when we control for the phonological composition of the syllables, as in the case with novel words, the first syllable bias for greater intensity peak in the Fp1 and Fp2 conditions does not disappear. A possible reason why the positive syllable intensity peak difference is smaller when the second syllable has primary stress is that primary stressed syllables tend to have more energy at high frequencies, as indicated by the spectral tilt, a correlate of word stress. This increased amplitude of high frequency harmonics, if large enough, can increase the overall intensity of the primary stressed second syllable vowel, relative to that of the first syllable, thereby decreasing the intensity peak difference between the two syllables. In order to know whether a positive syllable difference in intensity peak indicates first syllable or second syllable primary stress, we would have to know the contribution of mid to high frequencies to the overall amplitude. This however is a measure of spectral tilt, which we have shown to be a correlate of word stress. Thus knowledge of the syllable difference in intensity peak, which is positive in the Fp1 and Fp2 conditions, is not sufficient information to determine the primary stressed syllable.

2.5.3 Non-Correlates

Syllable differences in the parameters H1*-H2*, an approximation of open quotient, and H1*-A1*, an approximation of F1 bandwidth, did not correlate to either word stress or pitch accent. Figures 15 and 16 show that in none of the three pitch accented conditions (i.e., Fa, Fp1, and Fp2) did syllable differences in either H1*-H2* or H1*-A1* consistently distinguish the primary stressed syllable from the non-primary full vowel syllable. Thus it seems that syllable differences in open quotient and F1 bandwidth, approximated as H1*-H2* and H1*-A1* respectively, are not parameters that native speakers of American English consistently use to convey prosodic information, at least at the word level.

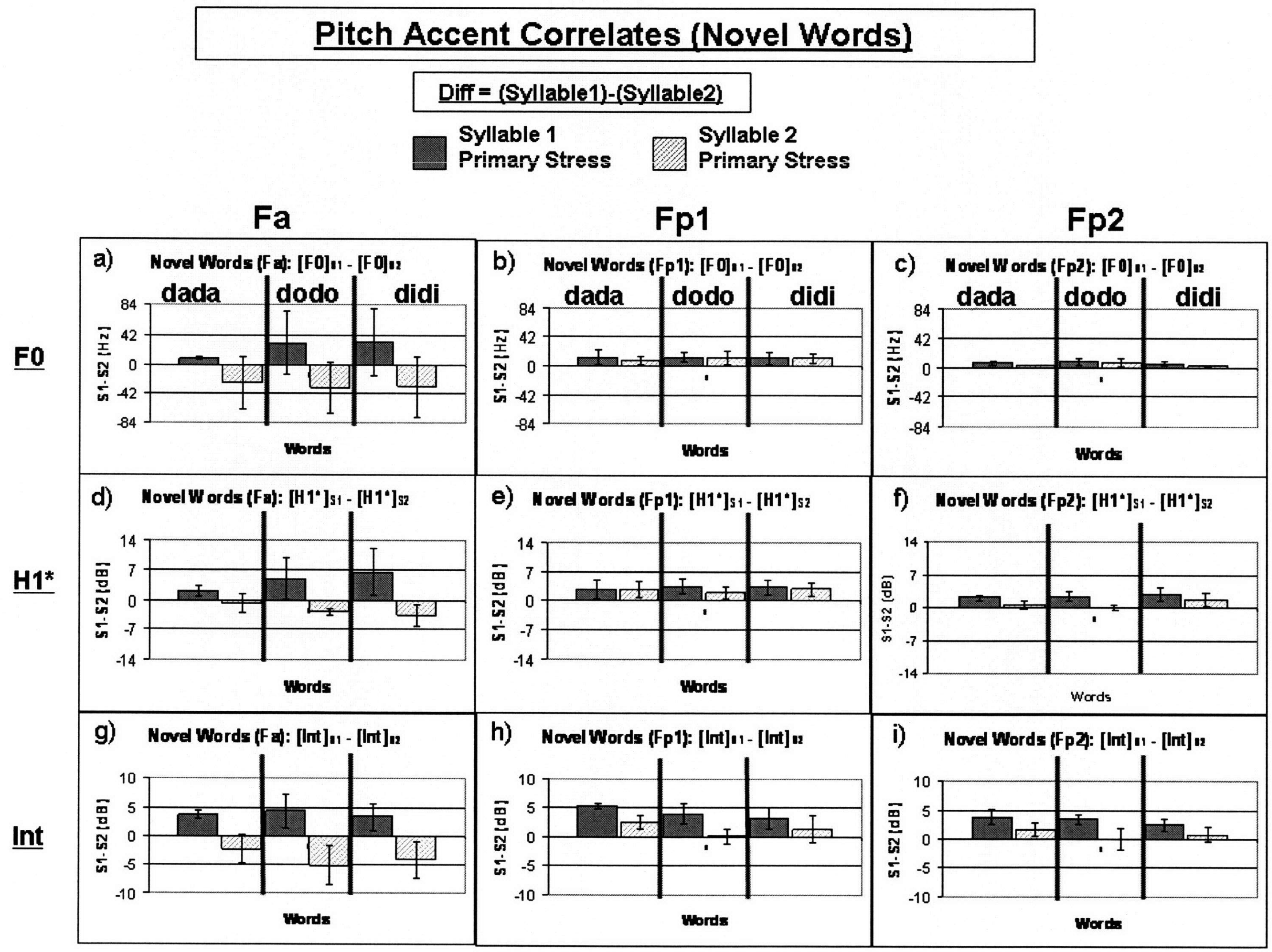


Figure 13. *Pitch accent correlates for novel words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

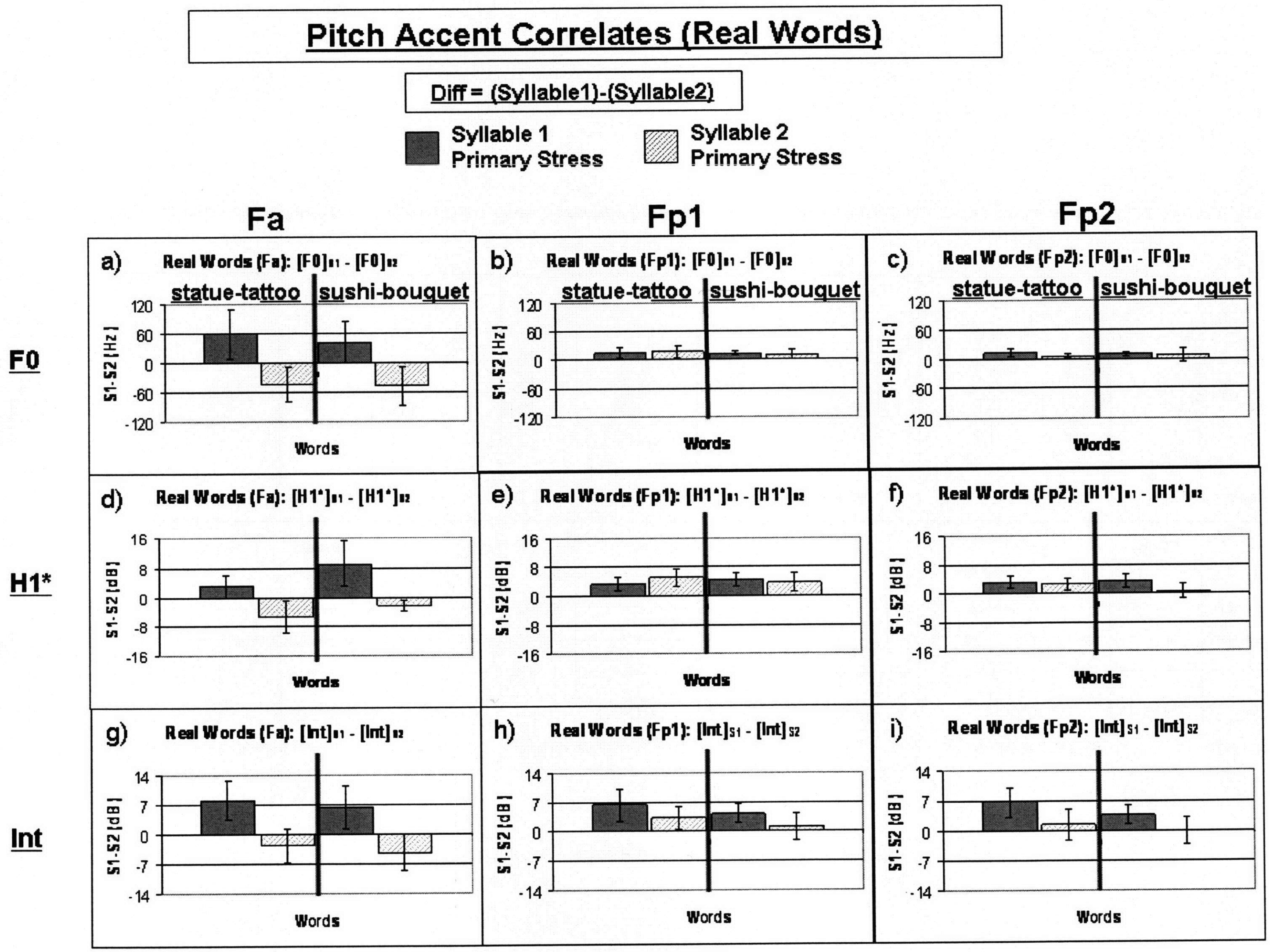


Figure 14. *Pitch accent correlates for real words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

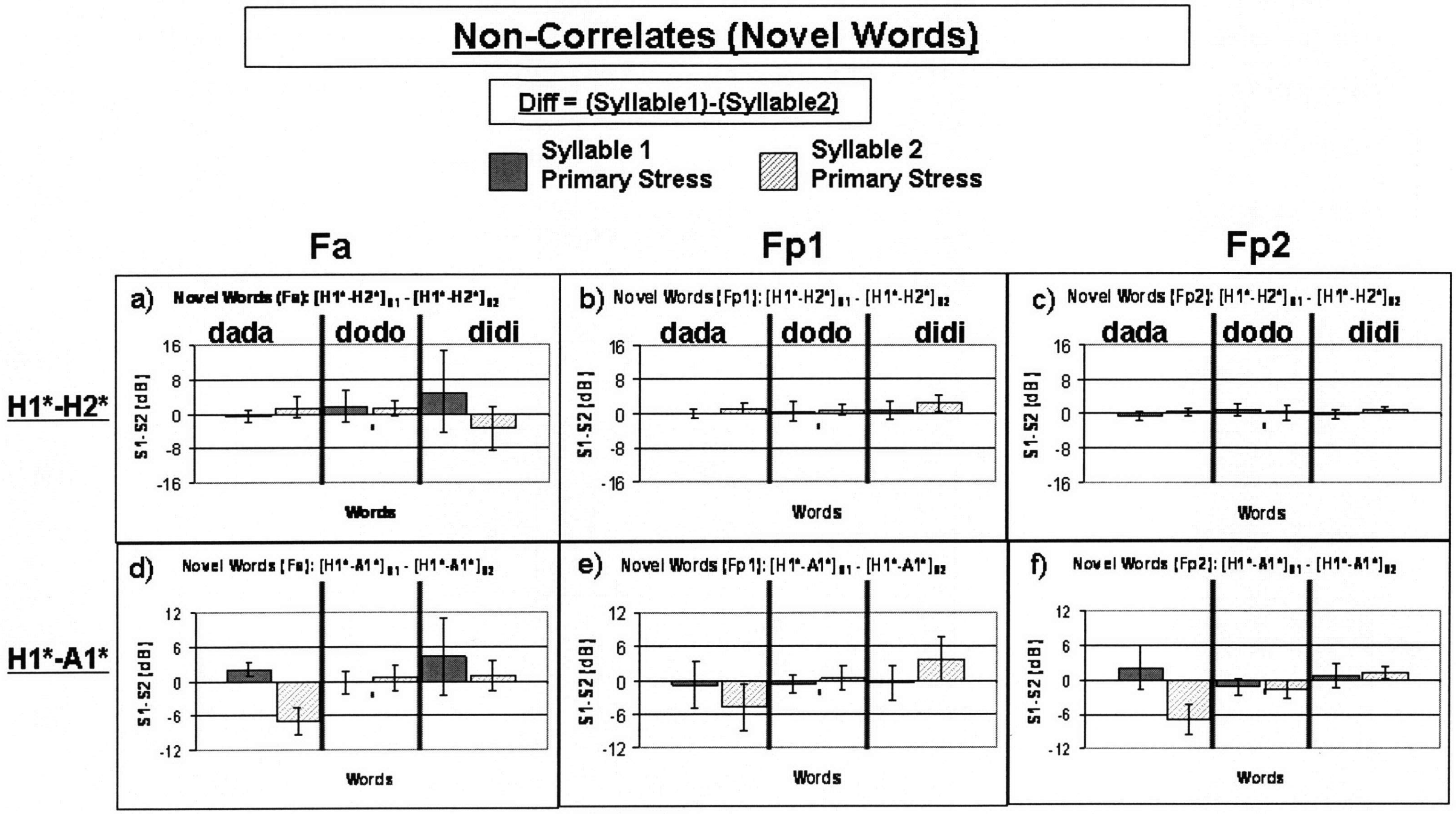


Figure 15. *Non-correlates for novel words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

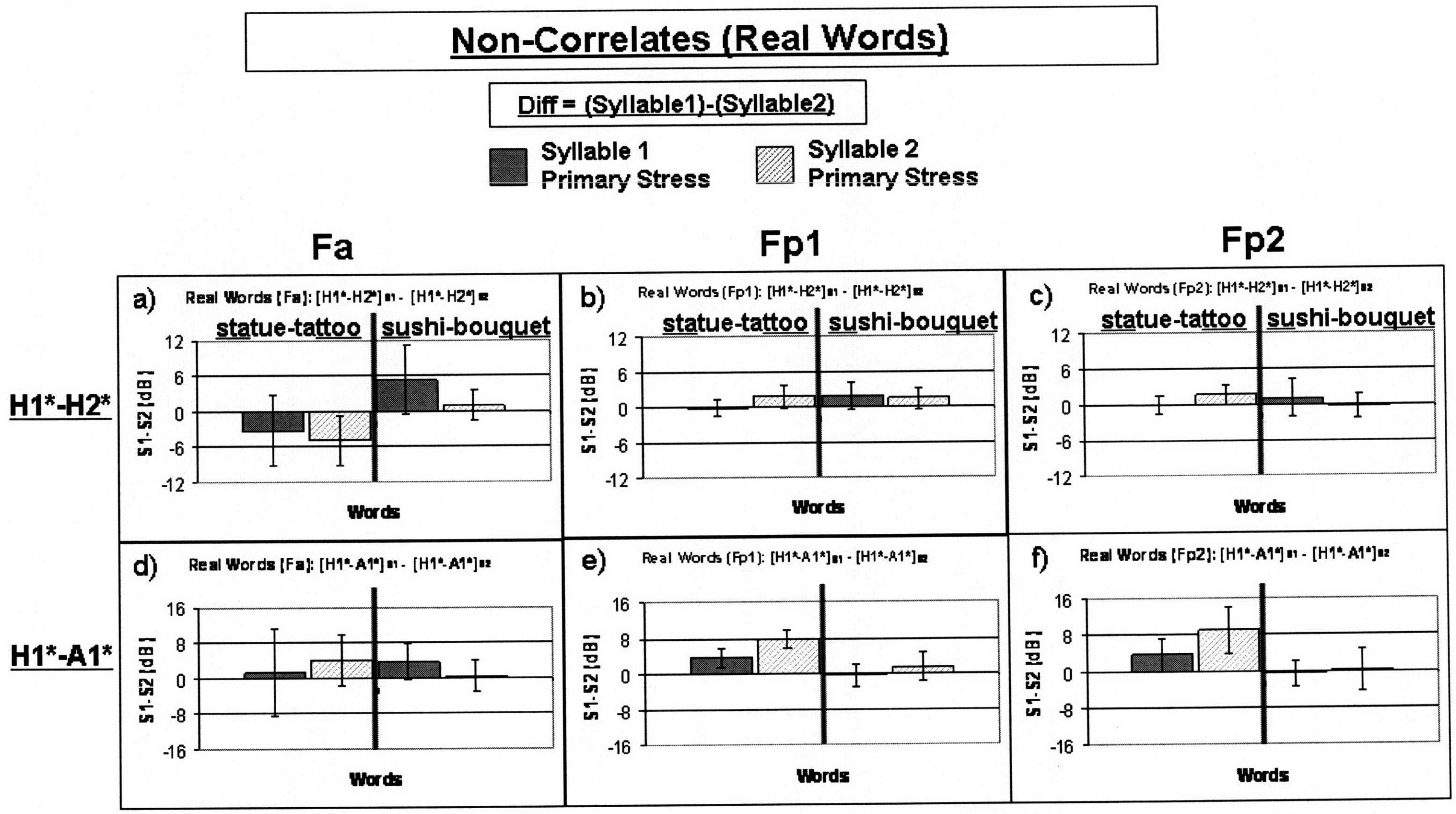


Figure 16. *Non-correlates for real words. The difference between the average first syllable value and the average second syllable value (S1-S2).*

2.6 High Focal Pitch Accent Effect on H1* Value

2.6.1 Changes in H1 Due to Pitch Accent Location and Proximity*

The results from Section 2.5 indicate that syllable difference in H1*-A3* is a correlate of word stress, even though H1* is by itself a correlate of pitch accent. Is it possible that the word stress distinction between primary stressed syllables versus non-primary full vowel syllables, in terms of H1*-A3*, is due to a combination of changes in H1* and A3*? Or is it that changes in H1*, which correlate to pitch accent, is some how confounding the spectral tilt measurement H1*-A3*, that correlates to word stress? How can we determine the part or parts of the measurement H1*-A3* that are contributing to the word stress syllable difference in spectral tilt, measured as H1*-A3*?

If we just look at H1* measurement differences between focal pitch accented primary stressed vowels and unaccented non-primary full vowels, we should see that accented vowels have higher values of H1*, since H1* is a correlate of pitch accent. This seems to be the case, as is shown in Figure 13d. When neither the primary stressed nor the non-primary full vowel was accented, no consistent difference in H1* was observed based on the primary word stress status of the vowel, since H1* is not a correlate of word stress. This is shown in Figures 13e-f. Interestingly, if we look at the change in H1* value of a particular syllable of a target novel word (i.e., the first or second syllable) as a function of pitch accent location, we find that H1* does not remain constant. Figure 17a shows the change in average H1* of the full vowel in the second syllable of all the novel target words, as a function of focal pitch accent position. As Figure 17a clearly shows that the average H1* value decreases as the distance of the focal pitch accent from the target word syllable of interest increases. The pattern is relatively consistent for all the novel target words. The value of H1* seems to stabilize when the high focal pitch accent is located about two syllables before the syllable vowel of interest and remain relatively unchanged when the focal pitch accent is three syllables in front of the syllable vowel of interest.

From Figure 17a, we can see that on average the high focal pitch accent increases the H1* value of a full vowel about 8dB from the base value observed when the focal accent is located three syllables preceding the full vowel syllable of interest. When the high focal pitch accent is located one syllable in front of the syllable of interest, that syllable's H1* value is about 2dB greater than the average base value of 32.7 dB. These results agree with findings from Stevens (1994) and Hanson (1997b), which showed that non-reduced vowels had reduced amplitudes following a nuclear pitch accent compared to when the vowels were themselves pitch accented. Figure 17b illustrates the effect of high focal pitch accent on the fundamental harmonic (H1) as a function of distance from the target word syllable. Thus the pattern of H1* differences shown in Figure 13 for the novel words can mostly be accounted for by the proximity and location of the focal pitch accent. It is also possible that the number of consonants or types of consonants between the syllables would affect the rate of decline of the focal accent effect on H1. Nevertheless, this finding rules out H1* as the cause of the spectral tilt difference observed in Figures 11d-f.

Change in H1* Value with Pitch Accent Location

a)

Accent Location:	Fps3	Fps2	Fps1	Fa
H1* Value for di:	32.7	32.8	35.1	40.5
	my	blue	di	di	drove here.

b)

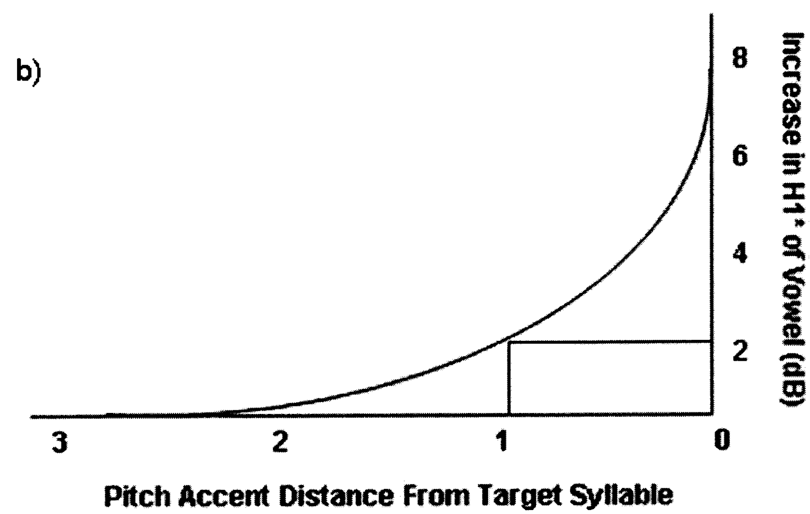


Figure 17. *Change in H1* value with pitch accent location. a) shows the change in average H1* of the full vowel in the second syllable of all the novel target words, as a function of focal pitch accent position from the second syllable of interest. b) shows a schematic of the effect of high focal pitch accent on the fundamental harmonic (H1) as a function of distance from the target syllable in terms of syllable distance. Fps3, Fps2, and Fps1are the conditions when the pitch accent is thre, two, and one syllable from the target second syllable, respectively. The target second syllable is pitch accented in the Fa condition.*

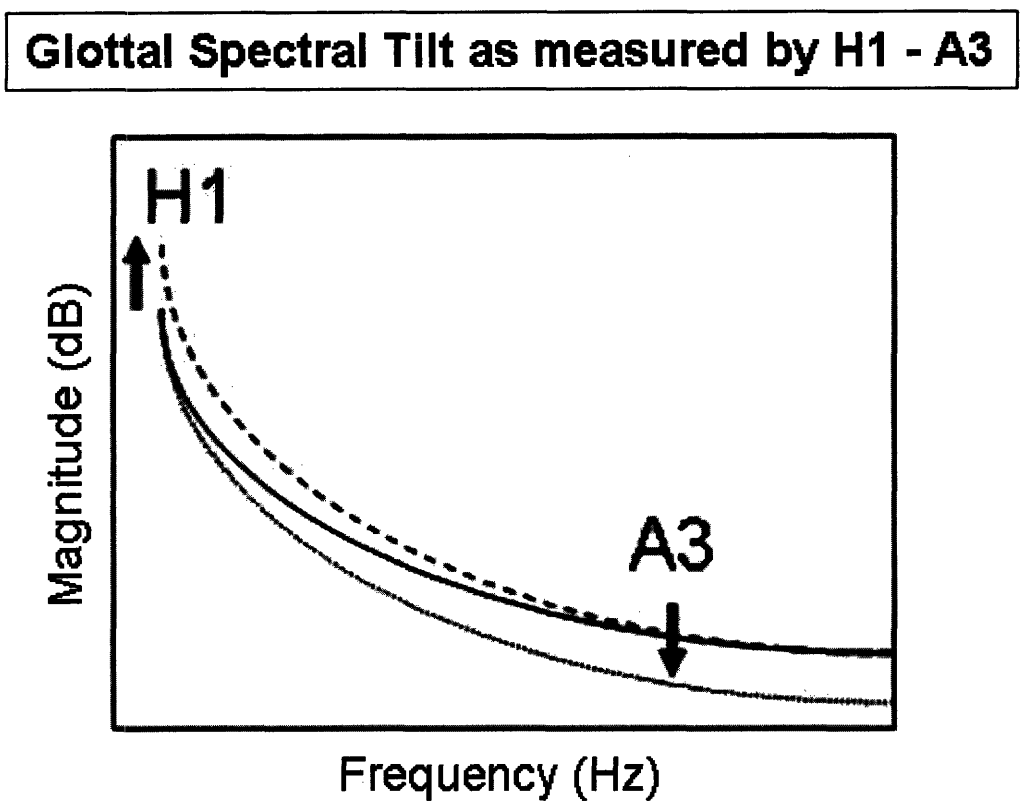


Figure 18. ***A change in H1*-A3*, can be due to either change in the H1* value or changes in A3*.***

Since H1* has been ruled out as the cause of the glottal spectral tilt (H1*-A3*) difference between primary stressed and non-primary full vowel syllables, how can it be determined that the difference is due to a decrease in A3* (Figure 18)? As discussed in Chapter 1, non-abrupt closure of the vocal folds during phonation causes the amplitude of the harmonics at higher frequencies to decrease, resulting in the increased presence of noise at those frequencies. Thus lower values of A3*, for non-primary full vowels, should result in greater evidence of noise in the region of the third formant (F3) for all three focal pitch accented conditions tested in the novel word. Figure 12g-i shows the results of the waveform noise rating for all three conditions for the novel words. This suggests that the measurement H1*-A3* can and should be corrected for the effect of high focal pitch accent on H1* in order to use it to more accurately differentiate between the primary stressed and non-primary full vowels in a two-syllable word.

