

ACOUSTIC REALIZATION AND PERCEPTION OF ENGLISH LEXICAL
STRESS BY MANDARIN LEARNERS

BY

Yuwen Lai

M.A. University of Kansas, 2004
M.S. National Taiwan University, 1998
B.S. National Taiwan University, 1996

Submitted to the graduate degree program in linguistics and the Graduate Faculty of
the University of Kansas in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

Chairperson

Committee members

Date defended:

The Dissertation Committee for Yuwen Lai certifies that this is the approved version
of the following dissertation:

ACOUSTIC REALIZATION AND PERCEPTION OF ENGLISH LEXICAL
STRESS BY MANDARIN LEARNERS

Chairperson

Committee members:

Date approved: _____

Abstract

The acquisition of English lexical stress by Mandarin L2 learners was examined through production and perception studies. An acoustic study focusing on the implementation of mean F0, max F0, duration, intensity, and F2 in stressed and unstressed vowels in noun-verb word pairs contrasting in stress location (e.g. *object* and *object*) was conducted. The results from native English speakers (n=10) showed that all correlates were utilized to signal stress in nouns. In verbs, however, mean and max F0 were not utilized and duration cues were amplified. Implementation patterns for Mandarin L2 learners (beginning=9; advanced=9) were similar to native speakers in nouns. However, in verbs learners used mean and max F0 as well. Reduction of unstressed vowels was found to be inconsistent in learners when compared to native speakers. A perceptual study utilizing the disyllabic nonword ‘dada’, with resynthesized max F0, duration, and vowel quality, was conducted in order to evaluate the perceptual relevance of those cues in stress perception. Results from an identification task indicate that full vowels induce significantly stronger stress perception in all listener groups. In terms of max F0 and duration, beginning listeners (n=25) relied mainly on duration, advanced listeners (n=25) focused more on max F0, while native listeners (n=25) made use of both duration and max F0 in perception. These findings are discussed in terms of the similarities and differences in prosodic systems between Mandarin and English, as well as the possible discrepancies in production and perception data from second language learning research.

Acknowledgments

This dissertation would not have been possible without my advisor, Professor Allard Jongman, and his invaluable assistance, guidance and most importantly, his long-term patience with my progress. I would also like to thank Professor Joan Sereno, for her unconditional support. Her steady encouragement inspired me to keep moving forward. I have benefited tremendously from being her teaching assistant, and hope to someday become half as good a mentor as she has been to me. Professor Jie Zhang has been instrumental in broadening my research interests and through many and varied projects has provided me with a wide range of experience. These projects took me to many new places to gather data and attend conferences, of which I am very appreciative.

I am grateful to Professor Arianne Dwyer, for providing a precious opportunity to work with the DoBeS (Dokumentation Bedrohter Sprachen) Project. The training I received for the project at the Max Plank Institute for Psycholinguistics, and my fieldwork in Qinghai, China, helped me to establish my interest in archiving endangered languages.

I would also like to thank Professor Alison Gabriele for her helpful comments regarding second language acquisition and its relation to my project, Professor Sanae Eda for all of her efforts on early drafts of the paper, and Professor Mike Vitevitch for agreeing to help me out at such short notice. I would also like to thank Professor Christen Jensen for all of her help in the planning and structuring of my dissertation, my good friend Maria del Carmen Parafita Couto for her support and encouragement, and L'Angelo Misterioso for his tireless editing and proofreading.

Table of Content

Abstract	iii
Acknowledgments	iv
Table of Content	v
CHAPTER 1: INTRODUCTION	1
1.1 Literature review	3
1.1.1 Second language learning	3
1.1.2 Prosodic features of the target languages in the present study	6
American English.....	6
Mandarin Chinese	6
1.1.3 Comparison between American English and Mandarin Chinese.	9
1.1.4 Acquisition of English lexical stress by non-native speakers	9
1.2 The present study	13
1.2.1 Acoustic study 1: acoustic realization of English stress in word pairs ..	14
1.2.2 Acoustic study 2: F0 contour realization in English stress	15
1.2.3 Perception study: Roles of max F0, duration, and vowel quality in stress perception.....	16
CHAPTER 2: ACOUSTIC STUDIES	17
2.1 Rating experiment.....	17
2.1.1 Methodology	18
Participants.....	18
Stimuli.....	18
Procedure	19
2.1.2 Results.....	19
2.2 Acoustic study I – lexical stress produced in <i>contract-contract</i> word pairs	22
2.2.1 Methodology	22
Participants.....	22
Stimuli.....	23
Procedure	23

Acoustic analyses.....	24
2.2.2 Results.....	27
Mean F0	27
Max F0	31
Intensity.....	34
Duration	37
Formant analysis	40
2.2.3 Summary	43
Mean F0, max F0, intensity, and duration	43
Formant analysis	46
2.3 Acoustic study II - F0 contours in stressed syllables.....	47
2.3.1 Methodology	47
2.3.2 Results.....	48
Stressed syllables	48
Unstressed syllables	50
2.3.3 Summary	54
2.4 General discussion	56
2.4.1 Overall summary.....	56
Experiment 1	56
Experiment 2.....	56
2.4.2 Comparison among speaker groups	57
Extrinsic factors at play	57
2.4.3 Comparison with previous studies.....	64
CHAPTER 3: PERCEPTION STUDY.....	67
3.1. Methodology	68
3.1.1 Stimulus construction.....	68
Vowel quality	68
Max F0 and duration resynthesis	71
3.1.2 Participants.....	75

3.1.3 Procedure	75
3.2 Results.....	76
3.2.1 Overall analysis.....	76
3.2.2 Max F0, duration, and proficiency.....	78
Main effects	78
Interactions.....	80
3.2.3. Cue weighting in each listener group.....	83
Native speakers	84
Advanced learners.....	87
Beginning learners	90
3.3 Discussion.....	93
3.3.1 Summary	93
Vowel quality.....	94
Duration and max F0 weighting in stress perception.....	95
3.3.3 Conclusion	98
CHAPTER 4: GENERAL DISCUSSION	99
4.1 Summaries.....	99
4.1.1 Experiment 1: Acoustic study on minimal word pairs.....	99
4.1.2 Experiment 2: F0 contours in English stressed syllables.....	102
4.1.3 Experiment 3: Perception study	103
4.2 Bridging stress production and perception	105
4.2.1 Native production and perception	105
Max F0 in production and perception.....	105
Duration in production and perception	107
Congruence between acoustic correlates and perceptual cues.....	109
4.2.2 Mandarin L2 learners of English	110
Beginning learners	110
Advanced learners.....	111
4.2.3 Prosodic transfer from Mandarin L1.....	113

4.2.4 The present study and current theoretical framework.....	114
4.2.5 An alternative view	115
4.3. Conclusion	117
References.....	119
Appendices.....	123
Appendix A - Target words in context sentences.	123
Appendix B - Duration resynthesis.....	124
Appendix C - Questionnaires.....	127
1: Questionnaire for native English speakers.....	127
2: <i>Questionnaire for Mandarin L2 learners of English (in Mandarin)</i>	128
3: <i>Questionnaire for Mandarin L2 learners of English (English translation)</i> .	129
Appendix D - Sample mean test	131
Appendix E - Results overall analysis (four-way ANOVA)	132
Appendix F - F0 contour from each individual speaker	135
Appendix G - Results of perception study from individual participant.....	136
Appendix H - Perception results in all cue combinations across three proficiencies	138

CHAPTER 1: INTRODUCTION

The cognitive assembly required of a person for the production and perception of a human language is highly sophisticated. Research in second language acquisition provides a unique window into this complex process, in allowing an examination of the ways in which the human brain adapts to and copes with the added complications of a second language, and the difficulties associated with the various fundamental differences which are likely to exist between the first and second languages. In so doing, such research adds to our understanding of the ways in which the human mind deals with specific languages in specific contexts, and provides insight into the nature of language itself.

First language transfer occurs when speakers or writers carry over the knowledge from their first language (L1) to a second language (L2) (Ellis, 1994, Odlin, 1989). Most research on L2 acquisition focuses on the acquisition at the segmental level, i.e. production and perception of individual consonants or vowels (Flege, 1995).

In the past three decades, suprasegmental features (prosody) have begun to receive more attention due to the integral role they play in foreign accentedness and their impact on comprehensibility of L2 speech (Anderson-Hsieh et al., 1992, Munro and Derwing, 1998, Trofimovich and Baker, 2006). Previous studies of L2 prosody acquisition have mostly focused on its impact on communication; little research has focused on examining the interference of L1 prosodic characteristics on second language learning (McGory, 1998, Ueyama, 2001).

Using 'prosodic pattern' as the categorization criterion, the world's languages can be grouped into three major categories - stress, tone, and pitch accent languages (Archibald, 1997). Stress languages adopt both segmental composition and stress (realized through pitch, amplitude, duration, and spectral composition) to signal different lexical meanings. Tone languages utilize pitch heights and contours lexically,

while pitch accent languages use pitch accent patterns. Notice that this distinction is largely based on the pitch domain of languages.

Previous research regarding prosodic learning in L2 has mostly focused on the interlanguage phonology of stress placement. Three primary categories of results have been reported in these types of studies - L2 learners either apply their L1 parameters, L2 parameters, or neither (Altmann, 2006, Archibald, 1998).

The current study focuses on the learning process of people whose native language is Mandarin (a tonal language), and who are learning English (a stress language). Previous research in this arena has primarily focused on speech production, and has revealed a general positive transfer effect of L1 prosody in L2 learning (see section 1.1.2 for a review). The current study adopts both production and perception approaches to examine this issue, and considers the role of language experience in the learning process.

The general outline of this dissertation is as follows: A literature review on the background information regarding second language acquisition at suprasegmental levels is provided in Chapter 1; Chapter 2 introduces two acoustic studies, which compare lexical stress production of beginning and advanced second language learners with that of native speakers, in order to explore the influence of a tonal first language background on the stress features of a second language. A rating experiment was conducted in order to assign the proficiency levels of the learners, and is also included in this chapter. A perceptual identification study, with resynthesized max F0, duration, and spectral composition stimuli, is described in Chapter 3. Chapter 4 contains a general discussion of findings from the production and perception studies and relates the findings from the current study to existing perceptual models for L2 learning.

1.1 Literature review

In the study of second language learning, it is recognized that learners start with their first language (L1) system and establish a correspondence between L1 and second language (L2) phonological systems (Ioup, 1984).

1.1.1 Second language learning

In the study of second language learning, it is recognized that at the phonological level second language learners start with the system of their first language and establish a correspondence between it and the new phonological system (Ioup, 1984). The acquisition process is very likely to be influenced by the first language, which is generally referred to as language transfer, and which can be either positive or negative. Positive transfer takes place when the first language facilitates second language learning, while negative transfer interferes with L2 learning.

First language transfer has been demonstrated in numerous studies, but has been primarily focused on segmental acquisition (for a detailed review, see Flege, 1987). Several models have been proposed to account for the transfer phenomenon (Bohn and Flege, 1992, Bohn and Flege, 1993).

The Speech Learning Model proposed by Flege and colleagues (1987, 1988, 1992, 1995) posits that acquiring a new phonetic category in L2 depends highly on the perceptual distance of the new sound from the sound inventory of L1. They argue that L1 and L2 categories exist in the same phonetic system, and that three possible relationships between L2 and L1 sounds occur in the shared system. First, when the L2 sound is distinctly different from the counterpart in L1 (this is referred to as the 'new' condition), learners have no difficulties in adding a new sound in their phonetic system. Second, an L2 sound could be identical to an existing sound in L1, and is therefore 'old' to the phonetic system. The third possible relationship, and the one

which causes the greatest difficulty for learners, is when an L2 sound is perceptually similar to an L1 sound, but not identical to it. This is referred as the ‘similar’ condition. In the similar condition, learners are most likely to identify the L2 sound as its counterpart in their L1. In production, they are also more likely to use the L1 sound to replace the L2 sound.

The Perceptual Assimilation Model proposed by Best and colleagues (1994, 2007) posits that perception of an L2 sound is highly affected by the articulatory-gesture similarity between the new phoneme and native phonemes. That is, when an L2 sound is similar to an L1 sound in its articulatory properties, learners will mostly assimilate the L2 sound to its counterpart in L1 sound inventory. Four types of similarities are proposed by this model. ‘Two categories’ refers to a pair of L2 sounds that assimilate two L1 categories respectively. This is posited to be least problematic for the L2 learners since each new sound has its own counterpart to assimilate to in the learners’ L1. ‘Category goodness’ refers to a pair of L2 sounds assimilating to one L1 category, but one is better assimilated to the L1 sound than the other. ‘Nonassimilable type’ refers to the situation in which an L2 sound is too distinct from L1 categories to be assimilated to any category in L1. Finally, ‘single category’ refers to when a pair of L2 sounds assimilate to one L1 category. It is this situation which causes the greatest difficulty for the learners.

The Native Language Magnet model proposed by Kuhl and colleagues (1989, 1991, 1992, 1994, 1998) was established to account for the development of the perception system during the early period of first language acquisition. This model proposes that L1 speech perception goes from a language-universal to a language-specific developmental sequence. To be more specific, L1 phonetic categories function like prototypes, which become perceptual magnets and attract nearby members of the same phonetic category. This model was further applied to account for perceptual

learning in second languages- an L2 sound is posited to be pulled to the magnet of its counterpart in the learners' L1 which makes it difficult for the learners to perceive.

While the Speech Learning Model, Perceptual Assimilation Model, and Native Language Magnet Model are the three leading models of L2 sound acquisition, they all focus on the acquisition of segmental distinctions. It is not clear if and to what extent, any of these models can account for the acquisition of suprasegmental distinctions.

It has been shown that stress involves a complex interaction of at least four basic physical properties of the acoustic signal: fundamental frequency (F0), duration, amplitude, and vowel quality. Typically, stress refers to the prominence given to a syllable in a word (lexical stress), or to a word in a sentence (sentential stress). A stressed syllable/word generally has higher F0, longer duration, greater amplitude and non-reduced vowels (Lehiste, 1970). However, the relative importance of these correlates is tuned to the phonological structures of different languages. For example, duration is frequently employed to signal stress, such that stressed syllables tend to be longer than unstressed syllables. However, in some languages, unstressed syllables are longer than stressed syllables (i.e. Estonian, Lehiste (1970)). The acoustic correlates mentioned above give rise to the perceptual parameters of pitch, duration loudness, and vowel quality, respectively. Nevertheless, there is no one-to-one correspondence between perceived stress and any single correlate. As a result, listeners perceive stress as a complex configuration of these correlates (Lieberman, 1960).

The use of different acoustic cues to signal lexical stress becomes especially interesting in second language acquisition when a speaker adopts acoustic correlates/perceptual cues to signal stress/pitch accent/tone that are different from those in the first language. In these cases one might expect prosodic transfer in both

the production and perception of suprasegmental features when acquiring a second language.

1.1.2 Prosodic features of the target languages in the present study

American English

Word stress refers to the prominence of syllables within a word. As a stress language, English has primary stress on one syllable in each content word. This prosodic information, however, is mostly not contrastive except in noun-verb word pairs such as 'object - ob'ject, which have near-identical segmental composition and only contrast in stress location, which affects the lexical category of the word. Word pairs of this nature have been utilized in earlier studies on the acoustic correlates of English word stress (Beckman, 1986, Fry, 1955, Lieberman, 1960). The results from these studies suggest that higher F0, longer duration, greater amplitude, and full vowel quality are associated with the syllable receiving primary stress. Perception results from Fry (1958) further pointed out that stimuli with linear fall or late fall F0 patterns are more likely to be perceived as stressed syllables.

Mandarin Chinese

There are four tones in Mandarin: Tone 1 – high level, tone 2 - mid rising, tone 3 – low dipping, and tone 4 – high falling. The F0 contours of the four tones, produced in isolation, are provided in Figure 1.1.

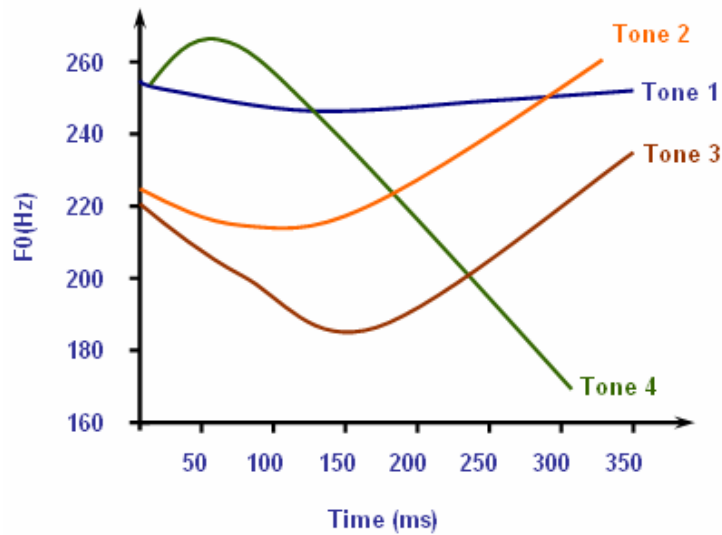


Figure 1. 1 Mandarin tones produced by a female native speaker. Tone 1 - high level, tone 2 - mid rising, tone 3 - low dipping, and tone 4 - high falling.

Research by Gandour (1983) has shown that five aspects are relevant for tone perception: 1.) average F0/F0 height; 2.) F0 contour; 3.) F0 slope; 4.) extreme endpoints; and 5.) tone duration. Previous research has shown that the primary acoustic parameters of Mandarin tones are F0 height and contour shape (Howie, 1976). Duration also differs among the four tones: Tone 2 and Tone 3 are the longest, while Tone 4 is the shortest (Nordenhake and Svantesson, 1983). Moore and Jongman (1997) have also shown that Tone 2 has an earlier turning point and a smaller F0 change between the onset and turning point than Tone 3.

Lin (1988) studied the production and perception of Mandarin tones in 20 native speakers. He found five patterns among the intensity contours: level, higher at onset, higher at offset, higher in the middle, and a double-peak amplitude contour. Duration also varied among tones, with production data showing tones 2 and 3 being the longest, while tone 4 was generally the shortest (see Jongman et al. (2006) for a detailed overview).

Studdert-Kennedy and Shankweiler (1970) proposed the notion that people can differentially tune their auditory systems to certain physical properties of a sound as a function of their linguistic experience. Stragray and Downs (1993) conducted a perception study to investigate whether differential sensitivity for frequency (125 Hz range and 1000 Hz range) differed between tone language speakers (Mandarin) and non-tone language speakers (English). In their study, subjects listened to two level tones for which the F0 difference was systematically manipulated, and performed a discrimination task on the presented tones. The results indicate that the limens for frequency were significantly larger for Mandarin speakers than for English speakers. They concluded that Mandarin speakers might have poorer differential sensitivity because they had learned to categorize sounds of similar frequency together to enhance their perception of tones.

1.3.2 Mandarin stress

A production study by Xu (1999) investigated how lexical tone and sentential stress contribute to the formation and alignment of F0 contours in Mandarin. The results from F0 contour and syllable alignment indicate that the intrinsic tone of a given syllable defines the local F0 contour. However, sentential stresses shape the global profile of the F0 contour of the entire sentence, which in turn affects F0 range and shape of local F0 contours.

The perception of Mandarin stress was investigated by Shen (1993). The results indicate that duration is the dominant cue for Mandarin speakers when perceiving sentential stress. Shen inferred that this phenomenon is the result of the “functional load” hypothesis, which claims that phonetic resources are available for indicating stress if and only if they are not used up in signaling lexical contrasts.

1.1.3 Comparison between American English and Mandarin Chinese.

Although American English and Mandarin both have lexically specified accents, they differ in terms of how they manipulate the acoustic correlates of stress (in English) and tone (in Chinese). American English utilizes all four acoustic correlates to cue word stress: F0, duration, amplitude, as well as vowel formant composition, while Mandarin tones are distinguished mainly by F0 height and contour.

Mandarin is also a more syllable-timed language and is generally thought of as showing no strong patterns of stress, with syllable duration remaining relatively constant across a sentence (van Santen and Shih, 2000). English, on the other hand, is categorized as a stress-timed language, in which the duration of inter-stress intervals is more consistent (Ramus et al., 1999). Given the rhythmic differences between these two languages, acquisition of the lexical stress correlates in English by Mandarin speakers may involve learning to reduce or shorten unstressed syllables.

Differences between these two languages in the perception of stress have also been found in previous research. Perception studies of English stress (Beckman, 1986, Fry, 1958, Lea, 1977) all show that listeners rely on differences in syllable duration, amplitude of the syllable, and the pitch of the syllable as cues for stress. On the other hand, Mandarin speakers rely on F0 height and contour tonal perception (for a detailed review, cf. Jongman et al. 2006).

1.1.4 Acquisition of English lexical stress by non-native speakers

French L1 to English L2

The perception of English lexical stress by native speakers and second language learners with French as their first language was investigated by (Fry, 1955, Fry, 1972) using synthesized stimuli. The stimuli consisted of word pairs contrasting in stress assignment – *subject* (noun) and *subject* (verb). The duration and intensity of the first syllable were resynthesized. The participants' task was to identify if they had heard a

noun or a verb. The results from the native English speakers indicated that their noun judgment changed from 19% to 90% as duration in the first syllable became greater. Their noun judgment changed from 46% to 75% when intensity in the first syllable became greater. French speakers less proficient in learning English as a second language were found to be equally sensitive to duration (32% to 75%) and intensity (33% to 77%), while more proficient learners patterned more like native English speakers, utilizing duration cues (13% to 90%) more than intensity cues (48% to 78%). Fry concluded that duration plays less of a role in French speakers' perception of English stress when they are less proficient in English. As proficiency advances, listeners acquire a native-like perception pattern.

Japanese L1 to English L2

The prosodic transfer between English and Japanese was examined thoroughly in a bidirectional second language learning study by Beckman (1986). The stress correlates in each language were investigated using word pairs contrasting in stress location, for example, English - *subject* and *subject*, and Japanese - *iken* (differing view) and *iken* (opinion). The results from first language users' production indicate that English native speakers utilize max F0, duration, and intensity to signal stress, while Japanese speakers only adopt max F0 to signal pitch accent.

Beckman (1986) then conducted a perception experiment to examine the perceptual salience of these cues in second language learning. Recordings from the production study were re-synthesized and manipulated in duration, F0, intensity, and spectral composition to serve as the stimuli. The participants were Japanese listeners (basic and advanced learners of English), as well as English listeners (basic and advanced learners of Japanese). All participants performed identification tasks to detect the stressed/pitch accented syllable in their native language and second language.

The results from first language perception indicate that native English listeners are sensitive to duration, F0, intensity, and spectral composition, with F0 being the most

important cue when listening to English tokens. Japanese listeners, on the other hand, were only sensitive to F0 cues when listening to Japanese stimuli. In the second language perception study, English learners of Japanese were found to be sensitive to F0 changes, while Japanese learners of English were only sensitive to F0 (and not duration, intensity, or spectral composition).

Japanese L1 and Korean L1 to English L2

The implementation of unstressed vowels by Japanese and Korean L2 learners of English was investigated in Lee et al. (2006). The participants included early and late bilinguals of Japanese (n=20; 10 early, 10 late) and Korean (n=20; 10 early, 10 late), as well as native English speakers (n=10). The early bilinguals in the study were exposed to an English speaking environment before age 6, and have resided in the United States on an average of 20 years, while late bilinguals have had English exposure after the age of 15, and on average lived in the United States for 10 years.

The recording stimuli included 10 trisyllabic English words featuring unstressed vowels in initial, medial, or final positions. They were produced in a carrier phrase “I said ___this time” by all the speakers. Care was taken to ensure that the bilingual were aware of the correct stress placement for the target words. The acoustic measurements included max F0, max intensity, duration, as well as F1 and F2 of both the stressed and unstressed vowels.

The duration results indicate that native English speakers produced unstressed vowels with nearly half of the duration of the stressed vowels (stressed-to-unstressed ratio .45). Both Japanese bilingual groups have similar ratios to the English speakers, while Korean bilingual groups have significantly less duration differences between the stressed and unstressed vowels (greater ratios).

The results from max F0 yielded no differences in the utilization of max F0 between the bilingual groups and the native speakers. Results from intensity (calculated by the decibel differences between the stressed and unstressed vowels) showed that Japanese bilinguals' performance is similar to native speakers, while the Korean bilinguals, although having greater intensity in the stressed vowels, showed significantly smaller intensity differences.

Vietnamese L1 to English L2

A similar study was conducted on native speakers of Vietnamese learning English as a second language (Nguyen and Ingram, 2005). Once again, word pairs contrasting in stress location were chosen as stimuli, but the location of stress in this study was marked on the stressed syllable. Beginning learners (n=10), advanced learners (n=10), and native English speakers (n=2) participated in the acoustic study targeting F0, vowel duration, syllable duration, and intensity. Nguyen and Ingram adopted a series of one-way ANOVAs to investigate the effect of speaker groups on the ratios of stressed-to-unstressed syllables. The results indicated that Vietnamese learners of English (both proficiencies) utilized F0 and intensity correlates similarly to native English speakers. A major difference was the lack of vowel and syllable duration cues in the beginning learners' production. The authors concluded that the learners' ability to encode F0 and intensity cues in English stress production originated from the tonal feature in their L1, i.e. positive L1 transfer. They also claimed that the beginning learners failed to incorporate duration due to negative L1 transfer, since duration is not phonemic in Vietnamese tonal distinctions.

Mandarin L1 to English L2

The acquisition of English sentential stress by Mandarin learners was examined in an acoustic study by Chen et al. (2001). In their experiment, participants read the sentence "I bought a cat there," in which sentential stress was assigned to either "I," "bought," "cat," or "there." The results from forty Mandarin learners of English and

forty English native speakers indicate that all speakers (learners and native speakers alike) differentiate stressed and unstressed words through F0, duration, and intensity – there are significantly greater values in these correlates when a word is stressed. However, learners showed a greater F0 difference and a smaller duration difference between stressed and unstressed words than native speakers. The results are interpreted as interference from the speakers’ tonal L1.

A handful of previous studies have examined the realization of stress correlates in second language learners with tonal native languages. Zhang et al. (2008) compared the realization of mean F0, max F0 location, intensity, duration, and spectral composition in word pairs contrasting in stress locations (e.g. *contract-contract*) between 10 Mandarin speakers and 10 native English speakers. The study adopted a long practice section to ensure that speakers placed stress on the correct syllable of the target word. Speakers were asked to produce the target words in context sentences which cue their lexical category, and then produce the target word in the carrier sentence “I said ___ this time” afterward. In addition, the experimenter explained the stress shift pattern of noun and verb readings for the target words. The speakers were then recorded producing the target words in isolation (with the context sentences as references). Only the word in isolation was analyzed. The statistical analysis conducted used native language and gender as between-subjects factors, and stressed vs. unstressed as the within-subjects factor. The results from Zhang et al. (2008) indicate that the learners and the native speakers behave similarly in their use of duration and intensity, but that learners produce stressed syllables with a higher F0 than native speakers.

1.2 The present study

The acquisition of English lexical stress by Mandarin L2 learners was examined through a series of production and perception studies. Mandarin Chinese is a tone language which adopts F0 height and contour to signal lexical meanings, and English

is a stress language which utilizes duration, F0, intensity, and formant composition to signal stress. Given the suprasegmental differences between the two target languages, I predict that Mandarin L2 learners may differ from native English speakers not only in acoustic realization of English lexical stress, but in perceptual relevance as well. Previous studies on acquisition at the suprasegmental level have focused on either production or perception. There is no study examining how the perception and production of these correlates affect one another. The present study includes a series of acoustic and perception experiments in order to more thoroughly explore the acquisition of English lexical stress by Mandarin learners at different proficiencies. The three experiments included in the current study are outlined below.

1.2.1 Acoustic study 1: acoustic realization of English stress in word pairs

In this production experiment, disyllabic word pairs (e.g. *contract* and *contract*) with similar segmental composition but contrasting in stress location were recorded by Mandarin learners of English with different proficiencies. Mean F0, max F0, intensity, duration, and F2 of stressed and unstressed vowels are measured and compared to the production of native English speakers. The phonemic feature of F0 in Mandarin is predicted to have an effect on the acoustic correlate implementation of English lexical stress by Mandarin second language learners.

Results from a few previous studies on English stress have shown that second language learners have difficulty modulating the inter-stress interval – the time interval between stressed syllables (e.g. Japanese, Anderson (1993); Korean, Koo (1997). Mandarin has been found to have syllable durations remaining relatively constant across a sentence (van Santen and Shih, 2000). In addition, Mandarin L2 learners of English have been shown to have smaller duration differences between stressed and unstressed words than native English speakers (Chen et al. 2001). Taken together, the Mandarin L2 learners are predicted to experience difficulty in reducing

the durations of unstressed vowels in English. As a result, they use duration cues to a lesser extent than native speakers.

Another correlate concerns the vowel quality of stressed and unstressed vowels. It has been shown that unstressed vowels are more reduced than stressed vowels (Gay, 1978). Reduction usually involves shifting F1 and F2 values for the vowels to mid-range frequencies, and is also known as vowel centralization. In this study I use F2 measurements as the main cue for vowel reduction. Given that vowel reduction patterns does not exist in Mandarin, I predict that Mandarin learners may not reduce unstressed English vowels.

1.2.2 Acoustic study 2: F0 contour realization in English stress

The goal of the second acoustic study is to examine the realization of F0 movements in English lexical stress by Mandarin L2 learners of English. The construction of word pairs as stimuli in the previous acoustic study provides well-controlled segmental conditions for the investigation of mean F0, max F0, intensity, duration and F2. However, most of the syllables in these stimuli do not comply with Mandarin syllable phonotactics. In order to better evaluate the realization of F0 contours in speakers' production, stimuli which comply with Mandarin syllable structures should be compiled.

The same speakers in the word-pair acoustic study participated in this experiment. The F0 contours of the target syllables in the stressed and unstressed positions were compared between native and second language learners. Mandarin L2 learners of English are predicted to use F0 in signaling stressed vowels. In addition, given that there are four phonemic F0 contour patterns in Mandarin (high-level, mid-rising, low-dipping and high-falling), it is predicted that learners may use the two high tones (high-level and high-falling) to implement stressed syllables in English. Native

English speakers, on the other hand, will have more diverse F0 contours in stressed syllables.

1.2.3 Perception study: Roles of max F0, duration, and vowel quality in stress perception

The results from the previous two acoustic studies will serve as the blueprint for stimulus construction in the perception study. The goal of this perception study is to compare the relative importance of perceptual cues for English stress between the Mandarin L2 learners of English and native English speakers. The target perceptual cues are going to be selected based on the most diverse utilization patterns between learners and native speakers in the acoustic studies. The stimuli for the perception study will be resynthesized from a naturally produced nonword 'dada'. The target cues will be resynthesized so that there are stepwise ratios of first-to-second vowels for those cues. A stress localization task, in which the listeners have to identify the stressed syllable, will be adopted to test the perceptual relevance of the selected cues. Cue weighting in stress perception by learner groups will be compared to that of native listeners.

CHAPTER 2: ACOUSTIC STUDIES

The acoustic implementation of English lexical stress by Mandarin L2 learners and native English speakers was examined through two acoustic studies. The first acoustic study aims to examine the realization of mean F0, max F0, intensity, duration, and F2 in disyllabic minimal word pairs contrasting in stress location (e.g. *contract-contract*). The near-identical segmental composition of these word pairs provides for a high degree of control, making it easier to compare a given syllable in stressed and unstressed positions.

The second acoustic experiment aims to investigate F0 contour patterns in English stressed syllables. Due to the fact that syllables in the first acoustic study mostly do not comply with Mandarin phonotactics, I utilized word pairs in the second study which contain CV(N) syllables, and which conform to legal syllable structures in Mandarin. Given the phonemic roles of F0 patterns (tones) in Mandarin, I predict that the learners may use F0 patterns with which they are most familiar in their native language when producing English syllables. I also predict that high tones (high-level and high-falling) are the most likely to be adopted by the beginning learners to signal stressed syllables in English. A rating experiment was also conducted to determine the proficiency levels of participants.

2.1 Rating experiment

The main goal of this dissertation is to investigate the influence of L1 on the prosody acquisition process in L2 production. In addition, I am also interested in how the L1 effect transforms as proficiency increases. As standard proficiency tests can sometimes prove to be unreliable indicators of speaking proficiency, a rating experiment was conducted, in which native English listeners determined the proficiency of the Mandarin L2 learners.

2.1.1 Methodology

Participants

Speakers – Twenty-six (10 males and 16 females) adult native speakers of Mandarin who learned English as a second language were recruited from Taiwan and the University of Kansas. All the speakers are Taiwanese/Mandarin bilinguals. They all started learning English at the age of 12 (in junior high) with the Grammar Translation method, which focuses mainly on vocabulary and grammar learning. Among them, some continued more communicative English education in college, and some are currently studying in America for postgraduate degrees. Ten adult native speakers of English (7 males, 3 females) were also recruited from the University of Kansas as the control group. None of the speakers have a history of hearing or speech difficulties.

Stimuli

All speakers were asked to read a short paragraph (see below) aloud. One long sentence (underlined in the paragraph) was chosen to be the stimulus for the rating experiment. This sentence was selected due to its length, content, and its medial position in the paragraph.

“Learning to speak a foreign language fluently and without an accent isn’t easy. In most educational systems, students spend many years studying grammatical rules, but they don’t get much of a chance to speak. Arriving in a new country can be a frustrating experience. Although they may be able to read and write very well, they often find that they can’t understand what people say to them. English is especially difficult because the pronunciation of words is not clearly shown by how they’re written. But the major problem is being able to listen, think, and respond in another language at a natural speed. This takes time and practice.”

Procedure

Each speaker was recorded separately in the anechoic chamber of the University of Kansas, or in a quiet room in Taiwan. All recordings were conducted with a Marantz PMD671 solid state recorder and Electro-Voice RE20 microphone, with a sampling rate of 22.05 kHz. The recording was then transferred to Praat. The target sentences from each speaker's recording were then extracted, to be the tokens for the rating experiment.

The raters were tested individually in a sound-treated room in the Kansas University Phonetics and Psycholinguistics Laboratory. The stimuli from all thirty-six speakers (26 learners and 10 native speakers) were presented to the raters in random order using SuperLab (Cedrus Corporation). During the experiment, the raters listened to each stimulus and were asked to use a 5-point scale to rate each production by pressing buttons on a response box. Four rating criteria were included in the experiment: foreign accentedness, on a scale ranging from "native" (1) to "heavily accented" (5); speaking rate, from "fast" (1) to slow (5); intonation, from "native-like" (1) to "not-at-all native-like" (5); and finally articulation, from "clear" (1) to "unclear" (5). Before beginning the rating session, four stimuli (2 from non-native and 2 from native speakers) were presented to the raters for practice.

2.1.2 Results

The inter-rater reliability was calculated using the Average Measure intra-class correlation coefficient. The absolute agreement correlation coefficient showed a high agreement among the raters: 0.842. Four one-way ANOVAs using speaker groups as a between subjects factor indicate that the native speakers receive significantly lower ratings (more native-like) in foreign accentedness, speaking rate, intonation, and articulation. These results are presented in Figure 2.1.

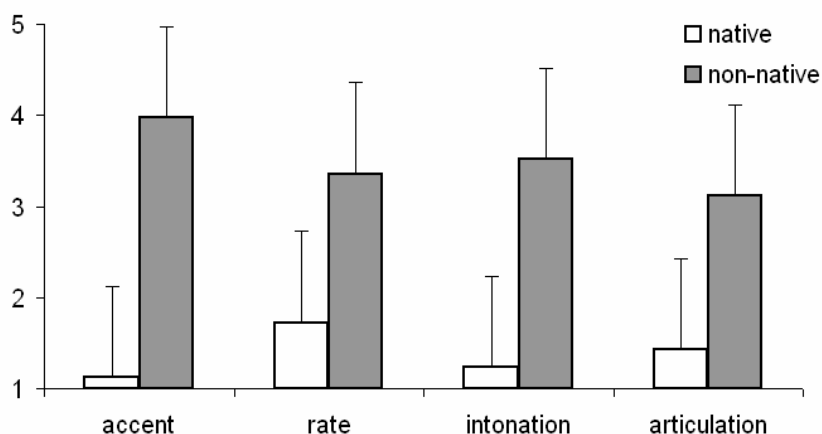


Figure 2. 1. Ratings on foreign accentedness, speaking rate, intonation, and articulation for native and non-native speakers, on a scale from 1 (native) to 5 (heavily accented).

A correlation analysis was also conducted on all four criteria, and yielded a positive correlation between accent, rate, intonation, and articulation. The statistical results are summarized in Table 2.1.

Correlations

			accent	rate	intonation	articulation
Kendall's tau_b	accent	Correlation Coefficient	1.000	.632**	.762**	.662**
		Sig. (2-tailed)	.	.000	.000	.000
		N	972	972	972	972
	rate	Correlation Coefficient	.632**	1.000	.599**	.524**
		Sig. (2-tailed)	.000	.	.000	.000
		N	972	972	972	972
	intonation	Correlation Coefficient	.762**	.599**	1.000	.626**
		Sig. (2-tailed)	.000	.000	.	.000
		N	972	972	972	972
	articulation	Correlation Coefficient	.662**	.524**	.626**	1.000
		Sig. (2-tailed)	.000	.000	.000	.
		N	972	972	972	972

** . Correlation is significant at the 0.01 level (2-tailed).

Table 2. 1. Correlations between foreign accent rating, speaking rate, intonation, and articulation.

Given the high correlations between the four criteria, the foreign accentedness rating was chosen to be the key criterion in proficiency placement for its encompassing nature. The 95% confidence interval data from the ratings for non-native speakers are plotted in the order of native (1) to heavily accented (5) in Figure 2.2. Based on the confidence interval error bars, the top nine ranked speakers were found to be significantly higher than the bottom nine ranked speakers (there was no overlap in the error bar ranges). The top nine participants were thus placed in the advanced learners group (green box), and the bottom nine participants in the beginning group (blue box). The beginning learners include speakers 1, 2, 5, 6, 8, 9, 20, 21, and 24. The advanced learners are speakers 7, 13, 14, 15, 17, 18, 22, 23, and 25.

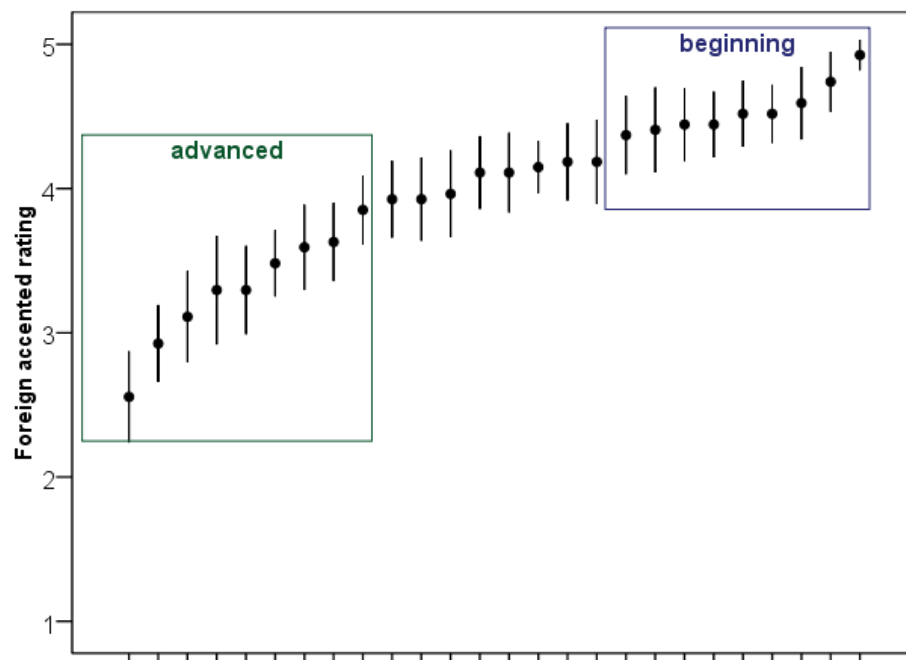


Figure 2. 2. The confidence intervals for foreign accentedness rating. The non-native speakers are sorted according to the received rating. In this graph, ‘1’ is native-like, and ‘5’ is heavily accented.

2.2 Acoustic study I – lexical stress produced in *contract-contract* word pairs

In this acoustic experiment, disyllabic word pairs (e.g. *contract* and *contract*) with similar segmental composition but contrastive stress location were recorded by Mandarin learners of English with different proficiencies. Mean F0, max F0, intensity, duration, and F2 of stressed and unstressed vowels are measured and compared to those in native English speakers' production. It is predicted that the tonal L1 will have a positive transfer effect on the stress L2, through facilitating the use of F0 correlates. As a result, Mandarin learners of English will utilize F0 to a greater extent than native English speakers to signal stressed syllables in English. As the proficiency level increases, learners will use F0 more similarly to native speakers.

In terms of duration, it is predicted that stressed and unstressed syllables will be similar in duration for beginning learners, due to the more syllable-timed rhythm pattern of Mandarin. The advanced learners, however, will perform more like native speakers, in showing greater duration differences between stressed and unstressed vowels. No difference in the use of intensity is predicted between native and nonnative speakers. Finally, no spectral composition change (vowel reduction) is expected in unstressed vowels for beginning learners, however advanced learners may pattern like native English speakers and exhibit vowel reduction.

2.2.1 Methodology

Participants

The data of those L2 learners who were placed in the advanced (2 males and 7 females) and beginning (5 males and 4 and females) groups by the rating experiment were included, as well as those of ten native English speakers (7 males and 3 females) who acted as a control group. None of the speakers have a history of hearing or speech difficulties.

Stimuli

The stimuli consist of 14 disyllabic word pairs with identical segmental construction, contrasting in stress location, and embedded in context sentences balanced for number of syllables in order to avoid possible rhythmic differences. Word pairs of this nature often correspond to nouns (trochaic stress pattern) and verbs (iambic stress pattern) – therefore the trochaic and iambic target words are labeled as nouns and verbs, respectively in the current study. A complete list of target words and their carrier sentences can be found in Appendix A.

Procedure

When presenting English words in isolation, second language learners and native speakers alike are sometimes unable to consistently produce the correct stress pattern for the target word's lexical category. Since the goal of this experiment is to examine the acoustic realization of stress, and not stress placement strategy, steps were taken to ensure that participants were well-informed of the stress locations of the target words. Information regarding the lexical categories of the target words was conveyed to the subjects by embedding the words in context sentences, with the target word highlighted and the stressed syllable underlined. For example, the noun *contract*: “The new **contract** needs to be signed.” And the verb *contract*: “Steel will **contract** when it is cooled.”

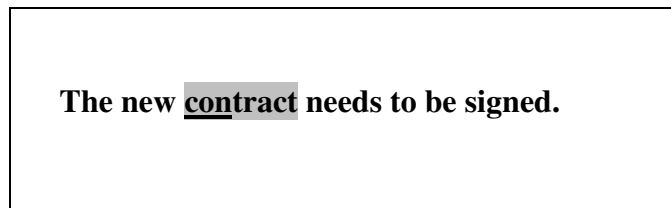


Figure 2. 3. Example of stimulus presented on computer screen.

Several recent acoustic studies on English stress correlates have demonstrated that the acoustic realization of stress varies as the pitch accent position changes (Beckman

and Edwards, 1994, Okobi, 2006). Furthermore, nouns and verbs occupy different phrase positions in sentences, and are subject to the influence of phrasal/sentential prosody. In order to minimize the effects of these two extrinsic factors on the implementation of the target words, the subjects were instructed to first read the entire sentence aloud, and then repeat the highlighted word in exactly the way it was read in the sentence. For example, the sentence in Figure 2.3 was read as “The new contract needs to be signed. contract.” Only the target word produced in isolation was analyzed. This method ensures that the syntactic category and stress pattern of the target word are known to the speaker, while reading the target word in isolation minimizes the effects of accentual positions and phrasal prosodic patterns.

Each speaker was recorded separately in the anechoic chamber of the University of Kansas, or in a quiet room in Taiwan. All recordings were conducted with a Marantz PMD671 solid state recorder, Electro-Voice RE20 microphone with a sampling rate of 22.05 kHz. The sentences were randomized and presented on a laptop screen one sentence at a time using Microsoft PowerPoint. The speakers were seated in front of the laptop during the recording. They controlled the speed of the slide transition by pressing the space bar. The recordings were arranged in three sections, each with one repetition of the stimuli, and a break between each section. Before the recording session began, the participants were presented with the list and asked to practice each sentence and the target word within until they were comfortable with the list.

Acoustic analyses

Acoustic analyses of the stimuli were performed using the speech analysis software program Praat 4.3.04 (By Paul Boersma and David Weenink) at the Phonetics and Psycholinguistics Laboratory at the University of Kansas. Only the target words produced in isolation were analyzed. The measurements were done through Praat scripts which measured mean F0, max F0, intensity, duration, and F2 for the vowels. For each production of each subject, a textgrid was generated and four interval

boundaries (onsets and offsets of stressed and unstressed vowels) were determined and set with the help of waveform, spectrogram, pitch contour, and intensity contour displays. A total of 2352 tokens were analyzed in this experiment (28 words × 3 repetitions × 28 speakers).

F0 measurements included ‘mean F0’- averaging the F0 measurements across each vowel, as well as ‘max F0’- the highest F0 measurement of each vowel. The duration of each vowel was measured from the F1 onset to the F2 offset. When syllable-final sonorants could not be reliably separated from the vowel, they were included in the duration measurement for both members of the minimal pair. The intensity of the vowel was measured by averaging the intensity measurements across each vowel.

Absolute acoustic measurements are sensitive to differences in speaking rate (affects duration), intrinsic F0 (affects F0), and amplitude (affects intensity) across different speakers. Therefore, I used the ratios of these measurements between stressed and unstressed vowels in the statistical analysis. Two types of vowel ratios were computed for different types of comparisons: first-to-second vowel ratios and stressed-to-unstressed ratios. Their roles are described below.

In order to examine if the speakers utilize the target correlates (mean F0, max F0, intensity, and duration) to differentiate noun (trochaic pattern) and verb (iambic pattern) readings, the first-to-second vowel ratios for each correlate were calculated. In theory, if a given correlate is encoded to signal stress in noun readings, the ratio for first-to-second vowel should be above 1, which is the assumed base line when the cue remains the same between the first and second vowels. The first-to-second vowel ratio should be below 1 if the cue is utilized in signaling stress in verb readings.

Ratios for stressed-to-unstressed vowels were computed in order to compare the extent to which a given correlate was utilized in signaling stress in nouns and verbs.

This ratio was computed from vowel 1 over vowel 2 ratio (V1/V2) in nouns, while in verbs it was calculated from vowel 2 over vowel 1 (V2/V1).

The second formant (F2) was measured at the midpoint of the vowel (midpoint of the interval between vowel onset and offset). Examples of the segmentation criteria for noun (*object*) and verb (*object*) readings are shown in Figures 2.4 and 2.5.

Noun reading

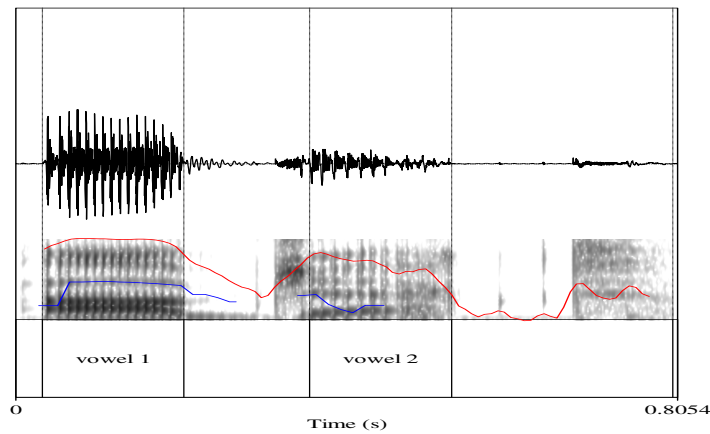


Figure 2. 4. An example of the acoustic measurements for the noun reading - *object*. The labeled intervals mark the duration of the stressed and unstressed vowels. Blue lines stand for F0 and red lines stand for intensity.

Verb reading

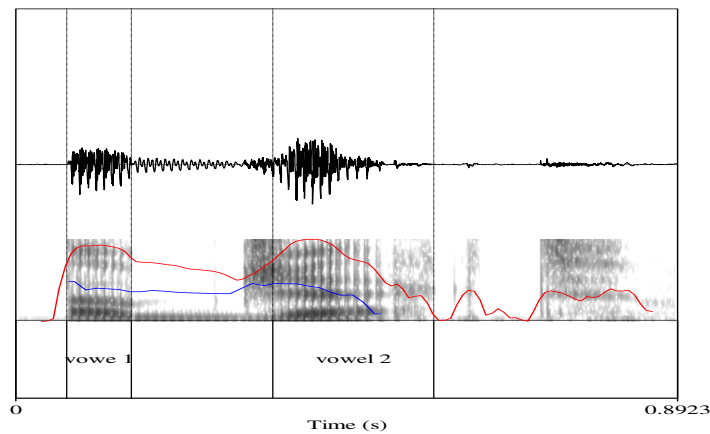


Figure 2. 5. An example of the acoustic measurements for the verb reading - *object*.

2.2.2 Results

In order to examine if speakers utilize mean F0, max F0, intensity, and duration to signal the difference between noun (trochaic pattern) and verb (iambic pattern) readings, the first to second vowel ratios for the acoustic correlates mentioned above were examined. After the implementation patterns for first-to-second vowel ratios were established, the vowel ratios for stressed-to-unstressed vowels were used as the dependent variables for the statistical analysis. Again, for noun readings, stressed-to-unstressed ratio refers to the ratio of vowel 1 over vowel 2, while in verb readings it is computed by vowel 2 over vowel 1.

Mean F0

First-to-second ratios

The 95% confidence interval graphs for the distribution of first-to-second vowel mean F0 ratios are plotted in Figure 2.6. The x axis lists the three proficiency levels - beginning, advanced, and native. The y axis represents the first-to-second vowel ratio. The dotted reference line, 1.0, is the assumed baseline for which mean F0 values are identical in the first and second syllables. The noun readings are coded in blue, and verb readings in green. As shown in the graph, noun ratios across all speaker groups are above 1. The distribution of mean F0 ratios for native English speakers, however, overlaps with the base line value 1.

A series of one-sample T-tests confirmed that the ratios for nouns are significantly greater than 1 for all speaker groups. The ratios for verbs are significantly smaller than 1, except for native speakers. These findings indicate that mean F0 in stressed vowels is significantly higher than in unstressed vowels in beginning and advanced learners' stress production. This pattern is consistent when learners produce both noun (trochees) and verb (iamb) readings. Native English speakers used greater mean F0 for stressed vowels in noun readings; however mean F0 was nearly identical between syllables in verb readings.

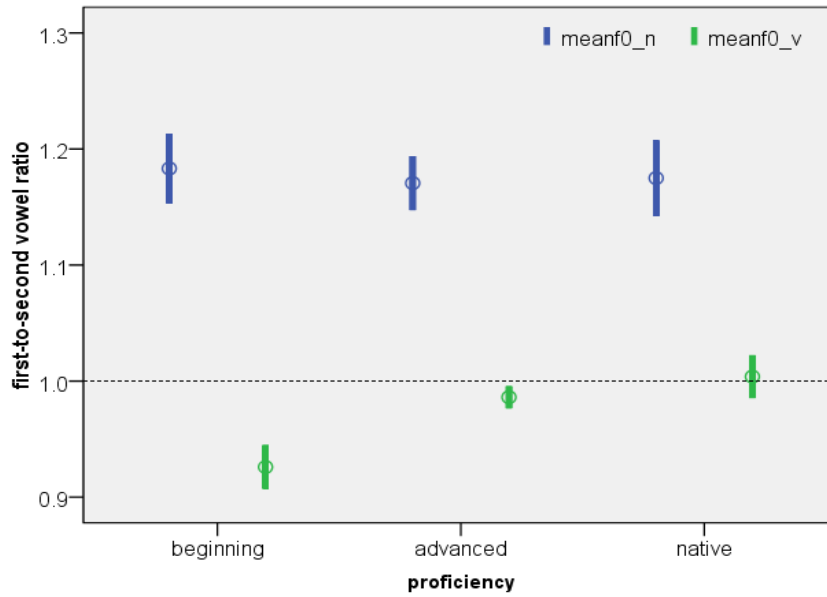


Figure 2. 6. Error bars at 95% confidence interval for first-to-second vowel mean F0 ratio across proficiencies.

Stressed to unstressed ratios

In order to evaluate how mean F0 is implemented in nouns and verbs in productions from different speaker groups, the stressed-to-unstressed ratios were calculated as the dependent variable for a two-way repeated measures analysis of variance (ANOVA). The within-subject factor is the stress pattern (noun and verb), the between-subject factor is the proficiency level (beginning, advanced, and native), while the dependent variable is the stressed-to-unstressed vowel ratio.

The statistical results are tallied in Table 2.2. The main effect of stress was significant - nouns (1.18) have greater ratios than verbs (1.03), indicating that the mean F0 cue is greater in nouns than in verbs. The main effect of proficiency was also significant. A posthoc analysis showed that beginning learners have a significantly greater mean F0 ratio (1.14) than advanced (1.09) and native speakers (1.09) (Figure 2.7).

mean F0	significance	F value	p value
stress	*	(1, 342) = 183.01	<.001
proficiency	*	(1, 342) = 9.92	<.001
stress × proficiency	*	(1, 342) = 4.93	=.008

Table 2. 2. The significance, F and p values for ANOVAs on mean F0.

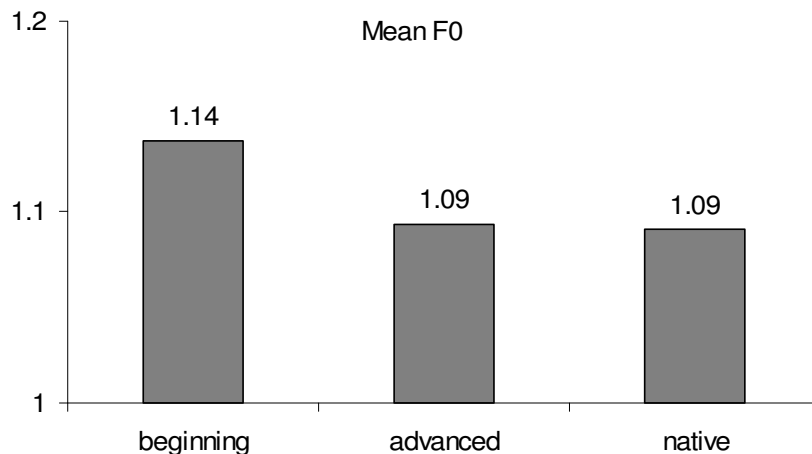


Figure 2. 7. Stressed-to-unstressed vowel ratios for mean F0 across three proficiencies.

The significant interaction of stress and proficiency is shown in Figure 2.8. Further analysis indicated that the effect of proficiency was not significant in nouns [$F(2, 345) = .18, p=.835$], but the effect was significant in verbs [$F(2, 388) = 22.6, p <.001$]. The results suggest that the speaker groups are similar in their use of mean F0 in noun readings but different in verb readings. The posthoc analysis on verb readings indicated that beginning learners have a greater ratio (1.09) than advanced (1.02) and native speakers (1.0). There was no significant difference between advanced learners and native speakers. In sum, all speaker groups utilize mean F0 cues similarly in noun readings, but in verb readings beginning learners show a greater use of mean F0.

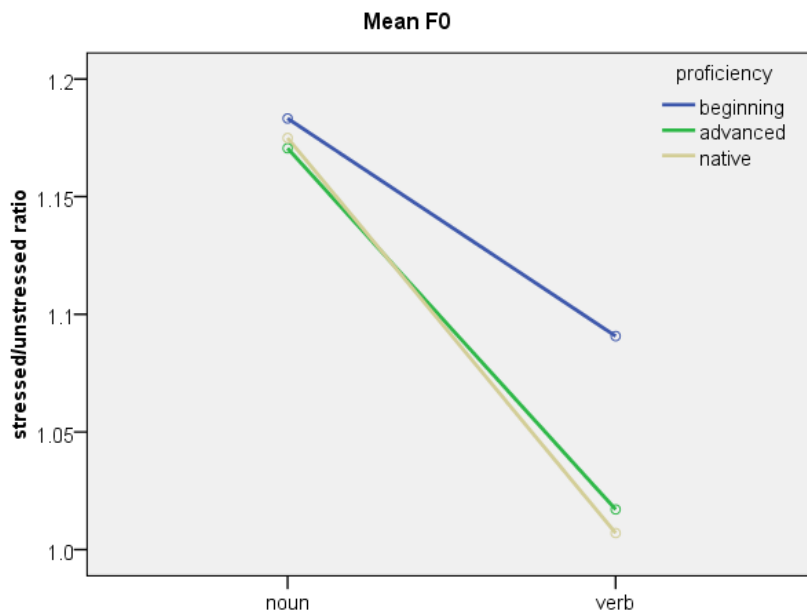


Figure 2. 8. The interaction between stress location and proficiency.