2.6.2 Correction for the Effect of High Focal Pitch Accent on Spectral Tilt Measurement

If we assume, according to Section 2.6.1, that the H1* differences between the primary stressed and non-primary full vowels (ΔH1*), as shown in Figure 14, are predominantly due to the presence, location, and proximity of the high focal accent, then we can correct for the effect of the high focal pitch accent on syllable difference in spectral tilt (ΔST, where ST = H1*-A3*) between the two vowels by subtracting from it ΔH1*. Equation 1 illustrate the ΔST correction for H1* difference due to high focal accent.

$$\Delta ST^{*} = \Delta ST - \Delta H1^{*} \qquad \text{Eq. 1}$$

where ΔST* is the corrected spectral tilt measurement.

A hypothesis arising from the correction of ΔST for the effect of focal accent is that, because of possible physiological constraints, the glottal events giving rise to the high focal pitch accent, such as increased pressure difference across the glottis and or increased open quotient, cannot be instantaneously stopped or reset. The result is that for the Fp1 and Fp2 conditions the residual effects of these events continue from the preceding vowel into the target word. A prediction of this hypothesis is that the first syllable of the target word would be the most affected, especially if it has primary stress and produced more modally. Another prediction would be that the effect of the events giving rise to the high focal pitch accent would decrease with increasing distance from the accent. Figure 13a-c supports this hypothesis.

Thus the ΔST correction should be applicable to all three focal accented conditions (i.e., Fa, Fp1, and Fp2). However, it should be most effective when the focal pitch accent is on the target word, since this is when the change in H1* from its "default" value is greatest. Implementation of Equation 1 on the spectral tilt difference results shown in Figures 13d-f and 14d-f, using the ΔH1* results shown in Figures 13a-c and 14a-c, respectively, is illustrated in Figures 19 and 20. Figures 19 and 20 shows that when the effect of the pitch accent on H1* is accounted for, spectral tilt differences between the vowels of a two syllable word can be better observed using the correction for the effect of high focal pitch accent on the spectral tilt measurement H1*-A3*.

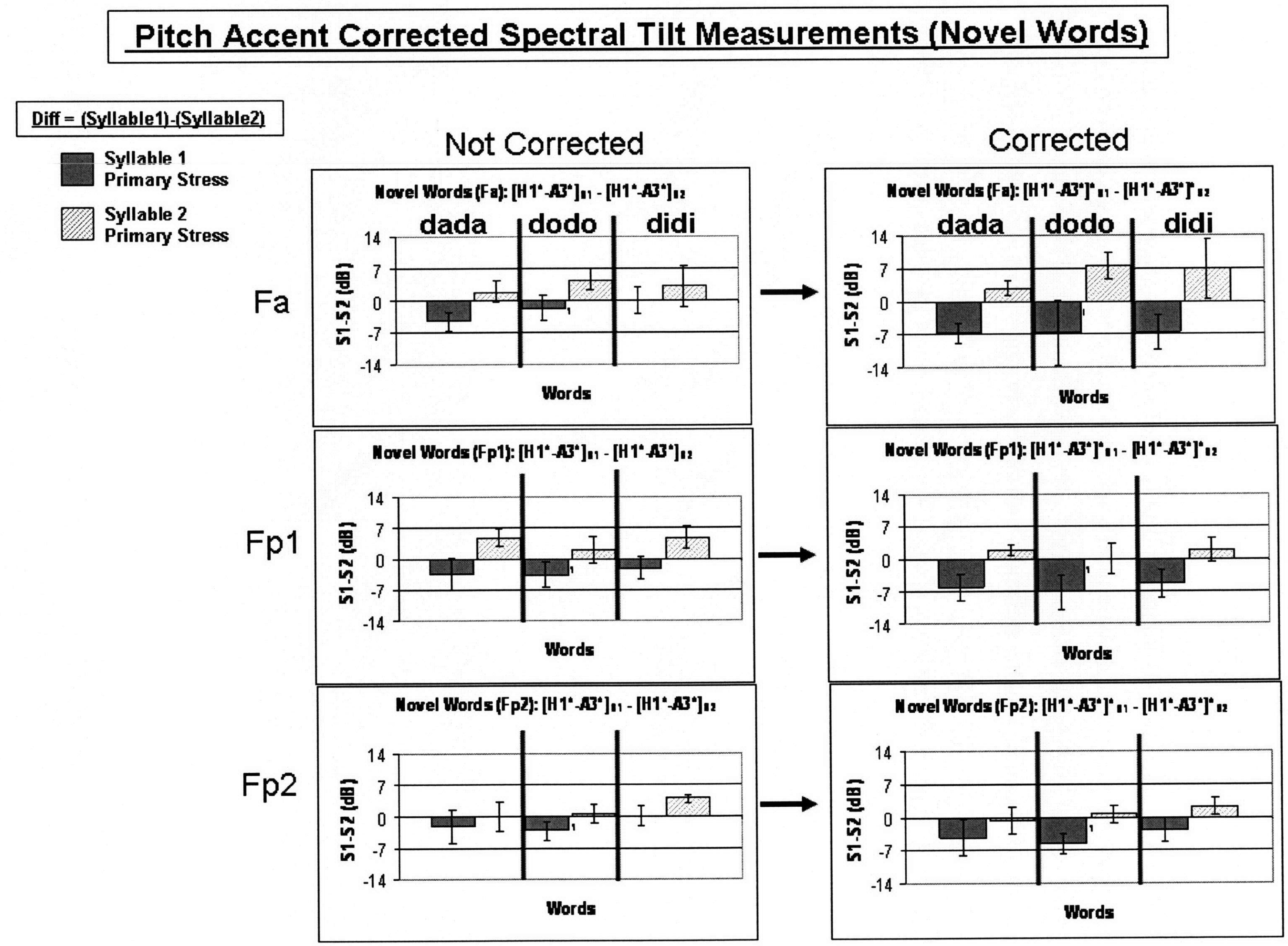


Figure 19. *Novel word spectral tilt measurements corrected for the effect of pitch accent on H1*. The difference between the average first syllable value and the average second syllable value (S1-S2).*

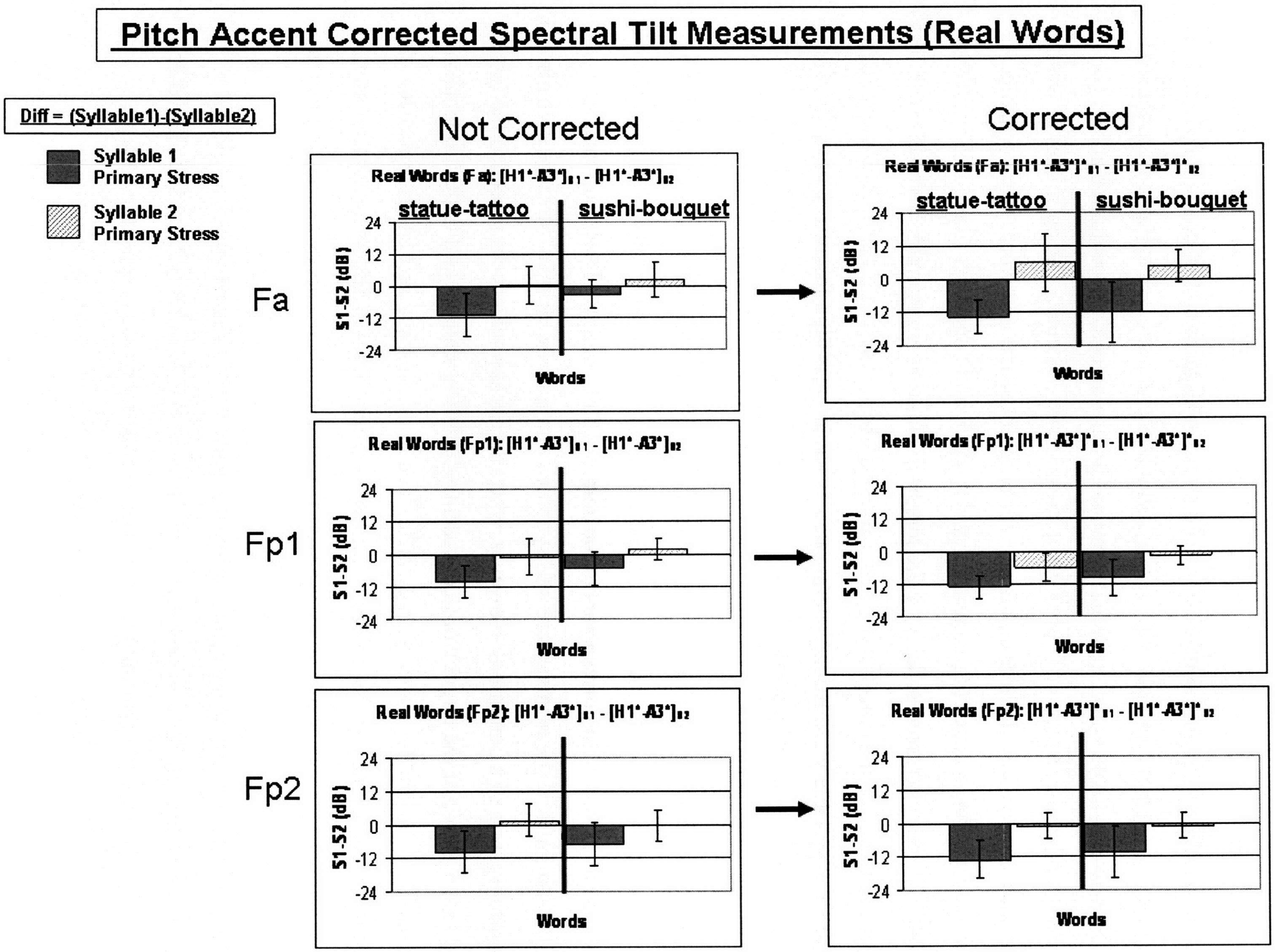


Figure 20. ***Real word spectral tilt measurements corrected for the effect of pitch accent on H1*. The difference between the average first syllable value and the average second syllable value (S1-S2).***

2.7 Discussion

From the results we observed that the parameters measured in this production study can be broken up into three groups: correlates of word stress in all three conditions, Fa, Fp1, and Fp2; correlates of pitch accent, that only distinguish the primary stressed syllable from the non-primary full vowel syllable when the target word has phrase-level high focal pitch accent (i.e., condition Fa); and non-correlates of either word stress or pitch accent. The correlates of word stress that were observed in all three conditions for both novel and real words were syllable differences in duration, spectral tilt, measured as H1*-A3*, and band-pass filtered F3 waveform noise ratings. Primary stressed syllables were longer in duration and contained vowels with less spectral tilt compared to the non-primary full vowel syllable in the same word. The vowel of a primary stressed syllable was also in general rated as having less high frequency noise than the non-primary full vowel syllable of the same word.

Correlates of word stress only when the target word was high focal pitch accented (i.e., pitch accent correlates) were found to be syllable differences in peak F0 and intensity within the vowel, as well as H1*, which corresponds to the amplitude of voicing. These parameters accurately distinguished the primary stressed syllable from the non-primary stressed syllable of a target word only in the Fa condition. However, when the focal pitch accent preceded the target word, the first syllable of the target word consistently had the greater peak F0, peak intensity, and H1* values. The smaller peak intensity difference in the Fp1 and Fp2 conditions, when the second syllable has primary stress, might be due to the effect of focal pitch accent proximity on H1* combined with the fact that primary stressed vowels have more energy at high frequencies.

To elaborate, a non-pitch accented primary stressed first syllable vowel would be expected to have more energy at high frequencies than the non-primary second syllable vowel. Depending on how large the spectral tilt difference between the two vowels, this energy difference at high frequencies can contribute to the overall peak intensity difference. Furthermore, since the first syllable is always closer in proximity to the focal pitch accent in the Fp1 and Fp2 conditions, it would be expected, according to section 2.6 and based on the results, to have a higher H1* value. This would further increase the intensity difference between the primary stressed first syllable and the non-primary second syllable, leading to the first syllable having a greater peak intensity (Figure 13h-i). If the second syllable has the primary stress, it would in general have less or equal spectral tilt as the non-primary first syllable in the same word, thereby neutralizing one of the two sources that gave the first syllable greater peak intensity when it had primary stress. Since the first non-primary first syllable will still have a greater H1*, because it is closer to the pitch accent in the Fp1 and Fp2 conditions, it is expected to still have the greater peak intensity, since energy at low frequencies contribute more to the overall amplitude than energy at high frequencies. However, the syllable difference, when the second syllable has primary stress, will not be as great, that is more positive.

The non-correlates of either word stress or pitch accent were the spectral approximations of open quotient, H1*-H2*, first formant bandwidth, H1*-A1, as well as formant differences between the primary stressed syllable and the non-primary full vowel syllable of a word. Thus it does not seem that, for the real words tested in this study, vowel formant differences, in this case for [u], allow us to determine the word stress pattern of the word. Interestingly, for one minimal stress pair of novel words, *'dada* and *da'da*, the primary stressed vowel consistently had a higher F1 frequency (See Appendix B). This larger F1 value for the primary stressed vowels of the novel words *'dada* and *da'da* is consistent with the effects of opening the mouth wider. It might have been easier for speakers to indicate the relationship between the vowels of the two syllables by opening the mouth wider, since the production of the vowel /a/ does not require rounding, as in the production of /o/ and /u/, or narrowing a region of the oral cavity, as in the production of the vowel /i/. Further explanation is given in Chapter 4.

The spectral approximations of open quotient, H1*-H2*, and first formant bandwidth, H1*-A1, do not clearly distinguish between primary stressed and non-primary full vowel syllables. Perhaps changes in H1* are confounding the results for H1*-H2* and H1*-A1*. However, analysis of the individual average speaker values and overall average H2* and A1* values shown the tables in Appendix B, suggest that in most cases changes in H2* and A1* do not correlate with either word stress or pitch accent. Overall, the differences between the primary stressed syllable and the non-primary full vowel syllable of a word, in terms of the correlates of word stress, were more distinct once the CV composition of the target words were controlled, as in the case with the novel words.

3. Perception Study: Individual and Co-variation of Word Stress Correlates

3.1 Listeners

A total of fourteen native speakers of American English participated in this perception study. Six of the participants were involved in both of the syllable prominence judgment tasks described below. A subset of the listeners were also involved in a naturalness rating task using the stimuli from the syllable prominence judgment tasks. All the participants were male and between 18 and 50 years of age, with no history of language disorder or speech therapy. Listeners were chosen to match the speakers who participated in the production study and some of them were also involved in the production study discussed in Chapter 2. As with the production study, listeners were compensated for their involvement in this perception study. All listeners were tested in the same sound insulated booth, where the production studies were conducted. Stimuli were presented through headphones at a sound level comfortable for each listener.

3.2 Synthesis of Stimuli

3.2.1 Stimuli for Individual Variation of Word Stress Parameters

The software application KLSYN88 was used to manipulate word stress acoustic parameters. In order to determine if listeners were influenced by syllable differences in the KLSYN88 parameters that corresponded to duration (KLSYN88 parameter DU), spectral tilt (KLSYN88 parameter TL), and aspiration noise (KLSYN88 parameter AH), a novel word "*dada*", with syllables that varied in these parameters, was synthesized and concatenated into the declarative carrier phrase "Your blue [dada] drove here." The carrier phrase was spoken by a male native speaker of American English, with high focal pitch accent on the first word of the phrase, as in the Fp2 condition discussed in Chapter 2. The novel word "*dada*" was copy synthesized from the same male speaker and was the only part of the carrier phrase that was synthesized. Figure 21 shows a comparison of the spectrum and waveform of the real vowel /a/ with the spectrum and waveform of the synthesized vowel.

The syllable difference in the word stress corresponding parameters, duration (approximated using the KLSYN88 parameter DU), spectral tilt (approximated using the KLSYN88 parameter TL), and noise at high frequencies (approximated using the KLSYN88 parameter for aspiration noise, AH) of the first and second syllables of the synthesized "*dada*" were individually manipulated such that there were differences between the two syllables. For each of the word stress corresponding parameters, the difference between the vowels of the first and second syllables of "*dada*" could have 1 of 17 values. When the parameter of a syllable was varied, the same parameter for the other syllable was kept constant at the designated minimum value.

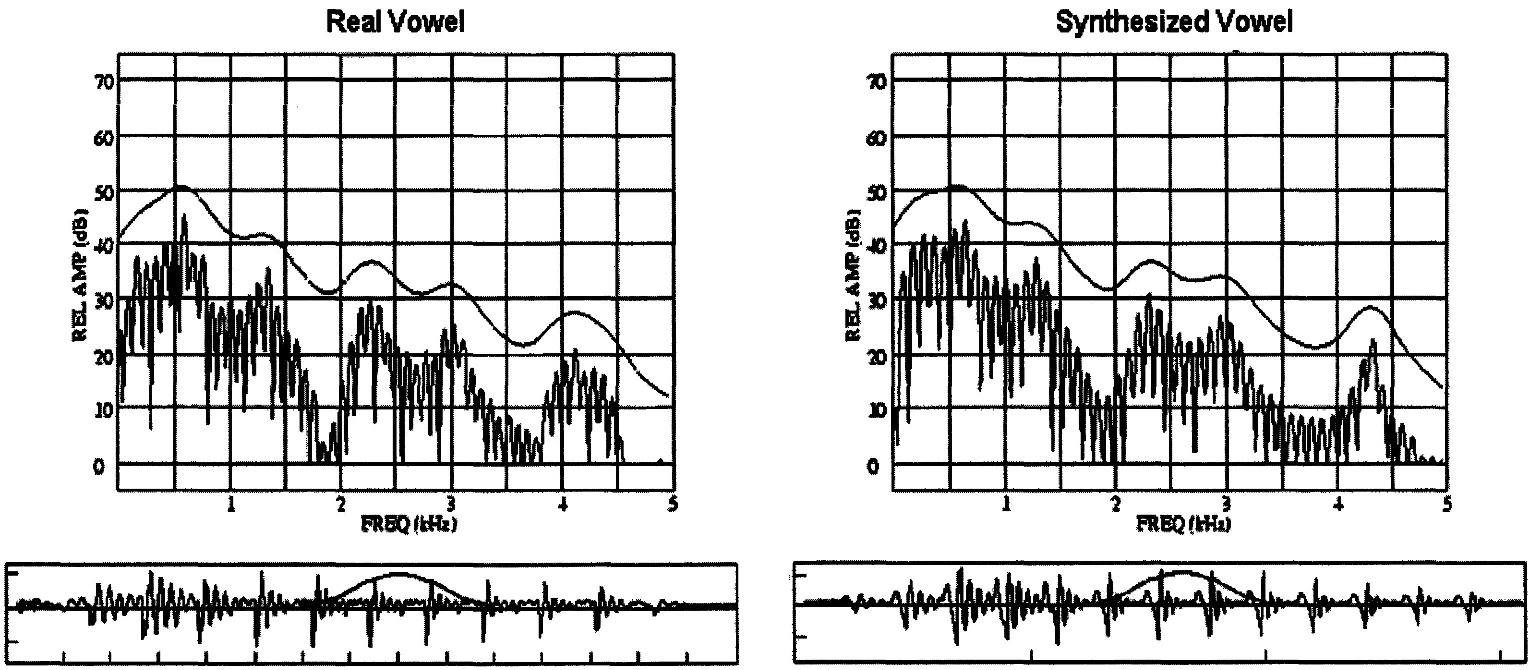


Figure 21. *Comparison of spectrum and waveform from the vowel of the first syllable of the real "dada" from the male carrier phrase speaker and the copy synthesized vowel spectrum. The copy synthesized vowel was waveform and spectrum matched to the real vowel property, before changing the KLSYN88 parameters corresponding to the word stress correlates (DU, TL, and AH)..*

The consonant-vowel (CV) composition of the synthesized "*dada*" was such that the first and second syllables had exactly the same acoustic production of the onset [d], while the vowels of the two syllables varied in one of the three acoustic parameters tested in this perception study. Acoustic parameters corresponding to the pitch accent correlates and non-correlates found in the production study of Chapter 2 were kept constant at the values observed for the male carrier phrase speaker during his production of the novel word *'dada* in the Fp2 condition of Chapter 2. The F0 started at 95Hz at the beginning of the vowel for the first syllable and dropped at a rate of 1Hz/20ms. The F0 for the second syllable started at 90Hz and declined at the same rate. Other parameters measured in the production study, such as formant values, H1 and H2 were also kept constant.

For the KLSYN88 parameter corresponding to duration, DU, the two syllables of "*dada*" could differ in DU by 0ms, 20ms, 30ms, 45ms, 60ms, 75ms, 90ms, 105ms, or 120ms. The minimum syllable duration was 150ms, which was the value both syllables had when the DU syllable difference was 0ms. The increase in DU of a syllable was accomplished by lengthening the vowel portion by 20ms for the first step, 10ms for the second step, and 15ms intervals afterwards. The 20ms was chosen as the minimum difference between syllables in order to insure that each incremental change in syllable DU also involved a change in the number of glottal pulses generated within the vowel of the syllable being manipulated. Thus given that the second syllable of the synthesized "*dada*" had a fundamental frequency (F0) starting at 90Hz and declined at a rate of 1Hz/20ms, 20ms DU increase from the minimum duration of 150ms insured that an additional glottal pulse was also generated. During changes in the duration (DU) difference between the two syllables of "*dada*", the syllable difference in the parameter TL was held constant with the second syllable having 2dB more TL then the first syllable. Syllable difference in the parameter AH was held constant at zero.

For the KLSYN88 parameter corresponding to spectral tilt, TL, the two syllables of "*dada*" could differ in TL by 0dB to 16dB, in 2dB steps. The minimum syllable TL was 0dB, which was the value both syllables had when their difference in TL was 0dB. The maximum TL a syllable could have was 16dB, because further increase in TL, using KLSYN88, resulted in changes in the overall amplitude of the vowel spectrum. During changes in the TL difference between the two syllables of "*dada*", the syllable difference in the parameter DU was held constant with the second syllable being 30ms longer, while syllable difference in the parameter AH was held constant at zero.

The KLSYN88 parameter corresponding to aspiration noise, AH, could differ between the two syllables of "*dada*" by 0dB to 16dB, in 2dB steps. The minimum syllable AH within the vowel region was 35dB, which was the value both syllables had when their difference in AH was 0dB. The maximum AH a syllable could have was 51dB, because further increase in AH, using KLSYN88, resulted in changes in the overall amplitude of the vowel spectrum. Further changes in AH also resulted in distinctly unnatural sounding speech. During changes in the AH difference between the two syllables of "*dada*", the syllable difference in the parameter DU was held constant with the second syllable being 30ms longer, while syllable difference in the parameter TL was held constant with the second syllable having 2dB more TL then the first syllable.

3.2.2 Stimuli for Co-variation of Word Stress Parameters

For the syllable prominence judgment task involving co-variation of the word stress parameters, the novel word "*dada*" was once again synthesized and concatenated into the declarative carrier phrase "Your blue [dada] drove here." The carrier phrase was identical to the one used for the individual variation of the word stress parameters and contained the high focal pitch accent on the first word of the phrase, as in the Fp2 condition. As before, the novel word "*dada*" was the only part of the carrier phrase that was synthesized.

The parameters corresponding to word stress correlates, duration (represented by DU), spectral tilt (represented by TL), and noise at high frequencies (represented by AH), were manipulated as described in Section 3.2.1. The KLSYN88 parameters corresponding to pitch accent correlates and non-correlates found in Chapter 2 were kept constant in the manner discussed in Section 3.2.1. However, for these syllable prominence judgment task stimuli, the KLSYN88 parameters were co-varied, such that there were a total of 343 possible unique tokens. For the KLSYN88 parameter corresponding to duration, DU, the two syllables of "*dada*" could differ in DU by 0ms, 30ms, 75ms, or 120ms. The minimum syllable duration was once again 150ms, which was the value both syllables had when the DU syllable difference was 0ms. These syllable differences in DU are a subset of the DU values used in Section 3.3.1.

For the KLSYN88 parameter corresponding to spectral tilt, TL, the two syllables of "*dada*" could differ in TL by 0dB, 2dB, 8dB and 16dB. The minimum syllable TL was 0dB, which was the value both syllables had when their difference in TL was 0dB. The maximum TL a syllable could have was 16dB, because of the effect of further increase in TL on the overall amplitude of the vowel spectrum. The KLSYN88 parameter corresponding to aspiration noise, AH, could differ between the two syllables of "*dada*" by 0dB, 2dB, 8dB and 16dB. The minimum syllable AH within the vowel region was 35dB and the maximum AH a syllable could have was 51dB, for the reasons discussed in Section 3.2.1.

3.3 Experiment Design

3.3.1 Syllable Prominence Judgment Tasks

The purpose of this portion of the perception study was to determine if the word stress correlates, found in the production study of Chapter 2, were perceptually realized as such by listeners when varied as individual parameters and when co-varied. Individual variation of the KLSYN88 parameters corresponding to the word stress correlates allowed us to determine how listeners' judgment of syllable prominence is influenced by syllable differences in these parameters, in an ideal hypothetical condition where all other word stress cues are held constant between the two syllables of the novel word "*dada*". Co-variation of the word stress corresponding parameters allowed us to determine which of the parameters was more perceptually salient relative to the other two parameters. For the syllable prominence judgment task involving individual variation of the parameters, listeners were asked during 4 trials to indicate which syllable of "*dada*" was more prominent. Each trial consisted of a practice session, during which listeners were exposed to the range of parameter manipulations using 4 tokens, and the test session, where a listener heard each of 17 possible tokens once.