Max F0

The first-to-second max F0 ratios for nouns and verbs across different speaker groups are plotted in Figure 2.9. The x axis lists the three proficiency levels - beginning, advanced, and native. The y axis represents the first-to-second vowel ratio. The dotted reference line, 1.0, is the assumed base line for which max F0 values are identical in the first and second syllable. A series of one-sample t-tests confirmed that the ratios for nouns are significantly larger than 1.0 for all speaker groups. The ratios for verbs are significantly smaller than 1.0 in all learners, but not in native speakers.

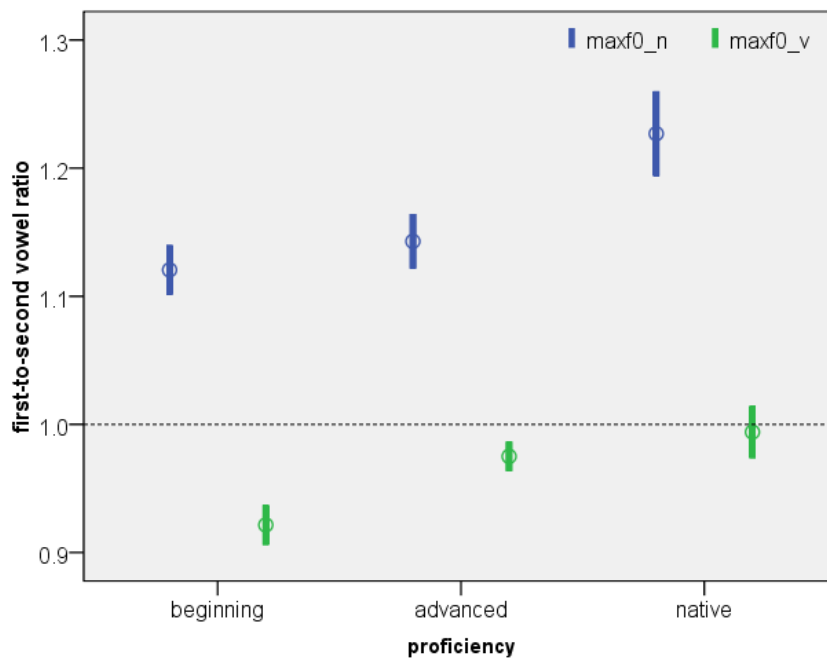


Figure 2. 9. Error bars at 95% confidence interval for first-to-second vowel max F0 across proficiencies.

Stressed-to-unstressed vowel ratios

In order to evaluate how max F0 is implemented in nouns and verbs in production from different speaker groups, the stressed-to-unstressed ratios were calculated as the dependent variable for a two-way repeated measures analysis of variance (ANOVA). The results are shown in Table 2.3.

There is a significant main effect of stress - nouns (1.16) have greater ratios than verbs (1.05), indicating that the max F0 cue is utilized to a greater extent in nouns than in verbs. The main effect of proficiency was also significant. As shown in Figure 2.10, native speakers have significantly greater max F0 than advanced learners, but not beginning learners. No difference between beginning and advanced learners was found.

max F0	significance	F value	p value
stress	*	(1, 377) = 159.37	<.001
proficiency	*	(1, 377) = 5.55	=.004
stress × proficiency	*	(1, 377) = 32.8	<.001

Table 2.3. The significance, F and p values for the ANOVA on max F0.

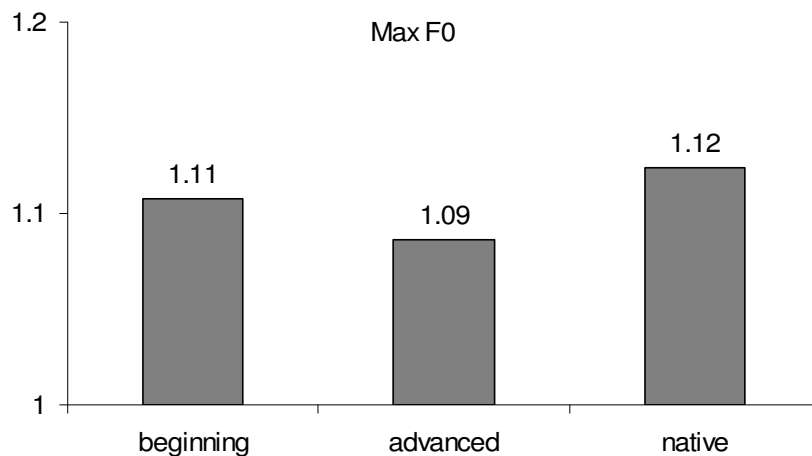


Figure 2. 10. Stressed-to-unstressed vowel ratios for max F0 across the three proficiencies.

The interaction of stress and proficiency is also significant (Figure 2.11). Further analysis indicated that the effect of proficiency was significant in nouns [$F(2, 379) = 18.96, p < .001$]. Native speakers use a greater max F0 ratio (1.23) than both beginning (1.12) and advanced learners (1.14) in nouns.

The effect of proficiency in verb readings is also significant [$F(2, 389) = 19.78, p < .001$]. The max F0 ratio for beginning learners (1.09) is significantly greater than that for both advanced learners (1.03) and native speakers (1.02). There is no significant difference between advanced learners and native speakers.

To summarize, native English speakers utilize max F0 to a significantly greater extent than both learner groups in noun readings. In verb readings however, max F0 plays no role in stress realization in native speakers' productions. Beginning learners, on the other hand, adopt a consistent max F0 difference in both noun (1.12) and verb readings (1.09).

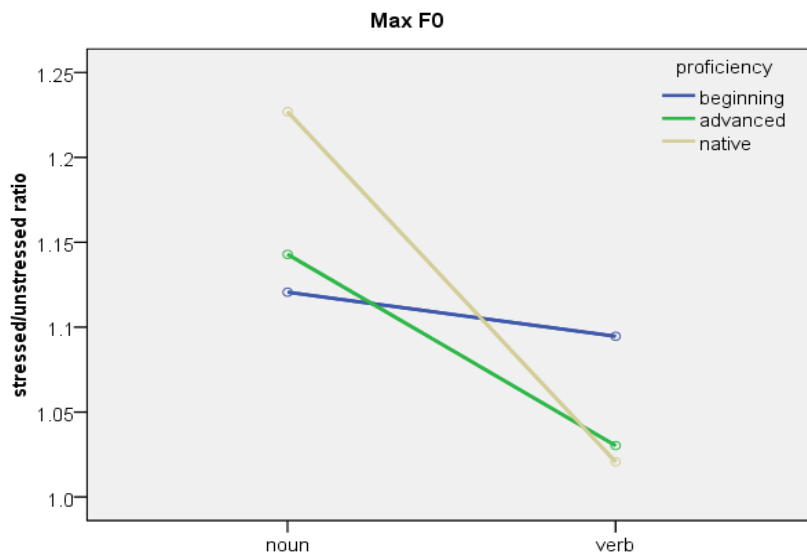


Figure 2. 11. The interaction between stress location and proficiency level for max F0.

Intensity

The first-to-second vowel intensity ratios for nouns and verbs across different speaker groups are plotted in Figure 2.12. The x axis lists the three proficiency levels - beginning, advanced, and native. The y axis represents the first-to-second vowel ratio. The dotted reference line, 1.0, is the assumed baseline for which intensity values are identical in the first and second syllables. A series of one-sample t-tests confirmed that the ratios for nouns are significantly larger than 1.0 for all speaker groups. The ratios for verbs are significantly smaller than 1.0.

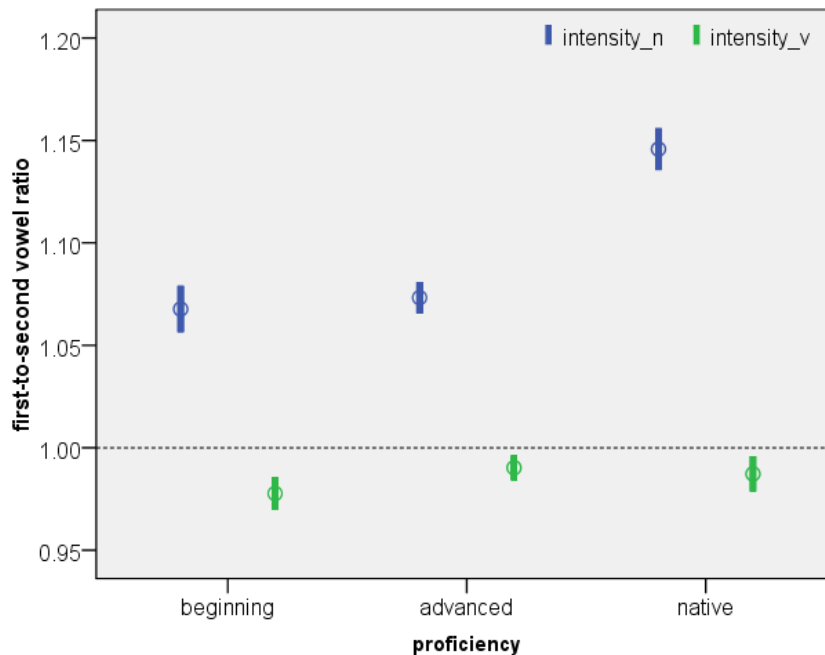


Figure 2. 12. Error bars at 95% confidence interval for intensity across the three speaker groups.

Stressed to unstressed ratios

Results from a two-way repeated measures ANOVA using the stressed-to-unstressed vowel intensity ratio as the dependent variable are shown in Table 2.4.

There is a significant main effect of stress. The ratio is greater in nouns (1.1) than verbs (1.02). The main effect of proficiency is also significant (see Figure 2.13). A posthoc analysis indicated that native speakers have a significantly greater ratio than advanced and beginning learners. The difference between advanced and beginning learners was not significant.

intensity	significance	F value	p value
stress	*	(1, 389) = 352.11	<.001
proficiency	*	(2, 389) = 41.5	<.001
stress × proficiency	*	(2, 389) = 63.15	<.001

Table 2. 4. The significance, F and p values for the ANOVA on intensity.

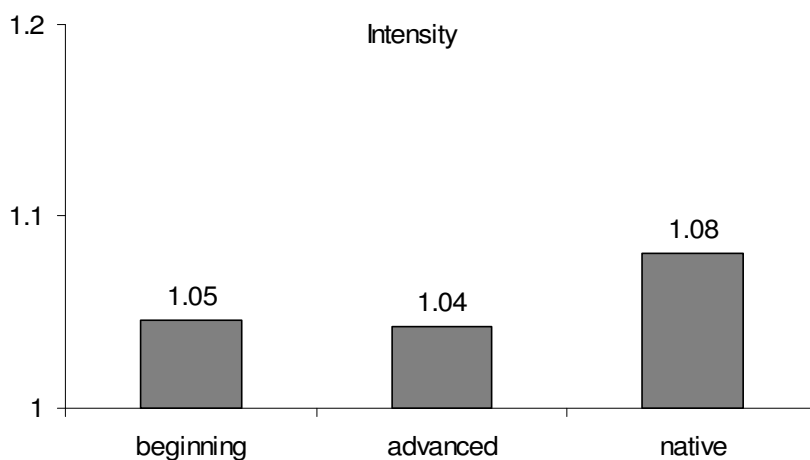


Figure 2. 13. Stressed-to-unstressed vowel ratios for intensity across three proficiencies.

The interaction between stress and proficiency is significant (Figure 2.14). Further analysis indicated that the effect of proficiency for noun readings is significant [F (2, 389) = 77.93, p<.001]. Native speakers have a greater ratio (1.15) than beginning

(1.07) and advanced learners (1.07). The difference between advanced and beginning learners was not significant.

The effect of proficiency in verb readings is also significant [F (2, 389) = 3.07, p= .048]. Beginning learners have a significantly greater intensity ratio (1.03) than advanced learners (1.01) but not native speakers (1.02). No difference was found between advanced learners and native speakers.

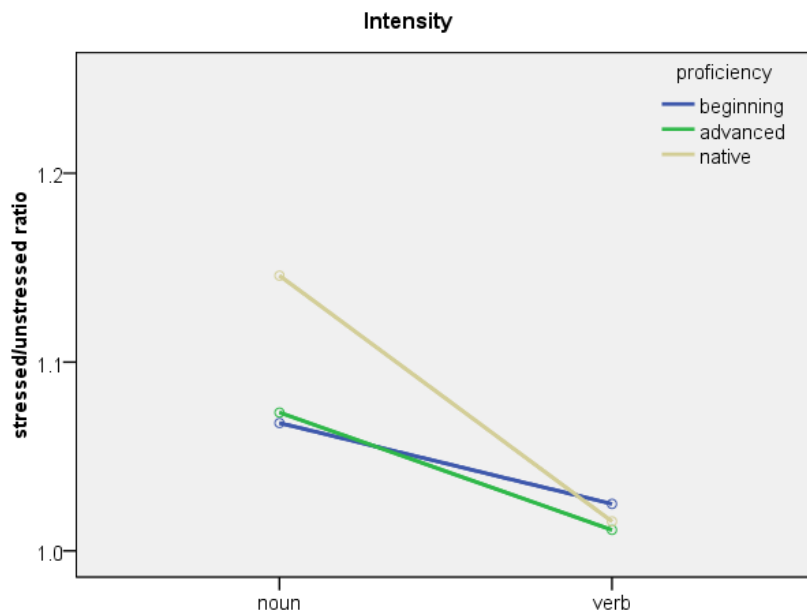


Figure 2. 14. The interaction between stress location and proficiency levels for intensity.

Duration

The first-to-second vowel duration ratios for nouns and verbs across different speaker groups are plotted in Figure 2.15. The x axis lists the three proficiency levels - beginning, advanced, and native. The y axis represents the first-to-second vowel ratio. The dotted reference line, 1.0, is the assumed baseline for which duration values are identical in the first and second syllables. A series of one-sample t-tests confirmed that the ratios for nouns are significantly larger than 1.0 for all speaker groups. The ratios for verbs are significantly smaller than 1.0.

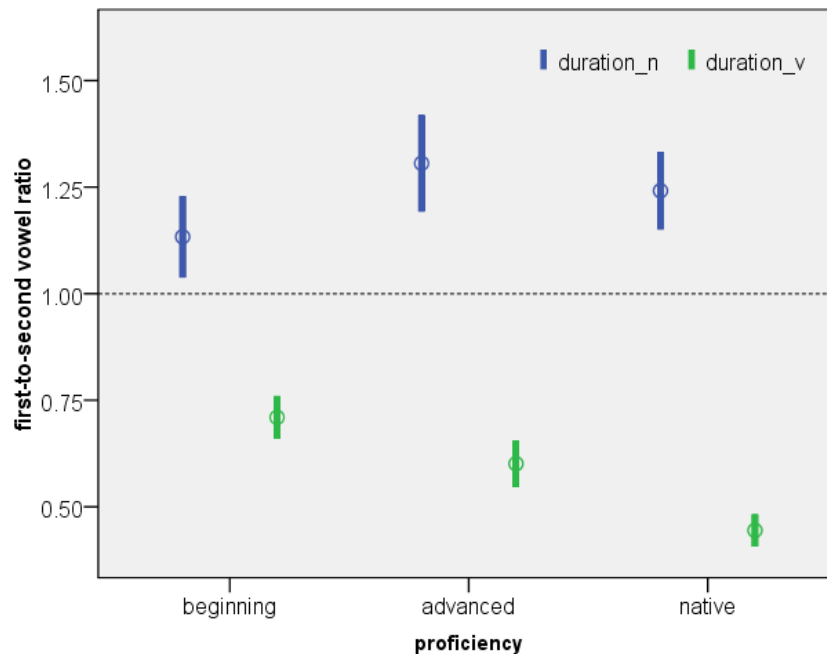


Figure 2. 15. Error bars at 95% confidence interval for duration ratios across three speaker groups.

Stressed to unstressed ratios

The stressed-to-unstressed ratios were calculated as the dependent variable for a two-way repeated measures analysis of variance (ANOVA). The results are shown in Table 2.5. The main effect of stress is significant - the duration ratio is greater in verbs (2.17) than nouns (1.23). The effect of proficiency is also significant (see Figure 2.16). A posthoc analysis indicated that native speakers have a significantly greater duration ratio than advanced learners, which have a significantly greater ratio than beginning learners.

duration	significance	F value	p value
stress	*	(1, 389) = 176.81	<.001
proficiency	*	(2, 389) = 25.37	<.001
stress × proficiency	*	(2, 389) = 68.80	<.001

Table 2. 5. The significance, F and p values for the ANOVA on duration.

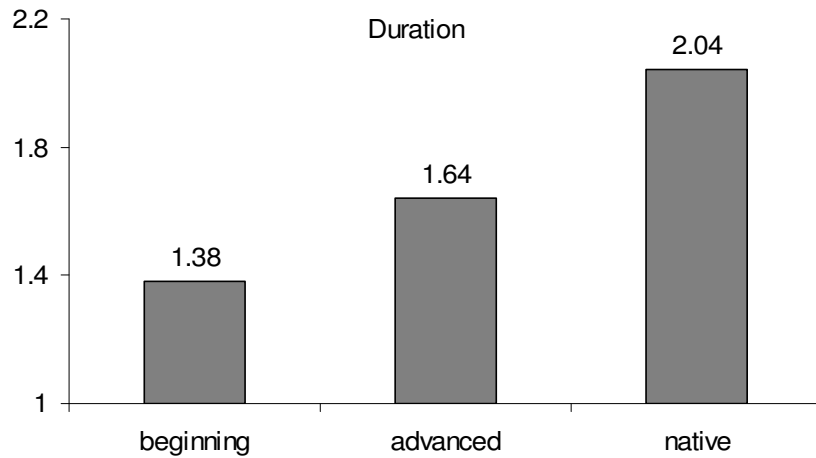


Figure 2. 16. Stressed-to-unstressed vowel ratios for intensity across three proficiencies.

The interaction between stress location and proficiency is also significant. Further analysis indicated that the effect of proficiency in noun readings is close to being significant [$F(2, 389) = 2.89, p = .057$]. The ratios are: native speakers (1.24), advanced learners (1.31), and beginning learners (1.13).

The effect of proficiency is also significant in verb readings [$F(2, 389) = 50.52, p < .001$]. Native speakers have a significantly greater duration ratio (2.84) than advanced speakers (1.97), who have a significantly greater duration ratio than beginning learners (1.63).

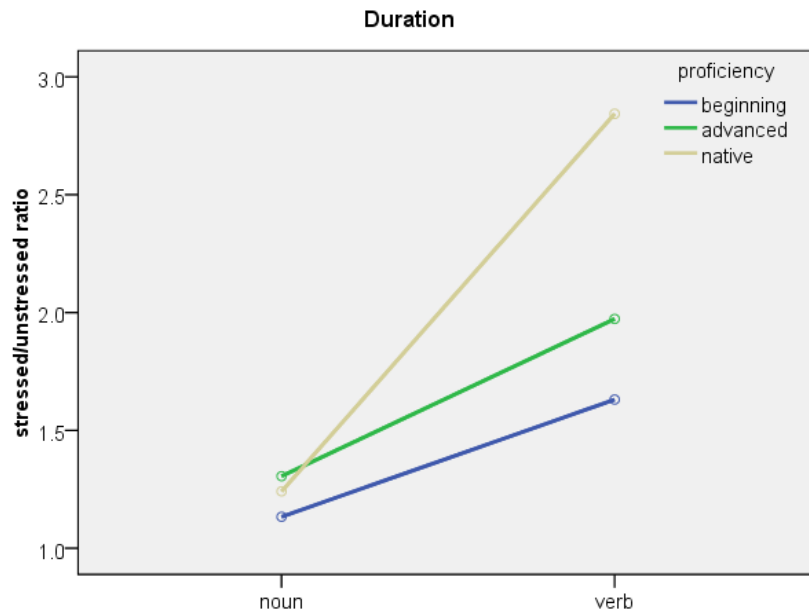


Figure 2. 17. The interaction between stress location and proficiency for duration.

Formant analysis

Previous acoustic studies on English vowels (e.g., Peterson and Barney, (1952) established that back vowels have a lower F2 than front vowels. Furthermore, the process of unstressed vowel reduction involves the centralization of F2(Lindblom, 1963). This centralization induces formant composition change in opposing directions for front and back vowels – a reduced front vowel has a lower F2 than the full vowel, while a reduced back vowel has a higher F2 than the full vowel. Due to this intrinsic F2 difference, and the different directions of formant shift reduction, the data from front and back vowels were analyzed separately.

The second formants of the stressed and unstressed vowels were measured to evaluate the degree of vowel reduction. The results from max F0, intensity, and duration have suggested that these correlates are not utilized in the same way in noun and verb readings.

The data from syllable 1 are plotted in Figure 2.18. Syllable 1 is stressed in noun readings, but is unstressed in verb readings. The stressed noun reading data are represented by the solid blue line, while the unstressed verb reading data are represented by the dotted green line. The arrows refer to the direction of formant shift for native speakers. The unstressed vowels in syllable 1 position (i.e. in verb readings) in native speaker productions demonstrate a classic vowel reduction - F2 is lower in unstressed front vowels (arrow goes down) but higher (arrow goes up) in unstressed back vowels. The degree of centralization is greater in back vowels, and relatively smaller in front vowels. Unstressed vowels for both learner groups, however, did not show any centralization in back vowels, and have opposite formant changes in front vowels when compared to native speakers.

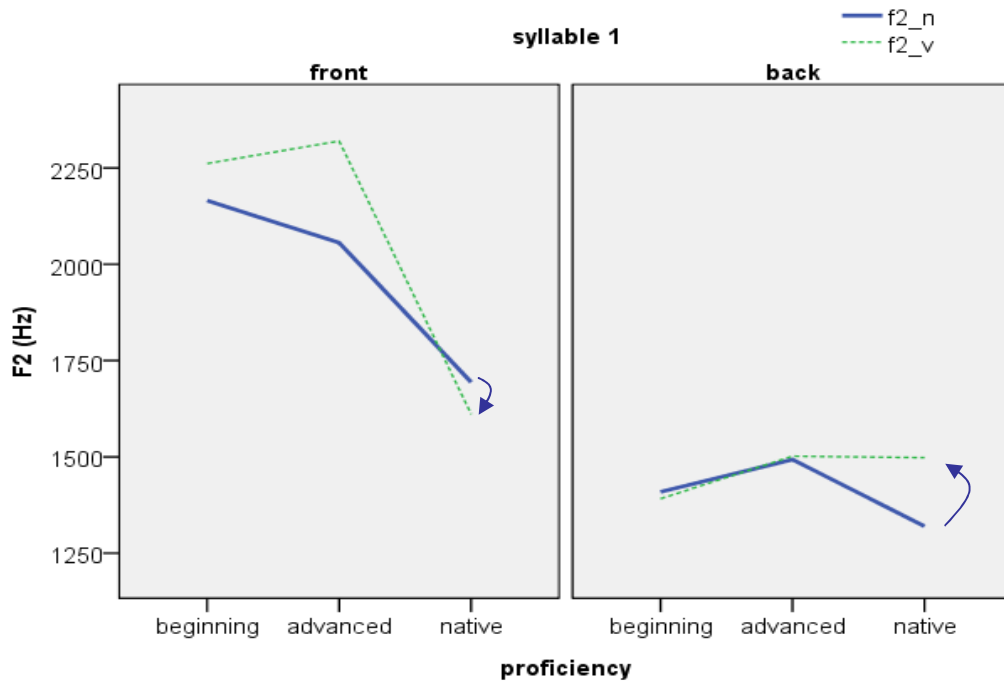


Figure 2. 18. Second formant for the first syllable in stressed (noun) and unstressed (verb) positions.

The data from syllable 2 are plotted in Figure 2.19. Syllable 2 is stressed in verb readings but is unstressed in noun readings. The stressed verb reading data are represented by the solid green line, while the unstressed verb reading data are represented by the dotted blue line. The arrows refer to the direction of formant shift for native speakers. The F2 of the unstressed vowels is centralized in all speakers' productions - lower in front vowels and higher in back vowels when compared to the same vowel in stressed position. The degree of centralization is again greater in back vowels, and relatively smaller in front vowels.

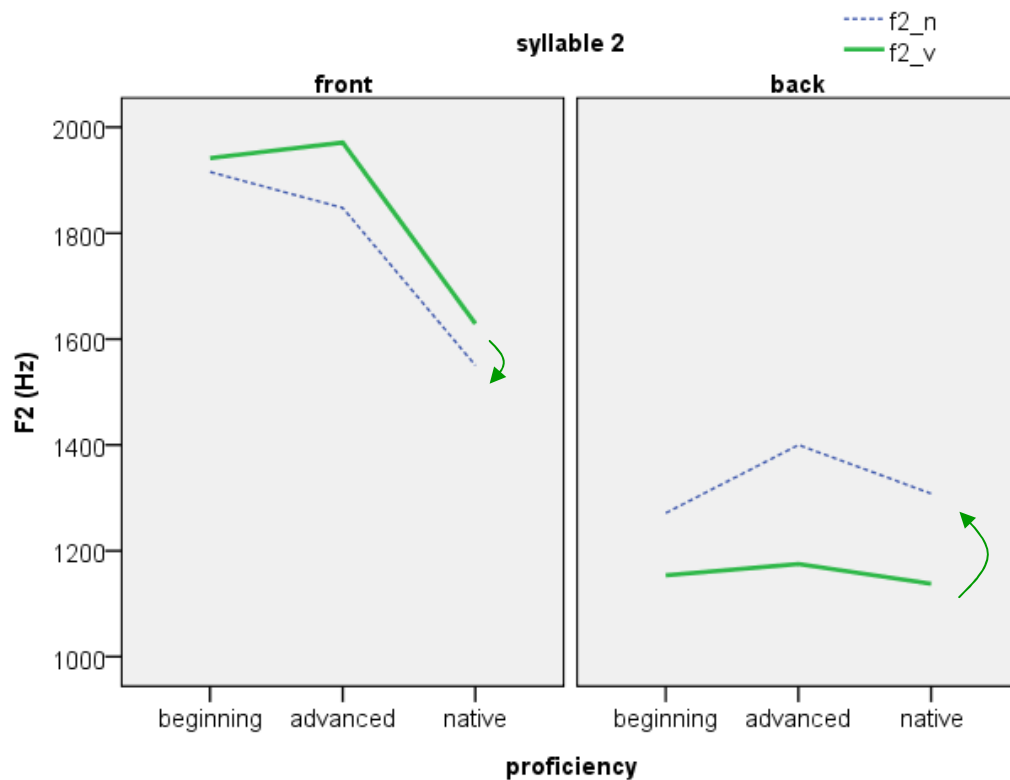


Figure 2. 19. Second formant at mid point of syllable 2 in stressed (verb) and unstressed (noun) position. The green solid line represents verbs while the blue dotted line represents nouns.