For the syllable prominence judgment task involving co-variation of the KLSYN88 parameters DU, TL and AH, syllable difference of a particular parameters could have 1 of 7 possible values, which were a subset of the 17 possible syllable difference values each parameter could have in the individual variation syllable prominence judgment task. Since the syllable difference in any of the three KLSYN88 parameters could have 1 of 7 possible values, there were 343 possible combinations of the three parameters. Thus the syllable difference values of a given parameter had 49 tokens in common. Listeners were given one trial, also consisting of a practice session and the test session, where listeners heard each of the 343 possible tokens once. As with the syllable prominence judgment task involving individual word stress variations, listeners were asked to determine which syllable of "*dada*", embedded in the carrier, was more prominent.

Results were obtained only from the test sessions of both syllable prominence judgment tasks. Listeners were given four choices: (1) the first syllable of "*dada*" was more prominent and they were certain; (2) they were uncertain, but if they had to guess they would guess that the first syllable was more prominent; (3) the second syllable of "*dada*" was more prominent and they were certain; (4) they were uncertain, but if they had to guess they would guess that the second syllable was more prominent. A subset of the listeners from both syllable prominence judgment tasks was also asked to rate the naturalness of the tokens used in the syllable prominence judgment tasks.

3.3.2 Naturalness Rating Tasks

Listeners were asked to rate the naturalness of each of the tokens used in the syllable prominence judgment tasks on a scale of 1 to 4, with 4 being natural and 1 being unnatural. The purpose of these tasks was to determine the range of syllable difference in the word stress correlates that is considered natural by native speakers of American English. This also allowed us to weight the results obtained from the syllable prominence judgment tasks, such that results from the more natural tokens are weighed greater in contributing to our knowledge of word stress than unnatural tokens.

For the syllable prominence judgment task involving individual variation of the word stress parameters (DU, TL, and AH), 7 listeners were asked to rate the naturalness of each token. They were asked to do it in 4 trials consisting of a practice and a test session. A token carrier phrase with the real "*dada*" was also included, as well as tokens containing "*dada*" with extreme syllable difference in parameter values and one token where the vowels in "*dada*" were replaced with broadband noise. For the syllable prominence judgment task involving co-variation of the parameters, 4 listeners were asked to rate the naturalness of each token. This was done in 1 trail, consisting of a practice and a test session. As with the syllable prominence judgment task involving individual word stress parameter variation, the real carrier phrase, as well as one token where the vowels in "*dada*" were replaced with broadband noise, were included to give listeners the full range of possible naturalness.

3.4 Syllable Prominence Judgment Results

3.4.1 Individual Word Stress Parameter Variation

For the syllable prominence judgment involving individually varied word stress corresponding KLSYN88 parameters duration (represented by DU), spectral tilt (represented by TL), and noise at high frequencies (represented by AH), the four choices given to listeners were categorized into 2 groups, response for first syllable prominence and response for second syllable prominence. Responses of each of the ten listeners for a particular token were averaged, such that a single number representing a listener's average response for a particular token during the 4 trials was obtained. An Analysis of Variance (ANOVA) statistical analysis was done on the average response of the ten listeners for the 17 tokens of each of the manipulated parameters DU, TL and AH. Changes in listeners' judgment of syllable prominence due to syllable difference in DU were found to be statistically significant ($p \ll 0.001$). This was also true for syllable difference in TL ($p \ll 0.001$). However, changes in syllable difference in AH did not significantly influence listeners' judgment of syllable prominence ($p = 0.664$). It should also be noted that there was significant differences between listeners in there responses (See Appendix E).

The average listener response to syllable difference in DU, shown in Figure 22, indicates that longer syllables, with greater value of DU, were perceived as having greater prominence. Interestingly, when the DU value was equal for the first and second syllable of "*dada*", listeners tended to perceive this as indicating first syllable prominence. This is in agreement with previous studies on duration (Fry, 1955; Oller, 1972; Klatt, 1976) and with the results obtained in the production study of Chapter 2. Figure 23 shows that syllable difference in the spectral tilt KLSYN88 equivalent parameter, TL, also cued for syllable prominence. The syllable with the greater TL value was perceived as being less prominent. AH results illustrated in Figure 24 show that syllable difference in AH had little effect on the response of native speakers of American English in this syllable prominence judgment task.

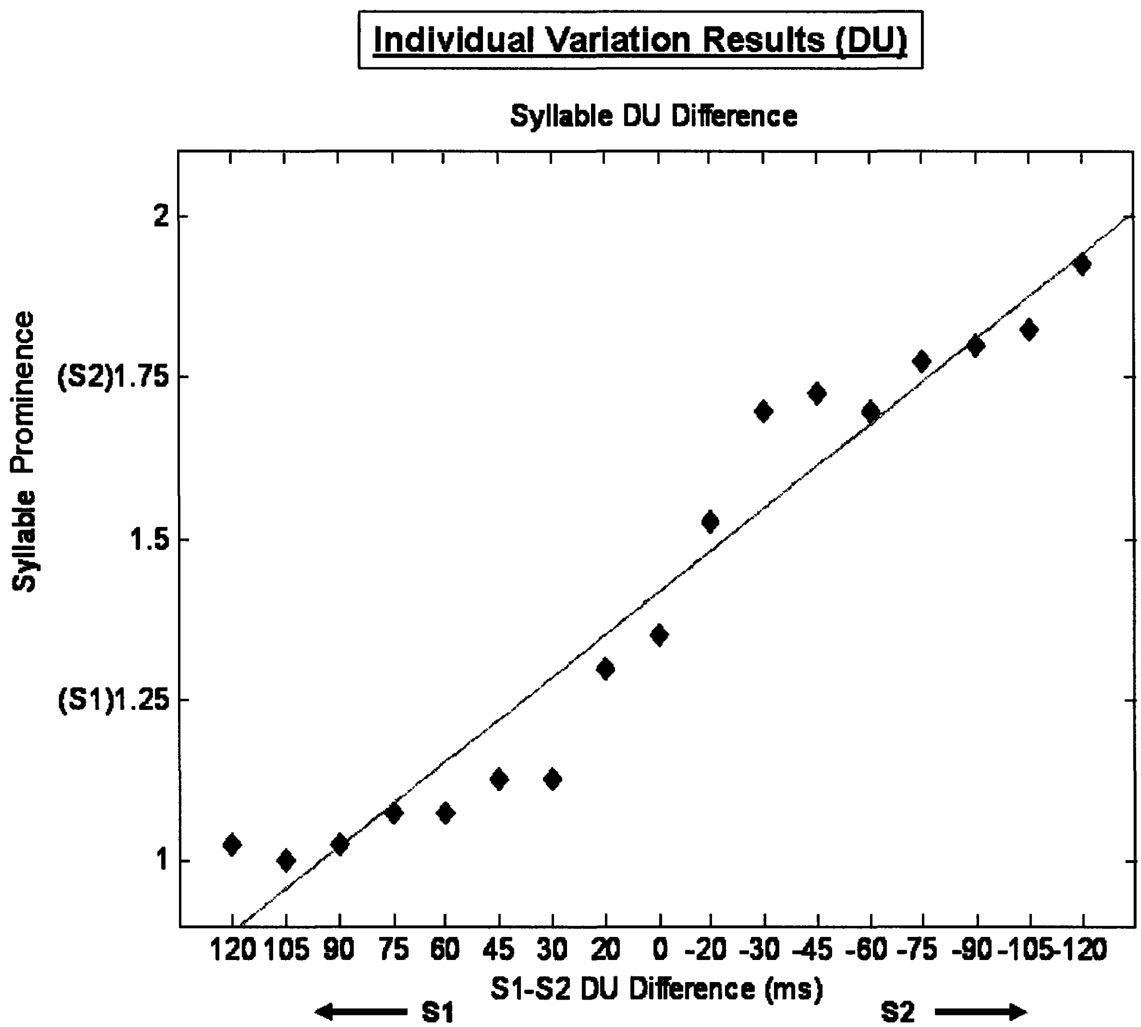


Figure 22: *The distribution plot of average response by 10 listeners, when the syllable difference in DU was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted line is just to aid in visualization of the response trend.*

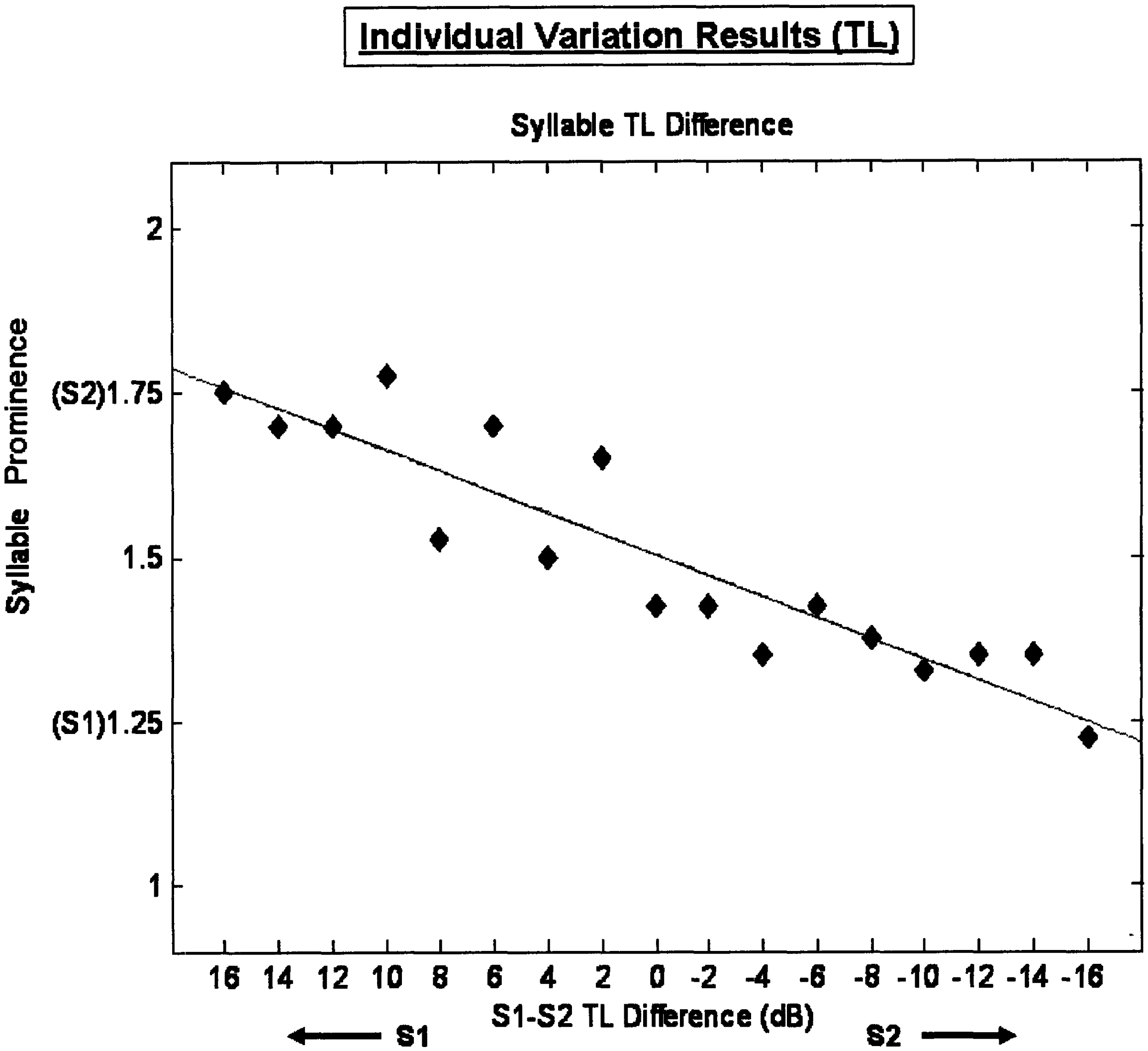


Figure 23: *The distribution plot of average response by 10 listeners, when the syllable difference in TL was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted line is just to aid in visualization of the response trend.*

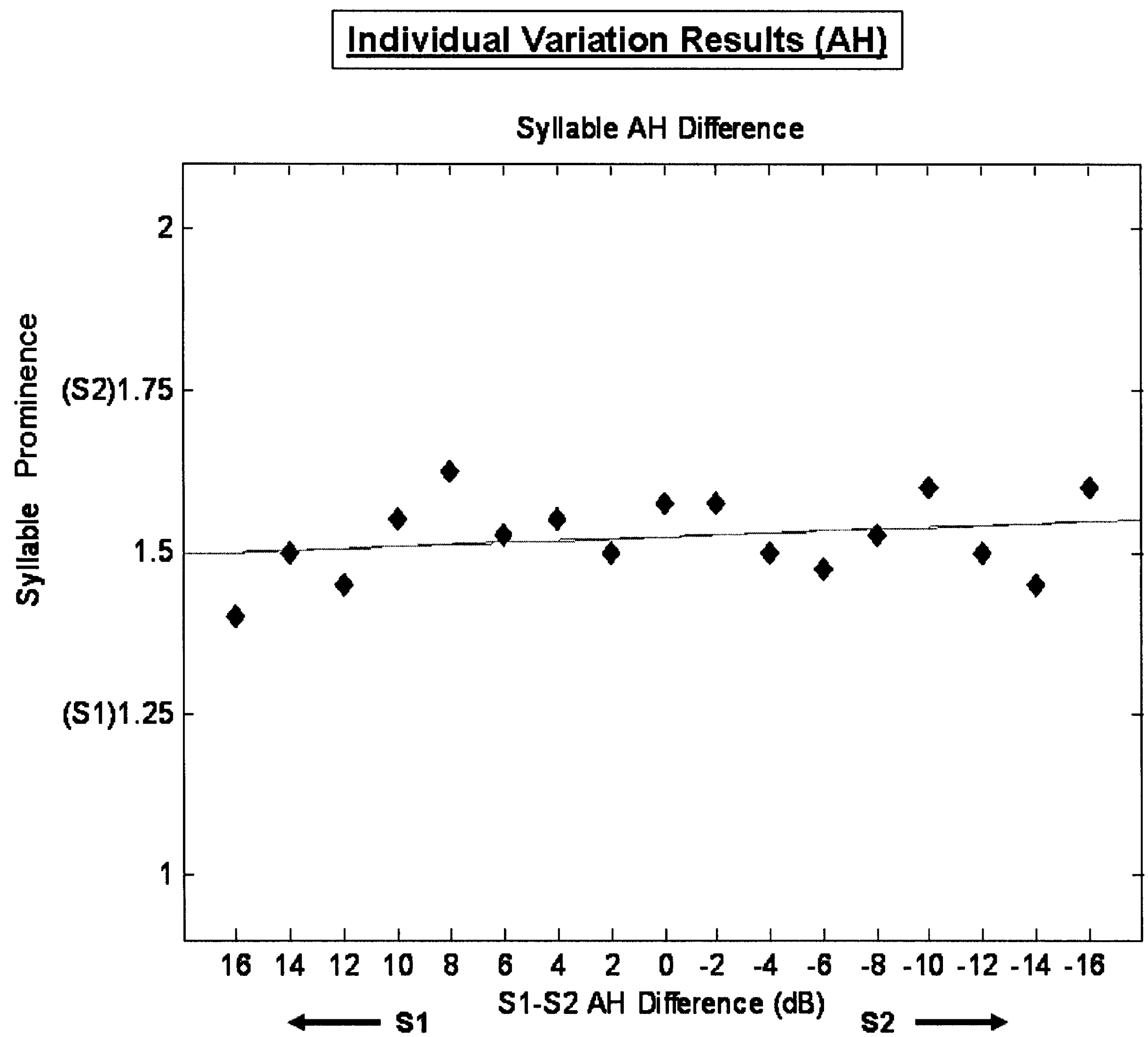


Figure 24: *The distribution plot of average response by 10 listeners, when the syllable difference in AH was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted line is just to aid in visualization of the response trend.*

3.4.2 Co-varied Word Stress Parameters

As with the syllable prominence judgment task involving individual word stress parameter variation, the four choices given to listeners for the co-varied KLSYN88 word stress parameters duration (represented by DU), spectral tilt (represented by TL), and noise at high frequencies (represented by AH) were categorized into 2 groups, response for first syllable prominence and response for second syllable prominence. Since each listener only heard each token once, there was no need to average. An Analysis of Variance (ANOVA) statistical analysis was done on the syllable prominence response of the ten listeners for the 343 tokens with respect to the individual co-varied parameters DU, TL, and AH. Changes in syllable difference in DU and TL significantly influenced listeners' judgment of which syllable of "*dada*" was more prominent ($p \ll 0.001$ and $p \ll 0.001$, respectively). Furthermore, a DU and TL interaction was present ($p = 0.002$), indicating that not only did syllable differences in DU and TL individually influence listener judgment, but that they also significantly affected each other's ability to influence the listener's judgment. As with the syllable prominence judgment task involving individual word stress parameter variation, changes in syllable difference in AH did not significantly influence listeners' judgment of syllable prominence ($p = 0.428$), nor was there significant interaction between it and the other parameters DU and TL ($p = 0.946$ and $p = 0.793$, respectively). It should also be noted that there was significant differences between listeners in there responses (See Appendix E).

Figure 25 shows that, as found with individually varied word stress parameters, longer syllables (i.e., with larger value of DU) were perceived as having the greater prominence. Although an interaction existed between syllable difference in DU and TL, changes in the parameter TL had little effect on listeners' use of syllable differences in DU as a cue for lexical prominence. When the DU duration value was equal for the first and second syllable of "dada", listeners on average perceived this as indicating first syllable prominence. As suggested by preliminary results, a "dada" with a second syllable longer than the first by about 30ms (i.e., -30ms) was perceived to be the most ambiguous syllable duration difference cue for native speakers of American English. Figure 26 shows that when the syllable difference in DU is small, that is when the first syllable is longer by 30ms or less and when the second syllable is longer by 30ms or less, syllable difference in the spectral tilt KLSYN88 equivalent parameter, TL, has the most influence on a listener's judgment of syllable prominence. As with the individual word stress parameter variation, the syllable with the greater TL value was perceived as being less prominent. As before the AH results illustrated in Figure 27, had little influence on listener judgment of syllable prominence ($p = 0.428$).

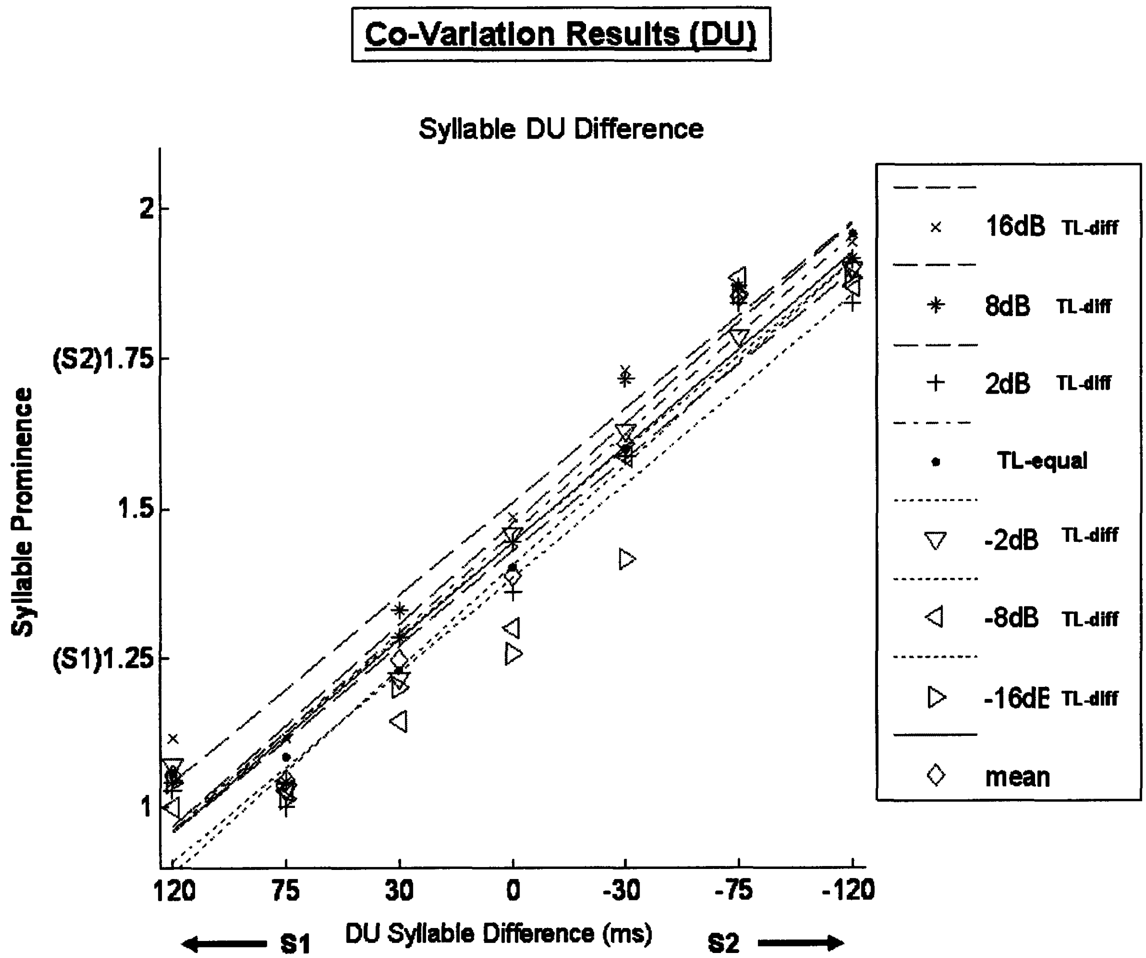


Figure 25: *The distribution plot of average response by 10 listeners, when the syllable difference in DU was co-varied with TL and AH for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted lines are just to aid in visualization of the response trends.*

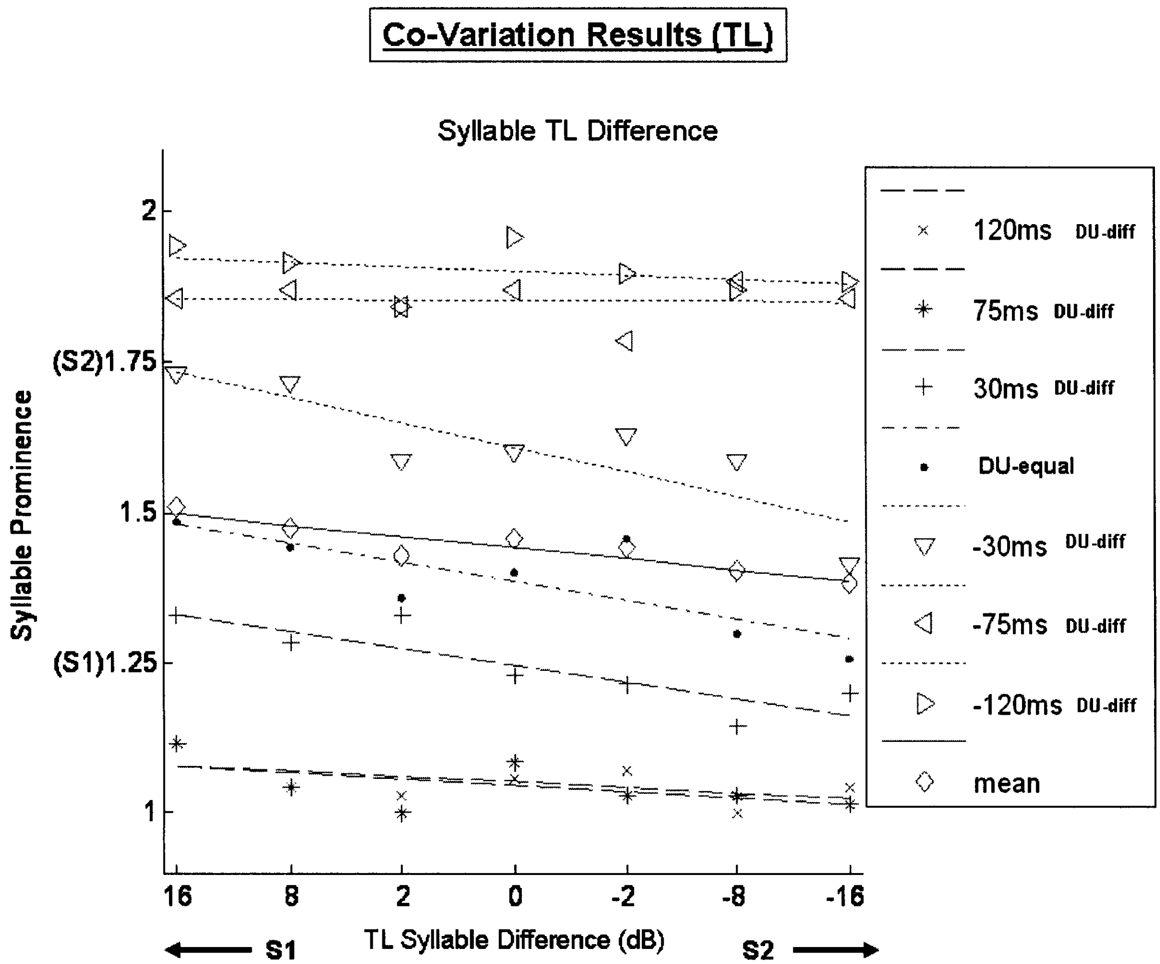


Figure 26: *The distribution plot of average response by 10 listeners, when the syllable difference in TL was co-varied with DU and AH for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted lines are just to aid in visualization of the response trends.*