The results from both syllable 1 and syllable 2 across front and back vowels are further summarized in Figure 2.20. In this figure noun readings are in blue, while verb readings are in green. Syllable 1 is stressed in noun readings, while syllable 2 is stressed in noun readings

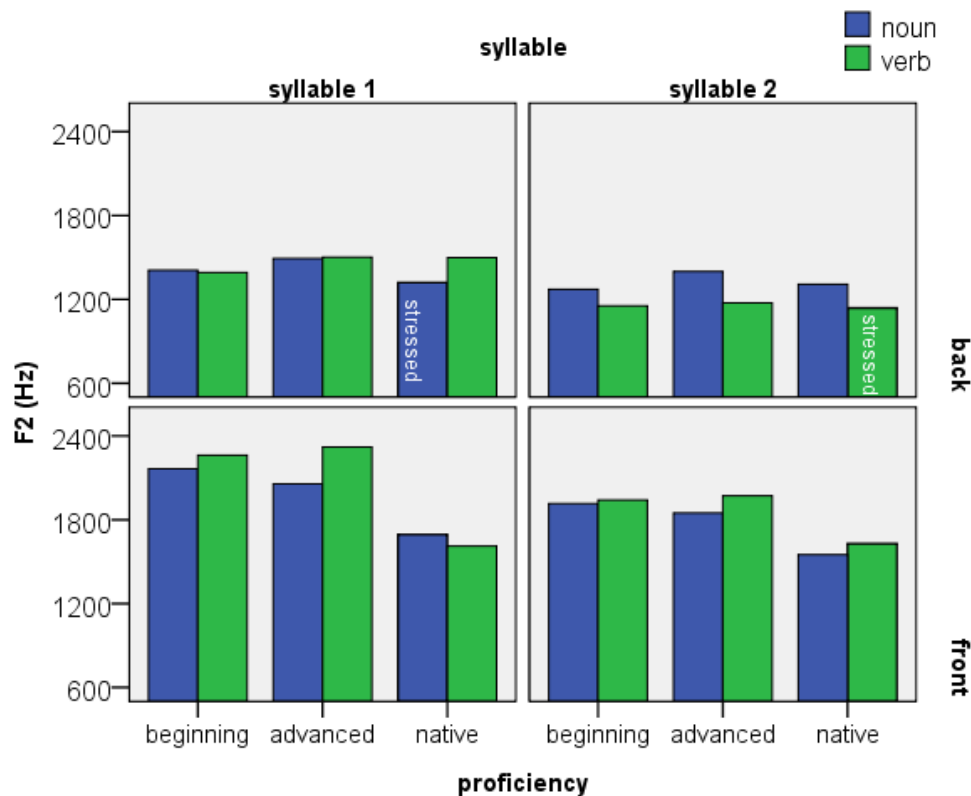


Figure 2. 20. Second formant across syllable positions, vowel types, and speaker groups.

As shown in Figures 2.18-2.19, native speakers centralized the unstressed syllable regardless of both word category and vowels backness. Learners patterned similarly to native speakers in syllable 2. However, in syllable 1 both advanced and basic learners exhibit higher F2 for unstressed position - the opposite of the native speakers' pattern.

2.2.3 Summary

Mean F0, max F0, intensity, and duration

In this acoustic study, disyllabic noun-verb word pairs (e.g. *contract* and *contract*) with similar segmental composition but contrasting in stress location were recorded by Mandarin learners of English with different proficiencies. Mean F0, max F0,

intensity, duration, and F2 of stressed and unstressed vowels were measured and compared to the production of native English speakers. Two types of dependent variables were computed from the acoustic measurements for the statistical analysis (cf. section 2.2.1). The first-to-second vowel ratios were used to evaluate if the speakers utilized the target correlates (mean F0, max F0, intensity, and duration) to signal the difference between nouns (trochaic pattern) and verbs (iambic pattern). On the other hand, stressed-to-unstressed vowel ratios were adopted to compare the magnitude of the acoustic correlates used to signal stressed vowels.

Results from first-to-second vowel ratios showed that, in general, all the target correlates were utilized to distinguish the stress contrast between nouns and verbs across the three proficiencies - the ratios are all significantly greater than 1.0 in noun readings (Figure 2.21), and mostly smaller than 1.0 in verb readings (Figure 2.22).

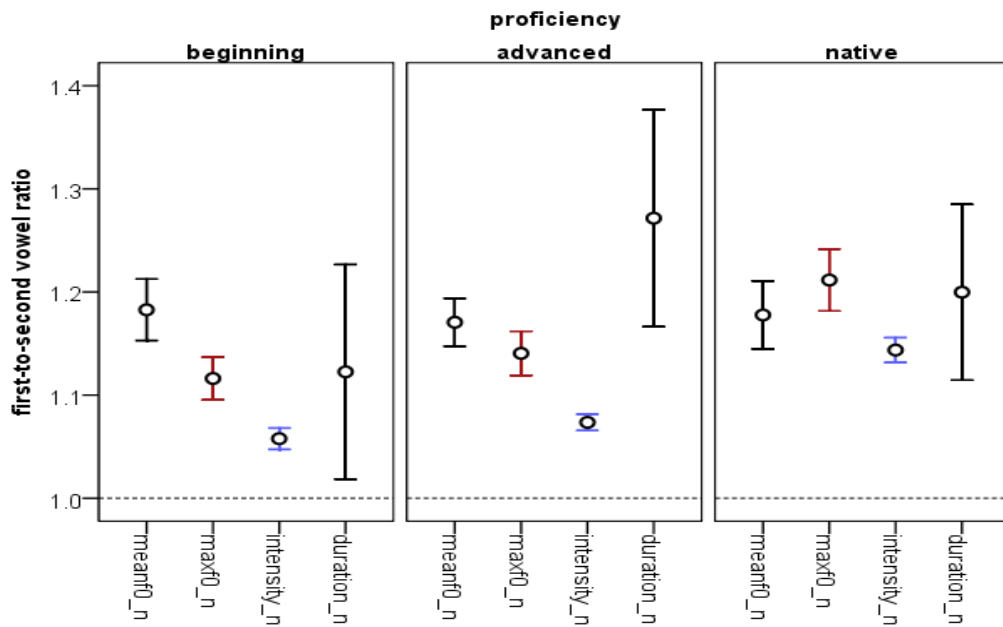


Figure 2. 21. 95% confidence interval of first-to-second vowel ratios for mean F0, max F0, intensity, and duration for nouns. The reference line (1.0) is the assumed baseline when a given correlate is identical in both syllables.

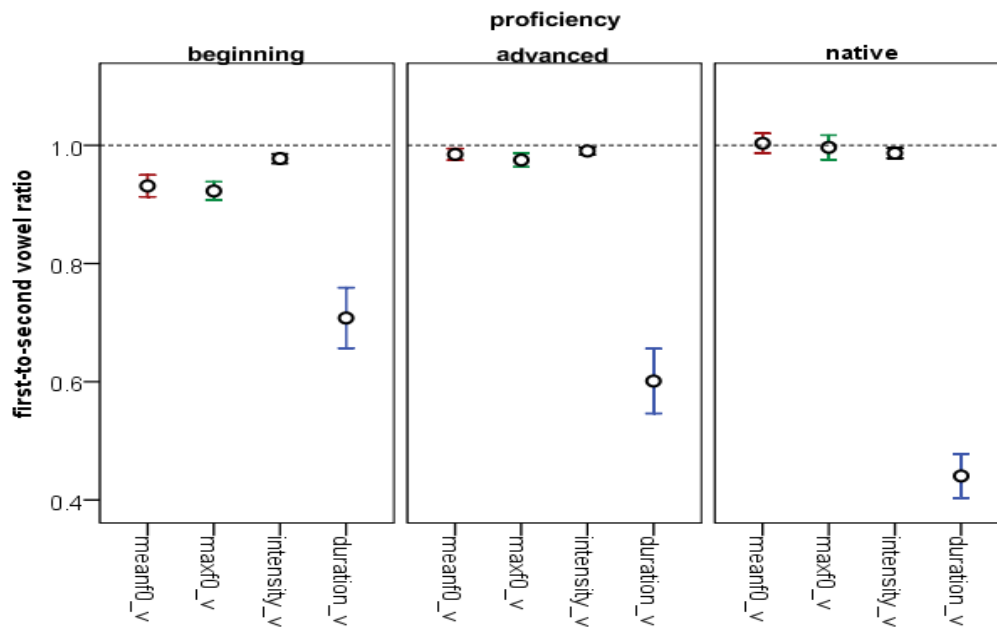


Figure 2. 22. 95% confidence interval of first-to-second vowel ratios for mean F0, max F0, intensity, and duration for verbs. The reference line (1.0) is the assumed baseline when a given correlate is identical in both syllables.

The magnitude of difference in the use of these correlates was evaluated through the statistical analysis, using stressed-to-unstressed vowel ratios as the dependent variable. This ratio was computed from vowel 1 over vowel 2 (V1/V2) in nouns, while in verbs it was from vowel 2 over vowel 1 (V2/V1). The results show that the implementation of these correlates was extremely different in noun and verb readings in native speakers. They utilized mean F0, max F0, intensity, and duration to signal stressed syllables in noun readings, but used only duration as the major cue for stressed syllables in verb readings. Mandarin L2 learners on the other hand showed a more consistent use of both correlates in noun and verb readings in two ways: the four correlates are all adopted in both conditions, and the magnitudes (differences between stressed and unstressed syllables) used were consistent across noun and verbs.

Formant analysis

The unstressed vowels in syllable 1 position (i.e. verb readings) as well as unstressed vowels in syllable 2 position (i.e. noun readings) were reduced/centralized in native speakers' production. This pattern was observed in both front and back vowels; however the degree of centralization (magnitude of reducing) is always greater in back vowels than in front vowels.

Unstressed vowel reduction is not as clear in the learner groups. The unstressed vowels in syllable 1 position (verb readings) were not reduced in either front or back vowels. However, when located in the second syllable, both beginning and advanced learners performed vowel reduction in both front and back vowels. The degree of centralization is again greater in back vowels and relatively smaller in front vowels. The magnitude of centralization is the smallest for beginning learners.

2.3 Acoustic study II - F0 contours in stressed syllables

The findings from the first acoustic study established that learners are less flexible in the use of F0 when compared to native English speakers. This effect may be triggered by the phonemic role of F0 in the learners' tonal L1 (Mandarin). To further examine the effect Mandarin L1 has on implementing F0 in stressed syllables in English, the second acoustic study was conducted using target syllables which are legal in Mandarin. The hypothesis is that the learners are less flexible in their realization of F0 contours, and will apply more fixed contours in stressed syllables. Based on the results in the first word pair experiment, in which Mandarin L2 learners of English produced higher F0 in stressed syllables, I hypothesize that learners will adopt the pattern of either of the high tones in Mandarin (high-level tone 1 and high-falling tone 4) to signal stress in stressed syllables.

2.3.1 Methodology

In this experiment, English syllables which obey Mandarin phonotactics - CV(N) - were chosen as stimuli to compare the F0 contours of the stressed and unstressed syllables in the production of Mandarin L2 learners of English. In addition, while stimuli in the first acoustic study provide a nice mapping of near-exact segmental composition of the minimal pairs, some of the target words may not be familiar to the second language learners. The present experiment included familiar words pairs, consisting of a monosyllabic word (e.g., key or knee), and a disyllabic word in which the monosyllabic word is embedded in unstressed position (e.g., monkey or money).

The target words are always located at the end of the sentence. All three types of syllables in the word pair were analyzed – the target syllable in stressed (e.g., 'key in 'key') and unstressed position (e.g., -key in 'monkey'), as well as the non-target syllable in a disyllabic word (e.g., mon- in 'monkey'). A complete list can be found in Appendix A.

The same speakers that were analyzed in the first experiment were also used in this experiment. The recording procedure is identical to the procedure in the first acoustic study. The participants read the entire sentence aloud and repeated the final target word. The F0 of rimes for stressed and unstressed syllables was measured at every 10% of the contours, yielding 10 measurements for each rime. The contours were then plotted using SPSS for further comparisons.

2.3.2 Results

Given that gender was not balanced in the three groups (more females in the advanced group, more males in the native group), the contours for males and females were plotted separately for this comparison. The F0 contours for the stressed and unstressed syllables are compared across three proficiencies. The F0 contour of individual speaker across all syllable types can be found in Appendix F.

Stressed syllables

First, monosyllables in stressed position (e.g. *key*) were examined. Results for the F0 contours are plotted in Figure 2.31 (top – females; bottom – males). The results from the females showed that both beginning and advanced speakers' productions showed high falling tonal contours with great F0 differences between the onset and offset (beginning learners particularly). In addition, their contours showed a rise/plateau at the onset, which resembles the profile of the Mandarin high falling tone. Native speakers, however, produced a relatively level contour. A similar pattern was also found in male data - both learner groups showed a greater F0 drop between the onset and offset, while native speakers produced more flattened contours.

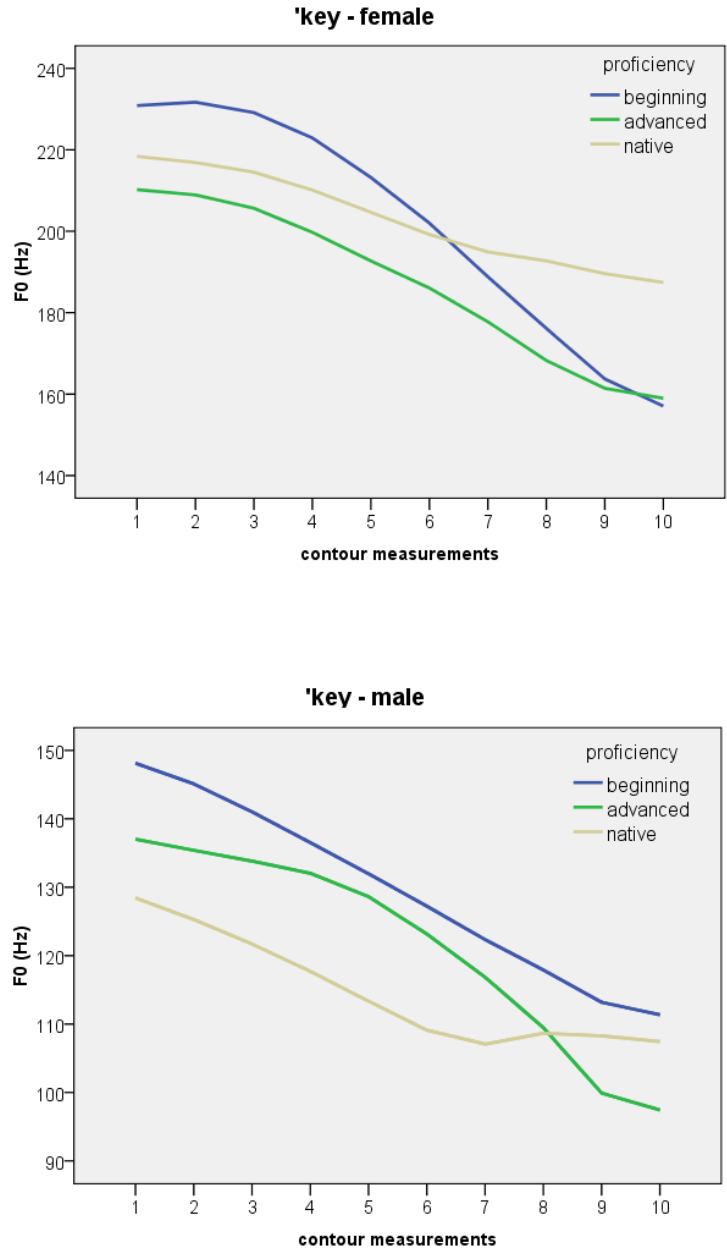


Figure 2. 23. F0 contours for 'key in stressed position across three proficiencies. Top – females. Bottom – males.

Unstressed syllables

The F0 contours on unstressed syllable (e.g. *-key* as in *monkey*) are plotted in Figure 2.24. The contours are more similar across proficiencies than in stressed syllables, and the contours all showed a gradual falling trend.

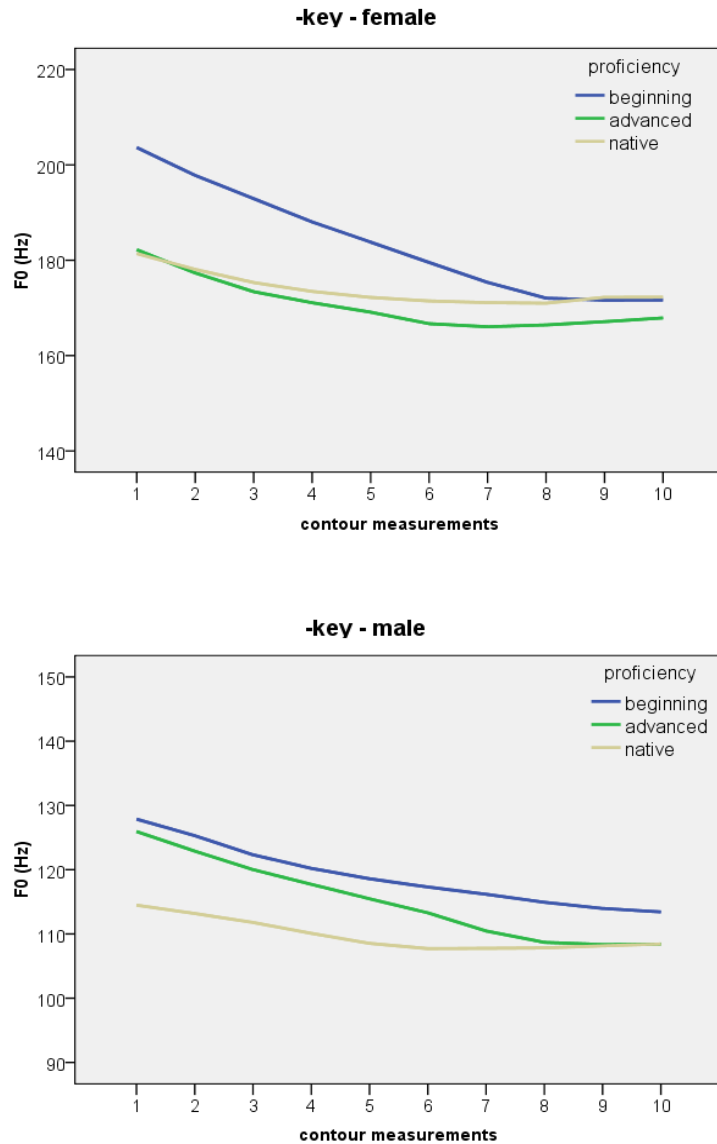


Figure 2. 24. F0 contours for *-key* in unstressed position across three proficiencies. Top – females. Bottom – males.

As the contours for *key* in the stressed position were found to resemble the high-falling tone 4 in Mandarin, stressed syllables in the disyllabic words in the stimuli list

(e.g. *mon* in monkey) were examined. F0 contours of the stressed syllables in non-final position are plotted in Figure 2.33. Results from both genders showed that learners at both proficiency levels utilize high level F0 contours to signal stressed syllables (females show a slight rising profile). The contours from the native speakers are relatively flat with a gradual fall.

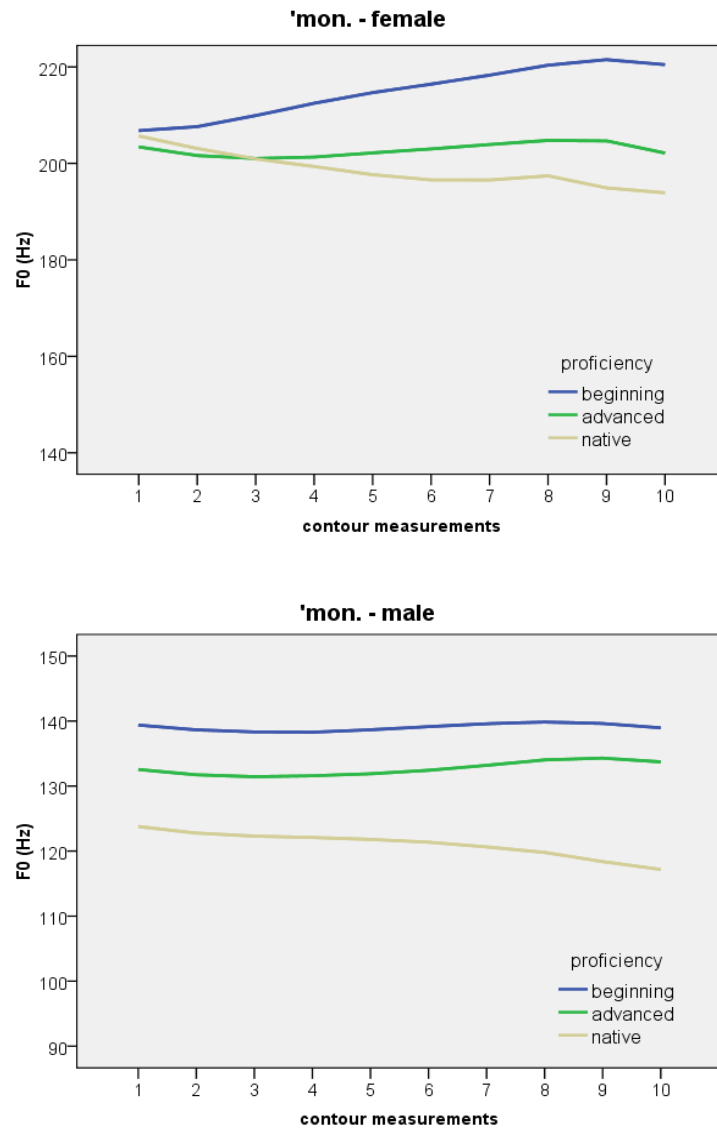


Figure 2. 25. F0 contours for *mon* (as in monkey) in stressed position across three proficiencies. Top – females. Bottom – males.

As shown in the previous section, the beginning and advanced learners' contours for stressed final syllables showed a high falling tonal contour, while the stressed non-final syllable carries a high level contour. The patterns are quite consistent across different speakers. The contour patterns of the native speakers are, on the other hand, more diverse. The F0 contours from the syllables examined are plotted for each proficiency separately in Figures 2.26-28.

Native speakers

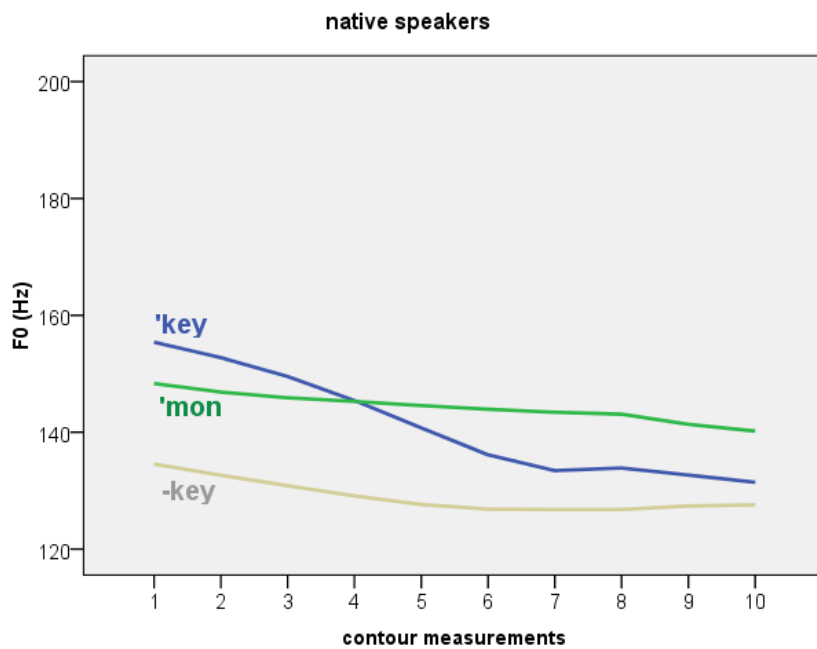


Figure 2. 26. F0 contours of three types of syllables from native speakers.

Advanced learners

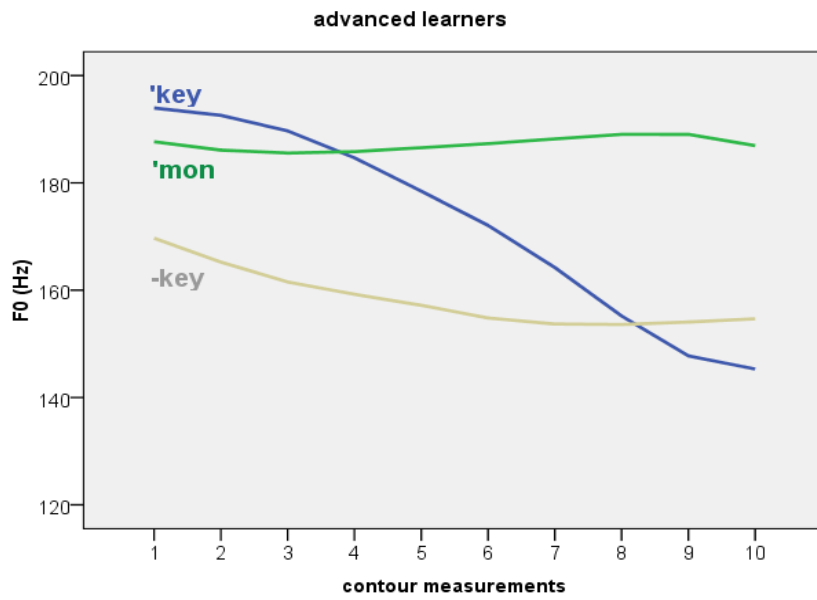


Figure 2. 27. The merged F0 contours of three types of syllables from advanced learners.

Beginning learners

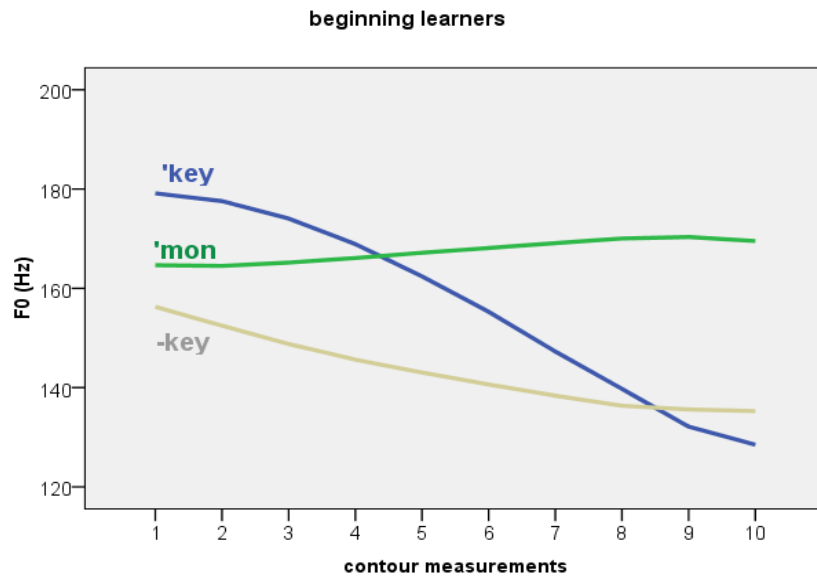


Figure 2. 28. The merged F0 contours (across male and female) of three types of syllables from beginning learners.

In sum, the beginning and advanced learners' contours for stressed final syllables showed a high falling tonal contour, while the stressed non-final syllable carries a high level contour. The patterns are quite consistent across different speakers. The contour patterns of native speakers are much more diverse (cf. Appendix F).

2.3.3 Summary

This experiment examined Mandarin L2 learners producing English stressed syllables which match Mandarin phonotactics, in order to examine the F0 contours in learners' productions. Although the stimuli in the previous acoustic study provide well-controlled segmental conditions for the investigation of mean F0, max F0, intensity, duration, and F2, most of the syllables in those stimuli are illegal according to Mandarin phonotactics. In order to better evaluate the realization of F0 contours in speakers' productions, stimuli which comply with Mandarin syllable structures were utilized.

In this experiment, the same participants who participated in the first acoustic study recorded four word pairs, such as key-monkey, D-daddy, knee-money, and bee-baby. These word pairs target the same syllable, in stressed position as a monosyllabic word, and in unstressed position in a disyllabic word. The F0 contour patterns for the two types of syllables in these two contexts were examined. The results indicate that F0 in unstressed syllables is lower when compared to stressed syllables. Interestingly, the F0 contours for stressed syllables in both beginning and advanced learners' productions resemble that of the high-falling tone 4 in Mandarin. Further examination of stressed syllables in the disyllabic words (e.g. mon-, da-, mon-, ba-) also showed that their F0 contours are similar to the high-level tone 1 in Mandarin. F0 contour results from native English speakers were, however, rather more diverse.

Mandarin learners of English, when utilizing higher F0 to signal stressed syllables, use the F0 patterns they are most familiar with in their first language. The two tones

in the high register region (high-level tone 1 and high-falling tone 4) fulfill the high F0 requirement for the purpose of signaling stressed syllables, and are thus appropriated by the learners.

These results are in line with the findings from the first acoustic study, which suggested that having F0 as a phonemic feature in their first language causes learners to demonstrate a more fixed/inflexible use of F0 as a stress cue. In the first acoustic study, learners demonstrated an inflexible use of F0 across stress locations while native speakers incorporated other prosodic cues, and adjusted their use of all the correlates.