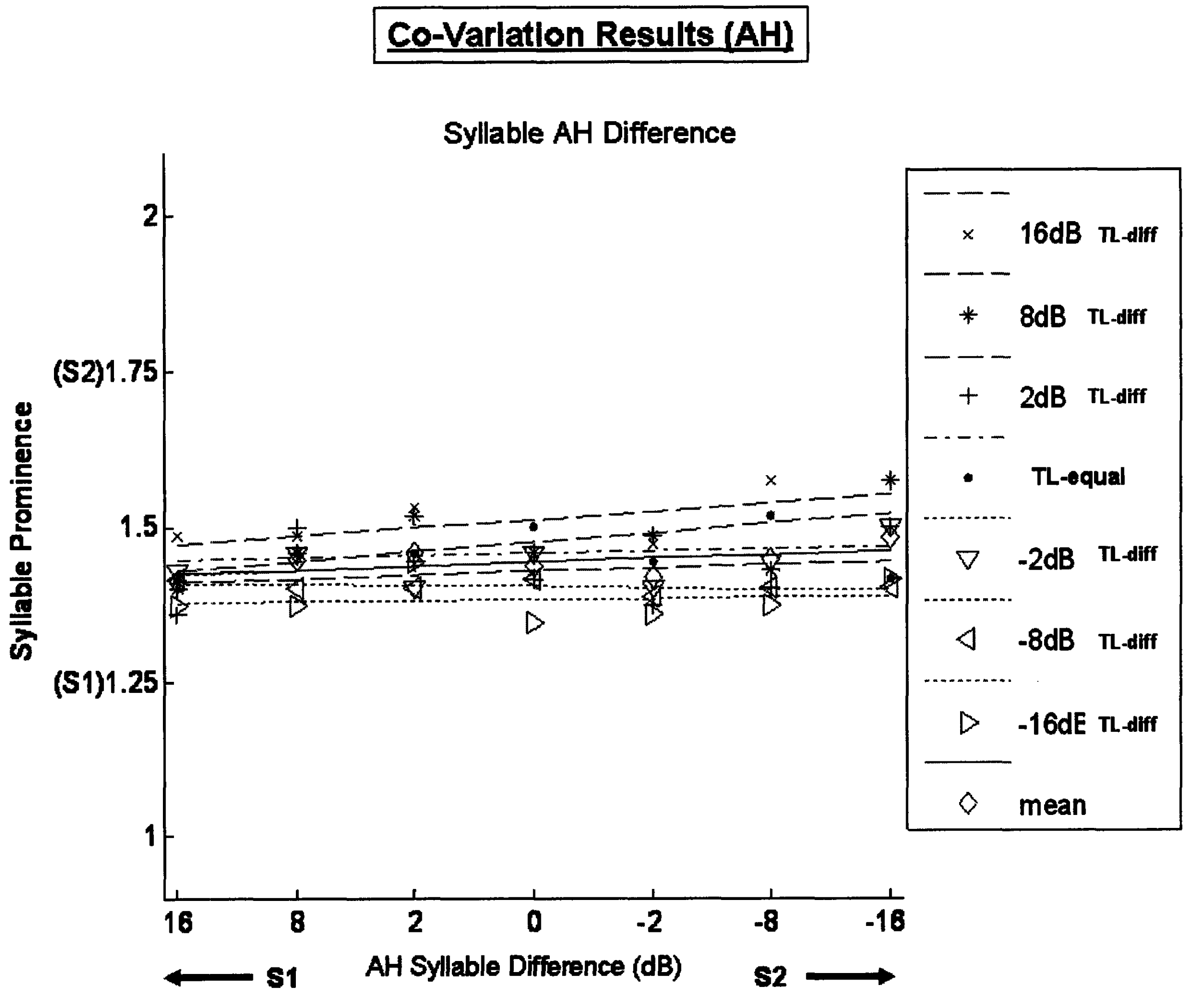


Figure 27: *The distribution plot of average response by 10 listeners, when the syllable difference in AH was co-varied with DU and TL for the novel word "dada" in the carrier phrase "Your blue dada drove here." S1 denotes the region for syllable one prominence and S2 denotes the region for syllable two prominence. The linear fitted lines are just to aid in visualization of the response trends.*

3.5 Naturalness Rating Results

3.5.1 Individual Word Stress Parameter Variation

The responses of the listeners who participated in the naturalness rating task for individually varied word stress parameters were averaged and used to construct a histogram indicating how native speakers of American English perceived the naturalness of the syllable differences in the KLSYN88 parameters DU, TL and AH. ANOVA was conducted on listeners' response to the co-varied parameters. Syllable differences in DU and TL influenced listeners' judgment of naturalness ($p = 0.002$ and $p = 0.006$). However, syllable differences in AH did not significantly influence listeners naturalness rating ($p = 0.103$). Responses to the extreme syllable difference values for the word stress KLSYN88 parameters were not included in the statistical analysis.

Figures 28-30 show that the majority of the synthesized "dada" were perceived as being fairly natural, regardless of which syllable had the greater value. However, Figure 28 shows that there is a slight preference in terms of naturalness of native speakers of American English for the second syllable, in the novel word "*dada*", to be slightly longer in duration, as indicated by the parameter DU. Likewise, Figure 29 shows that listeners perceived a second syllable of "*dada*" with slightly greater spectral tilt, as indicated by the parameter TL, to be more natural. Figure 30 shows that in general the range of AH values used in the prominence experiment were perceived as fairly natural.

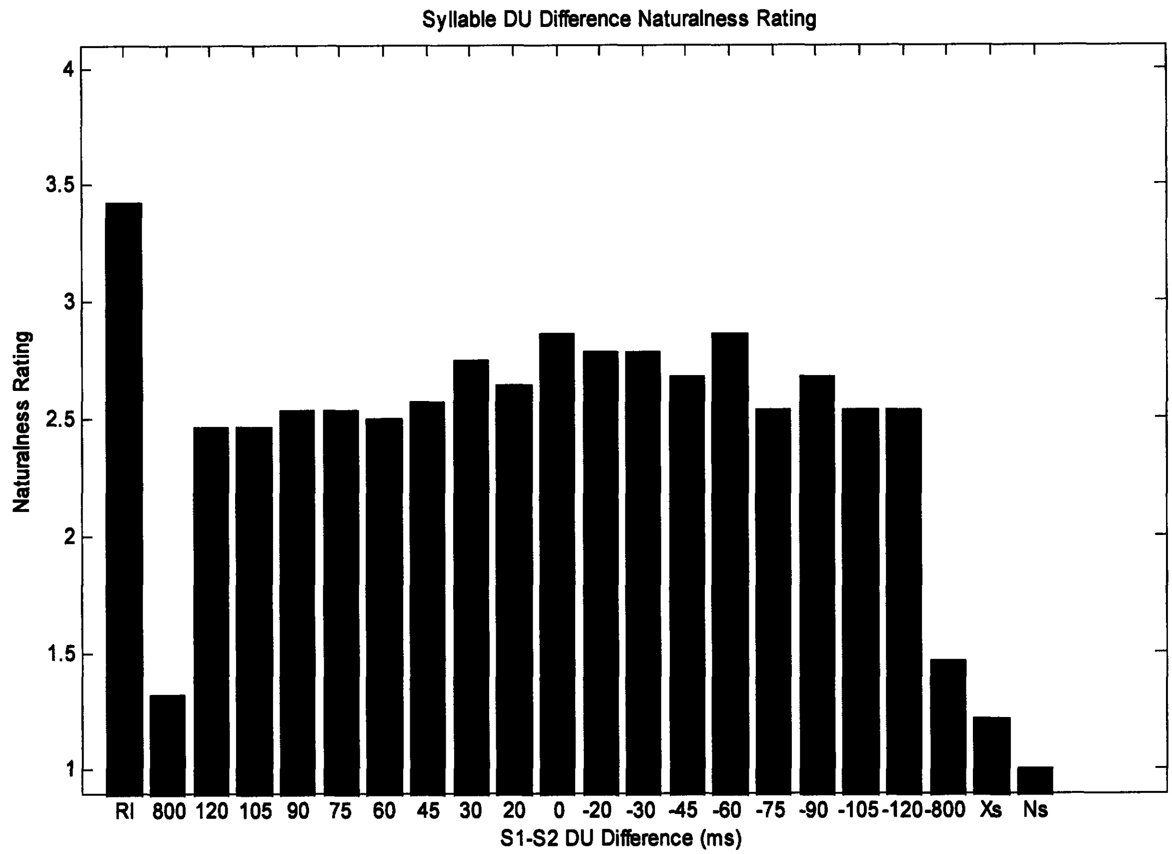


Figure 28: *Histogram plot of average naturalness rating by all 7 listeners, when the syllable difference in DU was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while syllable DU difference of 800 and -800 indicate that the first syllable and the second syllable were longer by 800ms, respectively. Xs indicates that both syllables were 950ms and Ns is an utterance token with broad band noise replacing the vowels in "dada."*

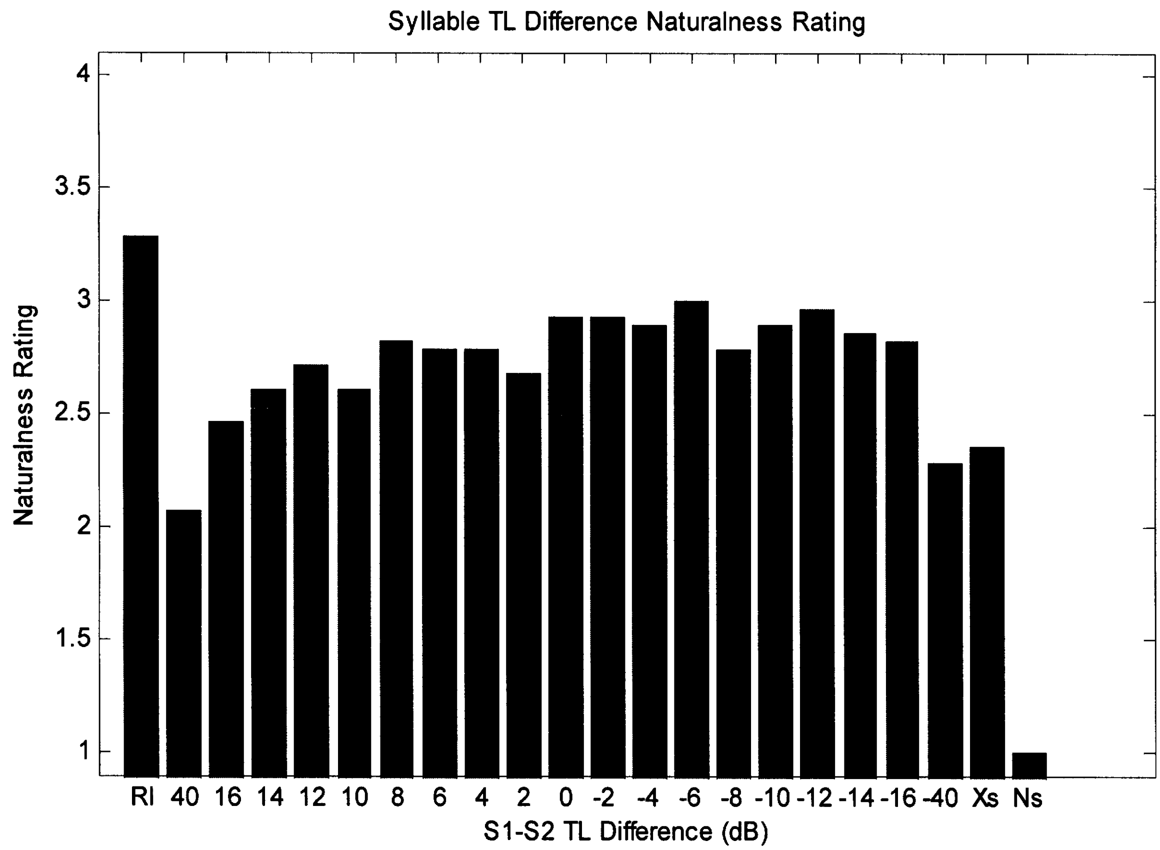


Figure 29: *Histogram plot of average naturalness rating by all 7 listeners, when the syllable difference in TL was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while syllable TL difference of 40 and -40 indicate that the first syllable and the second syllable were greater by 40dB, respectively. Xs indicates that both syllables had 40dB TL and Ns is an utterance token with broad band noise replacing the vowels in "dada."*

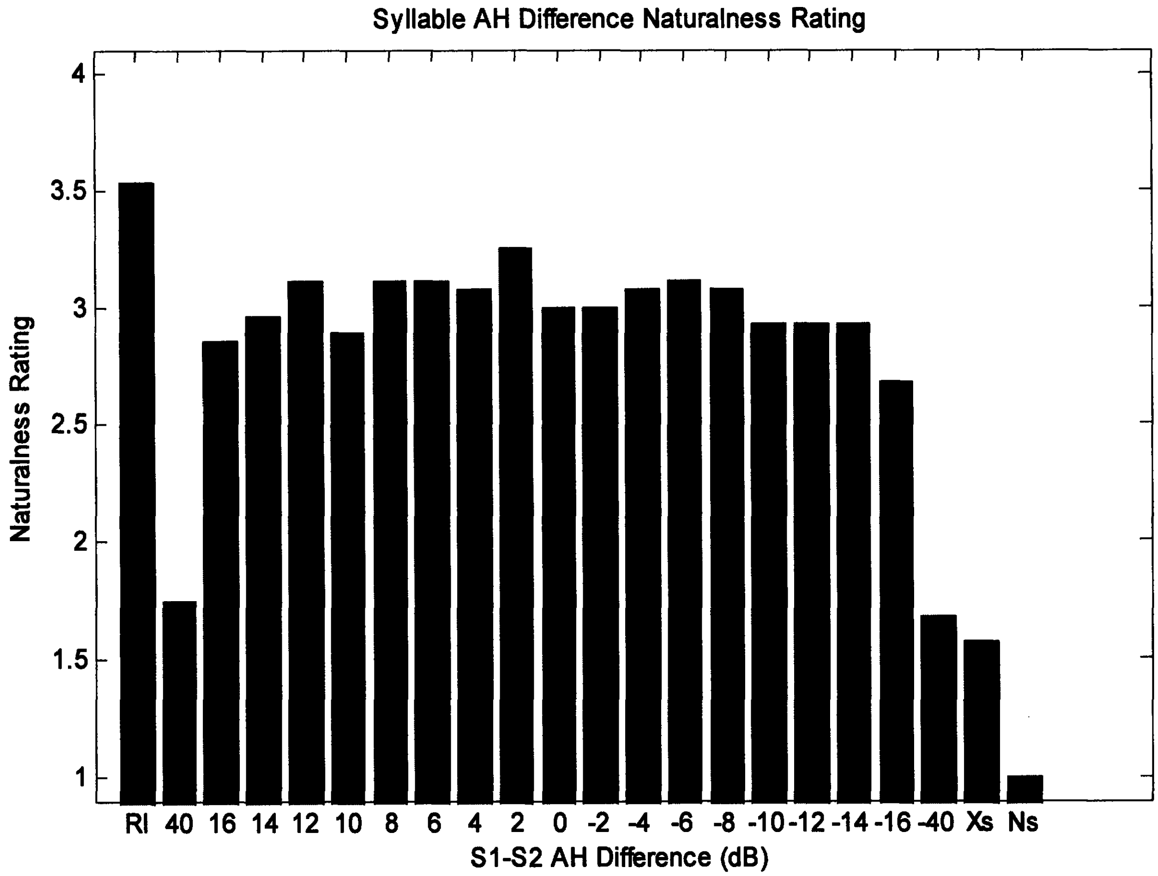


Figure 30: *Histogram plot of average naturalness rating by all 7 listeners, when the syllable difference in AH was varied for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while syllable AH difference of 40 and -40 indicate that the first syllable and the second syllable were greater by 75dB, respectively. Xs indicates that both syllables had 75dB AH and Ns is an utterance token with broad band noise replacing the vowels in "dada."*

3.5.2 Co-varied Word Stress Parameters

The responses of the listeners who participated in this naturalness rating task for the co-varied word stress parameters were averaged and used to construct a histogram indicating how native speakers of American English perception of the naturalness of speech was influenced by syllable differences in the KLSYN88 parameters DU, TL and AH. ANOVA was conducted on listeners' response to the co-varied parameters. Syllable differences in DU, TL, and AH in the novel word "*dada*" all influenced listeners' judgment of naturalness ($p << 0.001$ for all). Furthermore, the ANOVA indicated that an interaction between syllable differences in TL and AH existed ($p = 0.002$) and between syllable differences in DU and TL ($p = 0.042$). However, no statistically significant interaction was found between DU and AH ($p = 0.961$). Responses to the extreme syllable difference values for the word stress KLSYN88 parameters; the real utterance; and the utterance with the noise replacing the vowels of "*dada*" were not included in the statistical analysis.

As with the individual word stress parameter variations, listeners had a slight preference in terms of naturalness for slightly longer second syllables, as indicated by the parameter DU averaged over all the TL and AH values (Figure 31). Likewise, Figure 32 shows that listeners perceive a second syllable of "dada" with greater spectral tilt, as indicated by the parameter TL averaged over all DU and AH values, to be more natural. Figure 33 shows that in general the range of AH values averaged over all DU and TL values used in the syllable prominence judgment tasks, involving co-variation of word stress parameters were perceived as fairly natural. This agreed with the naturalness rating results from when the parameter AH was varied by itself.

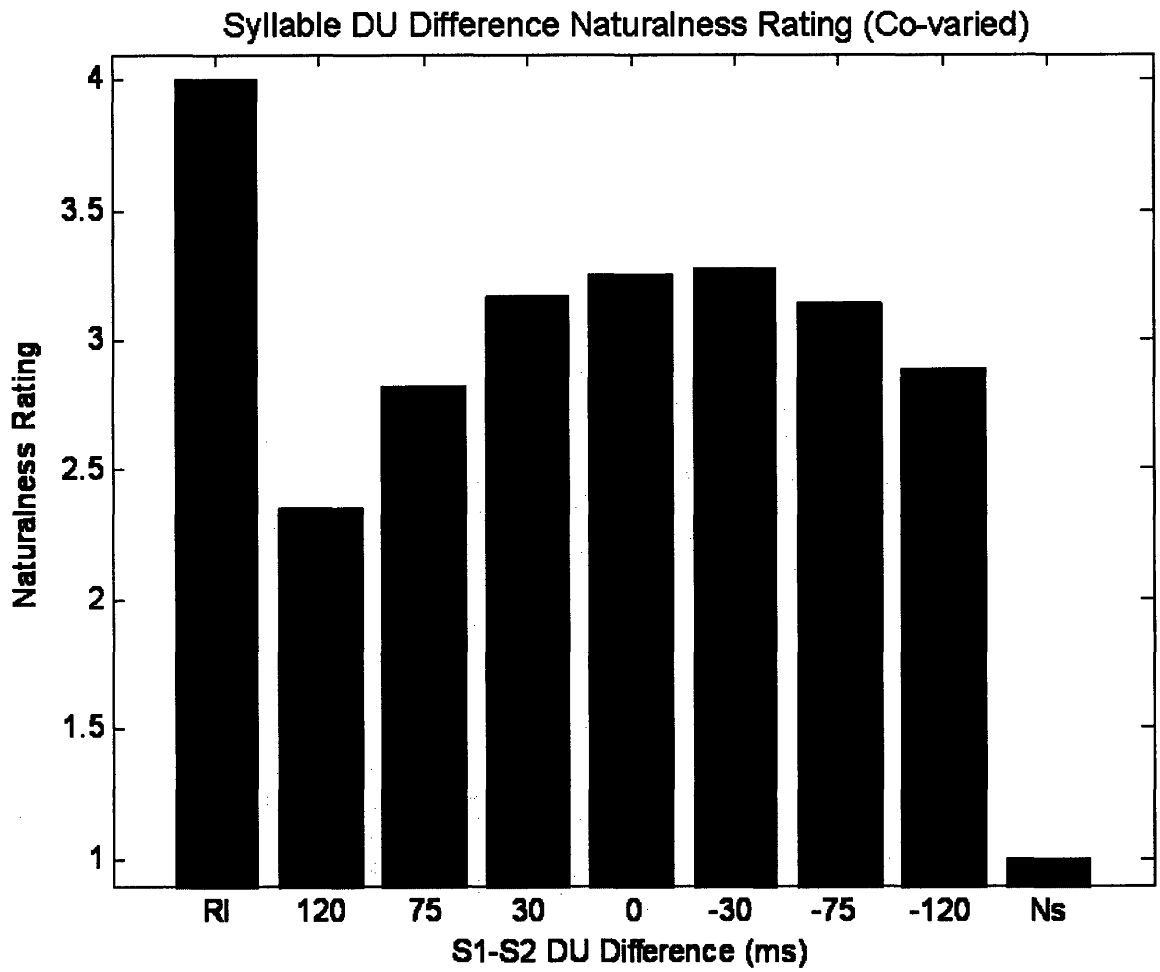


Figure 31: *Histogram plot of average naturalness rating by 5 listeners, when the syllable difference in DU was co-varied with TL and AH for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while Ns is an utterance token with broad band noise replacing the vowels in "dada."*

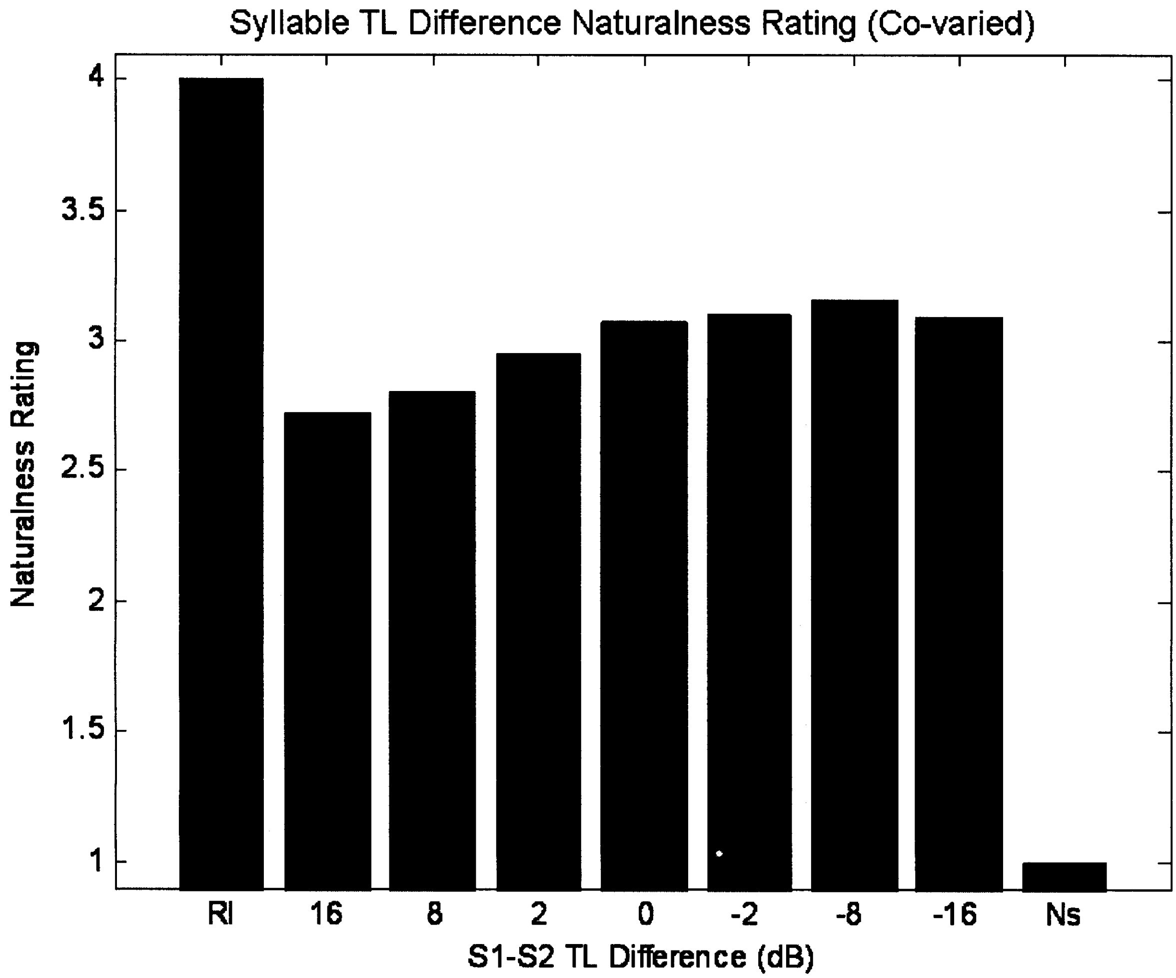


Figure 32: *Histogram plot of average naturalness rating by 5 listeners, when the syllable difference in TL was co-varied with DU and AH for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while Ns is an utterance token with broad band noise replacing the vowels in "dada."*

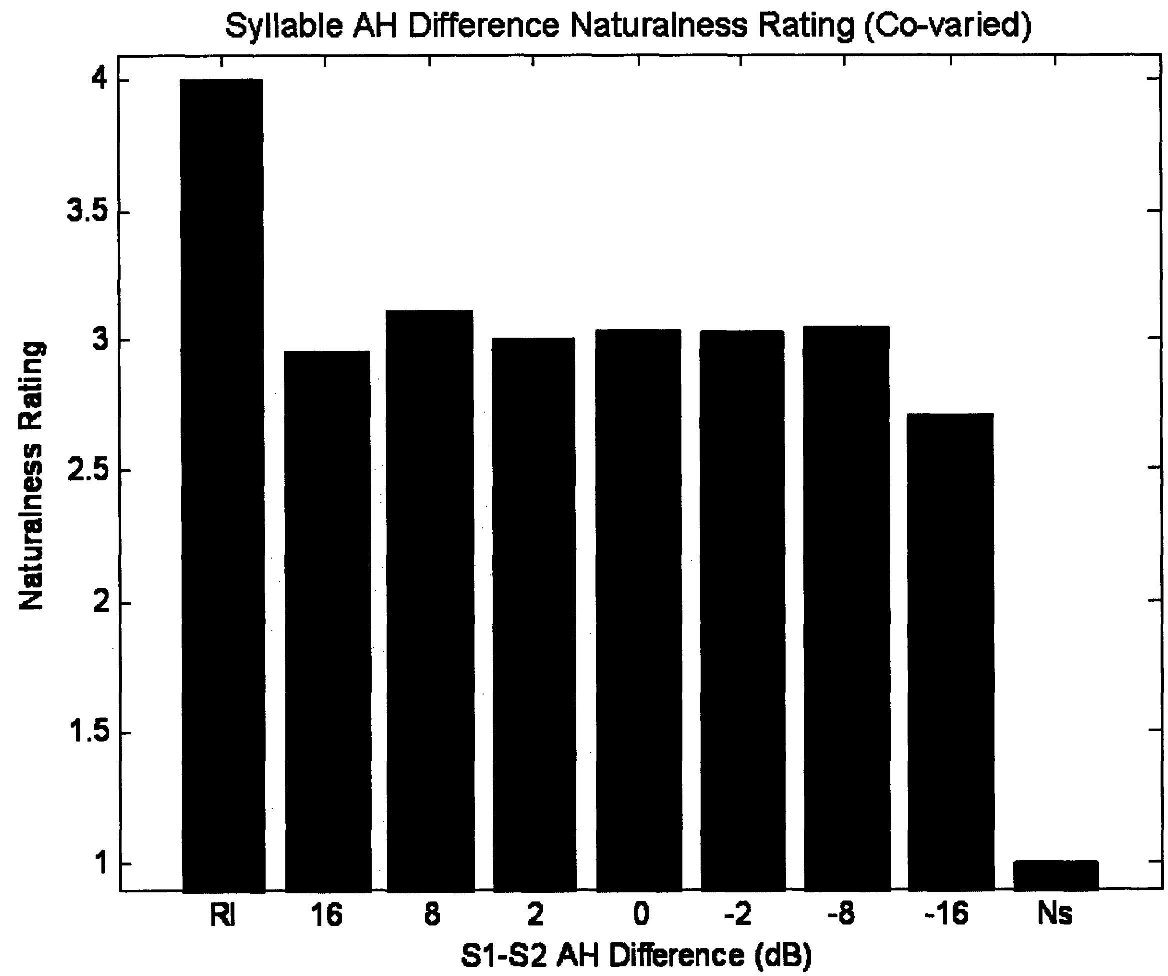


Figure 33: *Histogram plot of average naturalness rating by 5 listeners, when the syllable difference in AH was co-varied with DU and TL for the novel word "dada" in the carrier phrase "Your blue dada drove here." The scale is from 1-4, with 1 being unnatural and 4 being natural. Rl is the real utterance, while Ns is an utterance token with broad band noise replacing the vowels in "dada."*

3.6 Discussion

Results from this study indicate that two of the three correlates of word stress produced by speakers in Chapter 2 represented by the KLSYN88 parameter DU (corresponding to duration), TL (corresponding to spectral tilt), and AH (corresponding to aspiration noise) were cues for listeners in the syllable prominence judgment tasks. Syllable difference in DU was a very strong and robust cue for syllable prominence when it was the only word stress parameter that differed between the two syllables of the synthesized novel word "*dada.*" This was also true when syllable difference in DU was co-varied with syllable difference in the other two word stress parameters, TL and AH. In general, the syllable with the larger value of DU (i.e., longer in duration) was perceived as having the greater prominence. However, there seems to be an equal syllable duration bias towards first syllable prominence, with the second syllable having to be longer than about 30ms before being considered prominent. This finding agrees with the syllable duration difference results obtained in the production study of Chapter 2.

Listeners' use of syllable DU difference in choosing the more prominent syllable in "*dada*" was not influenced much by changes in the syllable difference of the other KLSYN88 parameters TL and AH. For both individual and co-varied word stress parameter syllable prominence judgment tasks, listener judgment of syllable prominence for the first syllable seems to reach saturation before the greatest syllable difference in DU tested in this study is achieved. However, it seems that listeners' judgment of longer second syllables as the more prominent syllable does not reach saturation, given the range of syllable difference in DU used in this study. This result, along with the naturalness rating for DU, indicates that the second syllable of "*dada*" can be longer before it is perceived as being unnatural. However, the syllable difference in DU might then indicate a phrasal boundary (Klatt, 1976, Shattuck-Hufnagel and Turk, 1996).

Syllable difference in the KLSYN88 parameter for spectral tilt, TL, also cued for prominence when it was individually varied and when it was co-varied with the other word stress parameters. However, syllable difference in TL was most influential as a prominence cue when the syllable difference in DU (i.e., duration), between two syllables with full vowels, is relatively small. According to the natural ness ratings for DU, small syllable differences in DU are perceived as being the most natural for a synthesized two-syllable novel word with two full vowels. In general, the syllable of "*dada*" that had the greater value of TL was perceived as being less prominent. However, there seems to be preference for the second syllable to have a slightly greater default value of TL, such that the second syllable TL value must be greater than the first syllable value by about 4dB before it is considered less prominent. The naturalness rating results also indicated that a significant interaction existed between syllable difference in DU and syllable difference in TL. These results all suggest that the duration and spectral tilt word stress correlates produced by speakers were intentional and natural for both the novel and real words.

Results for syllable difference in AH had the least influence on listeners' judgment of syllable prominence. Listeners' use of syllable difference in AH was slightly, but not significantly, influenced by syllable difference in TL. In general, syllable difference in AH did not influence listeners' judgments and thus was not perceived as a cue for syllable prominence. When syllable difference in AH was individually varied, listeners seemed to find the range of syllable difference in AH, for "*dada*" with a second syllable longer than the first by a DU value of 30ms and slightly more spectral tilt (TL value of 2dB), to be all within natural range. The syllable difference in AH naturalness ratings were overall high and varied little. However, syllable differences in AH did significantly influence listeners' judgment about the naturalness of the utterance containing the synthesized "*dada*". This was apparent when syllable difference in AH was co-varied with the other word stress parameters. This can serve as evidence that the listeners could perceive the syllable difference in AH in the syllable prominence judgment tasks, since the identical tokens were used for both prominence judgment and naturalness rating. It would be interesting to determine the nature of the interaction between syllable difference in TL and syllable difference in AH. Results from the production study, would suggest that their would be a positive correlation, such that listeners would find it more natural to find a syllable with greater spectral tilt, represented by TL, to also have greater noise at high frequencies, represented by AH.

Overall the naturalness ratings indicated that the range of syllable differences in DU used in the syllable prominence judgment task was fairly natural compared to the carrier phrase with the real "dada," except when the first syllable was 120ms longer than the second syllable. These results agree with the production study, where equal syllable duration was used by speakers to indicate first syllable primary stress. The range of syllable differences in TL and AH were all considered by listeners to be fairly natural. However, a slight preference for second syllables with greater TL was still observed, suggesting that a first syllable with a slightly greater spectral tilt would be considered enough to cue for second syllable prominence. Results from the real word production study seem to confirm this hypothesis.

4. Conclusion

Results from the production and perception studies reported in this thesis indicate that there are acoustic correlates of word stress, which consistently distinguish between primary stressed syllables from the non-primary full vowel syllables in all the pitch accented conditions tested. These correlates of word stress were spectral tilt, noise at high frequencies, indicated by ratings of band-pass filter F3 waveforms, and syllable duration. The production and perception studies indicate that duration is the strongest correlate and cue to word stress. These findings are is in agreement with studies by Klatt (1976), Beckman and Campbell (1997), and Sluijter *et al.* (1995, 1996a-b, and 1997). Nevertheless, when the syllable duration difference is small, listeners' judgment of syllable prominence is strongly influenced by syllable difference in spectral tilt, as found in the perception study of Chapter 3.

Although, speaker average syllable difference in band-pass filtered F3 waveform noise ratings correlated consistently with word stress patterns in the production study of Chapter 2, noise at high frequencies was not used by listeners to determine word prominence in the syllable prominence judgments. When the KLSYN88 parameter corresponding to aspiration, AH, was varied individually and in combination with the other consistent correlates of word stress, it did not significantly influence listeners' judgment of syllable prominence for the synthesized "*dada*". It seems that AH, in the range that it was varied in the perception studies, was not a cue for syllable prominence, but was a correlate of word stress brought about by spectral tilt. Increase in spectral tilt also decreases the ratio of the amplitude of high frequency harmonics relative to that of the amplitude of high frequency noise already present. This could be used as another evidence that increase in spectral tilt, as measured by H1*-A3*, is due to lowering of the amplitude of A3* not the increase of H1*. This seems like a more natural process, since increasing H1*, even by a small amount, could increase the overall spectral amplitude. Increase in overall amplitude was found to be correlated to pitch accent in Chapter 2, using syllable difference in peak intensity, as well as in other studies (Fry, 1955 and 1958; Lieberman, 1960; and Harrington *et al.*, 1998).

Furthermore, results from the production study indicate that word stress correlates are augmented in the Fa condition, when the high focal pitch accent was on the target word. However, this Fa condition also has the effect of masking the spectral tilt differences, as measured by H1*-A3*, between primary stressed and non-primary full vowel syllables. This effect of the high focal pitch accent on the H1*-A3* measurement can be corrected using Equation 1 of Chapter 2. Application of this focal pitch accent correction to measurements of H1*-A3* in conditions where the high focal pitch accent precedes the target word, such as in the Fp1 and Fp2 conditions also result in more accurate and clearer syllable differences in spectral tilt.

Vowel quality differences, such as increase in the first formant (F1), also seem to distinguish primary stressed syllables from non-primary stressed syllables for the vowel /a/ in the novel words *'dada* and *da'da* in all three conditions tested in the production studies. This was found to be consistent across the five speakers (See Appendix B, Tables 1-3). However it was not true for the other novel words containing the vowels /o/ and /i/, or for the real words. As demonstrated by the syllable prominence judgment tasks in Chapter 3, syllable differences in formant values are not essential for making judgments about syllable prominence. In the case of *'dada* and *da'da*, it might have been easier for speakers to indicate the relationship between the vowels of the two syllables by opening the mouth wider, since the production of the vowel /a/ does not require rounding, as in the production of /o/ and /u/, or narrowing of a region of the oral cavity, as in the production of the vowel /i/. Thus it seems that the goal of the speakers was to maintain the identity of the vowels, while simultaneously indicating the word stress relationship between these vowels, within the target words.

There are also acoustic correlates of pitch accent that only distinguish between primary stressed syllables from the non-primary full vowel syllables when the target word has phrase level high focal pitch accent. These pitch accent correlates were shown in the production study of Chapter 2 to be F0 prominence, intensity prominence and amplitude of the first harmonic (H1*). When the focal pitch accent preceded the target word, the first syllable of the target word consistently had the greater peak F0, peak intensity, and H1* values.

Preliminary experiments, not discussed in this thesis, indicated that in both the production and perception studies syllable differences in first formant bandwidth, as approximated by H1*-A1*, could also serve as a weak correlate of word stress in conditions where the target word does not have high focal pitch accent. There seems to be some evidence of this for the real words, Figure 14e-f. However, there is no evidence of H1*-A1* being a word stress or pitch accent correlate once syllable differences in consonant and vowels were controlled, as with novel words in Chapter 2. This might be because loss of energy at high frequencies did not spread to lower frequencies. H1*-H2* was also found not to correlate with either word stress or pitch accent. It is possible that the measurement technique used in the production study was not sensitive enough. Perhaps more direct means of measuring these parameters, such as laryngeal endoscopy with calibrated sizing function, are needed in order to determine if they do play a role in distinguishing primary stressed full vowel syllables from non-primary full vowel syllables.

The naturalness rating results showed that the synthesized tokens used in the perception study of Chapter 3 were in general perceived by native speakers of American English as being fairly natural, but still fell short of the real utterance. Furthermore, the ratings revealed that listeners had preferences for syllable differences in KLSYN88 parameters corresponding to the correlates of word stress (i.e., DU, TL, and AH). For example listeners seemed to find the range of syllable difference in AH, when individually varied with the second syllable being longer by a DU value of 30ms and having a TL value of 2dB, to be all within the natural range. However when the syllable difference in DU and TL were co-varied with AH, significant interaction between TL and AH was observed. Listeners seemed to favor second syllables with slightly longer or equal in duration than the first syllable, as well as second syllables that had slightly greater spectral tilt. These preferences might help shed light on why, for listeners and speakers, judgment of first syllable prominence and production of primary stressed syllables, respectively, are equated with equal syllable duration.

Naturalness ratings in the production study in Chapter 3 also confirmed that listeners could perceive the syllable differences in the KLSYN88 parameter for aspiration noise, AH, as could the author of this thesis and others not reported, but did not use it to assign word prominence. Suggesting that the higher waveform noise rating for non-primary full vowels observed in the production study was due to lowered amplitude of high frequency harmonics exposing noise already present, rather than active generation of noise by the speakers. However individual speaker differences exist (See Appendix B, Tables 7-12).

A general conclusion from the results obtained in this thesis research of two syllable novel and real words is that during speech production male native speakers of American English use changes in the shape of the vocal tract to distinguish between different vowel types. However, in order to distinguish between the primary stressed syllable and the non-primary full vowel syllable, speakers use duration and changes in glottal configuration during vowel phonation to lower or increase the amplitude of high frequency harmonics. For most of the vowels tested in this study (i.e., /o/, /i/, /u/), significant changes in the vocal tract shape in order to indicate the word level prosodic relationship between two syllables of a target word, could compromise the identity of the vowels. Thus the prosodic relationship between the two syllables of the target words used was indicated using duration and changes in the glottal region of the larynx that result in different degrees of spectral tilt, which also gave rise to syllable difference in noise at high frequencies. These word stress syllable differences were also observed for the novel words with the vowel /a/, however additional first formant (F1) syllable differences that correlated to word stress were observed. This is possibly because the vowel /a/ does not have the same vocal tract shape restrictions as /o/, /i/, and /u/, since changes in syllable differences in F1 did not correlate with word stress for the other novel and real target words.

Duration seems to be the more salient of the cues for word stress, for both production and perception. Perhaps, this is because syllable differences in duration is a more simple and robust means of relaying word stress prosodic information, since major adjustments of speech articulators are not needed. What is needed is to just maintain the speech action, such as phonation, for a period of time. According to Turk and Sawusch (1996), harmonic signals produced with longer duration are perceived as being louder. Such an effect would be applicable to vowels. Also associated with loudness are changes in the amplitude of high frequency harmonics around 3kHz, which is the region of lowest intensity threshold in human hearing (Fletcher and Munson, 1937). Thus it is possible that changes in syllable difference in duration and spectral tilt are a means of changing the perceived loudness of the primary stressed syllable.

Studies done by Turk and Sawusch (1996) and Kochanski *et al.* (2005) suggest that more research is needed to in order to understand the role of duration and spectral tilt in determining the loudness of linguistic units at the level of the syllable. Overall, the differences between the primary stressed syllable and the non-primary full vowel syllable of a word, in terms of the correlates of word stress, were more distinct once the phonological composition of the target words were controlled, as with the novel words. Furthermore, significant individual differences exist in the production and perceptual use of word stress correlates.

5. Future Work

In the future, a replication of this study with female native speakers of American English will be conducted. This will allow for comparison of word stress correlate production and perception across gender. The current hypothesis is that no differences should exist in the perception of word stress correlates. It is however possible that word stress correlate gender differences might exist for speech production, given that female native speakers of American English tend to have less energy at high frequencies (Klatt and Klatt, 1990). Closer look at individual differences would also be appropriate, since differences between speakers and listeners do exist.

A possible future addition to this study is a physiological component that could help to strengthen the validity of the acoustic production and perception results obtained. The physiological component of the study would involve the visualization of vocal fold configurations during vowel phonation. This can be accomplished by utilizing a laryngeal endoscope with calibrated sizing function to visualize the glottal region during phonation and to quantitatively measure changes is the glottal area that would be associated with increase or decrease of spectral parameters, such as open quotient, increases in first formant bandwidth and spectral tilt. Many of these measurements can also be accomplished using electroglottography (EGG). In either case, correlation between the acoustic and physiological findings that support the results obtained in this thesis would greatly increase the validity of these results, as well as expand the number of fields and disciplines in which this study has an impact.

Further research can also be done to determine the role of duration and spectral tilt with regards to word stress. Evidence from this thesis research suggests that it is possible that the syllable differences between primary stressed and non-primary full vowel, might be an attempt to change the perceived loudness of the primary stressed syllable. It would also be important to investigate the effect of neighboring consonants on the perceived prominence of a syllable. For example is there a difference in the high frequency energy of the burst of a stop-consonant onset of a primary stressed syllable compared to the burst of a matched stop-consonant onset of a non-primary syllable of the same word? There are still many interesting unanswered questions with regard to word stress. Results from this study have shed light on a few, but many more unanswered questions still remain, such that the field of prosody will remain interesting for decades to come.

APPENDICES

APPENDIX A: Correction of Spectral Measurements Using Inverse Filtering

During the production of vowels, such as /a/, /i/, /o/, and [u], airflow through the glottis, caused by pressure differences across the glottis, is modulated by the vocal fold vibrations. This modulation of airflow can be represented as changes in the volume velocity, Ug(t), as is shown in Figure 34a. For many speakers, there is an airflow bypass that is not modulated by the vocal folds and is represented as a DC flow. The derivative of Ug(t) with respect to time gives rise to the glottal waveform illustrated in Figure 34b. A Fourier transformation of the glottal waveform gives rise to the glottal source spectrum shown in Figure 34c.

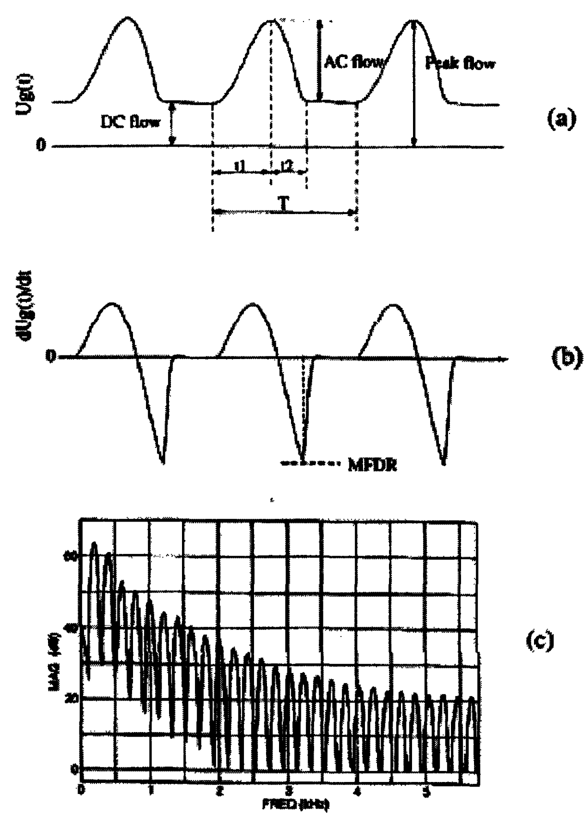


Figure 34: ***Glottal pulse (a), glottal waveform (b), and glottal source spectrum (c). (figures are from Hanson, 1995).***

The glottal source spectrum is then altered (i.e., filtered) by the supra-glottal region known as the vocal tract (See Figure 8). It is the configuration of the vocal tract during vowel production that gives rise to the poles and zeros that in turn filter the glottal source spectrum. Figure 35 shows the vocal tract filtered glottal source spectrum, with H1, H2, A1 and A3 indicating the amplitudes of F0, 2F0, F1 and the third resonant frequency F3, respectively.

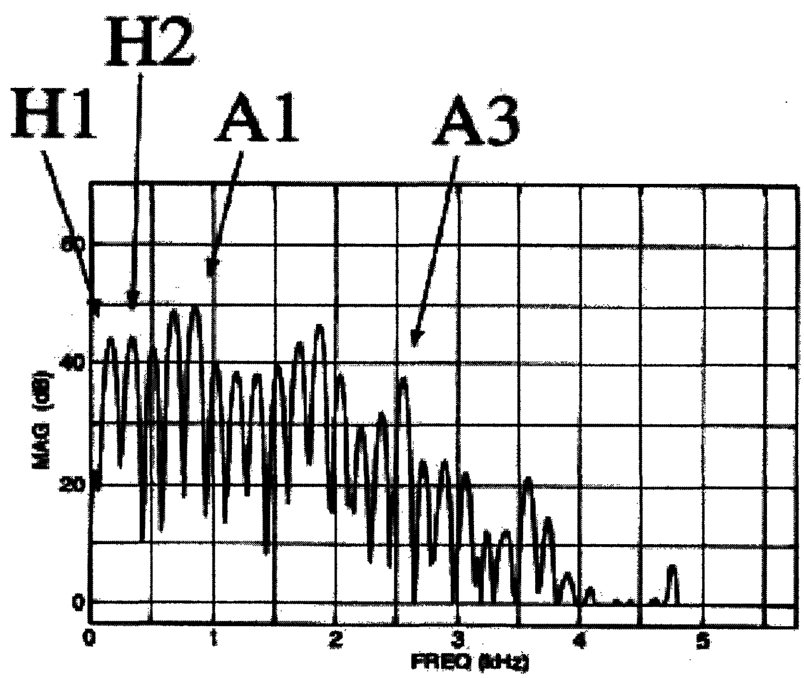


Figure 35: ***Vocal tract filtered glottal source spectrum. H1, H2, A1 and A3 indicating the amplitudes of F0, 2F0, F1 and the third resonant frequency F3, respectively. (from Hanson and***

Inverse filtering is used to remove the effect of the poles and zeros of the vocal tract transfer function that alter the amplitude of the glottal source spectrum in the frequency domain. Thus inverse filtering is done in order to obtain a more accurate measurement of the glottal source spectrum. It is done by measuring the amplitude of the harmonics of interest from vocal tract filtered glottal spectra, like the one illustrated in Figure 35. From these measurements is subtracted the influence of the vocal tract transfer function. This allows for the comparison of glottal characteristics, such as open quotient (approximated as H1-H2), across different vocal tract shapes.

If we model the vocal tract using an all-pole transfer function, then the complex function (T(ω)) can be represented by Equation 2.

$$T(\omega)=\left(\frac{s_1 s_1^*}{(s-s_1)(s-s_1^*)}\right)\left(\frac{s_2 s_2^*}{(s-s_2)(s-s_2^*)}\right)\cdots\left(\frac{s_n s_n^*}{(s-s_n)(s-s_n^*)}\right) \qquad \text{Eq. 2}$$

where $s = j\omega$. $s_n = (\alpha_n + \omega_n)$ and $s_n^* = (\alpha_n - \omega_n)$, while n is the number of the vocal tract resonant frequencies (i.e., formants).

Given Equation 2, the transfer function for just the first resonant frequency is given by Equation 3.

$$F_1(\omega)=\frac{(\propto_1 + j\omega_1)(\propto_1 - j\omega_1)}{(j\omega-(\propto_1 + j\omega_1))(j\omega-(\propto_1 - j\omega_1))} \qquad \text{Eq. 3}$$

where $\omega = 2\pi f$ and $\omega_1 = 2\pi F1$.

Equation 3 can be used to represent the influence of F1 on the amplitudes of F0 and 2F0, H1 and H2, respectively. In this case $\omega = 2\pi f$, where f = F0 (or f = 2F0, for correction to H2).

For a vowel like /a/, we can assume that the F1 pole in the S-plane is sufficiently close to the imaginary $j\omega$-axis and $\alpha_1 << \omega_1$, such that we can approximate $\alpha_1 \approx 0$. Thus Equation 3 can be reduced to Equation 4.

$$|F_1(f)| = \frac{F1^2}{F1^2 - f^2} \qquad \text{Eq. 4}$$

where f = F0, for H1 correction, or f = 2F0, for H2 correction.

Since the amplitudes H1 and H2 are in dB, we need to convert the magnitude of Equation 4 into the log domain. This gives rise to Equation 5.

$$dB[F_1(f)] = 20\log_{10}\left(\frac{F1^2}{F1^2 - f^2}\right) = 10\log_{10}\left(\frac{F1^2}{F1^2 - f^2}\right)^2 \qquad \text{Eq. 5}$$

where f = F0 or f = 2F0.