2.4 General discussion

2.4.1 Overall summary

Experiment 1

In the first acoustic experiment, disyllabic word pairs (e.g. *contract* and *contract*) with similar segmental composition but contrastive stress location were recorded by Mandarin learners of English with different proficiencies. Mean F0, max F0, intensity, duration, and F2 of stressed and unstressed vowels were measured and compared to those of native English speakers. Learners use the mean F0, max F0, intensity, and duration similarly to native speakers in noun readings. When producing stressed vowels in verb readings (2nd syllable stressed) however, native speakers encoded only duration to signal stress, while beginning and advanced learners utilized all of the correlates. Mandarin L2 learners demonstrated a more consistent use of these correlates in noun and verb readings in two specific ways: the four correlates were all adopted in both conditions, and the magnitudes (differences between stressed and unstressed syllables) used were also consistent across conditions. Results from formant analysis also indicate that the learners' vowel reduction patterns are much more similar to native English speakers in noun readings than in verb readings. This difference in the relative importance of acoustic correlates in different stress locations seems to have been overlooked in previous studies.

Experiment 2

The second acoustic experiment focused on the F0 contour patterns in English stressed and unstressed vowels produced by Mandarin L2 learners. The results revealed that learners at both proficiencies adopt a high-level F0 contour (when stress is located in non-final vowels) and a high-falling contour (when stress is located in the final syllable) to signal stressed syllables, while native speakers exhibit more diverse contour patterns. This effect may be triggered by the phonemic role of F0 in Mandarin.

2.4.2 Comparison among speaker groups

As stated in 2.4.1, the results from the minimal word pair experiment indicate that the acoustic correlates were implemented differently across speaker groups. Native speakers utilize mean F0, max F0, intensity, and duration to signal stressed syllables in noun readings, but used only duration as the major cue for stressed syllables in verb readings, while learners consistently used all correlates in both noun and verb readings. This divergence is also supported by the formant analysis data: vowel reduction patterns in the learner groups are more similar to native speakers when the unstressed vowel is located in the second syllable (noun readings) than when the unstressed vowel is located in the first syllable (verb readings).

Closer examination of this divergence shows that in verb readings, native speakers reduce the duration of the unstressed vowel markedly, generating a greater duration difference between the stressed and unstressed vowels. Mandarin L2 learners on the other hand use more consistent (fixed) differences between stressed and unstressed vowels to signal stress in both nouns and verbs. F0 contour patterns are also more fixed in the Mandarin L2 learners' production. I propose that this divergence may originate from the learners' inflexible use of F0 cues in Mandarin, i.e., the tone patterns are fixed in Mandarin; each syllable can only have one of the four tone patterns. As a result, Mandarin speakers are better at controlling features in producing F0. In addition, the phonemic status of F0 triggers a more consistent use of the cue. Consequently, Mandarin L2 learners are hypothesized to adopt a more fixed magnitude of F0 difference to signal the stressed vowel. Unlike the magnitude used by native English speakers which varies according to other prosodic cues.

Extrinsic factors at play

The statistical analysis of stressed-to-unstressed vowel ratios revealed another asymmetry. The magnitude of mean F0, max F0, and intensity was found to be greater in nouns than verbs. When averaged across all speaker groups, the ratios for

nouns and verbs respectively are mean F0 (1.18, 1.03), max F0 (1.16, 1.05), and intensity (1.1, 1.02). On the contrary, duration is used to a greater extent in verb readings (2.17) than noun readings (1.23). I believe that magnitudes of mean F0, max F0, and intensity for stressed syllables in nouns are reduced by down-stepping, while duration cues for stressed syllables in verbs are augmented by final lengthening. Down-stepping is known as the general trend of energy declination (which usually involves pitch and intensity) in natural speech (Ladefoged, 1982). Final lengthening refers to the elongation of the domain-final unit in utterances.

Down-stepping and mean F0, max F0, and intensity

Figure 2.29 below demonstrates the hypothetical interaction between the down-stepping effect and F0 and intensity. Assuming speakers adopt the same F0/intensity magnitudes to signal stressed vowels, the absolute difference between P1 and P2 should be identical in noun and verb readings. As shown in section (a), the F0/intensity peak of the first vowel (henceforth P1) is greater than P2 in nouns, while P2 is higher in verbs. The dotted lines in section (b) depict the end result after down-stepping. As a result of down-stepping, the F0/intensity differences are enhanced in noun readings but diminished in verb readings. This is in line with the acoustic data.

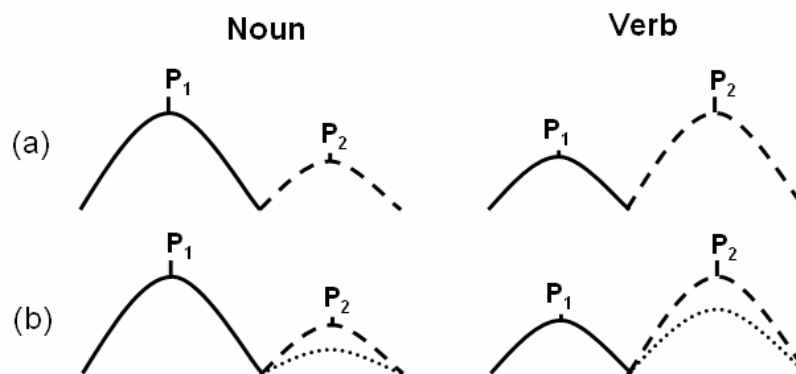


Figure 2. 29. Hypothetical scheme of how down-stepping interacts with F0/intensity in noun and verb readings.

The realization of mean F0, max F0, and intensity across three proficiencies was also further examined. The stressed-to-unstressed vowel ratios ($V1/V2$ for nouns and $V2/V1$ for verbs) for these correlates are plotted again in Figures 2.30-32. As shown in the figures, all speaker groups have greater ratios for nouns than for verbs, i.e. the difference in magnitude is greater in nouns than verbs. However, native English speakers (golden line) exhibit the greatest ratio differences between nouns and verbs (demonstrated by the greatest slope among the three proficiencies). Beginning learners, on the other hand, have closer stressed-to-unstressed vowel ratios between nouns and verbs. The ratio differences between nouns and verbs from the advanced learners are consistently in between those of native speakers and beginning learners.

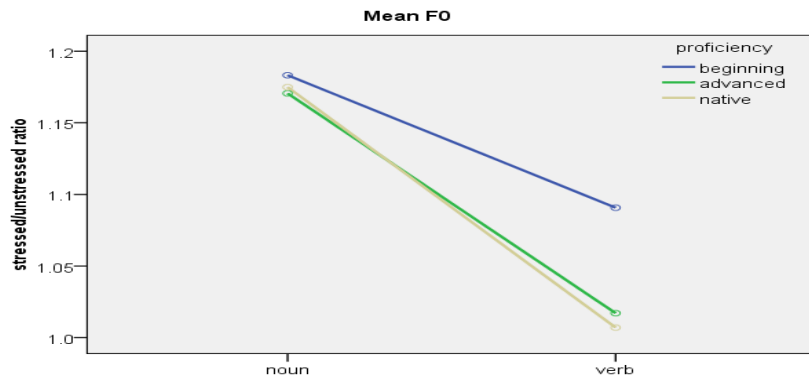


Figure 2. 30. The interaction between stress location and proficiency.

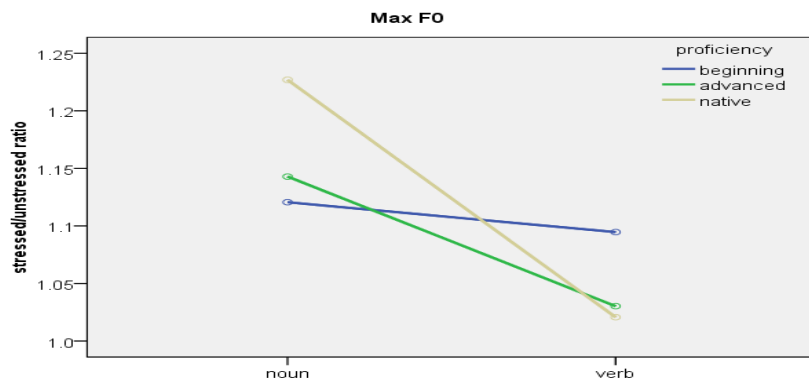


Figure 2. 31. The interaction between stress location and proficiency level for max F0.

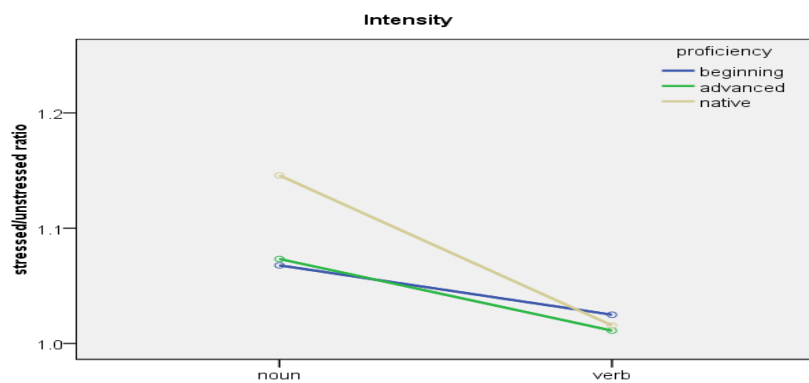


Figure 2. 32. The interaction between stress location and proficiency level for intensity.

Final lengthening and duration

I propose that final lengthening can account for the greater use of duration in verb readings. The effect of final lengthening is simulated in Figure 2.24. As shown in section (a), the duration of the first vowel (henceforth V1) is longer than V2 in nouns, while V2 is longer than V1 in verbs. Again, assuming the speakers adopt the same duration difference to distinguish stressed and unstressed vowels, the absolute difference between V1 and V2 should be identical in noun and verb readings.

The dotted lines in section (b) were added to the end of V2 to represent the final lengthening effect. It is shown that final lengthening reduces the duration difference between V1 and V2 in noun readings (Figure 2.33b), i.e. final lengthening increase the duration V2 hence reduced the duration differences V1 and V2. In verb readings, however, it increases the duration difference between V1 and V2. Consequently, final lengthening would predict that the duration differences between V1 and V2 in noun readings will be smaller when compared to those of verb readings, which is consistent with the acoustic results- when comparing stressed-to-unstressed vowel ratios, verb readings have significant greater duration ratios than noun readings (cf. ANOVA results in Table 2.5).

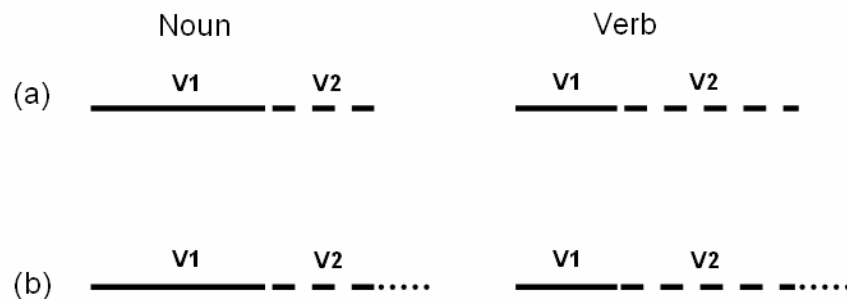


Figure 2. 33. Hypothesized final lengthening interaction with duration cues in nouns and verbs.

I further examine the interaction of duration cue with final lengthening in all three proficiencies. As shown in Figure 2.34, all speaker groups have greater stressed-to-unstressed vowel ratios for verbs than for nouns, i.e. the difference in magnitude is greater in verbs than in nouns. This confirmed our hypothesis that final lengthening enhanced the duration cue in general. However, compared to the learners, native English speakers use a significantly greater duration ratio for verb readings (as indicated by the steep slope of the golden line for native speakers and flatter lines for beginning and advanced learners).

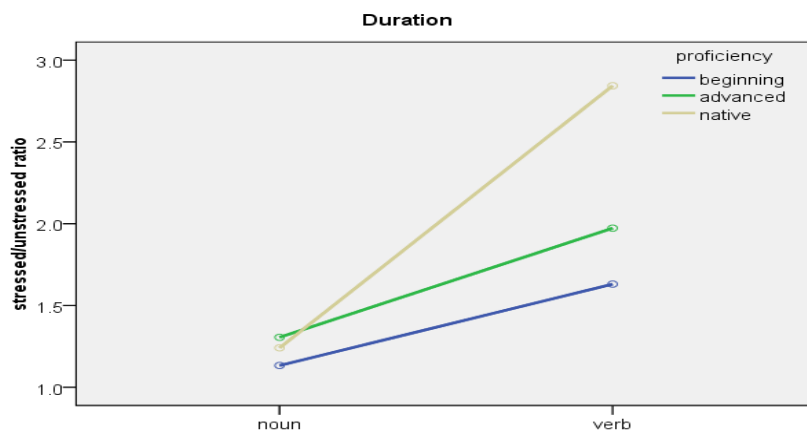


Figure 2. 34. The interaction between stress location and proficiency levels for duration.

Based on the discussion above, I propose that native English speakers' realization of F0 cues for word stress seems to interact with extrinsic prosodic factors, while the learners, beginning learners particularly, show a consistent use of F0 across different stress conditions. I argue that this is derived from the contrastive characteristic of F0 in the learners' L1, Mandarin. It appears that when the feature is phonemic in the L1, the speakers have more consistent control of utilizing the feature during production in a second language. I will discuss these results again in the general discussion (section 2.4).

Other factors

The fact that Mandarin has a similar prosodic template to the stress readings for nouns in English may contribute to the observed pattern between native English speakers and Mandarin L2 learners of English. Apart from the four lexical tones in Mandarin, there is also a neutral tone, which occurs mostly in utterance/phrase-final positions. Syllables carrying this tone are either grammatical particles or reduplicated syllables. Phonetically speaking, these syllables have been found to be lower in F0, or resemble the underlying tonal contours. The domain-final feature of neutral tone familiarizes Mandarin speakers with the semi-trochaic stress pattern in English, and further facilitates their acquisition of English trochaic words. The iambic stress pattern, however, has no similar prosodic structure existing in Mandarin.

Results from previous studies have shown that Mandarin speakers learning English as a second language have issues in the temporal domain. More specifically, the syllable durations in their production are not as varied as native speakers, which originate from their lack of reducing the unstressed vowels (Tseng, 2001). Furthermore, it has been reported that the sentential stress in Mandarin is manifested mostly by F0 followed by intensity. Duration was not found to be an important correlate for stress in Mandarin (Coster and Kratochvil, 1984).

The present study has established that stressed vowel realization in native English speakers is altered in different prosodic positions. I believe that this is due to the native speakers' flexibility in incorporating other extrinsic factors, such as down-stepping and final lengthening, in the implementation of stress. Mandarin L2 learners show a much more fixed utilization (fixed difference magnitude in nouns and verbs) of F0. I believe that this is due to the phonemic feature of F0 in Mandarin.

2.4.3 Comparison with previous studies

In previous research on the phonetic realization of stress at the word level by second language learners, researchers often merged the data from noun and verb readings, and overlooked the differences in initial and final stress. These differences have been revealed by the results of the current study.

These findings are supported by research at the sentential level - the phonetic realizations of a target word alter when located in different accentual positions in a sentence. McGory (1998) investigated F0 and duration realization in word pairs contrasting in the location of stress (e.g. *memorize-memorial*) in the production of native English speakers, Mandarin L2 learners of English, and Korean L2 learners of English. The target words were produced in statements and questions, as well as different sentence focus locations (nuclear, prenuclear, and postnuclear unaccented). Results indicated that native English speakers produce different F0 pitch patterns depending on the sentence type and accentual position. Both Mandarin and Korean L2 learners, however, produce the same F0 pitch patterns across sentence types and accentual positions.

Zhang et al. (2008) compared the realization of mean F0, max F0 location, intensity, duration, and spectral composition in word pairs contrasting in stress locations (e.g. *contract-contract*) between 10 Mandarin speakers and 10 native English speakers. Their results showed that the learners and the native speakers behave similarly in their use of duration and intensity, but that learners produce stressed syllables with a higher F0 than native speakers. A major issue arises from their statistical analysis: they used absolute measurements for each correlate as dependent variables. This makes the results more susceptible to differences in speaking rate and intrinsic F0 across speakers. The results of greater F0 from Mandarin learners' production could be due to the smaller laryngeal structure of Asian speakers (similar results were found

in Yang, 1996 and Chen et al., 2001). Secondly, the difference between nouns and verbs were overlooked.

Based on the differences in acoustic realization in this study, as well as the contextual effect found in McGory (1998), I believe that native speakers adjust their implementation of stress to contextual factors- location of the stress and extrinsic intonation patterns. To evaluate the degree of resemblance between performance by second language learners and native speakers, more contextual factors need to be incorporated in the experimental design. Most importantly, these measures need to be analyzed separately in order to evaluate the performance of second language learners. A design with multiple sentence types and accentual positions would allow for a more controlled comparison across different contexts.

In addition, for researchers who are interested in the implementation of stress at the phonetic level, care should be taken to ensure the speakers are indeed placing the stress in the correct location. Several methods have been utilized to accomplish this: informing speakers about the stressed syllables (extended practice in Lee et al. 2006, highlighting stressed syllables in the current study); adopting native listener ratings on correct stress placement, and eliminating those tokens rated as having the wrong stress placement (Zuraig, 2005). The results from this rating are greatly affected by the listeners' perceptual strategies, and may result in accidentally eliminating tokens which speakers produce with correct stress placement but with weaker perceptual cues for the listeners.

The results from the two acoustic studies described in this chapter provide a good grounding for selecting the cues which will be resynthesized as the stimuli for the following perception study. F0 (especially Max F0) and duration are selected due to the divergent use of these two correlates by different speaker groups. Intensity, although also adopted in all speaker groups, seems to correlate with max F0 in noun

readings, and is not significantly different between speaker groups in verb readings. Therefore, intensity will not be included as a cue. Finally, given the differences in the realization of vowel quality in stressed and unstressed vowels between speaker groups, vowel quality will also be included as one of the cues to be tested in the perception study.

CHAPTER 3: PERCEPTION STUDY

This chapter describes a perception experiment designed to explore the relative importance of max F0, duration, and vowel quality in stress perception by Mandarin L2 learners of English. These particular target cues were chosen based on their diverse roles in stress realization, and were determined by the previous two acoustic studies. Max F0 is realized differently by native English speakers and learners in both nouns and verbs, and the role F0 cues play in stress perception by Mandarin L2 learners of English, due to its phonemic characteristics in Mandarin, is of particular interest. Duration plays a primary role in native English verb production, but not in the production of beginning learners, and is therefore also of interest.

Unstressed vowel reduction has been shown to be difficult for second language learners. Vowel quality was chosen as the final cue, therefore, in order to determine if it affects the perception of stress for second language listeners as it does for native speakers of English. It is predicted that full vowels will attract stress localization responses in native English listeners, but not in beginning learners. Advanced learners may be quite sensitive to vowel quality, and may identify full vowels as stressed vowels more than learners.

Max F0 was chosen instead of mean F0, because mean F0 does not differ across speaker groups in noun readings (cf Figure 2.6). In addition, mean F0 is the mean of all F0 measurements within the vowel, and characteristics which might show up through F0 analysis could diminish or disappear in averaging. The implementation of intensity was found to diverge between speaker groups. However, due to its high correlation with max F0 (similar trends in stress implementation), it is not considered in the current study.

Beginning learners are predicted to be the most sensitive to max F0 cues, and the least sensitive to duration cues, while native listeners are expected to be sensitive to both

duration, and to a somewhat lesser extent, max F0. Advanced learners are predicted to pattern somewhere in between these two groups.

3.1. Methodology

3.1.1 Stimulus construction

The stimuli were resynthesized versions of the naturally produced, non-word disyllable “dada.” The non-word “dada” was chosen over real words to help ensure that responses were not biased by the frequency of occurrence of a frequent noun or verb form in a minimal pair. Two types of stimuli were synthesized: the first type is ‘dada’ with full vowels in both syllables, and the second type has a reduced vowel in the first syllable and a full vowel in the second syllable. These two types of stimuli are designed to investigate the listeners’ sensitivity to the reduced vowel quality of the first syllable. The resynthesis procedure for each type is described below.

Vowel quality

The source stimuli were first recorded in the anechoic chamber at the University of Kansas by a female native English speaker producing [ˈdadu] and [daˈdu]. The recording was conducted with a Marantz PMD671 solid state recorder and Electro-Voice RE20 microphone, with a sampling rate of 22.05 kHz. The recording was then transferred to Praat (v. 4.3.04, © P. Boersma and D. Weenink) for PSOLA resynthesis.

The stressed and unstressed [da] were then extracted to become the bases for further stimulus construction. To ensure that the reduced [da] (henceforth [də]) has the same acoustic attributes as the [da] syllable except spectral composition, the duration, F0 tier, and intensity tier of [ə] were rescaled to match those of the full vowel [a]. This procedure is shown in the flow chart in Figure 3.1. Numbers were attached to the tokens to encode the steps of resynthesis. [da0] and [də0] refer to the unprocessed tokens extracted from the recording.

The vowel durations of [da0] and [də0] were measured first. As expected, the duration of the reduced vowel in [də0] was found to be shorter (144 ms) than the full vowel in [da0] (197ms). Therefore, the duration of the reduced vowel in [də0] was lengthened to 197 ms.

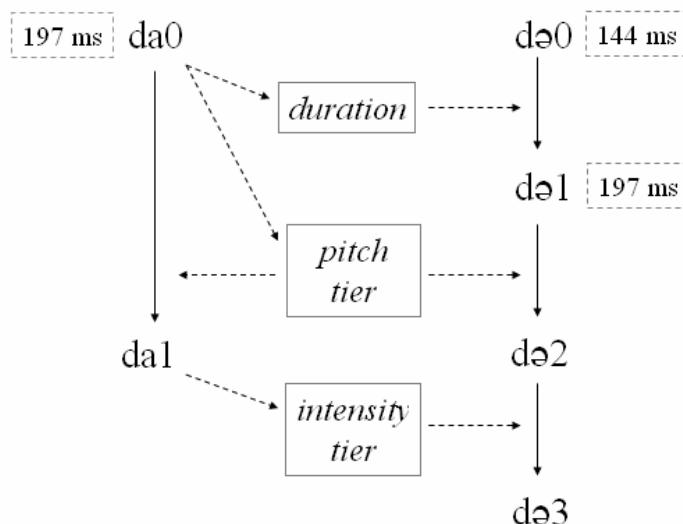


Figure 3. 1. Procedure of duration, pitch tier, and intensity tier rescaling on [da0] and [də0] syllables to generate the bases [da1] and [də3] for further stimulus manipulation.

Results from the second acoustic study on F0 contour realization in stressed syllables indicated that Mandarin L2 learners of English use different F0 contour patterns than native English speakers. It is reasonable to hypothesize that F0 contours may affect the stress perception of Mandarin L2 learners as well. Although the current study does not include F0 contour patterns as one of the testing correlates, care was taken to create target tokens with identical F0 contours, as described below.

The pitch points for onset (216 Hz) and offset (193 Hz) of the pitch tier of [da0] were first measured, and then all the pitch points on the pitch tier of [da0] were removed. A new pitch tier was generated by resetting the onset and offset pitch points to 216 Hz and 193 Hz respectively, and interpolating a new tier between the onset and offset pitch points. This pitch tier was then combined with the original sound and to

generate [da1]. The same procedure was applied to [də1] to generate [də2], which has the same F0 contour as [da1]. Intensity rescaling was done by replacing the intensity tier of [də2] with that of [da1].

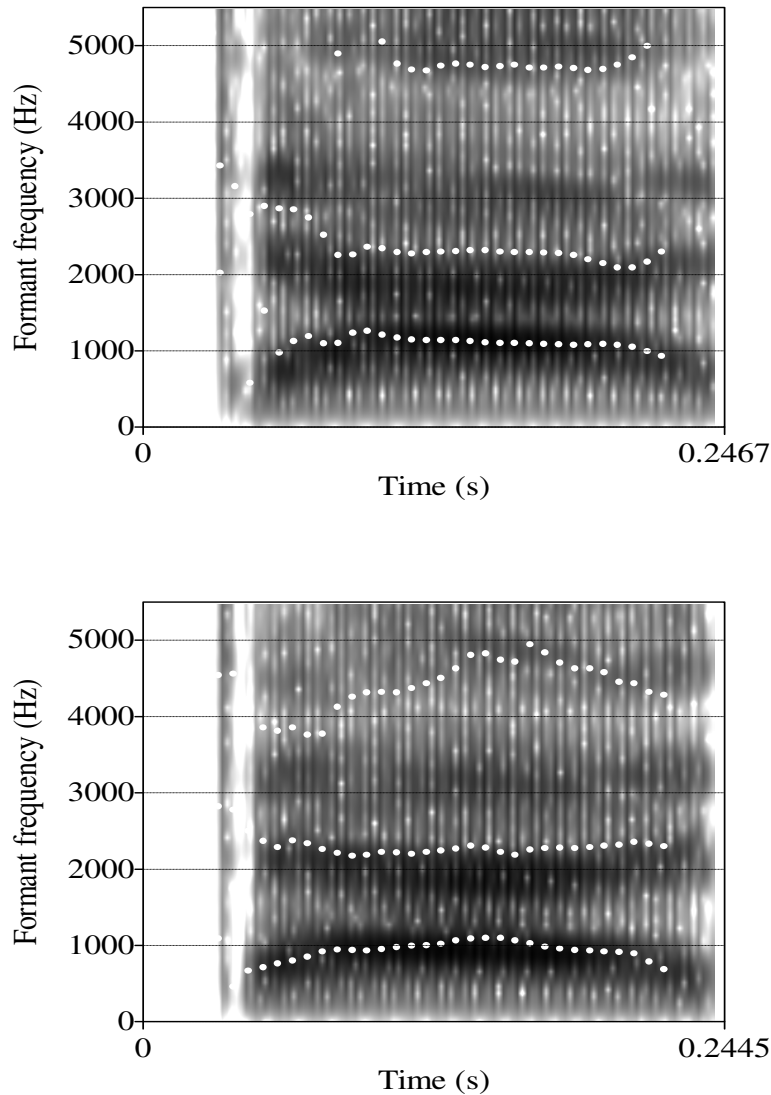


Figure 3. 2. The spectrograms and formants for the full vowel [da1] (top) and the reduced vowel [də3] (bottom).

Two disyllabic non-words were generated for the contrast of spectral composition. The first set of stimuli has full vowels in both syllables. This was accomplished by

appending [da1] after another [da1], generating the word “dada.” The interval between the offset of the first vowel and the release of closure in the second syllable was adjusted to be 60 ms, to simulate the closure of the second [d] sound in natural production. Similarly, [də3] was appended with a [da1], creating the second non-word [dəda], which has a reduced first vowel and a full second vowel.

Max F0 and duration resynthesis

The resynthesis values for max F0 and duration were determined based on the measurements obtained from the natural production data in the word pair acoustic experiment. The greatest absolute ratios for max F0 and duration across noun and verb readings in all speaker groups were calculated first. This ratio, as well as its reciprocal, became the end points for the resynthesis. For max F0, the greatest ratio is from the noun reading of native speakers, at 1.23. This ratio and its reciprocal, .813, became the end points of resynthesis. For duration, the greatest ratio is from the native speakers’ verb readings, at .45. This ratio and its reciprocal, 2.22, serve as the end points for duration resynthesis. After the endpoints were obtained, the intermediate ratios between the endpoints and the baseline ratio of 1.0 were calculated, in order to generate five target first-to-second syllable ratios for max F0 as well as duration. The resynthesis ratios for max F0 are (1.23, 1.115, 1, .897, and .813), and the ratios for duration are (2.22, 1.11, 1, .901, and .45).

In addition, the findings from the acoustic study indicated that when stressing a syllable, native speakers modify the max F0 and duration of both syllables, i.e., they not only raise the magnitude of these cues in the stressed syllable, but also lower the same cues in the unstressed syllable. In order to simulate actual production, the target ratios were achieved by modifying max F0 and duration for both syllables. The absolute values for syllables 1 and 2, as well as the corresponding ratios are shown in Figure 3.3 (max F0) and Figure 3.4 (duration).