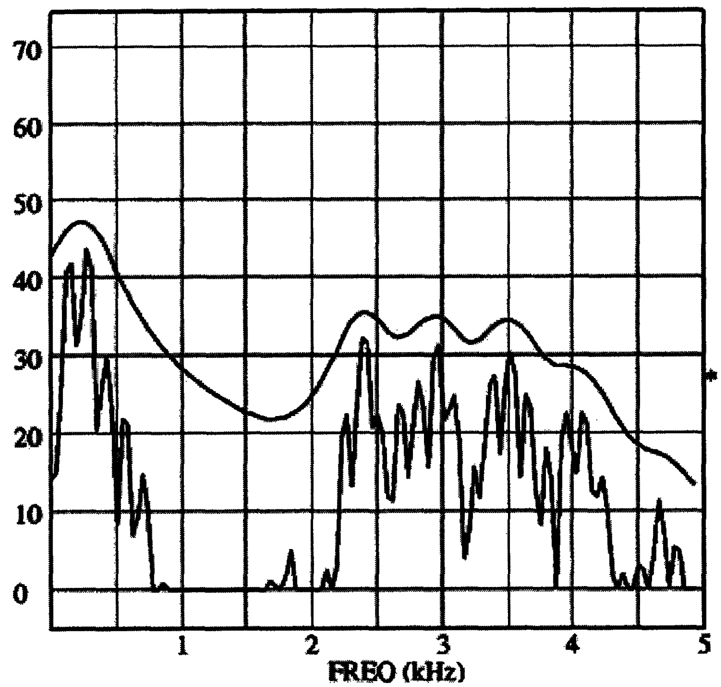


Figure 36: ***Vocal tract filtered spectrum of the vowel /i/ with the first formant centered around the second harmonic frequency.***

Although the above correction works, particularly for the vowel /a/, where F1 is far from F0 and 2F0, there is a problem. The problem with the above correction is that by approximating $\alpha \approx 0$ we also made the assumption that F1 has no bandwidth. However, as Figure 36 shows, if F1 is low enough in frequency, as in the case for the vowel /i/, the bandwidth B1 does have an influence on the amplitude of harmonics in the frequency range of F0 and 2F0. According to Hanson (1995) and Iseli and Alwan (2004), Equation 3.4 is most accurate only when F0 or 2F0 is at least a bandwidth away from F1. Thus for F1 close to or within the 0Hz – 500Hz frequency range, α_1 cannot be approximated as zero and must instead be estimated in Equation 3.

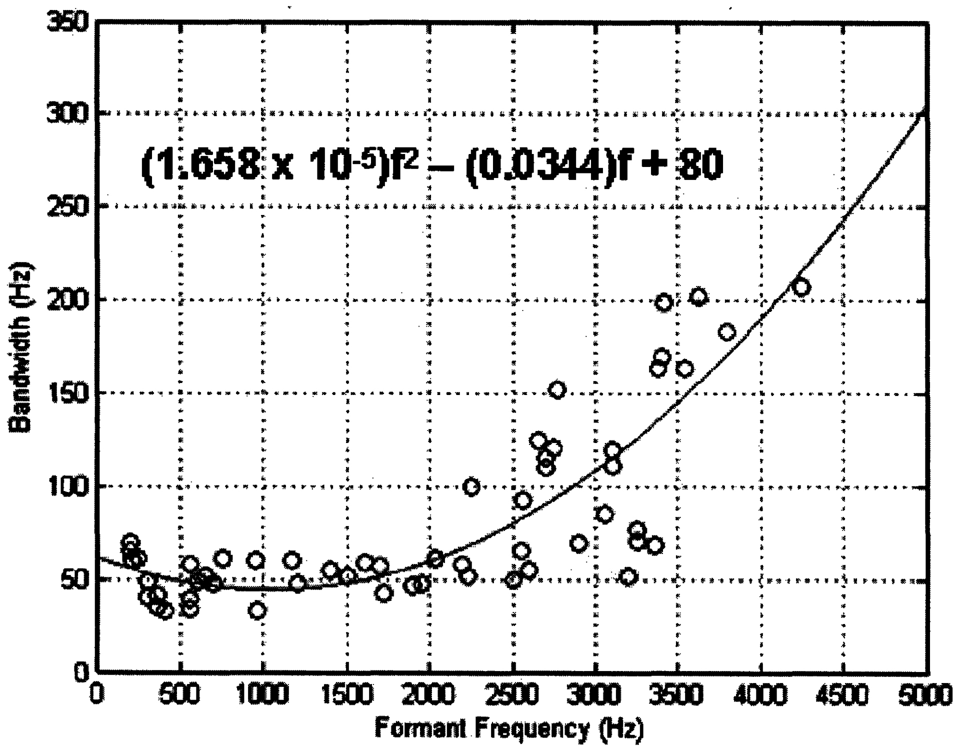


Figure 37: ***The average formant bandwidth as a function of frequency, obtained from sweep-tone measurements with the glottis closed. The data points are fitted with a 2nd order polynomial equation (Data obtained from Stevens, 1998).***

Figure 37 shows that the average formant bandwidth, obtained from sweep-tone measurements with the glottis closed, as a function of frequency (Stevens, 1998). Bandwidth (BW) values derived from the second order curve fitted to the data in Figure 37 were used to estimate α for Equation 2, where $\alpha = \pi BW$. For the production study of Chapter 2, the bandwidths of the formants were estimated using the second order equation from Figure 37 and then used to obtain a value from the transfer function of Equation 2. The resultant function was used to correct for the effects of formant locations on the amplitude of neighboring harmonic frequencies (i.e., H1 and H2), as well as the effect of neighboring formants on each others amplitude (i.e., F2 and F3 for the vowel /i/). This was accomplished by subtracting the transfer function quantity, in dB, from the measured parameter amplitudes (i.e., H1 and H2). In the case of the formant amplitudes, the quantity of the transfer function using the measured formant frequencies was subtracted from the measured formant amplitudes and the transfer function quantity using formant values for a neutral vocal tract of length 17.5cm was added. More detailed explanation of this process can be found in Hanson (1995) and Iseli and Alwan (2004). Equation 6 illustrates the vocal tract transfer function correction for H1.

$$H1^{*} = H1 - 20\log_{10}(T(\omega)) \qquad \text{Eq. 6}$$

where H1* is corrected for the effects of vocal tract transfer function (i.e., shape) on the measured H1 value and $T(\omega)$ is from Equation 2.

APPENDIX B: Individual Speaker Production Study Results

Word	Syllable	F1	F2	F3	F0pk	Int	Dur	Syllable	F1	F2	F3	F0pk	Int	Dur
dada1Fa	DM-S1'	609	1354	2360	105	65.5	0.167	DM-S2	560	1364	2425	94	63.0	0.154
	KL-S1'	622	1326	2562	144	81.8	0.160	KL-S2	487	1388	2531	133	78.1	0.114
	TM-S1'	620	1300	2431	92	69.4	0.208	TM-S2	510	1373	2490	85	65.4	0.178
	AM-S1'	614	1296	2459	91	69.3	0.210	AM-S2	513	1374	2495	85	65.3	0.183
	KF-S1'	721	1461	2405	86	72.4	0.223	KF-S2	543	1497	2474	78	68.7	0.116
	Ave:	**637**	**1347**	**2443**	**104**	**71.7**	**0.193**	Ave:	**523**	**1399**	**2483**	**95**	**68.1**	**0.149**
dada2Fa	DM-S1	521	1539	2360	95	64.9	0.120	DM-S2'	586	1346	2383	102	64.1	0.224
	KL-S1	356	1645	2539	122	76.1	0.094	KL-S2'	647	1280	2608	148	80.9	0.180
	TM-S1	482	1416	2513	80	67.0	0.149	TM-S2'	617	1275	2475	89	68.2	0.302
	AM-S1	440	1599	2438	113	73.8	0.091	AM-S2'	755	1331	2386	205	78.6	0.227
	KF-S1	417	1690	2595	86	71.0	0.094	KF-S2'	778	1449	2432	81	72.8	0.284
	Ave:	**443**	**1578**	**2489**	**99**	**70.6**	**0.109**	Ave:	**677**	**1336**	**2457**	**125**	**72.9**	**0.243**

Word	Syllable	F1	F2	F3	F0pk	Int	Dur	Syllable	F1	F2	F3	F0pk	Int	Dur
dodo1Fa	DM-S1'	422	1286	2112	107	65.7	0.162	DM-S2	435	1219	2096	99	64.7	0.175
	KL-S1'	404	1377	2396	175	82.4	0.178	KL-S2	411	1325	2347	163	79.6	0.138
	TM-S1'	445	1300	2331	93	71.7	0.205	TM-S2	443	1343	2337	88	67.0	0.200
	AM-S1'	482	1264	2409	221	82.9	0.217	AM-S2	471	1267	2278	111	74.1	0.190
	KF-S1'	478	1410	2556	92	76.8	0.174	KF-S2	453	1341	2596	80	73.4	0.138
	Ave:	**446**	**1328**	**2361**	**138**	**75.9**	**0.187**	Ave:	**443**	**1299**	**2331**	**108**	**71.7**	**0.168**
dodo2Fa	DM-S1	381	1484	2133	97	63.8	0.104	DM-S2'	420	1284	2151	104	66.3	0.222
	KL-S1	372	1680	2555	127	78.6	0.099	KL-S2'	427	1383	2464	185	82.1	0.198
	TM-S1	422	1319	2376	85	67.6	0.184	TM-S2'	440	1350	2362	89	70.1	0.307
	AM-S1	371	1680	2422	129	71.9	0.094	AM-S2'	462	1426	2513	212	82.0	0.289
	KF-S1	405	1439	2645	79	69.6	0.118	KF-S2'	513	1240	2583	94	76.8	0.269
	Ave:	**390**	**1520**	**2426**	**104**	**70.3**	**0.120**	Ave:	**452**	**1337**	**2415**	**137**	**75.5**	**0.257**

Word	Syllable	F1	F2	F3	F0pk	Int	Dur	Syllable	F1	F2	F3	F0pk	Int	Dur
didi1Fa	DM-S1'	271	2041	2646	108	61.9	0.150	DM-S2	281	2047	2565	101	61.1	0.154
	KL-S1'	284	2219	2781	163	77.3	0.150	KL-S2	276	2232	2687	152	75.0	0.133
	TM-S1'	278	2193	2505	95	63.1	0.197	TM-S2	285	2169	2492	91	60.7	0.231
	AM-S1'	451	2300	2773	238	79.7	0.206	AM-S2	326	2209	2612	121	72.9	0.162
	KF-S1'	291	2383	2830	99	71.6	0.189	KF-S2	296	2350	2805	79	67.1	0.136
	Ave:	**315**	**2227**	**2707**	**140**	**70.7**	**0.178**	Ave:	**293**	**2201**	**2632**	**109**	**67.4**	**0.163**
didi2Fa	DM-S1	310	1987	2588	101	61.6	0.119	DM-S2'	313	2008	2640	105	62.9	0.201
	KL-S1	323	2140	2635	128	74.7	0.121	KL-S2'	339	2190	2700	178	76.8	0.177
	TM-S1	276	2141	2475	93	62.1	0.184	TM-S2'	278	2174	2569	95	64.1	0.302
	AM-S1	365	2158	2503	122	72.7	0.095	AM-S2'	427	2211	2705	221	81.3	0.261
	KF-S1	309	2212	2682	97	65.2	0.128	KF-S2'	288	2373	2818	100	71.7	0.270
	Ave:	**317**	**2128**	**2577**	**108**	**67.3**	**0.129**	Ave:	**329**	**2191**	**2686**	**140**	**71.4**	**0.242**

Table 1: *Average Fa condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity (Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.*

dada1Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	563	1349	2318	94	64.0	0.152	DM-S2	537	1367	2383	86	59.3	0.153
KL-S1'	557	1416	2559	134	79.6	0.148	KL-S2	466	1400	2474	104	73.9	0.093
TM-S1'	540	1347	2467	80	65.6	0.178	TM-S2	531	1335	2493	72	60.2	0.186
AM-S1'	547	1345	2465	80	65.6	0.177	AM-S2	532	1330	2495	80	60.3	0.187
KF-S1'	663	1481	2384	87	70.9	0.174	KF-S2	482	1493	2433	74	65.1	0.099
Ave:	574	1388	2439	95	69.1	0.166	Ave:	510	1385	2456	83	63.8	0.143

dada2Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	508	1445	2271	93	64.2	0.125	DM-S2'	550	1318	2440	83	60.5	0.169
KL-S1	356	1650	2565	132	72.8	0.103	KL-S2'	577	1324	2565	114	70.2	0.173
TM-S1	469	1475	2445	82	64.9	0.138	TM-S2'	596	1328	2487	79	64.2	0.255
AM-S1	476	1461	2372	99	70.1	0.098	AM-S2'	562	1322	2456	90	68.1	0.178
KF-S1	432	1582	2619	80	71.9	0.097	KF-S2'	649	1415	2392	76	68.5	0.209
Ave:	448	1523	2454	97	68.8	0.112	Ave:	587	1341	2468	88	66.3	0.197

dodo1Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	430	1271	2057	98	65.8	0.145	DM-S2	433	1208	2031	88	62.3	0.149
KL-S1'	422	1427	2375	133	81.8	0.152	KL-S2	430	1318	2246	111	75.5	0.141
TM-S1'	415	1327	2336	85	68.2	0.183	TM-S2	443	1349	2393	82	65.7	0.198
AM-S1'	441	1289	2311	111	73.0	0.154	AM-S2	482	1194	2339	99	70.9	0.170
KF-S1'	478	1275	2499	80	72.7	0.148	KF-S2	450	1271	2645	68	67.1	0.130
Ave:	437	1318	2316	101	72.3	0.156	Ave:	447	1268	2331	90	68.3	0.157

dodo2Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	381	1318	2073	99	65.7	0.116	DM-S2'	430	1231	2145	90	64.4	0.185
KL-S1	352	1683	2565	144	78.7	0.104	KL-S2'	420	1435	2432	120	78.9	0.183
TM-S1	399	1427	2443	82	64.9	0.137	TM-S2'	452	1313	2408	81	67.0	0.230
AM-S1	449	1408	2303	110	72.8	0.116	AM-S2'	480	1209	2357	102	72.1	0.168
KF-S1	387	1465	2552	91	72.4	0.100	KF-S2'	477	1189	2541	76	72.1	0.196
Ave:	394	1460	2387	105	70.9	0.115	Ave:	452	1275	2376	94	70.9	0.193

didi1Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	281	2028	2651	98	61.2	0.125	DM-S2	273	2135	2583	87	57.7	0.135
KL-S1'	258	2167	2724	138	77.9	0.133	KL-S2	321	2156	2643	111	73.2	0.134
TM-S1'	318	2124	2451	87	64.0	0.167	TM-S2	247	2111	2477	81	60.9	0.172
AM-S1'	337	2156	2520	101	65.5	0.129	AM-S2	331	2165	2507	95	65.4	0.145
KF-S1'	287	2225	2793	87	66.2	0.153	KF-S2	282	2234	2818	80	61.5	0.115
Ave:	296	2140	2628	102	67.0	0.141	Ave:	291	2160	2606	91	63.7	0.140

didi2Fp1

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	296	2077	2572	102	61.8	0.109	DM-S2'	276	2015	2591	91	60.8	0.155
KL-S1	310	2086	2625	137	74.7	0.102	KL-S2'	313	2114	2633	117	73.8	0.163
TM-S1	280	2112	2452	84	62.0	0.156	TM-S2'	265	2121	2461	81	62.7	0.241
AM-S1	510	2166	2465	105	68.2	0.110	AM-S2'	339	2175	2511	98	67.7	0.159
KF-S1	294	2012	2565	91	69.6	0.094	KF-S2'	268	2240	2856	79	64.3	0.191
Ave:	338	2090	2536	104	67.2	0.114	Ave:	292	2133	2610	93	65.9	0.182

Table 2: *Average Fp1 condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity(Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.*

dada1Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	557	1362	2263	87	63.8	0.150	DM-S2	505	1427	2222	83	58.9	0.140
KL-S1'	563	1383	2531	116	77.3	0.151	KL-S2	435	1427	2518	106	71.7	0.094
TM-S1'	542	1345	2492	80	62.3	0.176	TM-S2	490	1327	2518	76	59.9	0.177
AM-S1'	529	1350	2482	83	63.0	0.177	AM-S2	518	1337	2497	77	60.1	0.175
KF-S1'	636	1435	2444	80	68.7	0.198	KF-S2	434	1576	2507	74	65.2	0.093
Ave:	565	1375	2442	89	67.0	0.171	Ave:	476	1419	2452	83	63.2	0.136

dada2Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	498	1471	2256	86	63.0	0.109	DM-S2'	557	1367	2367	82	59.9	0.172
KL-S1	413	1615	2526	109	71.3	0.083	KL-S2'	625	1328	2539	107	70.1	0.193
TM-S1	449	1504	2559	86	66.5	0.119	TM-S2'	521	1374	2702	82	66.4	0.239
AM-S1	443	1517	2311	96	67.8	0.094	AM-S2'	568	1317	2355	92	66.3	0.168
KF-S1	462	1634	2676	77	68.3	0.101	KF-S2'	573	1538	2523	75	66.1	0.215
Ave:	453	1548	2466	91	67.4	0.101	Ave:	569	1385	2497	88	65.8	0.197

dodo1Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	425	1180	2125	88	65.2	0.146	DM-S2	422	1180	2088	84	62.2	0.147
KL-S1'	422	1338	2347	125	76.9	0.151	KL-S2	430	1349	2274	110	72.6	0.132
TM-S1'	412	1362	2345	85	67.8	0.199	TM-S2	435	1379	2365	81	65.6	0.202
AM-S1'	454	1271	2295	101	69.6	0.151	AM-S2	447	1205	2361	94	66.1	0.139
KF-S1'	440	1311	2598	79	69.9	0.176	KF-S2	442	1264	2596	71	65.7	0.127
Ave:	431	1292	2342	96	69.9	0.165	Ave:	435	1275	2337	88	66.4	0.149

dodo2Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	387	1506	2100	92	64.4	0.103	DM-S2'	421	1276	2118	87	64.4	0.203
KL-S1	339	1576	2500	116	77.3	0.087	KL-S2'	443	1341	2422	109	75.6	0.182
TM-S1	400	1328	2405	85	66.0	0.167	TM-S2'	460	1301	2380	83	67.7	0.279
AM-S1	368	1595	2367	108	65.3	0.083	AM-S2'	465	1250	2451	90	67.3	0.177
KF-S1	408	1348	2513	77	69.8	0.111	KF-S2'	461	1180	2491	74	67.8	0.240
Ave:	380	1471	2377	96	68.6	0.110	Ave:	450	1270	2372	89	68.6	0.216

didi1Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	270	2039	2518	94	60.9	0.135	DM-S2	273	2034	2523	90	59.4	0.146
KL-S1'	326	2145	2585	120	74.0	0.127	KL-S2	306	2135	2637	112	70.1	0.121
TM-S1'	273	2193	2534	92	62.5	0.190	TM-S2	292	2156	2508	87	61.0	0.199
AM-S1'	339	2164	2570	112	68.9	0.113	AM-S2	328	2181	2533	107	66.8	0.164
KF-S1'	302	2271	2729	81	64.7	0.154	KF-S2	268	2250	2751	79	61.3	0.115
Ave:	302	2162	2587	100	66.2	0.144	Ave:	293	2151	2590	95	63.7	0.149

didi2Fp2

	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	286	2041	2432	92	61.2	0.110	DM-S2'	280	2018	2484	90	60.7	0.168
KL-S1	339	2138	2612	120	74.1	0.114	KL-S2'	313	2127	2646	116	72.1	0.131
TM-S1	328	2105	2446	86	62.7	0.145	TM-S2'	332	2105	2483	85	63.8	0.236
AM-S1	310	2121	2435	103	68.5	0.102	AM-S2'	298	2174	2505	99	66.8	0.165
KF-S1	287	2134	2632	77	63.5	0.101	KF-S2'	282	2230	2816	76	63.0	0.187
Ave:	310	2108	2511	95	66.0	0.114	Ave:	301	2131	2587	93	65.3	0.177

Table 3: *Average Fp2 condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity (Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.*

		F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
statue1Fa	DM-S1'	555	1565	2432	123	65.8	0.300	DM-S2	344	1584	1995	97	62.3	0.238
	KL-S1'	547	1677	2604	206	77.5	0.295	KL-S2	315	1774	2305	151	70.1	0.178
	TM-S1'	574	1527	2413	93	67.5	0.357	TM-S2	335	1694	2143	86	63.4	0.259
	AM-S1'	573	1569	2363	240	82.5	0.304	AM-S2	391	1419	2129	101	68.0	0.249
	KF-S1'	555	1735	2544	159	79.2	0.344	KF-S2	297	1831	2354	92	68.7	0.215
	Ave:	561	1615	2471	164	74.5	0.320	Ave:	336	1660	2185	106	66.5	0.228
tattoo2Fa	DM-S1	498	1523	2422	105	61.4	0.179	DM-S2'	297	1245	2065	127	64.1	0.276
	KL-S1	464	1659	2487	140	74.4	0.153	KL-S2'	414	1578	2419	213	73.3	0.208
	TM-S1	544	1568	2420	88	62.0	0.217	TM-S2'	360	1667	2138	93	65.1	0.314
	AM-S1	563	1618	2279	119	75.7	0.162	AM-S2'	410	1397	2233	210	84.7	0.269
	KF-S1	542	1725	2471	99	75.7	0.179	KF-S2'	301	1384	2397	128	75.2	0.285
	Ave:	522	1619	2416	110	69.8	0.178	Ave:	356	1454	2251	155	72.5	0.270

		F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
sushi1Fa	DM-S1'	281	1305	2154	145	67.7	0.279	DM-S2	286	2000	2287	103	61.5	0.243
	KL-S1'	482	1723	2431	177	73.1	0.213	KL-S2	300	2096	2366	145	71.0	0.152
	TM-S1'	305	1682	2214	93	67.4	0.314	TM-S2	324	2081	2365	86	65.9	0.273
	AM-S1'	409	1429	2271	215	81.1	0.261	AM-S2	350	2075	2365	101	67.7	0.222
	KF-S1'	314	1412	2478	153	78.9	0.305	KF-S2	290	2103	2539	141	70.0	0.280
	Ave:	358	1510	2309	157	73.6	0.274	Ave:	310	2071	2384	115	67.2	0.234
bouquet2Fa	DM-S1	323	951	2117	111	66.0	0.108	DM-S2'	383	1810	2323	135	66.2	0.318
	KL-S1	370	1279	2336	130	74.4	0.098	KL-S2'	440	1810	2409	171	77.7	0.215
	TM-S1	332	1259	2184	87	67.5	0.134	TM-S2'	396	2013	2396	91	68.6	0.346
	AM-S1	339	975	2259	116	70.3	0.096	AM-S2'	512	1710	2467	225	80.3	0.297
	KF-S1	365	997	2315	92	73.3	0.120	KF-S2'	456	2073	2500	150	81.0	0.331
	Ave:	346	1092	2242	107	70.3	0.111	Ave:	438	1883	2419	154	74.7	0.301

Table 4: *Average Fa condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity(Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.*

		F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
statue1Fp1	DM-S1'	505	1539	2388	101	63.7	0.273	DM-S2	339	1641	2010	94	61.0	0.203
	KL-S1'	537	1623	2458	136	76.5	0.252	KL-S2	323	1727	2177	117	69.3	0.154
	TM-S1'	540	1553	2427	79	65.3	0.325	TM-S2	350	1699	2145	73	60.9	0.231
	AM-S1'	499	1572	2470	109	70.9	0.229	AM-S2	344	1642	2139	100	65.8	0.182
	KF-S1'	560	1697	2565	109	75.5	0.307	KF-S2	289	1706	2335	76	62.6	0.148
	Ave:	528	1597	2462	107	70.4	0.277	Ave:	329	1683	2161	92	63.9	0.184
tattoo2Fp1	DM-S1	469	1506	2396	103	62.0	0.171	DM-S2'	321	1280	2018	96	59.7	0.246
	KL-S1	463	1615	2494	149	75.2	0.165	KL-S2'	333	1615	2351	124	71.0	0.199
	TM-S1	488	1549	2450	82	59.9	0.191	TM-S2'	352	1723	2250	82	60.9	0.235
	AM-S1	488	1619	2476	129	74.7	0.162	AM-S2'	319	1417	2201	108	69.0	0.178
	KF-S1	462	1660	2608	105	71.6	0.148	KF-S2'	293	1605	2321	75	66.2	0.216
	Ave:	474	1590	2485	114	68.7	0.167	Ave:	324	1528	2228	97	65.4	0.215

		F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
sushi1Fp1	DM-S1'	302	1333	2049	104	62.2	0.246	DM-S2	273	1948	2274	94	60.2	0.194
	KL-S1'	334	1658	2400	128	74.1	0.207	KL-S2	326	2135	2374	110	67.6	0.166
	TM-S1'	335	1664	2168	88	64.8	0.275	TM-S2	326	2146	2380	82	61.9	0.242
	AM-S1'	337	1550	2191	114	67.6	0.190	AM-S2	310	2152	2406	101	64.6	0.165
	KF-S1'	304	1445	2435	96	73.4	0.234	KF-S2	289	2105	2559	84	66.1	0.198
	Ave:	322	1530	2249	106	68.4	0.231	Ave:	305	2097	2399	94	64.1	0.193
bouquet2Fp1	DM-S1	381	917	2114	101	64.2	0.118	DM-S2'	375	1740	2222	95	62.3	0.255
	KL-S1	372	1231	2357	128	77.6	0.094	KL-S2'	411	1791	2370	112	72.9	0.197
	TM-S1	345	1206	2122	90	66.9	0.125	TM-S2'	412	1858	2435	86	63.2	0.274
	AM-S1	353	1076	2245	113	69.3	0.111	AM-S2'	456	1692	2331	111	72.2	0.190
	KF-S1	341	1052	2464	123	74.4	0.102	KF-S2'	409	2057	2536	96	75.8	0.241
	Ave:	358	1096	2260	111	70.5	0.110	Ave:	413	1827	2379	100	69.3	0.231

Table 5: ***Average Fp1 condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity (Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.***

statue1Fp2	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	531	1534	2266	95	63.1	0.283	DM-S2	349	1628	2023	91	61.6	0.188
KL-S1'	498	1633	2471	130	74.9	0.250	KL-S2	323	1748	2190	118	68.1	0.157
TM-S1'	579	1517	2441	87	66.3	0.347	TM-S2	330	1711	2155	80	60.7	0.253
AM-S1'	518	1587	2521	120	73.6	0.230	AM-S2	370	1529	2150	97	65.6	0.191
KF-S1'	568	1669	2463	91	73.6	0.315	KF-S2	285	1859	2448	75	62.6	0.154
Ave:	539	1588	2432	104	70.3	0.285	Ave:	331	1695	2193	92	63.7	0.188

tattoo2Fp2	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	524	1505	2237	98	61.1	0.188	DM-S2'	334	1370	2284	94	61.0	0.204
KL-S1	479	1618	2458	125	73.1	0.153	KL-S2'	336	1634	2292	120	71.2	0.168
TM-S1	511	1563	2422	89	60.1	0.199	TM-S2'	335	1694	2139	88	64.3	0.320
AM-S1	521	1511	2389	108	68.4	0.165	AM-S2'	371	1345	2171	99	65.1	0.172
KF-S1	485	1620	2634	88	71.1	0.147	KF-S2'	300	1453	2515	79	65.8	0.185
Ave:	504	1563	2428	101	66.8	0.170	Ave:	335	1499	2280	96	65.5	0.210

sushi1Fp2	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1'	307	1406	2047	99	62.6	0.252	DM-S2	302	2065	2365	90	60.2	0.211
KL-S1'	326	1641	2383	135	70.9	0.214	KL-S2	332	2109	2370	123	67.5	0.151
TM-S1'	317	1701	2189	83	64.9	0.272	TM-S2	311	2086	2345	81	63.0	0.250
AM-S1'	342	1541	2310	109	67.1	0.175	AM-S2	329	2241	2494	99	64.1	0.184
KF-S1'	326	1372	2459	94	71.6	0.260	KF-S2	296	2132	2505	80	64.3	0.235
Ave:	323	1532	2277	104	67.4	0.235	Ave:	314	2127	2416	95	63.8	0.206

bouquet2Fp2	F1	F2	F3	F0pk	Int	Dur		F1	F2	F3	F0pk	Int	Dur
DM-S1	373	932	2117	96	66.1	0.113	DM-S2'	388	1805	2224	97	63.9	0.271
KL-S1	352	1250	2331	124	76.9	0.111	KL-S2'	443	1771	2240	122	74.9	0.194
TM-S1	332	1241	2190	86	67.9	0.135	TM-S2'	402	1938	2393	86	66.1	0.286
AM-S1	397	1094	2315	94	64.3	0.103	AM-S2'	459	1728	2451	92	69.4	0.206
KF-S1	323	1036	2477	121	71.4	0.104	KF-S2'	404	2044	2555	92	73.1	0.240
Ave:	355	1111	2286	104	69.3	0.113	Ave:	419	1857	2373	98	69.5	0.240

Table 6: ***Average Fp2 condition measurement values for the first three formants (F1,F2, andF3); the peak fundamental frequency value (F0pk) within the vowel; the peak intensity(Int) within the vowel; and the syllable duration (Dur) of each syllable for all speakers. Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1, and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.***

dada1Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1'	32.5	36.3	40.8	27.7	3.2	DM-S2	29.2	34.5	39.3	18.1	3.0
KL-S1'	47.4	48.1	57.5	49.8	2.4	KL-S2	45.4	46.3	58.7	45.9	3.0
TM-S1'	32.0	41.8	47.5	30.8	2.6	TM-S2	29.2	37.6	45.7	23.8	3.0
AM-S1'	31.8	41.4	47.1	30.4	2.0	AM-S2	29.3	37.8	45.6	24.4	3.0
KF-S1'	34.1	43.3	47.6	32.4	3.4	KF-S2	34.1	42.0	51.5	25.3	3.8
Ave:	35.6	42.2	48.1	34.2	2.7	Ave:	33.4	39.7	48.2	27.5	3.2

dada2Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1	30.6	36.0	43.8	25.0	2.6	DM-S2'	32.6	36.0	40.5	26.3	3.4
KL-S1	46.3	42.6	60.4	47.2	3.0	KL-S2'	46.5	48.0	56.7	49.2	2.8
TM-S1	29.7	38.6	50.5	25.9	2.8	TM-S2'	31.6	41.0	45.1	31.2	2.2
AM-S1	42.6	42.8	56.1	37.2	5.5	AM-S2'	45.4	47.3	48.5	39.9	4.3
KF-S1	35.3	42.8	56.3	29.8	4.0	KF-S2'	32.8	42.6	45.7	32.2	3.0
Ave:	36.9	40.6	53.4	33.0	3.6	Ave:	37.8	43.0	47.3	35.7	3.1

dodo1Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1'	33.1	35.3	48.9	22.2	3.8	DM-S2	30.6	35.0	46.5	20.9	4.0
KL-S1'	50.6	50.1	65.8	50.5	3.8	KL-S2	47.9	47.5	61.1	45.3	4.3
TM-S1'	33.0	43.2	56.8	27.6	3.2	TM-S2	29.4	39.2	51.4	20.2	3.6
AM-S1'	52.3	50.5	64.3	48.5	5.6	AM-S2	38.7	44.8	53.5	29.9	5.8
KF-S1'	38.1	48.3	61.1	36.2	4.2	KF-S2	35.5	44.9	58.6	34.6	5.0
Ave:	41.4	45.5	59.4	37.0	4.1	Ave:	36.4	42.3	54.2	30.2	4.5

dodo2Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1	30.5	34.7	47.9	20.1	3.2	DM-S2'	32.8	36.1	49.4	24.5	3.6
KL-S1	49.1	46.5	63.8	43.2	4.3	KL-S2'	50.8	50.2	65.0	49.2	3.8
TM-S1	29.0	38.6	54.1	23.1	3.6	TM-S2'	32.2	41.8	56.2	28.8	2.8
AM-S1	48.9	43.2	56.0	41.2	5.0	AM-S2'	52.5	50.6	64.3	51.9	5.0
KF-S1	33.9	41.4	54.2	30.8	5.3	KF-S2'	37.3	46.8	58.6	40.9	4.5
Ave:	38.3	40.9	55.2	31.7	4.3	Ave:	41.1	45.1	58.7	39.1	3.9

didi1Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1'	32.8	34.3	49.3	26.1	3.2	DM-S2	29.1	33.3	49.7	19.3	3.6
KL-S1'	49.6	46.1	65.2	43.2	3.0	KL-S2	46.8	44.0	62.7	40.6	3.0
TM-S1'	31.7	40.0	52.5	16.0	5.2	TM-S2	28.2	36.0	48.8	9.5	5.4
AM-S1'	55.4	38.4	57.8	39.9	4.2	AM-S2	39.5	44.5	57.7	27.6	5.2
KF-S1'	40.8	48.7	62.7	33.1	3.8	KF-S2	34.3	42.3	57.8	28.4	4.4
Ave:	42.1	41.5	57.5	31.6	3.9	Ave:	35.6	40.0	55.3	25.1	4.3

didi2Fa

	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1	30.9	34.8	48.7	22.8	3.2	DM-S2'	32.9	36.7	50.0	24.5	3.6
KL-S1	49.0	44.6	59.1	38.8	3.2	KL-S2'	51.5	43.1	62.3	39.5	3.0
TM-S1	31.2	39.7	51.7	14.1	4.8	TM-S2'	32.9	40.6	52.0	18.7	4.4
AM-S1	47.9	47.3	53.6	27.4	4.8	AM-S2'	55.7	43.3	62.5	42.2	4.0
KF-S1	36.2	43.0	53.8	22.0	5.3	KF-S2'	40.3	47.3	63.6	34.6	3.8
Ave:	39.0	41.9	53.4	25.0	4.2	Ave:	42.7	42.2	58.1	31.9	3.8

Table 7: ***Average Fa condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

Word	Syllable	H1*	H2*	A1*	A3*	Nw	Syllable	H1*	H2*	A1*	A3*	Nw
dada1Fp1	DM-S1'	29.5	33.6	38.9	25.6	3.0	DM-S2	26.2	30.4	36.1	14.5	3.8
	KL-S1'	45.5	44.1	56.7	46.5	3.0	KL-S2	39.1	39.0	53.9	41.8	2.8
	TM-S1'	30.1	38.5	45.3	25.5	2.6	TM-S2	28.4	36.4	38.9	20.6	3.0
	AM-S1'	30.0	38.6	46.0	25.5	2.2	AM-S2	28.5	36.3	38.9	21.4	3.0
	KF-S1'	33.2	41.8	47.5	29.8	4.5	KF-S2	32.7	39.9	48.8	24.0	6.3
	Ave:	33.7	39.3	46.9	30.6	3.1	Ave:	30.9	36.4	43.3	24.5	3.8
dada2Fp1	DM-S1	29.7	33.7	43.4	21.7	3.4	DM-S2'	26.0	30.6	35.6	21.7	3.2
	KL-S1	44.7	41.1	56.6	44.0	3.0	KL-S2'	40.5	40.1	54.3	46.9	2.3
	TM-S1	29.9	38.8	48.1	21.7	4.0	TM-S2'	30.4	39.2	40.8	24.0	2.5
	AM-S1	37.8	44.4	54.3	35.6	5.4	AM-S2'	34.6	42.4	45.8	37.9	4.8
	KF-S1	35.7	43.7	56.9	24.7	5.4	KF-S2'	32.4	40.9	45.3	27.0	4.8
	Ave:	35.5	40.3	51.9	29.5	4.2	Ave:	32.8	38.6	44.4	31.5	3.5

Word	Syllable	H1*	H2*	A1*	A3*	Nw	Syllable	H1*	H2*	A1*	A3*	Nw
dodo1Fp1	DM-S1'	31.0	35.3	48.2	18.6	3.6	DM-S2	28.0	31.9	44.8	7.4	5.4
	KL-S1'	50.0	49.9	64.2	48.1	3.6	KL-S2	43.5	41.8	58.0	37.5	3.0
	TM-S1'	29.9	37.8	54.4	25.0	4.0	TM-S2	27.8	37.4	50.2	22.4	3.6
	AM-S1'	43.0	46.0	55.8	37.6	4.7	AM-S2	39.5	46.4	53.5	32.4	5.0
	KF-S1'	34.5	43.5	56.2	31.1	5.0	KF-S2	32.3	40.2	51.6	25.0	5.4
	Ave:	37.7	42.5	55.8	32.1	4.2	Ave:	34.2	39.5	51.6	25.0	4.5
dodo2Fp1	DM-S1	30.5	34.8	49.4	20.4	4.3	DM-S2'	29.4	33.2	46.3	17.0	4.3
	KL-S1	48.4	48.0	62.8	46.8	3.0	KL-S2'	44.5	44.0	62.2	43.5	2.8
	TM-S1	28.7	38.2	50.3	20.9	3.8	TM-S2'	28.9	39.2	51.0	24.2	3.0
	AM-S1	42.5	46.1	57.0	35.8	5.3	AM-S2'	40.0	46.0	55.5	36.8	5.0
	KF-S1	36.5	43.0	58.5	32.7	5.6	KF-S2'	33.7	41.6	54.8	35.0	4.4
	Ave:	37.3	42.0	55.6	31.3	4.4	Ave:	35.3	40.8	54.0	31.3	3.9

Word	Syllable	H1*	H2*	A1*	A3*	Nw	Syllable	H1*	H2*	A1*	A3*	Nw
didi1Fp1	DM-S1'	29.4	34.4	49.7	20.0	3.2	DM-S2	26.0	30.3	45.0	12.9	4.0
	KL-S1'	48.6	46.6	65.3	48.6	3.0	KL-S2	42.9	43.7	58.5	39.3	2.6
	TM-S1'	30.4	38.9	52.5	19.4	2.8	TM-S2	27.1	36.4	50.5	16.2	4.3
	AM-S1'	38.6	44.4	48.7	26.0	4.4	AM-S2	36.0	44.7	49.7	24.4	4.4
	KF-S1'	34.7	43.3	56.3	29.1	5.0	KF-S2	34.0	40.3	50.8	23.5	5.4
	Ave:	36.3	41.5	54.5	28.6	3.7	Ave:	33.2	39.1	50.9	23.3	4.1
didi2Fp1	DM-S1	31.2	35.0	49.3	18.7	4.0	DM-S2'	28.5	32.7	48.6	21.8	3.3
	KL-S1	47.5	42.8	58.9	39.8	4.2	KL-S2'	43.9	43.7	59.6	38.5	2.4
	TM-S1	29.1	37.3	50.5	19.2	3.6	TM-S2'	29.0	38.4	52.8	21.8	3.0
	AM-S1	41.6	45.0	46.5	21.8	5.0	AM-S2'	37.9	45.4	52.9	26.7	4.7
	KF-S1	38.4	45.6	59.1	29.8	5.0	KF-S2'	33.9	42.5	53.9	29.7	4.2
	Ave:	37.6	41.1	52.9	25.9	4.4	Ave:	34.6	40.5	53.6	27.7	3.5

Table 8: ***Average Fp1 condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

dada1Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1'	27.3	32.7	38.7	23.7	2.8	DM-S2	24.8	30.0	36.4	13.7	3.0
	KL-S1'	42.0	40.8	53.4	45.1	2.4	KL-S2	40.2	36.8	52.5	42.4	2.4
	TM-S1'	29.5	38.6	41.7	21.3	4.0	TM-S2	27.4	36.2	40.4	21.6	3.8
	AM-S1'	30.0	38.9	42.8	23.1	3.4	AM-S2	27.3	36.3	39.9	20.3	3.6
	KF-S1'	34.6	41.8	43.2	27.2	5.0	KF-S2	33.3	40.1	50.5	21.1	5.6
	Ave:	32.7	38.6	44.0	28.1	3.5	Ave:	30.6	35.9	43.9	23.8	3.7

dada2Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1	28.0	32.7	43.3	19.2	3.5	DM-S2'	26.4	31.0	35.8	19.5	3.3
	KL-S1	39.5	36.3	54.9	46.1	3.0	KL-S2'	39.2	36.6	49.0	41.2	3.0
	TM-S1	28.2	38.2	51.5	26.6	2.0	TM-S2'	28.6	38.6	46.7	26.0	3.0
	AM-S1	36.8	43.7	51.4	32.9	6.3	AM-S2'	35.1	43.5	43.0	35.7	4.8
	KF-S1	33.6	42.2	53.2	25.1	5.3	KF-S2'	33.1	40.9	40.8	24.2	5.3
	Ave:	33.2	38.6	50.9	30.0	4.0	Ave:	32.5	38.1	43.1	29.3	3.9

dodo1Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1'	28.3	33.2	47.9	20.1	3.2	DM-S2	26.9	30.9	43.7	12.9	3.8
	KL-S1'	44.8	40.1	59.2	42.9	2.8	KL-S2	41.1	38.1	55.7	34.2	3.0
	TM-S1'	30.2	38.6	53.3	26.5	3.0	TM-S2	28.7	38.2	49.5	23.3	3.2
	AM-S1'	40.1	45.8	52.5	37.4	5.0	AM-S2	36.7	44.5	49.3	32.8	5.0
	KF-S1'	35.0	42.9	55.6	28.4	4.6	KF-S2	32.6	40.5	51.9	24.1	5.0
	Ave:	35.7	40.1	53.7	31.1	3.7	Ave:	33.2	38.4	50.0	25.5	4.0

dodo2Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1	29.4	32.8	47.9	17.4	3.3	DM-S2'	29.3	33.0	47.1	19.1	3.3
	KL-S1	43.9	41.5	59.9	40.4	3.0	KL-S2'	43.2	39.5	58.1	39.7	3.0
	TM-S1	28.9	38.1	52.3	24.7	4.2	TM-S2'	29.8	39.1	52.0	26.1	3.4
	AM-S1	36.8	42.5	53.6	30.9	6.0	AM-S2'	36.3	45.1	52.1	33.4	4.5
	KF-S1	31.7	40.7	55.5	30.1	5.2	KF-S2'	32.0	39.6	51.1	28.1	5.0
	Ave:	34.1	39.1	53.8	28.7	4.3	Ave:	34.1	39.3	52.1	29.3	3.8

didi1Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1'	28.6	32.3	50.0	16.7	3.0	DM-S2	26.9	31.3	48.1	17.1	3.4
	KL-S1'	45.5	44.1	59.0	37.7	2.8	KL-S2	41.5	39.5	56.3	36.7	3.0
	TM-S1'	31.9	40.5	52.7	12.2	4.8	TM-S2	29.5	37.4	49.7	7.9	5.0
	AM-S1'	45.0	47.7	53.9	30.9	4.3	AM-S2	40.6	44.2	53.6	24.9	5.0
	KF-S1'	34.3	41.8	54.6	25.0	4.8	KF-S2	32.8	38.6	52.3	22.4	5.6
	Ave:	37.0	41.3	54.1	24.5	3.9	Ave:	34.3	38.2	52.0	21.8	4.4

didi2Fp2		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1	29.5	32.0	49.3	17.5	3.3	DM-S2'	27.4	31.7	47.3	20.3	3.0
	KL-S1	47.4	44.9	58.5	37.9	3.6	KL-S2'	44.1	42.2	57.9	38.2	2.8
	TM-S1	30.0	38.2	50.2	13.0	3.7	TM-S2'	30.2	38.9	52.0	18.2	2.7
	AM-S1	40.9	45.8	55.7	27.1	5.4	AM-S2'	38.5	44.6	53.9	27.6	5.6
	KF-S1	34.4	41.5	52.5	22.7	5.0	KF-S2'	33.6	41.0	52.7	25.1	4.3
	Ave:	36.5	40.5	53.2	23.6	4.2	Ave:	34.7	39.7	52.7	25.9	3.7

Table 9: ***Average Fp2 condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

statue1Fa		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1'	34.8	36.4	43.7	28.0	3.8	DM-S2	28.9	33.7	47.3	9.5	4.2
	KL-S1'	46.5	44.8	57.0	42.6	2.8	KL-S2	43.9	39.0	56.4	32.8	5.2
	TM-S1'	32.3	41.2	45.8	29.0	2.6	TM-S2	29.4	37.2	51.4	18.7	3.6
	AM-S1'	33.3	52.4	60.6	42.0	4.5	AM-S2	35.4	41.7	47.6	20.0	6.0
	KF-S1'	42.2	46.6	58.1	36.8	4.6	KF-S2	37.0	38.8	54.4	28.4	5.8
	Ave:	37.8	44.3	53.0	35.7	3.7	Ave:	34.9	38.1	51.4	21.9	5.0

tattoo2Fa		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1	31.8	34.8	39.2	22.7	4.4	DM-S2'	34.9	34.5	51.0	20.9	5.6
	KL-S1	47.2	42.0	54.7	40.2	4.0	KL-S2'	50.6	35.2	54.4	40.9	6.0
	TM-S1	30.3	37.7	40.3	23.3	3.0	TM-S2'	32.1	40.6	52.4	18.9	3.8
	AM-S1	42.9	45.8	55.5	31.9	5.5	AM-S2'	55.9	52.0	71.2	49.1	6.5
	KF-S1	40.5	48.8	56.9	33.8	3.8	KF-S2'	46.3	48.7	65.0	50.5	2.8
	Ave:	38.5	41.8	49.3	30.4	4.1	Ave:	44.0	42.2	58.8	36.1	4.9

sushi1Fa		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1'	35.6	37.5	56.7	29.7	4.4	DM-S2	29.8	34.0	49.5	19.6	3.4
	KL-S1'	48.6	39.9	52.0	29.8	5.0	KL-S2	42.5	37.2	54.7	28.4	5.0
	TM-S1'	34.5	41.6	55.5	27.7	4.0	TM-S2	31.4	39.5	54.0	23.2	4.2
	AM-S1'	55.9	46.7	65.9	47.2	5.3	AM-S2	37.3	43.4	53.5	18.3	4.8
	KF-S1'	47.0	50.2	68.3	46.1	5.4	KF-S2	34.6	42.2	59.0	30.1	4.4
	Ave:	44.3	43.2	59.7	36.1	4.8	Ave:	35.1	39.3	54.1	23.9	4.4

bouquet2Fa		H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
	DM-S1	35.9	34.5	51.2	28.0	4.8	DM-S2'	38.4	36.9	49.6	27.0	3.8
	KL-S1	47.3	42.7	58.5	31.7	4.4	KL-S2'	47.3	46.5	60.3	41.8	4.0
	TM-S1	32.1	39.3	54.2	23.2	3.8	TM-S2'	34.0	42.0	54.0	27.7	2.4
	AM-S1	46.7	42.6	55.4	30.9	6.0	AM-S2'	50.2	48.9	62.7	41.6	4.7
	KF-S1	38.6	45.7	56.1	32.0	5.0	KF-S2'	41.9	46.6	62.5	30.5	3.8
	Ave:	40.1	41.0	55.1	29.1	4.8	Ave:	42.4	44.2	57.8	33.7	3.7

Table 10: ***Average Fa condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

statue1Fp1	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1'	28.8	34.9	41.4	24.5	3.4	DM-S2	27.5	32.5	45.8	5.8	3.8
KL-S1'	45.1	42.0	54.7	41.0	2.8	KL-S2	40.2	36.8	54.7	30.2	5.0
TM-S1'	30.4	39.0	44.7	26.5	3.3	TM-S2	28.5	36.7	47.8	12.9	4.5
AM-S1'	40.6	44.1	52.8	34.0	5.2	AM-S2	37.4	43.0	50.9	18.8	5.4
KF-S1'	37.1	43.1	53.6	30.8	5.3	KF-S2	32.4	37.1	49.9	23.4	5.3
Ave:	36.4	40.6	49.5	31.3	4.0	Ave:	33.2	37.2	49.8	18.2	4.8

tatoo2Fp1	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1	31.5	34.4	40.9	21.7	4.0	DM-S2'	27.5	30.8	45.1	7.7	4.7
KL-S1	45.6	41.7	54.8	39.5	4.5	KL-S2'	40.8	36.8	57.0	36.3	4.5
TM-S1	30.7	37.5	39.3	19.5	4.7	TM-S2'	28.9	36.8	46.2	12.2	5.3
AM-S1	46.9	47.0	56.8	34.8	5.0	AM-S2'	41.4	44.2	56.0	33.2	5.3
KF-S1	40.6	44.5	54.8	32.0	5.0	KF-S2'	32.3	40.5	56.3	29.3	5.0
Ave:	39.0	41.0	49.3	29.5	4.6	Ave:	34.2	37.8	52.1	23.7	5.0

sushi1Fp1	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1'	29.9	33.3	48.0	17.4	3.8	DM-S2	27.4	32.0	48.0	18.4	4.4
KL-S1'	45.2	43.4	59.4	33.2	4.3	KL-S2	37.6	37.6	51.5	23.9	4.0
TM-S1'	31.2	38.5	52.6	23.2	4.0	TM-S2	28.7	37.1	49.7	13.0	4.2
AM-S1'	43.1	42.5	54.4	30.4	4.3	AM-S2	38.6	43.6	50.5	18.1	4.3
KF-S1'	38.8	46.8	62.8	42.9	6.7	KF-S2	33.9	41.2	53.3	25.9	4.3
Ave:	37.6	40.9	55.4	29.4	4.6	Ave:	33.3	38.3	50.6	19.9	4.2

bouquet2Fp1	H1*	H2*	A1*	A3*	Nw		H1*	H2*	A1*	A3*	Nw
DM-S1	30.9	34.2	46.9	20.0	4.6	DM-S2'	28.4	32.6	46.5	18.5	3.2
KL-S1	48.4	46.1	62.8	37.9	6.0	KL-S2'	40.6	39.1	56.1	37.2	3.5
TM-S1	31.9	40.4	53.4	22.3	4.2	TM-S2'	30.7	38.4	48.0	24.4	3.0
AM-S1	43.5	42.7	53.7	31.5	6.0	AM-S2'	40.4	44.0	54.2	31.1	4.8
KF-S1	40.8	44.4	58.0	34.5	4.8	KF-S2'	37.8	42.7	58.9	27.3	3.8
Ave:	39.1	41.6	54.9	29.3	5.1	Ave:	35.6	39.4	52.7	27.7	3.7

Table 11: ***Average Fp1 condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