The ratios were also assigned codes – a combination of a letter (F for max F0 ratios and D for duration ratios) and a numeral ('5' for the greatest 1st-to-2nd vowel ratio, and '1' for the smallest ratio) to facilitate further description. Ratios for max F0 are 1.23 (F5), 1.115 (F4), 1 (F3), .897 (F2), and .813 (F1). Ratios for duration are 2.22 (D5), 1.11(D4), 1 (D3), .901 (D2), and .45 (D1). Notice that, in general, greater numerals should induce greater trochee responses (1st syllable stressed, noun readings) in stress perception (e.g. F5 and D5 refer to the greatest first-to-second vowel ratios in max F0 and duration). The first syllable, then, should have a greater chance of being perceived as stressed. In addition, F3 and D3 both have ratios of 1, in which absolute measures of max F0 and duration are equal between first and second vowels respectively.

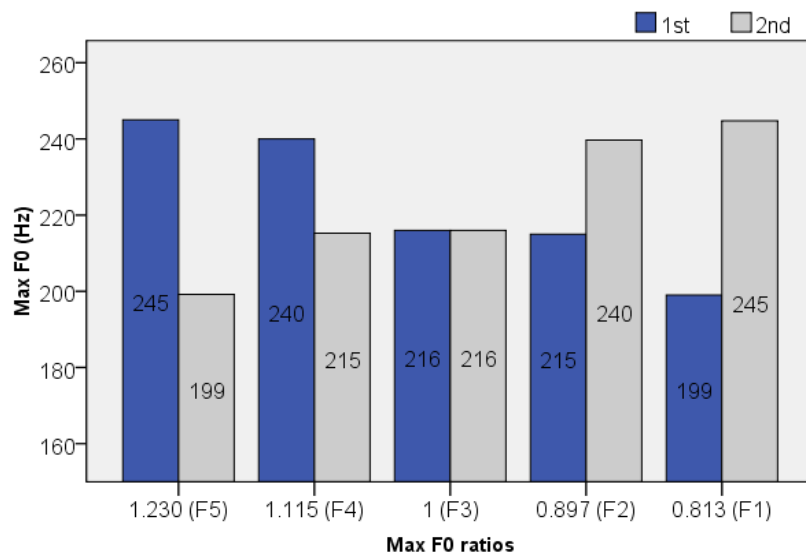


Figure 3. 3. The first-to-second syllable max F0 ratios and the absolute measures.

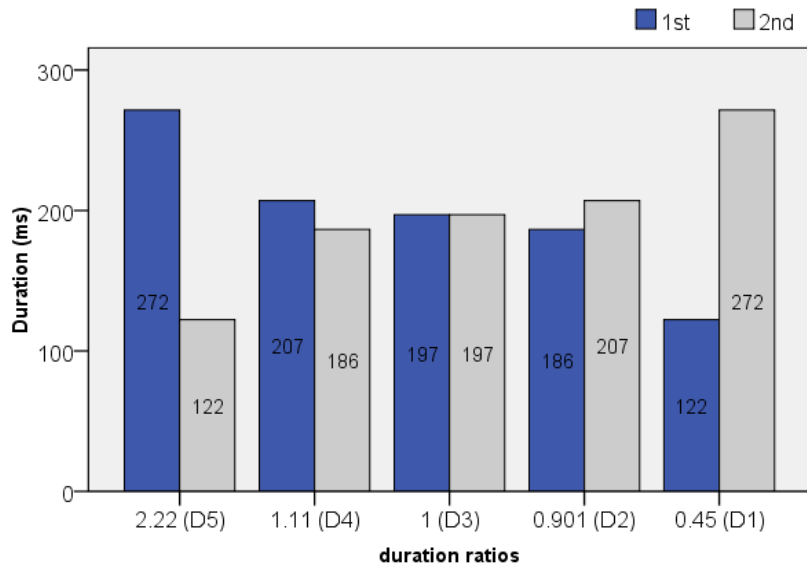


Figure 3. 4. The first-to-second syllable duration ratios and the absolute measures.

Further resynthesis used the *dada* and *dəda* disyllables (cf. vowel quality section) as the bases of duration and max F0 manipulation. Firstly, duration was resynthesized using Time-Domain Pitch-Synchronous Overlap-and Add (TD-PSOLA) manipulation method. The manipulation window of TD-PSOLA contains sound, pulses, pitch tier and duration tier. The duration ratios were implemented by adjusting the durations of syllables in the disyllabic words using duration manipulation (see Appendix B for a thorough manual). Figure 3.5 shows an example for duration tier manipulation. The adjusted duration tier was then combined with the original sound to generate a new disyllable which will have the target first-to-second vowel duration ratio.

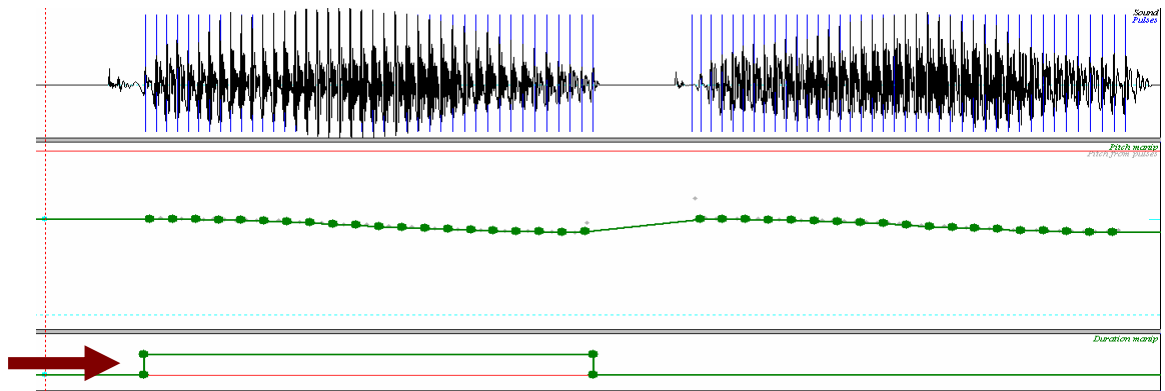


Figure 3. 5. Duration resynthesis using TD-PSOLA. From top to bottom: sound/pulses, pitch tier and duration tier.

Max F0 resynthesis was done by reassigning pitch points of the syllables. Figure 3.6 shows the procedure of the resynthesis. The old pitch points (as shown in grey dots) were first removed. The onset pitch point was assigned based on the target max F0 calculated in the previous section. Noted that max F0 in natural production by the native English speaker happen to locate at the beginning of the syllable, hence the resynthesis target max F0 point in onset position. The offset pitch point is always set to be identical as the original sound.

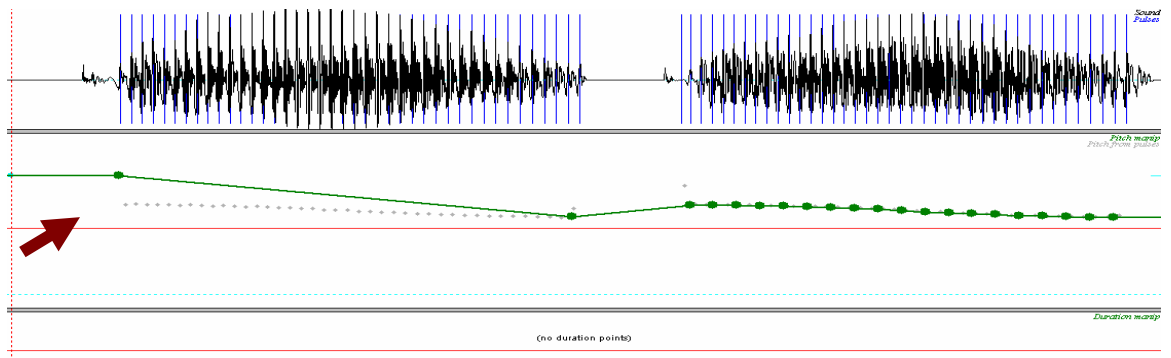


Figure 3. 6. Max F0 resynthesis using TD-PSOLA. From top to bottom: sound/pulses, pitch tier and duration tier.

3.1.2 Participants

Fifty native speakers of Mandarin who learned English as a second language were recruited from the University of Kansas, National Taiwan University, and Richmond, Virginia. They are grouped into two proficiency levels - beginning and advanced - based on a language background questionnaire (in Mandarin, see Appendix C). The questionnaire elicited biographical information including age, gender, age when the participant first started learning English, number of years of English learning, daily usage of English, immersion in an English-speaking environment, and the self-rating of their English speaking and listening proficiency. Twenty-five native English speakers from the University of Kansas participated in this experiment as the control group. They also filled out a questionnaire (Appendix C) regarding their exposure to other languages.

3.1.3 Procedure

The perceptual reliance of Mandarin L2 learners of English was tested through a stress localization experiment administered using Paradigm software (Perception Research Systems Inc.). The participants were tested in a quiet room. The instructions and prompts were presented on a computer monitor, and the stimuli were presented over headphones. Each participant listened to 150 tokens [max F0 ratios (5) × duration ratios (5) × vowel type (2) × 3 repetitions] in one block in randomized order. The listeners were instructed to first click a circle on the screen to center the mouse and to initiate stimuli. A latency of 250ms occurred between clicking the circle and the onset of the target stimulus. After listening to the token, the listeners' task was to indicate which syllable was stressed (syllable 1 or syllable 2) by clicking the corresponding text - DAda (first syllable stressed) or daDA (second syllable stressed) on the screen. There was a practice session composed of 16 trials before the experiment. The experiment took about 15-20 minutes per subject. Due to the high attention demand of the task, the participants could take breaks during the experiment by delaying clicking the circle at the beginning of each trial.

3.2 Results

3.2.1 Overall analysis

To examine the effect of all the factors included in the current study, a four-way ANOVA with vowel quality (full and reduced in the first syllable), max F0 (5 ratios), duration (5 ratios), and proficiency (3 levels), was performed. The dependent variable is the responses provided by the listeners: ‘1’- first syllable is stressed or ‘2’- second syllable is stressed. A lower response (closer to 1) refers to a more noun-like perception, while a higher response (closer to 2) refers to a more verb-like perception. The significance, F and p values of the ANOVA are tallied in Table 3.1.

	significance	F	P
Vowel	*	(1, 3600)=101.7	<.001
max F0	*	(4, 3600)=204.58	<.001
Duration	*	(4, 3600)=195.39	<.001
Proficiency	*	(2, 3600)=11.83	<.001
vowel*duration	n.s.	(4, 3600)=1.43	.221
vowel*max F0	*	(4, 3600)=2.67	.030
vowel*proficiency	n.s.	(2, 3600)=.27	.759
duration*max F0	n.s.	(16,3600)=.60	.881
duration*proficiency	*	(8, 3600)=39.47	<.001
max F0*proficiency	*	(8, 3600)=104.37	<.001
vowel*duration*max F0	n.s.	(16, 3600)=.572	.907
vowel*duration*proficiency	n.s.	(8, 3600)=1.51	.148
vowel*max F0*proficiency	n.s.	(8,3600)=1.44	.174
duration*max F0*proficiency	n.s.	(32, 3600)=.85	.715
vowel*duration*max F0*proficiency	n.s.	(32, 3600)=.66	.929

Table 3. 1 The significance, F, and p values for the four-way ANOVA testing the effect of vowel quality, max F0, duration, and proficiency on stress identification.

The results indicated that the main effect of vowel quality is significant. The response for words with a full vowel in the first syllable (1.42) is lower than for words with a reduced vowel (1.52), which indicates that a reduced first vowel triggers more verb-like responses (values closer to 2). Surprisingly, the interaction between vowel

quality and proficiency is not significant, which indicates that the three proficiency levels do not differ in their utilization of the vowel quality cue. Further analysis of the two sets of vowel type words confirmed that the three groups are very similar to each other in both sets. In order to more closely examine the effect of proficiency, max F0, and duration, further discussion will focus on a three-way ANOVA on the full vowel set - dada. (see Appendix E for the results of the four-way ANOVA).

3.2.2 Max F0, duration, and proficiency

Main effects

A 5 (max F0) × 5 (duration) × 3 (proficiency) analysis of variance (ANOVA) of the responses showed a main effect of max F0 [$F(4, 1800) = 93.76, p < .001$] (see Figure 3.7). Note that a greater first-to-second syllable ratio should induce more noun (1st syllable stressed- smaller mean) responses. The posthoc analysis indicated that the mean response for ratios F5 and F4 was smaller than for F3, which was smaller than F2 and F1. In other words, ratios F5 and F4 received significantly more noun responses than others. The ratios F2 and F1 received more verb responses. F3 (ratio 1) was found to have more noun responses, which indicates that there may be a noun bias when no other cues are provided.

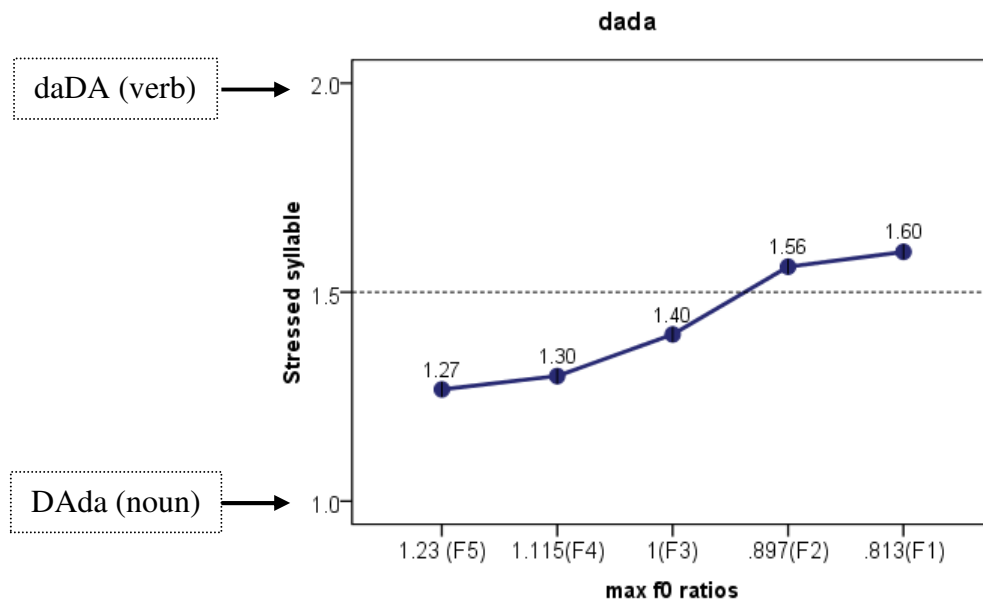


Figure 3. 7. Responses for first-to-second max F0 ratios. The response ‘1’ refers to 1st syllable stressed (noun response) while ‘2’ refers to 2nd syllable stressed (verb response). The dotted reference line, 1.5, is the assumed chance level performance.

The main effect of duration was found to be significant [$F(4, 1800) = 111.42, p < .001$] (Figure 3.8). The posthoc analysis showed that responses for ratios D5 and D4 were significantly smaller than those for other ratios, but were not significantly different from each other. Responses for the baseline D3 were significantly smaller than both D2 and D1. Responses for D1 were significantly higher than all other ratios.

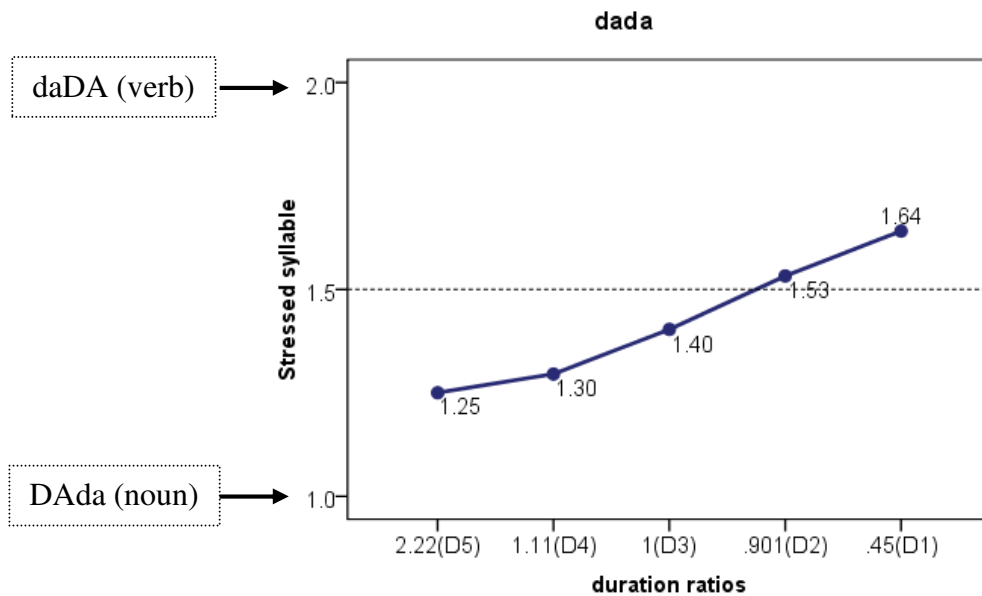


Figure 3. 8. Responses for first-to-second vowel duration ratios. The response '1' refers to 1st syllable stressed (noun response) while '2' refers to 2nd syllable stressed (verb response). The dotted reference line, 1.5, is the assumed chance level performance.

The main effect of proficiency is also significant [$F(2, 1800) = 7.55, p = .001$] (Figure 3.9). The post hoc analysis indicated that there were significantly more noun responses from advanced learners and native speakers than from beginning learners. No significant differences were found between advanced learners and native speakers.

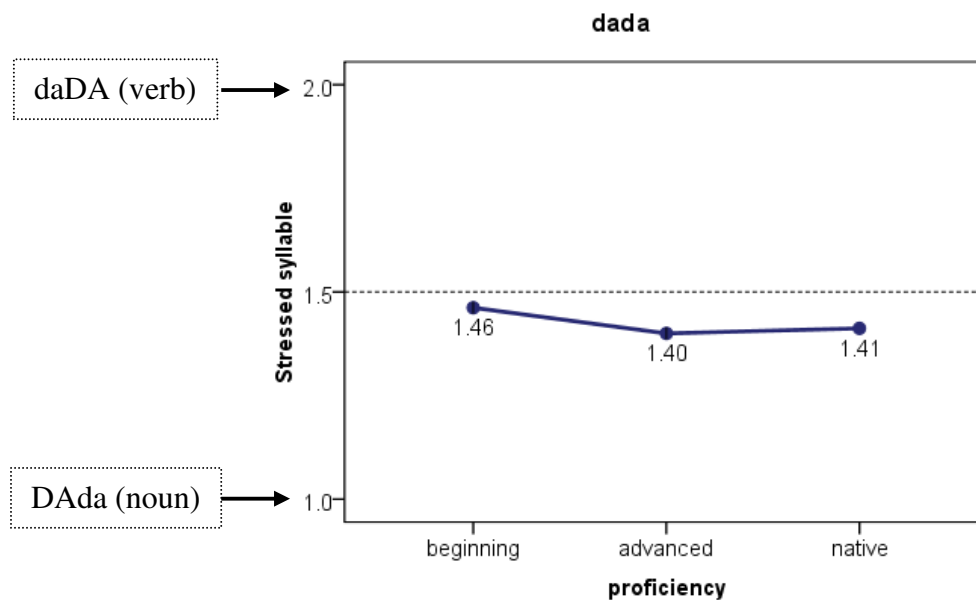


Figure 3. 9. Responses from beginning and advanced learners as well as native speakers of English. The response ‘1’ refers to 1st syllable stressed (noun response) while ‘2’ refers to 2nd syllable stressed (verb response). The dotted reference line, 1.5, is the assumed chance level performance.

Interactions

The interaction between max F0 and proficiency was found to be significant [$F(8, 1800) = 51.39, p < .001$] (Figure 3.8). Responses from the beginning learners do not differ much across the five max F0 ratios. The max F0 ratios have a gradual effect on the native speakers’ responses – a greater max F0 ratio entails more noun responses, while a smaller max F0 ratio entails more verb responses. Results from the advanced learners show a near-categorical pattern across the ratios. More noun responses were associated with ratios F5 and F4, while more verb responses were associated with ratios F2 and F1.

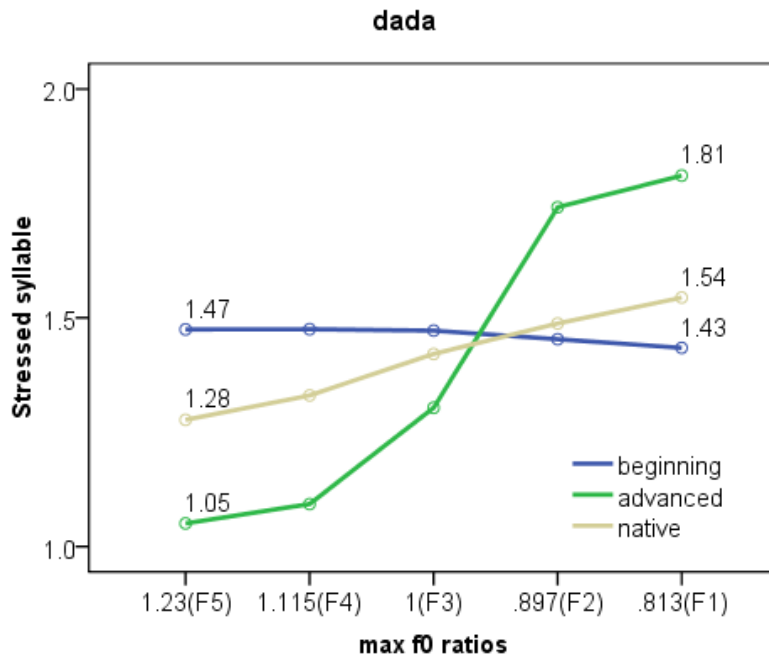


Figure 3. 10. Responses for five max F0 ratios from three proficiencies. The response ‘1’ refers to 1st syllable stressed (noun response), while ‘2’ refers to 2nd syllable stressed (verb response).

Impact degree

The first-to-second max F0 vowel ratios have a gradient effect on the responses of native speakers – a greater max F0 ratio means more noun responses. Native speaker responses to different max F0 ratios changed from 1.28 (for F5) to 1.54 (for F1). Advanced learners exhibit a near-categorical effect across max F0 ratios. Their responses changed from a high percentage of noun responses (1.05 for F5), to a high percentage of verb responses (1.81 for F1). Finally, beginning learners’ stress perception did not seem to be affected by different max F0 ratios - their responses remain near chance level across all five ratios.

The max F0 cue has the greatest impact on advanced learners, followed by native speakers, with no discernible effect on the stress perception of beginning learners.

The interaction between duration and proficiency was also significant [$F(8, 1800) = 20.41, p < .001$] (Figure 3.9). The results from the beginning learners demonstrate a near-categorical effect of the duration ratio on their responses: ratios D5 and D4 received a high percentage of noun responses, while D2 and D1 received a high percentage of verb responses. The results from the advanced learners and native speakers showed a more gradient effect of duration ratios - verb responses gradually increase as the first-to-second vowel ratio decreases.

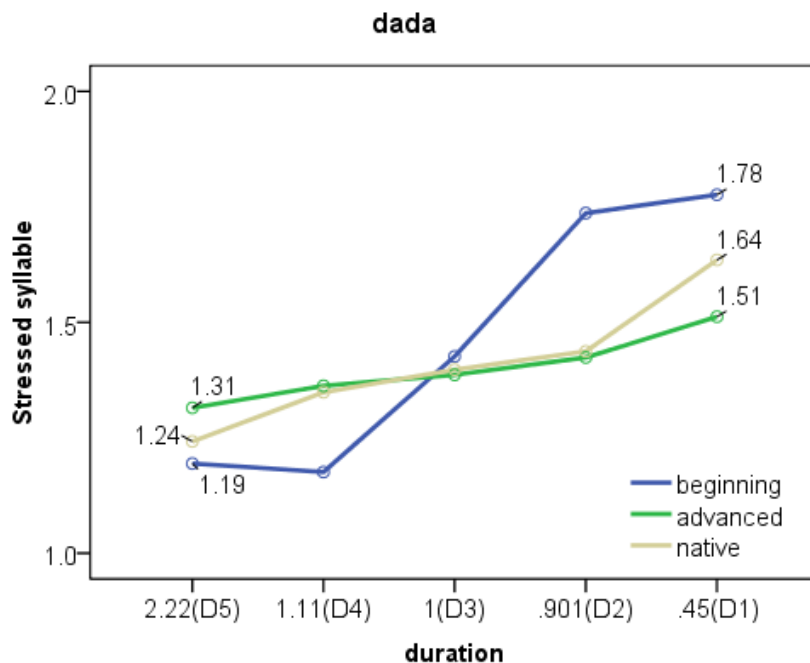


Figure 3. 11. Responses for five duration ratios from beginning learners, advanced learners and native listeners. The response '1' refers to 1st syllable stressed (noun response) while '2' refers to 2nd syllable stressed (verb response).

Apart from vowel quality, duration was also found to play a role in stress perception in all listener groups, in which greater first-to-second vowel ratios entailed more noun responses (Figure 3.14). Nevertheless, the magnitude of the effect from each listener group is different: results from the beginning learners demonstrate a near-categorical

effect of duration ratio. Their responses changed from 1.19 (for D5) to 1.78 (for D1). The magnitude of change is smaller for native speakers. Their responses changed from 1.24 (for D5) to 1.64 (for D1). Finally, the responses of advanced learners changed from 1.31 (for D5) to 1.51 (for D1).

In short, the duration cue has the greatest impact on the responses of beginning learners, followed by native speakers and then advanced learners.

3.2.3. Cue weighting in each listener group

To further examine the consistency of the effect of max F0 for each duration ratio (and vice versa), the responses for five first-to-second vowel max F0 ratios (x-axis) across five duration ratios (separate lines) are plotted for each proficiency level. Parallel graphs show first-to-second vowel duration ratios along the x-axis and max F0 as separate lines (see Figures 3.12). Results from individual listeners can be found in appendix G. Responses for all cue combinations across all speaker groups are plotted in appendix H.

Native speakers

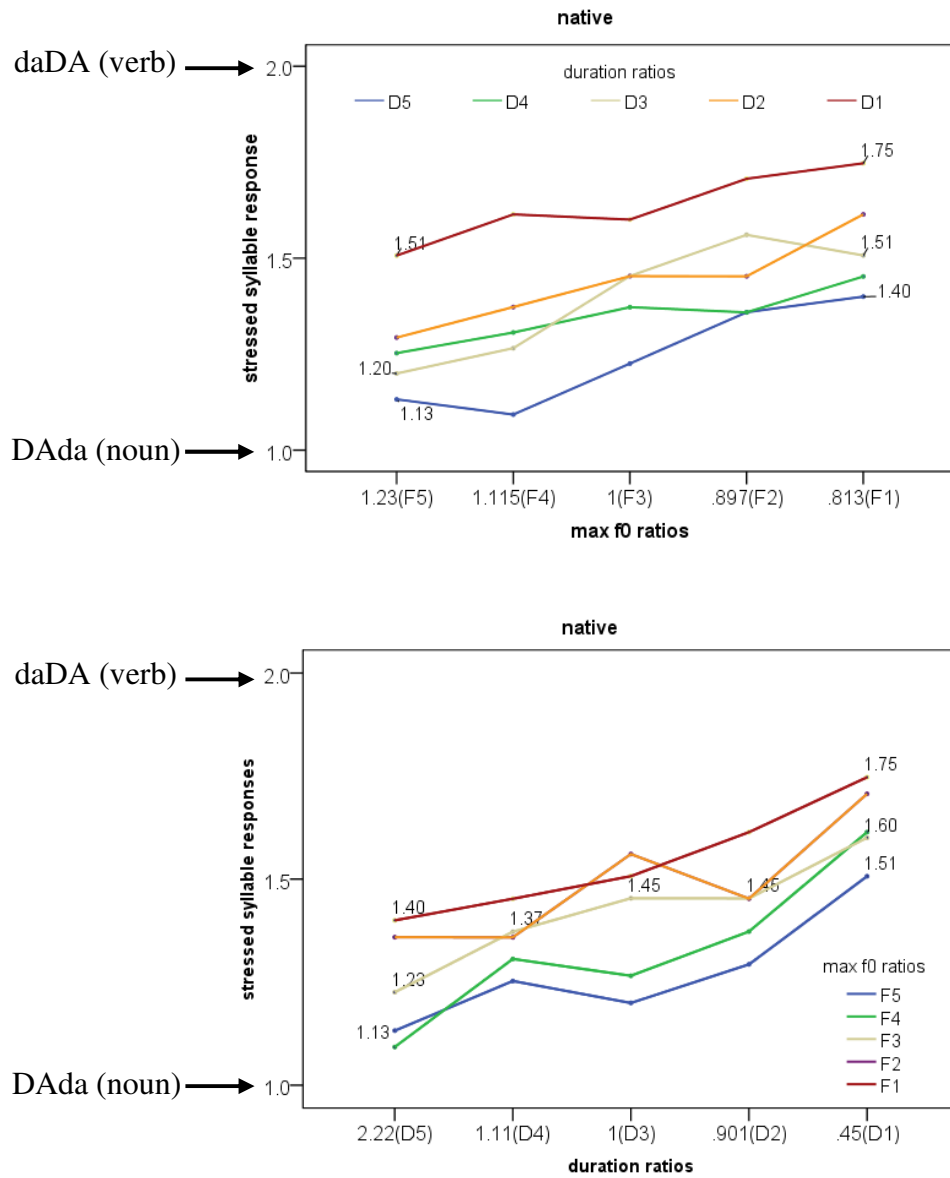


Figure 3. 12. Responses from native listeners plotted with max F0 ratios along the x-axis (top) and duration ratios along the x-axis (bottom). The response '1' refers to 1st syllable stressed (noun response), while '2' refers to 2nd syllable stressed (verb response).

As shown in Figure 3.12, both graphs demonstrate that the smaller the first-to-second vowel ratios (D5 to D1) are, the more verb responses are elicited from native speakers. In order to evaluate the relative importance of max F0 and duration, native listener responses for each max F0 and duration ratio combination are plotted in Figure 3.13. The responses for the following combinations were compared to evaluate the effects of max F0 and duration: 1.) max F0 and duration signal the same syntactic category, 2.) max F0 and duration signal opposite syntactic categories, 3.) max F0 cue present only, and 4.) duration cue present only.

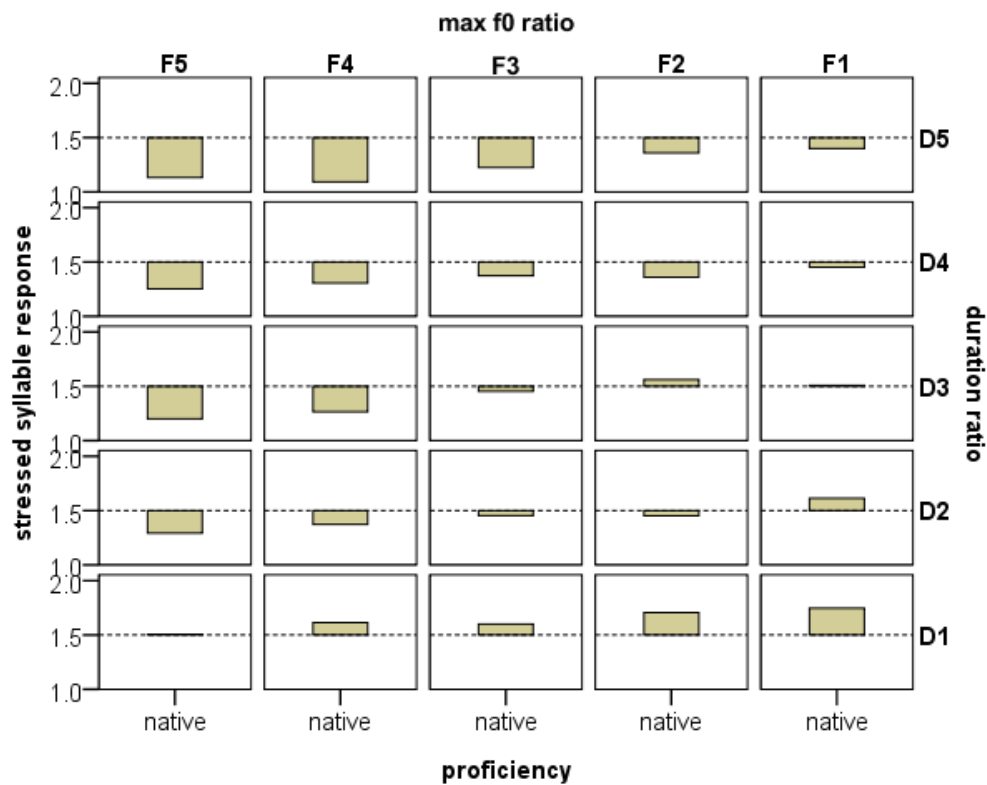


Figure 3. 13. Responses from native English speakers for all cue combinations.

Matched cues

When duration and max F0 both signal nouns (F5, F4) × (D5, D4) or both signal verbs (F2, F1) × (D2, D1), the responses generally match the cues - more noun responses for (F5, F4) × (D5, D4), and more verb responses for (F2, F1) × (D2, D1).

Conflicting cues

When duration and max F0 are in conflict with each other, in the combinations of (F5, F4) × (D2, D1), in which F0 cues noun but duration cues verb, only F5D2 and F4D2 have more noun responses. F5D1 has close to chance level performance, and F4D1 has more verb responses. The results indicate that duration cues can cancel out F0 cues (F5D1), or outweigh F0 cues (F4D1).

In other cue-conflicting combinations - (F2, F1) × (D5, D4), in which F0 cues verb but duration cues noun, native speakers have more noun responses in all four conditions, indicating that duration cues outweigh F0 cues in these contexts.

Max F0 cue-only and duration cue-only

In duration cue-only conditions (F3 column), native speakers have more noun responses for D5 and F4, close to chance level for D3 and D2, and more verb responses for D1. In max F0 cue-only conditions (D3 row), native speakers have more noun responses for F5 and F4, but performed at chance level for F3, F2, and F1. The results suggest that when duration is not present, even though F0 cues verb, the tokens were still not perceived as verbs.

Advanced learners

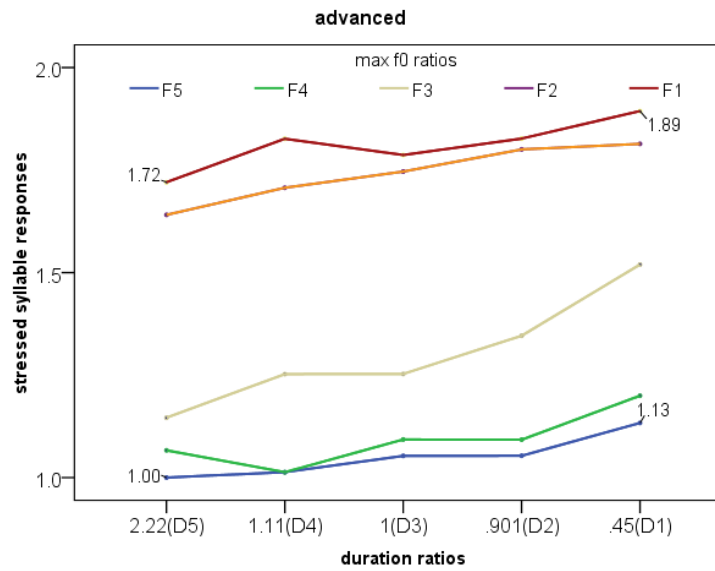
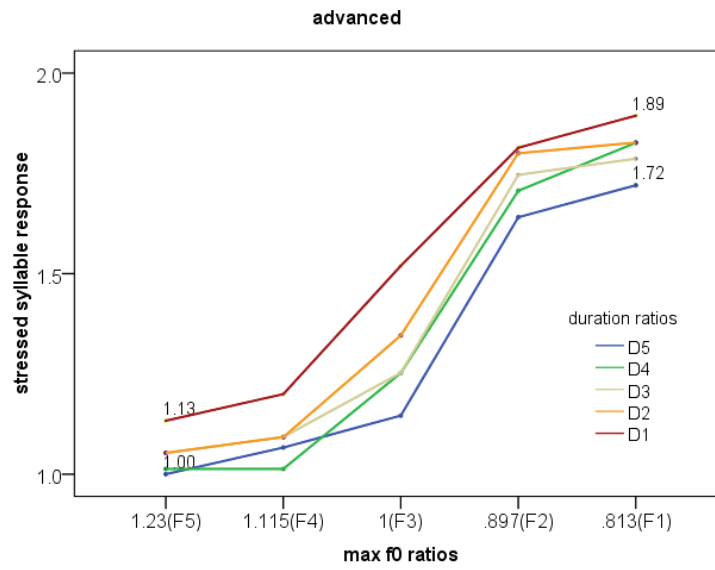


Figure 3. 14. Responses from advanced learners plotted with max F0 ratios along the x-axis (top) and duration ratios along the x-axis (bottom).

In Figure 3.14, both graphs demonstrate that as the first-to-second vowel ratios decrease (D5 to D1), advanced Mandarin L2 learners of English have more verb responses. The categorical effect found across max F0 suggested that max F0 plays a major role in stress location judgment.

To further confirm this observation, the responses for each max F0 and duration ratio combination are plotted in Figure 3.15. The responses for the following combinations were compared to evaluate the effects of max F0 and duration: 1.) max F0 and duration signal the same syntactic category, 2.) max F0 and duration signal opposite syntactic categories, 3.) max F0 cue present only, and 4.) duration cue present only.

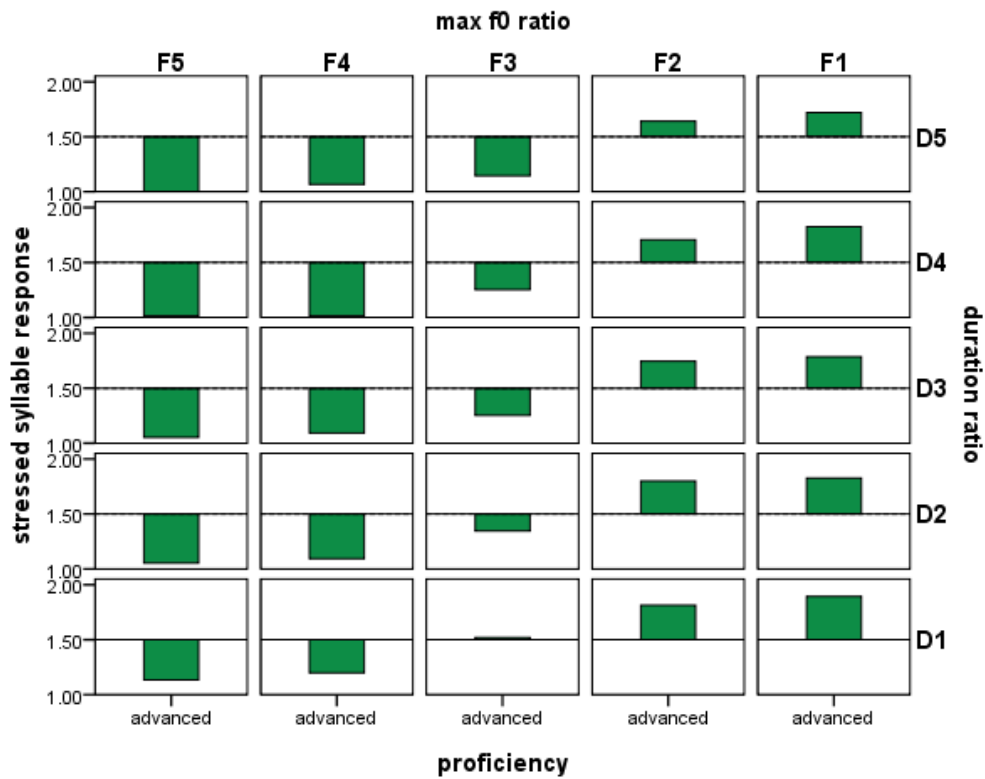


Figure 3. 15. Responses from advanced Mandarin L2 learners of English for all cue combinations.

Matched cues

When duration and max F0 both signal nouns (F5, F4) × (D5, D4) or both signal verbs (F2, F1) × (D2, D1), the responses generally match the cues - more noun responses for (F5, F4) × (D5, D4) and more verb responses for (F2, F1) × (D2, D1).

Conflicting cues

When duration and max F0 are in conflict with each other, in the combinations of (F5, F4) × (D2, D1), in which F0 cues noun but duration cues verb, advanced learners have more noun responses, suggesting that F0 plays a more important role in perception. In parallel, results from (F2, F1) × (D5, D4), in which F0 cue verb but duration cues noun, advanced learners have more verb responses in all four conditions, indicating that F0 cues outweigh duration cues in these contexts.

Max F0 cue-only and duration cue-only

In duration cue-only conditions (F3 column), advanced learners have noun responses in almost all duration ratios, however the deviation from 1.5 decreases from D5 to D1. Responses for D1 are close to chance level. Interestingly, for F3D3 (all cues neutral), advanced learners have more noun responses, suggesting that their responses may be noun-biased.

In max F0 cue-only conditions (D3 row), advanced learners have more noun responses for F5 and F4, and more verb responses for F2 and F1. The results suggest that the judgments of advanced learners are primarily based on F0 cues.

Beginning learners

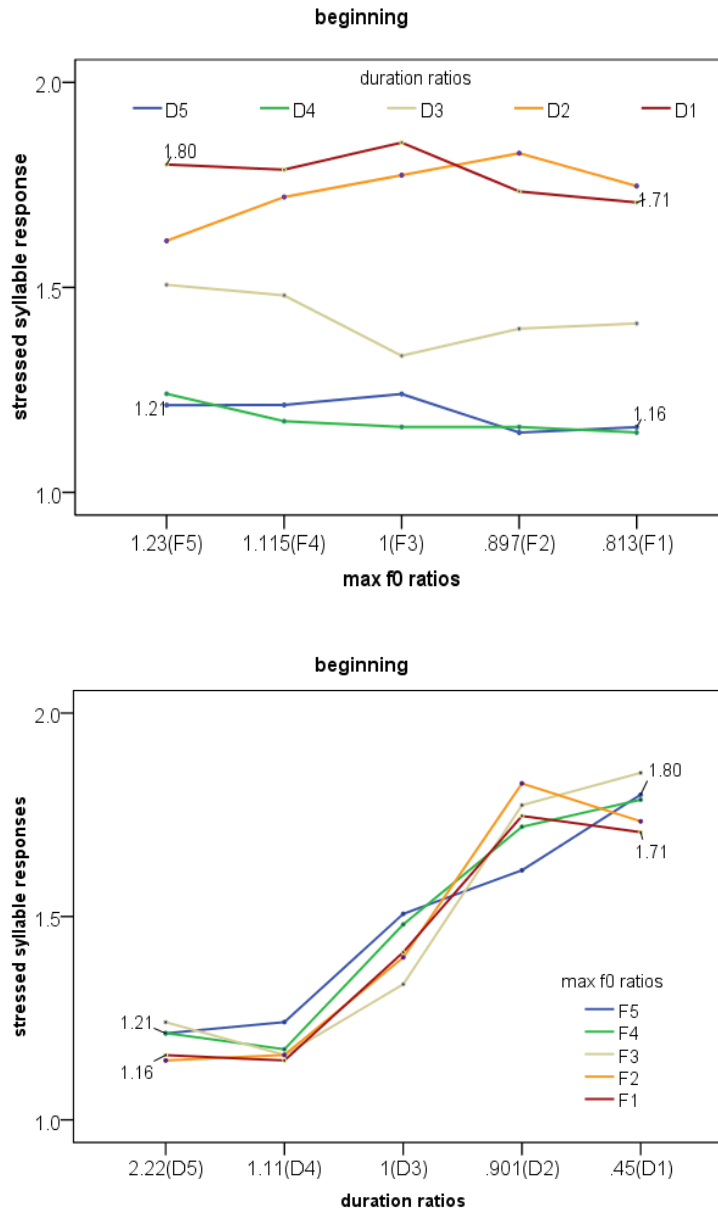


Figure 3. 16. Responses from the beginning learners plotted with max F0 ratios along the x-axis (top) and duration ratios along the x-axis (bottom).

In Figure 3.16, the top graph with the max F0 ratios along the x-axis demonstrates that the responses from the beginning Mandarin L2 learners of English do not change across different max F0 ratios (as seen by the flat lines for all duration ratios). On the other hand, the bottom graph demonstrates a near-categorical effect across different duration ratios, which suggests that duration plays a major role in stress location judgment.

To further confirm this observation, the responses for each max F0 and duration ratio combination are plotted in Figure 3.17. The responses for the following combinations were compared to evaluate the effects of max F0 and duration: 1.) max F0 and duration signal the same syntactic category, 2.) max F0 and duration signal opposite syntactic categories, 3.) max F0 cue present only, and 4.) duration cue present only.

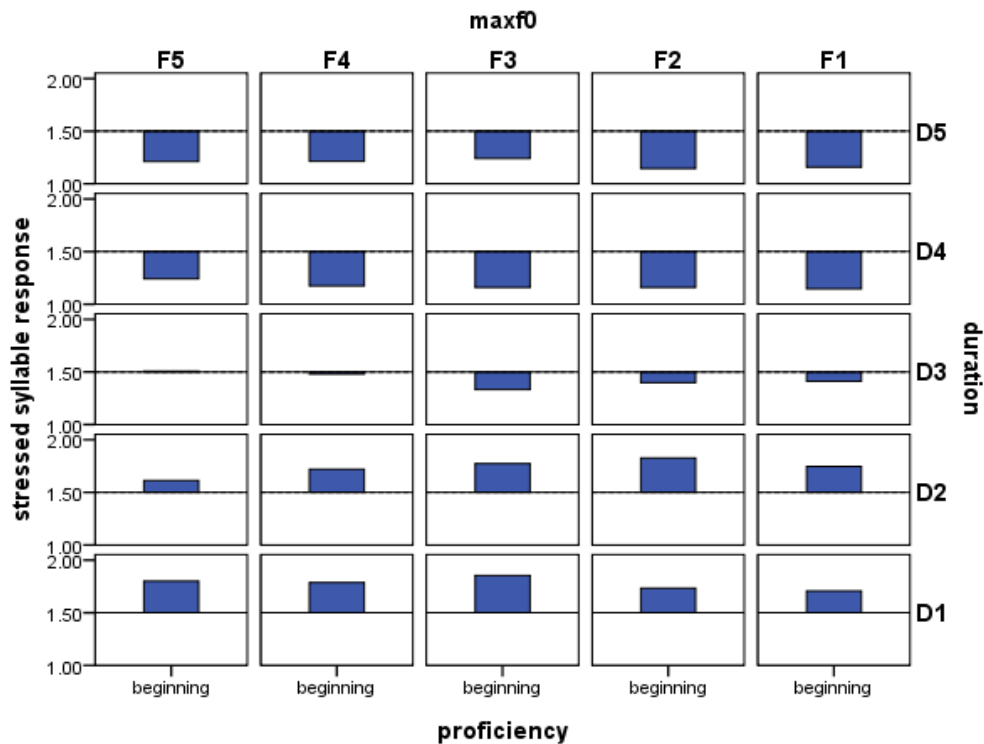


Figure 3. 17. Responses from beginning Mandarin L2 learners of English for all cue combinations.

Matched cues

When duration and max F0 both signal nouns (F5, F4) × (D5, D4) or both signal verbs (F2, F1) × (D2, D1), responses from the beginning learners match the cues - more noun responses for (F5, F4) × (D5, D4) and more verb responses for (F2, F1) × (D2, D1).

Conflicting cues

When duration and max F0 are in conflict with each other, in the combinations of (F5, F4) × (D2, D1), in which F0 cues noun but duration cues verb, beginning learners have more verb responses, suggesting that duration plays a more significant role in the perception of stress. In parallel, results from (F2, F1) × (D5, D4), in which F0 cues verb but duration cues noun, beginning speakers have more noun responses in all four conditions, indicating that duration cues outweigh F0 cues in these contexts.

Max F0 cue-only and duration cue-only

In duration cue-only conditions (F3 column), beginning learners have more noun responses in D5 and D4, and more verb responses for D2 and D1. Results from F3D3 (all cues neutral) have more noun responses, suggesting that beginning learners may also be noun-biased. In max F0 cue-only conditions (D3 row), beginning learners show close to chance level responses, which suggests that their stress perception is not affected by max F0 ratios.

3.3 Discussion

3.3.1 Summary

This perception study compared the relative importance of perceptual cues for English stress between the Mandarin L2 learners of English and native English speakers. The stimuli were resynthesized versions of a naturally produced non-word disyllable, “dada.” The target perceptual cues tested were max F0, duration, and vowel quality. Vowel quality in the first vowel of the disyllable was resynthesized to be either full or reduced. Resynthesis first-to-second vowel ratios for max F0 [(1.23(F5), 1.115(F4), 1(F3), .897(F2), and .813(F1))] and duration [2.22(D5), 1.11(D4), 1(D3), .901(D2), and .45(D1)] were generated based on the results of the word pair experiment (see section 2.2).

The results from a stress localization test indicate that all three proficiency levels of listeners - beginning, advanced, and native speakers - are sensitive to the vowel quality of the first vowel, i.e. full vowels in the first syllable attracted more noun responses, while reduced vowels in the same location received significantly more verb responses. No significant difference was found between the three groups of speakers.

Results from ANOVA (section 3.2.2) and cue weighting analysis (3.2.3) indicate that native speakers are sensitive to both duration and max F0 cues, but focus more on duration. The advanced learners’ data suggest that they are primarily affected by max F0 cues, while still reacting to duration cues in stress perception to some extent. In addition, different max F0 ratios have a near-categorical effect on advanced learners’ responses: a high percentage of noun response for ratios F5 and F4, and a high percentage of verb response for ratios F2 and F1. Beginning learners are very sensitive to duration cues, and do not seem to react to max F0 cues at all. A near-categorical effect from different duration ratios was also observed in beginning learners’ responses.

3.3.2 The roles of vowel quality, max F0, and duration in stress perception

Vowel quality

The formant analysis data from the first acoustic study on minimal word pairs indicated that native English speakers reduced unstressed vowels, while no strong effect of unstressed vowel reduction was found in Mandarin L2 learners (beginning learners particularly). Based on this finding, as well as the fact that vowel reduction does not exist in Mandarin, I hypothesized that vowel quality might not play a role in the learners' perception of English. Results from the perception study, however, do not support this hypothesis. All listener groups had more noun responses (1st syllable stressed) when the first vowel was full, and more verb responses when the first vowel was reduced.

This could originate from intensity rescaling for the stimuli. Although the mean intensity, as well as the intensity tier were recalled to match those of the full vowel [a]'s, there may still have been some intensity differences between the reduced and full vowels in specific frequency bands. Results from Sluijter and van Heuven (1996) on stressed and unstressed Dutch syllables indicate that in natural production, the intensity differences are concentrated in the higher parts of the spectrum, i.e. intensity differences in the lower part of the spectrum (< 500Hz) were negligible. Sluijter et al. (1997) later conducted a perception study using stimuli in which intensity differences were concentrated in the higher frequency bands. These stimuli were found to provide stronger stress cues than uniformly increased intensity difference stimuli. When manipulating intensity then, care should be taken to consider not only the mean intensity and intensity tier, but also the distribution of intensity in different frequency ranges.

Duration and max F0 weighting in stress perception

Results from the statistical analysis and cue weighting analysis suggest that when perceiving word-level stress, native English speakers utilize duration as the primary cue, and max F0 as a lesser cue in identifying the stressed syllable in non-words. Advanced Mandarin L2 learners of English rely mainly on max F0 cues, while beginning learners focus more on duration cues.

The results of the current study differ from the findings of previous studies in that duration was found to be more important than F0. Previous studies had claimed that F0 is the most salient cue in English stress perception. A short review of these studies is provided below.

A series of experiments was conducted to test the perceptual cues for stress in English by Fry (1955, 1958). In Fry (1955), the duration and intensity of the first syllable were resynthesized in word pairs contrasting in stress assignment – *subject* (noun) and *subject* (verb). The participants' task was to identify if they had heard a noun or a verb. Results from native English speakers indicated that duration provided an overriding cue for stress in native speakers' perception. In a follow-up study, Fry (1958) compared the roles of duration and F0 in stimuli with these two cues manipulated in a step-wise manner. His findings suggested that a syllable higher in F0 was more likely to be perceived as stressed, and in some cases can override duration cues.

Bolinger also used both natural and synthesized tokens to test the perceptual cues for English word stress in a series of experiments (see Bolinger (1958) for a detailed review). He concluded that pitch prominence is the primary cue for stress. Furthermore, Morton and Jassem (1965) adopted synthetic non-words (e.g. /sisi/, /sasa/) to explore the role of F0, intensity, and duration in stress perception. Their

results indicated that F0 differences play a far greater role than duration and intensity in stress perception.

Beckman (1986) also conducted a perception experiment in order to examine the perceptual salience of these cues in English stress perception. She used a hybrid resynthesis process which manipulated *subject-subject* word pairs in duration, F0, intensity, and spectral composition. Native English speakers performed an identification task to detect the stressed syllable. The results indicated that native English listeners are sensitive to duration, F0, intensity, and spectral composition, with F0 being the most important cue.

One possible reason for the discrepancy between the current study and previous studies may be due to the fact that the target F0 cue manipulated in the current study is the max F0, while previous studies manipulated the mean F0 of the syllables. It may be the case that mean F0 provides stronger perceptual salience than max F0, thus generating a stronger effect of stress perception. The present study selected max F0 instead of mean F0 for two reasons. Firstly, it was observed that the acoustic realization of mean F0 is more similar between native English speakers and Mandarin L2 learners of English, while max F0 data are more diverse between these two groups. The assumption was that the effect of max F0 in the perception of stress would also be more diverse between the two groups.

On the other hand, Fry (1958) has pointed out that stimuli with linear fall or late fall F0 patterns are more likely to be perceived as stressed syllables. Since F0 movement is not a target factor of the current study, it was important to control F0 movement in the target syllables. Precision in controlling the location of max F0 can exclude F0 alignment as a possible cue. This was done by consistently placing max F0 at the beginning of the syllables for all tokens (as a simulation of the naturally produced base for resynthesis, see 3.1.1 for the details of max F0 resynthesis).

Asymmetry

To investigate if the same absolute magnitude of max F0 and duration across first and second syllables would generate the same degree of noun and verb responses, some cue combination conditions were plotted in Figure 3.18, and will be discussed below.

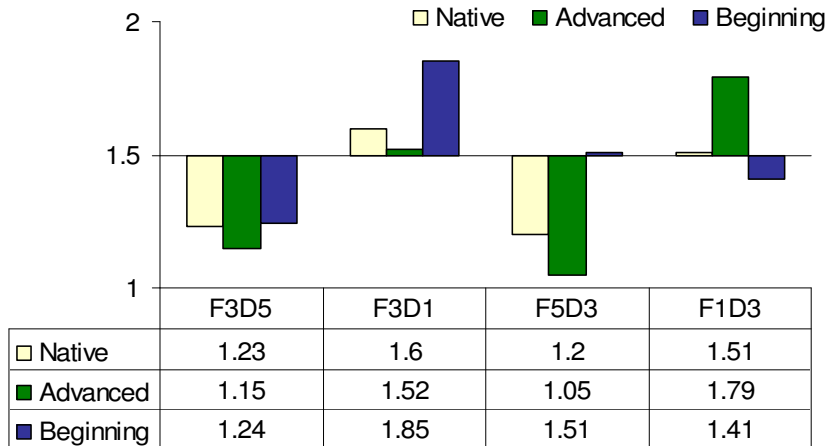


Figure 3. 18. Responses for duration-neutral (F3D1, F3D5), and max F0-neutral (F1D3, and F5D3) cue combinations.

Results from native speakers indicate that in duration cue-only contexts, the same absolute duration differences between the two vowels in the disyllable induce more noun responses than verb responses, as can be seen by the greater deviation from 1.5 in F3D5, and the smaller deviation in F3D1. In addition, native speakers display another asymmetry in utilizing max F0. In max F0 cue-only conditions (e.g. F1D3 and F5D3), the same absolute max F0 differences between the two vowels in the disyllable induce a large percentage of noun responses in F5D3, but significantly fewer verb responses in F1D3.

Interestingly, results from both native and advanced listeners show a strong noun bias. This effect can be observed through the max F0 F3 (ratio of 1), for which responses from both listener groups are mostly below chance level 1.5. In addition, responses

for ratios D5 and D4 for both groups are much closer to 1 (noun responses) than ratios D2 and D1.