Word	Speaker	H1*	H2*	A1*	A3*	Nw	Speaker	H1*	H2*	A1*	A3*	Nw
statue1Fp2	DM-S1'	28.1	33.6	40.3	20.8	3.6	DM-S2	27.0	31.9	45.8	6.5	4.6
	KL-S1'	43.8	38.0	53.5	41.8	2.8	KL-S2	39.3	35.6	52.9	29.4	4.4
	TM-S1'	31.3	39.1	43.5	27.4	2.4	TM-S2	28.1	35.5	47.6	17.9	4.0
	AM-S1'	39.2	45.1	55.2	39.3	5.8	AM-S2	36.6	43.1	50.8	16.3	5.8
	KF-S1'	37.1	40.6	52.1	25.7	4.4	KF-S2	32.3	34.0	49.3	20.1	5.4
	Ave:	35.9	39.3	48.9	31.0	3.8	Ave:	32.7	36.0	49.3	18.0	4.8

Word	Speaker	H1*	H2*	A1*	A3*	Nw	Speaker	H1*	H2*	A1*	A3*	Nw
tattoo2Fp2	DM-S1	29.8	34.3	38.0	17.3	4.2	DM-S2'	27.8	32.6	46.4	14.9	3.8
	KL-S1	45.2	39.7	53.3	37.9	4.8	KL-S2'	42.5	38.2	57.7	34.7	4.3
	TM-S1	30.9	36.7	38.1	24.2	3.0	TM-S2'	30.0	37.5	52.7	23.1	3.3
	AM-S1	38.0	44.0	49.5	31.3	5.0	AM-S2'	36.1	42.4	49.6	26.3	4.5
	KF-S1	39.7	42.4	51.1	29.5	5.3	KF-S2'	34.3	41.4	55.6	36.1	5.8
	Ave:	36.7	39.4	46.0	28.0	4.4	Ave:	34.1	38.4	52.4	27.0	4.3

Word	Speaker	H1*	H2*	A1*	A3*	Nw	Speaker	H1*	H2*	A1*	A3*	Nw
sushi1Fp2	DM-S1'	29.3	33.7	48.1	19.4	3.8	DM-S2	26.3	31.6	46.8	18.6	3.4
	KL-S1'	44.8	39.5	56.3	35.6	6.0	KL-S2	41.0	35.1	50.8	23.3	4.5
	TM-S1'	31.2	39.0	52.3	19.5	5.2	TM-S2	30.7	38.3	50.6	17.7	5.0
	AM-S1'	41.9	42.2	52.3	32.1	5.3	AM-S2	36.5	43.0	49.2	16.6	5.0
	KF-S1'	37.5	45.9	60.4	41.0	5.5	KF-S2	32.8	39.7	51.4	19.1	5.8
	Ave:	36.9	40.0	53.9	29.5	5.2	Ave:	33.5	37.5	49.7	19.1	4.7

Word	Speaker	H1*	H2*	A1*	A3*	Nw	Speaker	H1*	H2*	A1*	A3*	Nw
bouquet2Fp2	DM-S1	30.2	34.8	49.4	22.7	4.6	DM-S2'	28.6	33.3	47.8	17.8	3.6
	KL-S1	48.0	46.2	63.2	37.7	5.0	KL-S2'	44.5	41.0	57.1	37.6	3.0
	TM-S1	32.8	40.6	55.0	23.7	4.2	TM-S2'	33.0	40.6	51.4	27.4	2.4
	AM-S1	39.4	41.9	46.0	26.9	5.5	AM-S2'	38.8	42.8	53.2	29.4	4.5
	KF-S1	34.3	40.0	55.4	29.8	6.3	KF-S2'	36.3	40.9	56.6	23.5	4.0
	Ave:	36.9	40.7	53.8	28.1	5.1	Ave:	36.3	39.7	53.2	27.1	3.5

Table 12: ***Average Fp2 condition values for the vocal tract filter corrected (*) spectral parameter measurements: amplitude of the first harmonic (H1*); amplitude of the second harmonic (H2*); amplitude of the first formant (A1); amplitude of the third formant (A3*); and waveform noise rating of 600Hz band-pass filtered third formant (Nw).***

Legend: The number after the target word indicates which syllables of the target word bears primary stress. Fa, Fp1 and Fp2 indicate the focal accent condition in which the target word was produced. The average measurement values of the respect acoustic parameters where obtained for the first (S1) and second (S2) syllables of each target word. They are presented adjacent horizontally form each other. An accent mark, ', indicates S1 or S2 as the primary stressed syllable of the target word.

Average H1*For Each Speaker					
		Fa	Fps1	Fps2	Fps3
dada:	DM-S2	32.6	29.2	26.2	24.8
	KL-S2	46.5	45.4	39.1	40.2
	TM-S2	31.6	29.2	28.4	27.4
	AM-S2	45.4	29.3	28.5	27.3
	KF-S2	32.8	34.1	32.7	33.3
	Ave:	**37.8**	**33.4**	**30.9**	**30.6**
dodo:	DM-S2	32.8	30.6	28.0	26.9
	KL-S2	50.8	47.9	43.5	41.1
	TM-S2	32.2	29.4	27.8	28.7
	AM-S2	52.5	38.7	39.5	36.7
	KF-S2	37.3	35.5	32.3	32.6
	Ave:	**41.1**	**36.4**	**34.2**	**33.2**
didi:	DM-S2	32.9	29.1	26.0	26.9
	KL-S2	51.5	46.8	42.9	41.5
	TM-S2	32.9	28.2	27.1	29.5
	AM-S2	55.7	39.5	36.0	40.6
	KF-S2	40.3	34.3	34.0	32.8
	Ave:	**42.7**	**35.6**	**33.2**	**34.3**
Total Ave:		**40.5**	**35.1**	**32.8**	**32.7**

Table 13: ***Effect of focal pitch accent on the fundamental harmonic amplitude (H1) of a full vowel in post-nuclear position. Other than in the Fa condition, the syllable of interest is the non-primary full vowel of the second syllables of the novel target words. Fa is the condition where the target second syllable is focal pitch accented (syllable distance = 0). Fps1 is the condition where the first syllable of the2-syllable word containing the target second syllable is focal pitch accented (syllable distance = 1). Fps2 is when two syllables preceding the target syllable is focal pitch accented (syllable distance = 2). Fps3 is when three syllables preceding the target syllable is focal pitch accented (syllable distance = 3).***

APPENDIX C: **Tables 14 - 16: Listener Syllable Prominence Judgment Responses (Ind. Variation)**

	Average DU Syllable DifferenceResponse of Individual Listeners (Individual Variation)									
Syll. Diff. (DU):	CL	JH	BK	MD	AK	PB	CS	CT	KN	JM
120ms	1.00	1.00	1.00	1.00	1.00	1.25	1.00	1.00	1.00	1.00
105ms	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
90ms	1.00	1.00	1.00	1.00	1.00	1.25	1.00	1.00	1.00	1.00
75ms	1.00	1.25	1.25	1.25	1.00	1.00	1.00	1.00	1.00	1.00
60ms	1.00	1.00	1.00	1.00	1.00	1.50	1.00	1.25	1.00	1.00
45ms	1.00	1.00	1.25	1.25	1.00	1.50	1.25	1.00	1.00	1.00
30ms	1.00	1.25	1.00	1.25	1.25	1.50	1.00	1.00	1.00	1.00
20ms	1.00	1.25	1.25	2.00	1.00	1.25	1.50	1.25	1.25	1.25
0	1.00	2.00	1.00	2.00	1.25	1.50	1.00	1.75	1.00	1.00
-20ms	1.25	1.50	1.75	2.00	1.75	1.25	1.50	1.50	1.50	1.25
-30ms	1.75	1.50	1.75	2.00	1.50	1.75	1.50	1.75	2.00	1.50
-45ms	2.00	2.00	1.75	2.00	1.75	1.50	1.25	1.50	1.75	1.75
-60ms	1.75	2.00	2.00	2.00	1.50	1.25	1.25	1.50	2.00	1.75
-75ms	2.00	2.00	2.00	2.00	2.00	1.50	1.25	1.50	1.75	1.75
-90ms	2.00	2.00	1.75	2.00	2.00	1.25	1.75	1.50	2.00	1.75
-75ms	2.00	2.00	2.00	2.00	1.75	1.25	1.75	1.50	2.00	2.00
-120ms	2.00	2.00	2.00	2.00	2.00	2.00	1.75	1.50	2.00	2.00

	Average TL Syllable Diff Response of Individual Listeners (Individual Variation)									
Syll. Diff. (TL):	CL	JH	BK	MD	AK	PB	CS	CT	KN	JM
16dB	1.75	1.50	2.00	2.00	1.75	1.75	1.75	1.50	1.50	2.00
14dB	1.00	1.50	2.00	2.00	1.50	1.50	1.75	1.75	2.00	2.00
12dB	1.75	1.50	2.00	2.00	1.25	1.50	1.75	1.25	2.00	2.00
10dB	1.75	1.75	2.00	2.00	1.75	1.75	1.75	1.25	1.75	2.00
8dB	1.25	1.00	1.25	2.00	1.75	1.75	1.25	1.25	1.75	2.00
6dB	1.50	1.50	2.00	2.00	1.25	1.75	1.75	1.50	2.00	1.75
4dB	1.50	1.25	1.75	2.00	1.25	1.25	1.25	1.25	1.50	2.00
2dB	1.50	1.50	2.00	2.00	1.50	1.50	1.75	1.25	1.75	1.75
0	1.25	1.25	1.25	2.00	1.00	1.50	1.50	1.00	2.00	1.50
-2dB	1.25	1.25	1.50	2.00	1.25	1.25	1.50	1.25	1.75	1.25
-4dB	1.25	1.25	1.00	2.00	1.25	1.00	1.50	1.50	1.50	1.25
-6dB	1.25	1.25	1.50	1.67	1.50	1.00	1.50	1.50	1.75	1.25
-8dB	1.25	1.50	1.25	2.00	1.25	1.00	1.50	1.00	1.75	1.25
-10dB	1.00	1.25	1.25	1.67	1.50	1.25	1.75	1.00	1.50	1.00
-12dB	1.00	1.50	1.25	1.67	1.00	1.25	1.50	1.50	1.75	1.00
-14dB	1.25	1.00	1.25	1.67	1.50	1.25	1.50	1.50	1.50	1.00
-16dB	1.25	1.25	1.25	1.00	1.00	1.25	1.50	1.00	1.50	1.00

	Average AH Syllable Difference Response of Individual Listeners (Individual Variation)									
Syll. Diff. (AH):	CL	JH	BK	MD	AK	PB	CS	CT	KN	JM
16dB	1.00	1.25	1.00	1.67	2.00	1.75	1.50	1.25	1.25	1.25
14dB	1.50	1.50	2.00	2.00	1.25	1.50	1.75	1.25	1.00	1.25
12dB	1.00	1.75	1.00	1.67	1.75	1.25	1.50	1.25	1.75	1.50
10dB	1.75	1.75	1.75	2.00	1.25	1.50	1.50	1.25	1.50	1.25
8dB	1.25	1.75	1.50	1.67	1.75	1.75	1.75	1.50	2.00	1.25
6dB	1.50	1.75	1.75	1.67	1.25	1.75	1.25	1.25	1.50	1.50
4dB	1.50	1.50	1.75	2.00	1.25	1.50	1.50	1.25	1.75	1.50
2dB	1.50	1.50	1.75	2.00	1.25	1.50	1.25	1.25	1.50	1.50
0	1.50	1.75	1.50	2.00	1.50	1.50	2.00	1.25	1.50	1.25
-2dB	1.25	1.75	2.00	2.00	1.25	1.50	2.00	1.25	1.50	1.25
-4dB	1.50	1.50	2.00	2.00	1.50	1.00	1.25	1.25	1.50	1.50
-6dB	1.25	1.50	1.50	2.00	1.25	1.25	1.75	1.25	1.50	1.50
-8dB	1.75	1.50	1.50	2.00	1.50	1.00	1.50	1.25	2.00	1.25
-10dB	1.50	1.50	2.00	2.00	1.50	1.50	1.50	1.25	1.75	1.50
-12dB	1.25	1.25	1.75	2.00	1.25	1.50	1.50	1.25	1.50	1.75
-14dB	1.25	1.25	1.25	2.00	1.50	1.25	1.50	1.25	1.75	1.50
-16dB	1.25	1.75	1.75	2.00	1.50	1.75	1.75	1.50	1.50	1.25

APPENDIX D: **Tables 17: Listener Syllable Prominence Judgment Responses (Co-variation)**

	Average DU Syllable DifferenceResponse of Individual Listeners (Co-variation)									
Syll. Diff. (DU):	CL	JT	KL	MD	SL	PB	CS	CT	KN	AM
120ms	1.00	1.24	1.02	1.00	1.00	1.16	1.04	1.04	1.00	1.00
75ms	1.00	1.14	1.04	1.04	1.00	1.14	1.02	1.04	1.02	1.00
30ms	1.18	1.35	1.06	1.98	1.00	1.37	1.18	1.16	1.10	1.08
0	1.27	1.61	1.18	2.00	1.04	1.53	1.33	1.22	1.29	1.39
-30ms	1.55	1.86	1.53	2.00	1.67	1.45	1.59	1.27	1.71	1.45
-75ms	1.86	1.98	1.94	2.00	2.00	1.55	1.92	1.39	1.96	1.94
-120ms	1.98	2.00	2.00	1.98	2.00	1.59	1.98	1.49	2.00	2.00

	Average TL Syllable Diff Response of Individual Listeners (Co-variation)									
Syll. Diff. (TL):	CL	JT	KL	MD	SL	PB	CS	CT	KN	AM
16dB	1.47	1.69	1.57	1.71	1.41	1.67	1.51	1.16	1.47	1.43
8dB	1.41	1.63	1.45	1.73	1.41	1.41	1.51	1.29	1.49	1.41
2dB	1.41	1.69	1.37	1.71	1.39	1.39	1.31	1.08	1.43	1.49
0	1.45	1.57	1.43	1.71	1.41	1.43	1.41	1.29	1.47	1.41
-2dB	1.41	1.63	1.33	1.67	1.39	1.41	1.43	1.29	1.47	1.39
-8dB	1.39	1.49	1.31	1.73	1.41	1.20	1.47	1.22	1.39	1.41
-16dB	1.31	1.47	1.33	1.71	1.31	1.29	1.43	1.29	1.37	1.33

	Average AH Syllable Difference Response of Individual Listeners (Co-variation)									
Syll. Diff. (AH):	CL	JT	KL	MD	SL	PB	CS	CT	KN	AM
16dB	1.37	1.61	1.33	1.69	1.41	1.29	1.45	1.22	1.37	1.37
8dB	1.43	1.55	1.37	1.71	1.41	1.43	1.47	1.18	1.51	1.41
2dB	1.43	1.53	1.43	1.69	1.39	1.51	1.45	1.29	1.43	1.41
0	1.39	1.63	1.39	1.71	1.39	1.39	1.37	1.18	1.47	1.43
-2dB	1.37	1.57	1.39	1.71	1.35	1.29	1.43	1.31	1.41	1.35
-8dB	1.41	1.57	1.41	1.73	1.35	1.41	1.47	1.18	1.49	1.45
-16dB	1.45	1.71	1.47	1.73	1.43	1.49	1.43	1.24	1.41	1.45

APPENDIX E: Table 18: ANOVA Results for Syllable Prominence Judgment Task

Individual Word Stress Parameter Variation Syllable Prominence Judgment

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Listener (DU)	1.552	9	0.172	3.90	0.000
DU	18.699	16	1.169	26.43	0.000
Listener (TL)	5.196	9	0.577	12.34	0.000
TL	5.113	16	0.320	6.83	0.000
Listener (AH)	4.817	9	0.535	11.45	0.000
AH	0.611	16	0.038	0.82	0.664

Co-Varied Word Stress Parameter Syllable Prominence Judgment

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Listener	52.316	9	5.813	60.06	0.000
Listener*DU	75.698	54	1.402	14.48	0.000
Listener* TL	11.755	54	0.218	2.25	0.000
Listener*AH	5.351	54	0.099	1.02	0.428
Listener*DU*TL	39.988	324	0.123	1.28	0.002
Listener*DU*AH	27.249	324	0.084	0.87	0.946
Listener*TL*AH	29.192	324	0.090	0.93	0.793

APPENDIX F: **Table 19: ANOVA Results for Naturalness Rating Task**

Individual Word Stress Parameter Naturalness Rating

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Listener (DU)	25.930	6	4.322	87.03	0.000
DU	2.035	16	0.127	2.56	0.002
Listener (TL)	29.293	6	4.882	85.33	0.000
TL	2.129	16	0.133	2.33	0.006
Listener (AH)	41.667	6	6.945	88.54	0.000
AH	1.927	16	0.120	1.54	0.103

Co-Varied Word Stress Parameter Naturalness Rating

Source	Sum Sq.	d.f.	Mean Sq.	F	Prob>F
Listener	336.062	4	84.016	332.62	0.000
Listener*DU	82.975	24	3.457	13.69	0.000
Listener*TL	46.061	24	1.919	7.60	0.000
Listener*AH	24.902	24	1.038	4.11	0.000
Listener*DU*TL	44.903	144	0.312	1.23	0.042
Listener*DU*AH	28.748	144	0.200	0.79	0.961
Listener*TL*AH	51.719	144	0.359	1.42	0.002

Bibliography

Beckman, M. E. and Edwards, J. (1994). Articulatory evidence for differentiating stress categories. In P. Keating (ed.) *Phonological structure and phonetic form. Papers in Laboratory Phonology III.* Cambridge: CUP.

Beckman, M.E. and Pierrehumbert, J.B. (1986). *Intonational structure in Japanese and English.* Phonology Yearbook 3, 255-309.

Bickley, C. (1982). *Acoustic analysis and perception of breathy vowels.* MIT Speech Comm. Working Papers, Cambridge, MA, 77-82.

Boersma, P. and Weenink, D. (2005): *Praat: doing phonetics by computer (Version 4.3.14)* [Computer program].

Bolinger, D.L. (1958). *A theory of pitch accent in English.* Word, 14, 109-149.

Campbell, N. and Beckman, M.E. (1997). *Stress, prominence, and spectral tilt.* In Intonation: theory, models and applications (Edited by A. Botinis, G. Kouroupetroglou, and G. Carayiannis). ESCA, Athens, 67-70.

Cappa, S.F., Marina, N., Ielasi, W. and Miozzo, A. (1997). *The representation of stress: Evidence from an aphasic patient.* Cognition, 65, 1-13.

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

Fletcher, H. and Munson, W.A. (1937). *Relationship between loudness and masking.* J. Acoust. Soc. Am. 9, 1-10.

Fry, D.B. (1955). *Duration and intensity as physical correlates of linguistic stress.* J. Acoust. Soc. Am. 27, 765-768.

Fry, D.B. (1958). *Experiments in the perception of stress.* Language and Speech, 1, 126-152.

Fujisaki, H., Hirose, K., and Sugito, M. (1986). *Comparison of acoustic features of word accent in English and Japanese.* J. Acoust. Soc. Jpn. 7, 57-63.

Hanson, H.M. (1995). *Glottal Characteristics of Female Speakers.* Doctorate Dissertation, Harvard University.

Hanson, H.M. (1997). *Glottal characteristics of female speakers: Acoustic correlates.* J. Acoust. Soc. Am., 101, 466-481.

Hanson, H.M. and Chuang, E.S. (1999). *Glottal characteristics of male speakers: Acoustic correlates and comparison with female data.* J. Acoust. Soc. Am., 106, 1064-1077.

Harrington, J., Beckman, M., Fletcher, J., and Palethorpe, S. (1998). "An electropalatographic, kinematic, and acoustic analysis of supralaryngeal correlates of word and utterance-level prominence contrasts in English," in *Proceedings of the 5th International Conference on Spoken Language Processing*, Sydney, Australia, 1851–1854.

Hillis, A.E. and Caramazza, A. (1991). *Mechanisms for accessing lexical representations for output: Evidence from a category-specific semantic deficit.* Brain and Language, 40, 106-144.

Holmberg, E.B., Hillman, R.E., Perkell, J.S., Guiod, P.C., and Goldman, S.L. (1995). *Comparisons among aerodynamic, electroglottographic, and acoustic spectral measures of female voice*. J. Speech and Hearing Research, 38, 1212-1223.

Huffman, M.K. (1987). *Measures of phonation types in Hmong.* J. Acoust. Soc. Am., 81, 495-504.

Huss, V. (1978). *English word stress in post-nuclear position.* Phonetica, 35, 86-105.

Iseli, M. and Alwan, A. (2004). *An Improved Correction Formula for the Estimation of Harmonic Magnitudes and Its Application to Open Quotient Estimation.* Proc. ICASSP, May (2004), Montreal, Canada.

Kager, R. (1995), "The metrical theory of word stress," in J. Goldsmith (ed.), *The Handbook of Phonological Theory*, Oxford, England, 367-402.

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

Klatt, D.H. and Klatt, L.C. (1990). *Analysis, Synthesis and Perception of Voice Quality Variations among Female and Male Talkers.* J. Acoust. Soc. Am., 87, 820-857.

Kochanski, G., Grabe, E., Coleman, J., and Rosner, B. (2005). *Loudness predicts prominence: Fundamental frequency lends little.* J. Acoust. Soc. Am., 118, 1038-1054.

Ladefoged, P. (1983). *The linguistic use of different phonation types.* Vocal fold physiology: Contemporary research and clinical issues (Ed. by D.M. Bless and J.H. Abbs). College Hill, San Diego, 351-360.

Levelt, W.J.M., Roelofs, A., and Meyer, A.S. (1999). *A theory of lexical acess in speech production.* Behavioral and Brain Sci., 22, 1-75.

Levelt, W.J.M. (2001). *Spoken word production: A theory of lexical access.* Proceedings of the National Academy of Sciences, 98 (23), 13464-13471.

Lieberman, P. (1960). *Some acoustic correlates of word stress in American English.* J. Acoust. Soc. Am., 32, 451-454.

Mannell, R.H. (1998). *Formant diphone parameter extraction utilizing a labeled single-speaker database*. Proc. ICSLP, November 30, Sydney, Australia.

Mattys, S. and Samuel, A.G. (2000). *Implications of stress-pattern differences in spoken-word recognition.* J. Memory and Lang., 42 (4) 571-596.

Miceli, G. and Caramazza A. (1993). *The assignment of word stress in oral reading: Evidence from a case of acquired dyslexia.* Cog. Neuropsych., 10 (3), 273-296.

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

Okobi, A.O. and Hirose, K. (2003). *Acoustic Analysis of English Lexical-Prosody Reproduction by Japanese Speakers.* IEICE Tech. Report, 103, 37-42.

Shattuck-Hufnagel, S. (1994). "Stress shift" as early placement of pitch accents: a comment on Beckman and Edwards. In P. Keating (ed.) *Phonological structure and phonetic form. Papers in Laboratory Phonology III.* Cambridge: CUP.

Shattuck-Hufnagel, S. and Turk, A.E. (1996). *A prosody tutorial for investigators of auditory sentence processing.* J. Psycholingusitic Research, 25, 193-247.

Shih, C. and Kochanski, G. (2002). *Prosody and Prosodic Models.* ICSLP Meeting, September 16, Denver, CO.

Sluijter, A.M.C., Shattuck-Hufnagel, S., Stevens, K.N., and van Heuven, V.J. (1995). *Supralaryngeal resonance and glottal pulse shape as correlates of stress and accent in English.* ICPhS Meeting, 2, Stockholm, Sweden.

Sluijter, A.M.C. and van Heuven, V.J. (1996a). *Acoustic correlates of linguistic stress and accent in Dutch and American English.* ICSLP Meeting, October 3-6, Philadelphia, PA.

Sluijter, A.M.C. and van Heuven, V.J. (1996b). *Spectral balance as an acoustic correlate of linguistic stress.* J. Acoust. Soc. Am., 100, 2471-2485.

Sluijter, A.M.C., van Heuven, V.J., Pacilly, J.A. (1997). *Spectral balance as a cue in the perception of linguistic stress.* J. Acoust. Soc. Am., 101, 503-513.

Stevens, K.N. (1994). *Prosodic influences on glottal waveform: Preliminary Data.* Internat. Symp. Prosody, September 18, Yokohama, Japan.

Stevens, K.N. (1998). *Acoustic phonetics.* Cambridge, MA: MIT Press.

Terken, J. (1991). *Fundamental frequency and perceived prominence of accented syllables.* J. Acoust. Soc. Am., 89, 1768-1776.

Turk, A.E. and Sawusch, J.R. (1996). *The processing of duration and intensity cues to prominence.* J. Acoust. Soc. Am., 99, 3782-3790.

Turk, A.E. and White (1999). *Structural influences on accentual lengthening in English.* J. Phonetics, 27, 171-206.

Wang, C. and Seneff, S. (2001). *Lexical Stress Modeling for Improved Speech Recognition of Spontaneous Telephone Speech in the Jupiter Domain.* Eurospeech Meeting, September 2-7, Aarlborg, Denmark.

Wang, Y.T., Kent, R.D., Duffy, J.R., and Thomas, J.E. (2005). *Dysarthria associated with traumatic brain injury: speaking rate and emphatic stress.* J. Comm. Disorders, 38, 231-260.

Wingfield, A., Lindfield, K.C., and Goodglass, H. (2000). *Effects of Age and Hearing Sensitivity on the Use of Prosodic Information in Spoken Word Recognition.* J. Speech, Language, and Hearing Research, 43, 915-925.