3.3.3 Conclusion

The current study is, to the author's knowledge, the first attempt at testing stress perception in Mandarin L2 learners of English. A systematic resynthesis approach was established to ensure that the only differences between the two vowels in the disyllabic non-words were in vowel quality, max F0, and duration. Results from native English speakers indicate that duration is the primary cue in stress perception in non-words, followed by max F0. Advanced Mandarin L2 learners of English rely mainly on max F0 cues, and exhibit a near-categorical pattern across max F0 ratios. Beginning learners focus mostly on duration cues, and are not affected by max F0 cues.

CHAPTER 4: GENERAL DISCUSSION

The current study aimed to explore the effect of F0 on the acoustic realization and perception of English lexical stress in second language learners. More specifically, the focus has been on the acquisition of English lexical stress by Mandarin L2 learners of differing proficiencies. The research goals were to examine the acoustic realization of stress, as well as the perceptual relevance of the stress. Sixty-eight Mandarin L2 learners of English and thirty five native English speakers participated in two acoustic studies and one perception study.

4.1 Summaries

4.1.1 Experiment 1: Acoustic study on minimal word pairs

In this acoustic experiment, disyllabic word pairs with similar segmental composition but contrastive stress location (e.g. *contract* and *contract*) were recorded by Mandarin learners of English with two proficiencies (beginning and advanced learners). Mean F0, max F0, intensity, duration, and F2 of stressed and unstressed vowels were measured and compared to those in the production of native speakers of English.

The results of this experiment demonstrate that the implementation of these correlates varies greatly between noun and verb readings in the production of native English speakers. Native speakers utilize mean F0, max F0, intensity, and duration to signal stressed syllables in noun readings (Figure 4.1), but use only duration as a major cue for stressed syllables in verb readings (Figure 4.2). Mandarin L2 learners, on the other hand, demonstrate a more consistent use of correlates in noun and verb readings in two ways: the four correlates are all adopted in both conditions, and the magnitudes (difference between stressed and unstressed syllables) used were consistent across nouns and verbs.

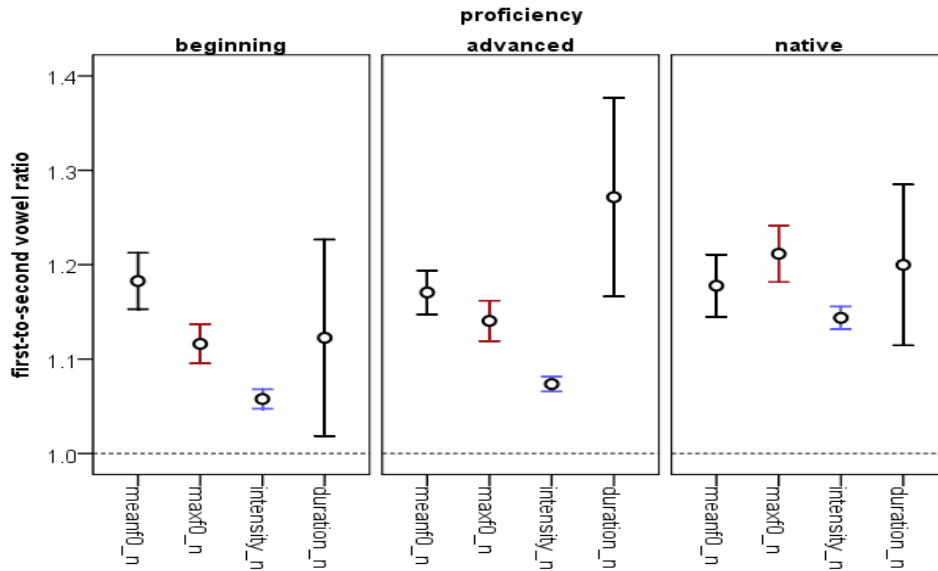


Figure 4. 1. 95% confidence interval of first-to-second vowel ratios for mean F0, max F0, intensity, and duration for nouns. The reference line (1.0) is the assumed baseline when a given correlate is identical in both syllables.

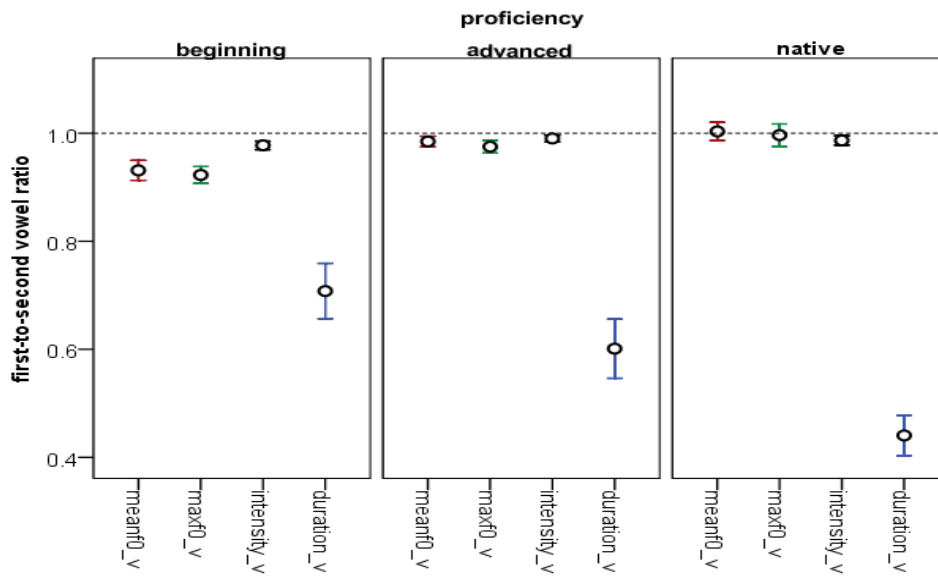


Figure 4. 2. 95% confidence interval of first-to-second vowel ratios for mean F0, max F0, intensity, and duration for verbs. The reference line (1.0) is the assumed baseline when a given correlate is identical in both syllables.

Furthermore, the results from formant analysis demonstrate that unstressed vowel reduction was also affected by the location of stress. Native speakers exhibit a greater degree of reduction when the first vowel is unstressed. In addition, this result is in line with the duration results- there is a greater duration difference across stressed and unstressed vowel when the first vowel is unstressed.

The results from the first acoustic study demonstrate that different stress locations cast different levels of challenge for learners. When the stress is located on the first syllable, learners showed a similar-to-native utilization pattern of the four correlates. The major differences across the three groups were found when stress was located on the second syllable. Beginning learners demonstrated significantly smaller duration differences (Figure 4.3), but greater F0 differences (Figure 4.4) when compared to native speakers. Advanced learners, unsurprisingly, were found to pattern in between beginning learners and native speakers.

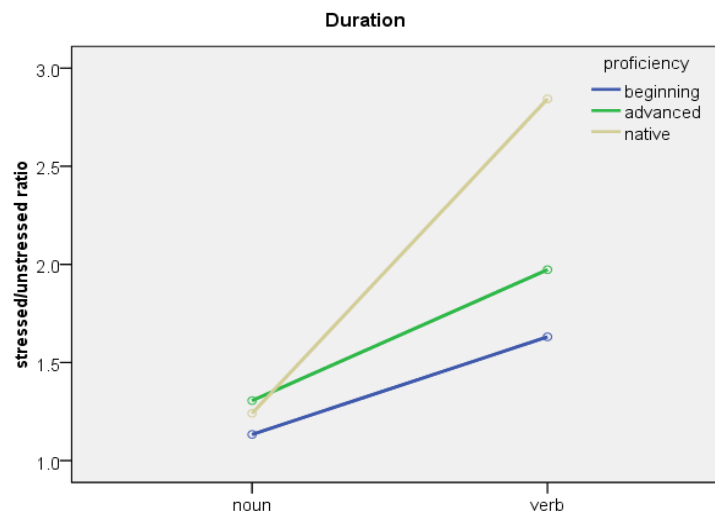


Figure 4. 3. The interaction between stress location and proficiency level for duration.

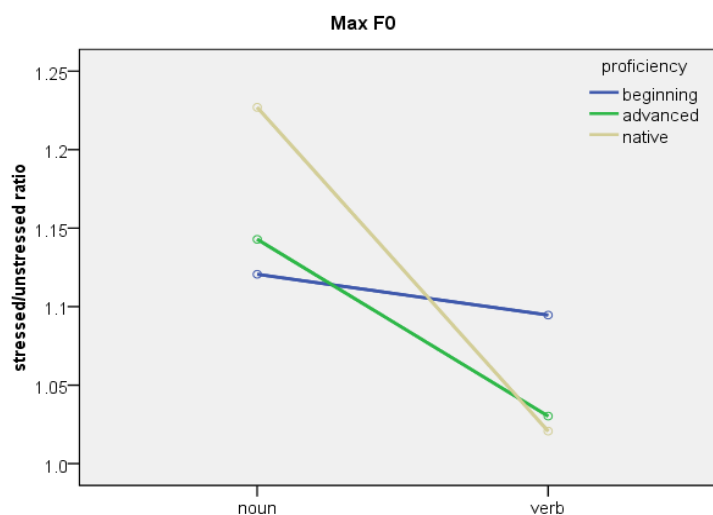


Figure 4. 4. The interaction between stress location and proficiency level for max F0.

4.1.2 Experiment 2: F0 contours in English stressed syllables

This experiment utilized English syllables which obey Mandarin phonotactics, in order to more closely examine F0 contours in learner production. While the stimuli in the previous acoustic study provide well-controlled segmental conditions for the investigation of mean F0, max F0, intensity, duration and F2, however most of the syllables are illegal according to Mandarin phonotactics. In order to better evaluate the realization of F0 contours in speakers' production, stimuli which comply with Mandarin syllable structures were compiled.

Learners at both proficiencies adopt a high-level F0 contour (when stress is located in non-final vowels) and a high-falling contour (when stress is located in the final syllable) to signal stressed syllables, while native speakers exhibit more diverse contour patterns (an example from female speakers was shown in Figure 4.5). This effect may be triggered by the phonemic role of F0 in Mandarin.

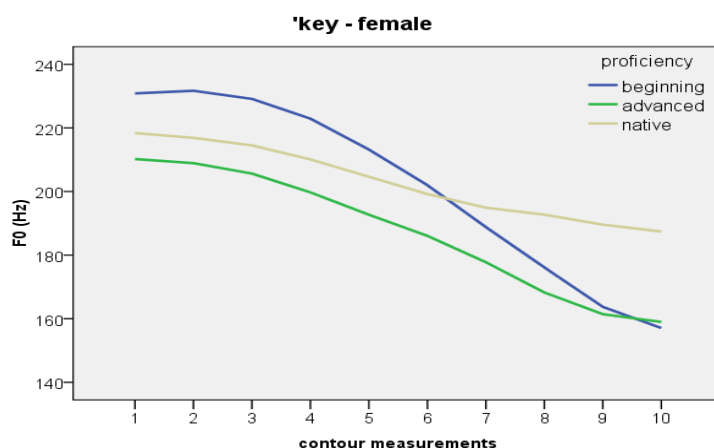


Figure 4. 5. F0 contours for *key* in stressed position across three proficiencies.

Based on the results from the acoustic study, I believe that this difference across stress locations has been overlooked by previous studies, and is crucial in investigating the effect of tone in L1 on L2 stress realization.

4.1.3 Experiment 3: Perception study

This perception study compared the relative importance of perceptual cues for English stress between Mandarin L2 learners of English and native English speakers. The stimuli were resynthesized versions of a naturally produced non-word disyllable, “dada.” The target perceptual cues tested were max F0, duration, and vowel quality. Vowel quality in the first vowel of the disyllable was resynthesized to be either full or reduced. Resynthesized first-to-second vowel ratios for max F0 [(1.23(F5), 1.115(F4), 1(F3), .897(F2), and .813(F1)] and duration [2.22(D5), 1.11(D4), 1(D3), .901(D2), and .45(D1)] were generated based on the results of the word pair experiment (see section 2.2).

The results from a stress localization test indicate that all three proficiency levels of listeners - beginning, advanced, and native speakers - are sensitive to the spectral composition of the first vowel, i.e. full vowels in the first syllable attracted more noun

responses, while reduced vowels in the same location received significantly more verb responses.

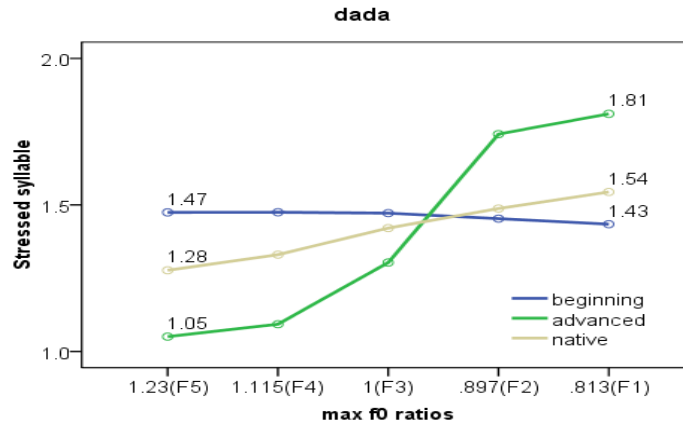


Figure 4. 6. Responses for five max F0 ratios from three proficiencies. The response ‘1’ refers to 1st syllable stressed (noun response) while ‘2’ refers to 2nd syllable stressed (verb response).

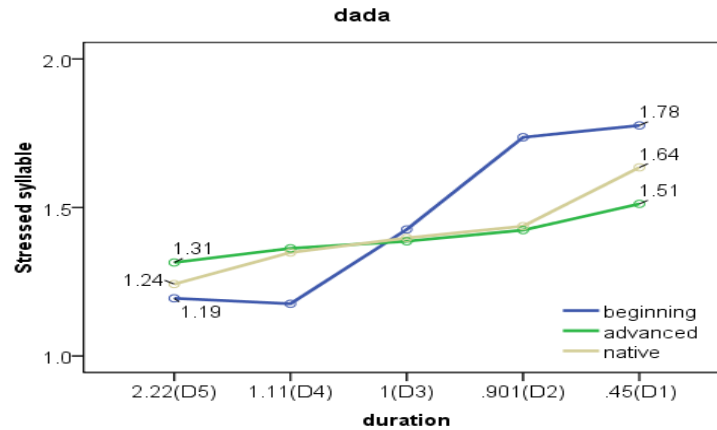


Figure 4. 7. Responses for five duration ratios from beginning learners, advanced learners, and native listeners. The response ‘1’ refers to 1st syllable stressed (noun response) while ‘2’ refers to 2nd syllable stressed (verb response).

The results indicate that native speakers are sensitive to both duration and max F0 cues, but focus more on duration than max F0. Advanced learners’ responses are more affected by the max F0 cue, but to a lesser extent are still affected by the duration cue. Beginning learners are highly sensitive to duration cue, and do not seem to be affected by max F0 in any way.

4.2 Bridging stress production and perception

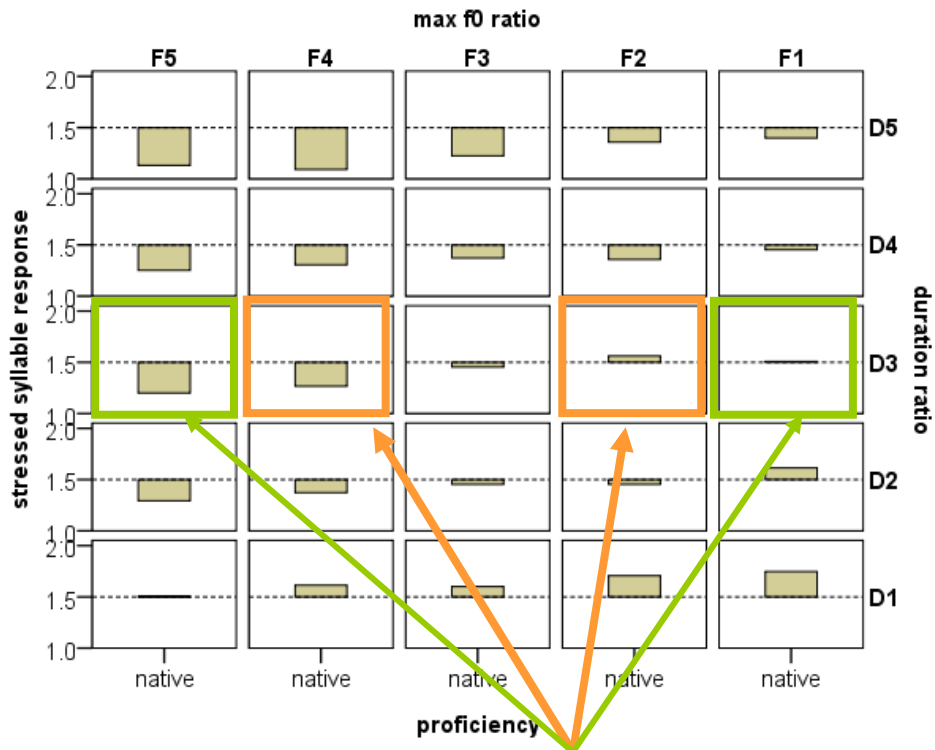
4.2.1 Native production and perception

An asymmetry of acoustic correlate realization in nouns and verbs was found in native speakers' production: native speakers utilize mean F0, max F0, intensity, and duration to signal stressed syllables in noun readings, but use only duration in cueing stressed syllables in verb readings. I believe this production asymmetry originates from the flexibility of native speakers in incorporating extrinsic prosodic factors in stress implementation (cf. 2.4.2).

Perceptual results from native English speakers indicate that their stress perception is affected by vowel quality, duration, and max F0 cues, but that duration sometimes outweighs max F0, and affects the perception of stress to a relatively greater extent. Inspired by the asymmetry of correlate realization in native speakers' production, I further investigated the perceptual roles of max F0 and duration in nouns and verbs.

Max F0 in production and perception

The responses for each max F0 ratio when the duration cue is neutral were closely examined. Two pairs of cue combinations (F4D3 and F2D3; F5D3 and F1D3) were compared (Figure 4.8). These two pairs both have neutral duration cues and the same absolute max F0 difference between vowel 1 and vowel 2 (cf 3.1.1 for manipulation of max F0). If max F0 has the same effect in noun and verb contexts, the same max F0 difference should induce the same response deviations from the 1.5 chance level performance.



- **production:** use F0 in nouns not in verbs
- **perception:** use F0 cues in nouns not in verbs
same degree of F0 causes many **more** responses in noun than in verb context

Figure 4. 8. Responses from native English speakers for all cue combination conditions.

The responses for F4D3 & F2D3 are marked with orange boxes, and F5D3 & F1D3 are in green boxes in Figure 4.8. Results from both comparisons showed that responses in noun contexts F5 and F4 are much greater than in the verb contexts F2 and F1 (both close to chance level 1.5). In other words, the perceptual role of max F0 is greater in noun perception, and plays almost no role in verb perception.

When further compared with the production data, a striking pattern emerges: the role of max F0 is parallel in production and perception in native speakers' performance. That is, in production, native speakers utilize max F0 to encode stress in noun readings but not in verb readings. In perception, they can perceive the max F0 difference in the noun context but in the verb context the same magnitude of difference does not induce the same degree of verb responses. Taken together, max F0 data from stress production and perception seem to imply that when an acoustic correlate is utilized in a certain context in production, the perceptual system will most likely only be sensitive to that cue in the same context.

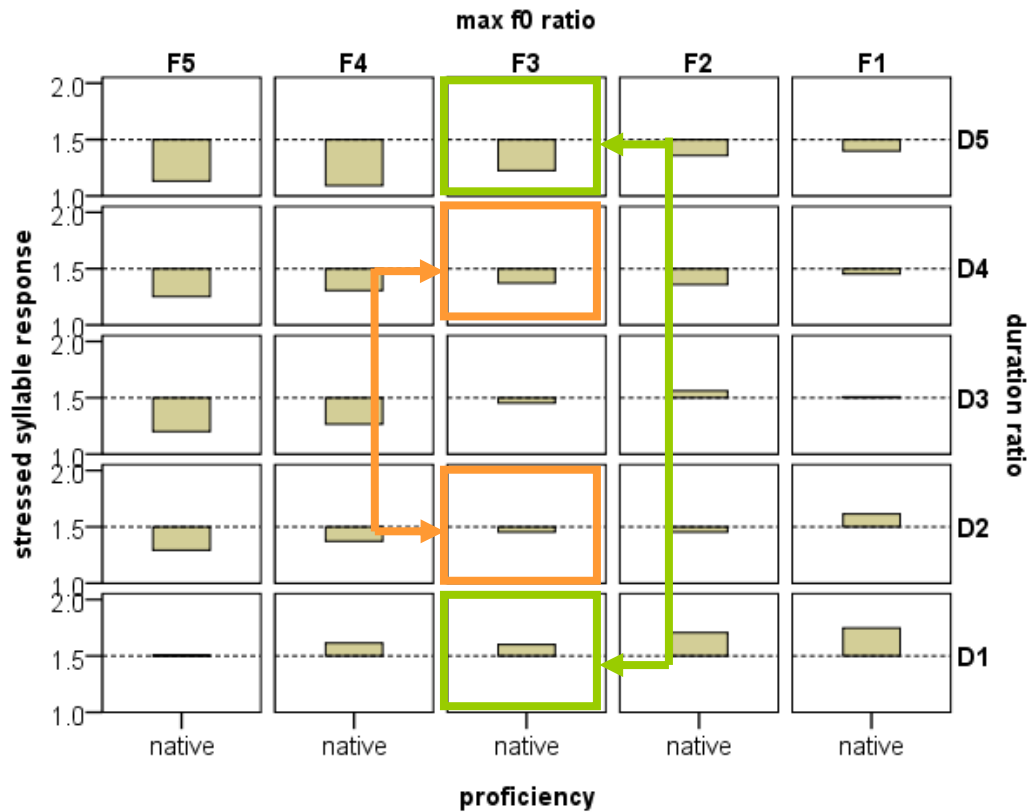
Duration in production and perception

The perception patterns for same absolute duration differences in noun and verb contexts are examined through Figure 4.19; the orange boxes show the results for F3D5 & F3D1 and F3D4 & F3D2 in which two pairs, max F0 cue is neutral but the same duration difference cues nouns for F3D5 and F3D4 but cues verbs for F3D1 and F3D2. The comparisons of these two pairs revealed that the same absolute duration difference induces greater responses in the noun context compared to the verb context: greater noun responses in F3D5 and F3D4 than the responses in F3D2 and F3D1.

Different from max F0, duration was encoded to signal stress in both noun and verb readings in native speakers' production. Following the hypothesis built upon the role of max F0 in production and perception, duration should play a role in stress perception for noun and verb contexts. The asymmetry observed here may correlate with the magnitude of the duration difference between stressed and unstressed vowel in production.

In production, duration is manifested in both noun and verb readings but with a substantially greater magnitude in verb readings (Figure 4.1-2). Compared to the greater noun responses results from the perception study, the magnitude used in

production in different prosodic positions (initial or final) seems to affect the magnitude required to perceive stress in the corresponding context. In other words, if a greater magnitude is used for a given context in stress production, a greater magnitude is required for the stress perception in that context (hence the greater noun responses for F3D5 than verb responses for F3D1).



- **production:** use duration in nouns and verbs, verbs with greater difference
- **perception:** verbs require a greater duration difference, i.e. the same degree of duration causes more responses in noun than in verb contexts

Figure 4. 9. Responses from native English speakers for all cue combination conditions.

Congruence between acoustic correlates and perceptual cues

Taken together, results from max F0 revealed that perception seems to be parallel to production cue realization- when a cue is utilized in production, it will be used in perception in the same prosodic position (e.g. initial or final stress). Results from duration data carry the observation even further and showed that the greater the magnitude of a cue in production, the greater the magnitude required for that cue to be perceived.

Based on these findings from max F0 and duration, I propose that the acoustic correlates and perceptual cues native speakers adopt are parallel in their production and perception. That is, the magnitude of a correlate required to obtain stressed syllable perception is parallel to the magnitude of its realization in production. Also, this pattern appears to be position-specific.

4.2.2 Mandarin L2 learners of English

Beginning learners

Results from the perception study showed that beginning learners focus exclusively on duration (Figure 4.10). As shown in the figure, nearly in all cue combinations, beginning learners were observed to rely mostly on duration cue.

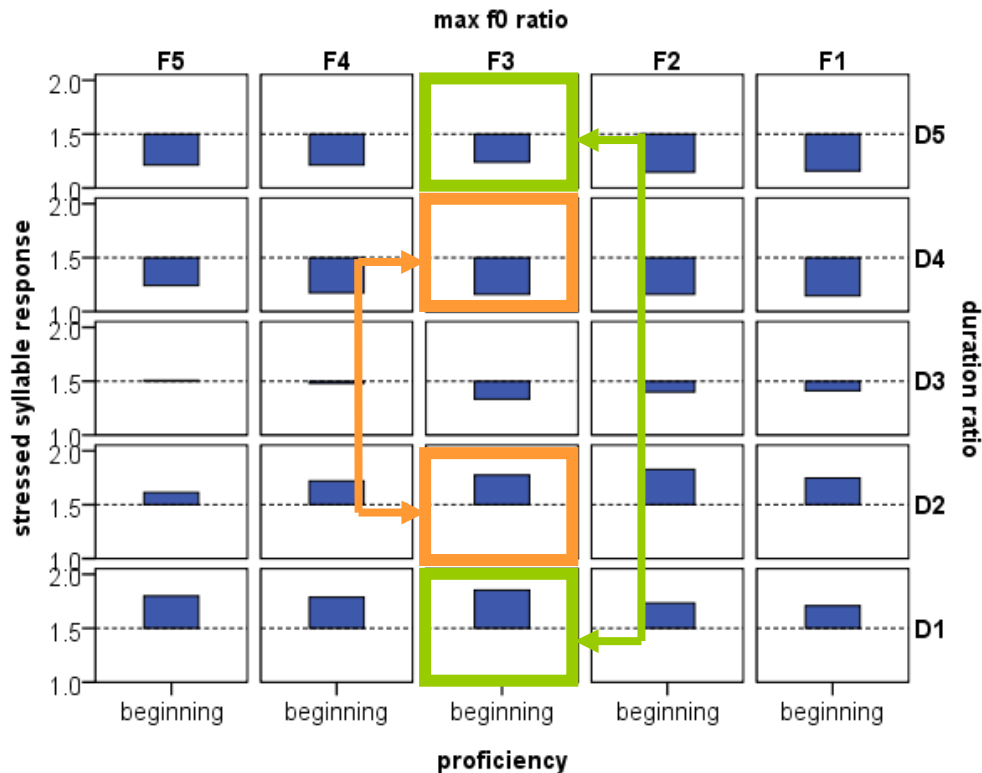


Figure 4. 10. Responses from beginning learners for all cue combination conditions.

In production, beginning learners are found to be not as flexible as native English speakers in shifting the relative weight of correlates in stress implementation. As a result, they use the same cue combinations (mean F0, max F0, intensity, and duration all used in nouns and verbs) and the same cue magnitudes to signal stressed syllables in different prosodic positions. The data from F0 contour analysis also support the inflexibility of F0 cue - learners use fixed F0 contours to signal stressed syllables. It is

argued that the ‘inflexibility’ of modifying the magnitudes and shapes of F0 originates from the tonal feature of the learners’ L1- Mandarin.

It is concluded that Mandarin L2 learners’ performance is greatly influenced by the phonemic role of F0 in Mandarin. In beginning learners’ production, they are found to consistently use the same magnitude of mean F0 and max F0 across noun and verb readings (cf. Figure 4.3 as shown by the similar stressed-to-unstressed vowel ratios). The phonemic feature of F0 in the learners’ L1 forces them to use this feature in a highly consistent manner. In term of perception, I propose that beginning learners carry over their insensitivity to F0 height (Gandour, 1983) from the perceptual habit in Mandarin L1. Hence beginning learners are not sensitive to the max F0 changes. Instead, they rely solely on duration in identifying stress.

Advanced learners

Furthermore, I argue that the production and perception patterns alter when the learners’ proficiency advances. In terms of max F0, advanced learners utilize it in both noun and verb readings. If the congruence hypothesis proposed in 4.2.1 is correct, a greater magnitude is required for advanced learners to perform the same degree of responses in the noun context than in the verb context. However, the results are not as clear as observed from native speakers (Figure 4.10). Comparisons between F5D3 & F1D3 as well as F4D3 & F2D3 showed that the same max F0 triggers more noun responses in the noun context (F5D3 and F4D3) than verb responses in the verb context (F2D3 and F1D3).

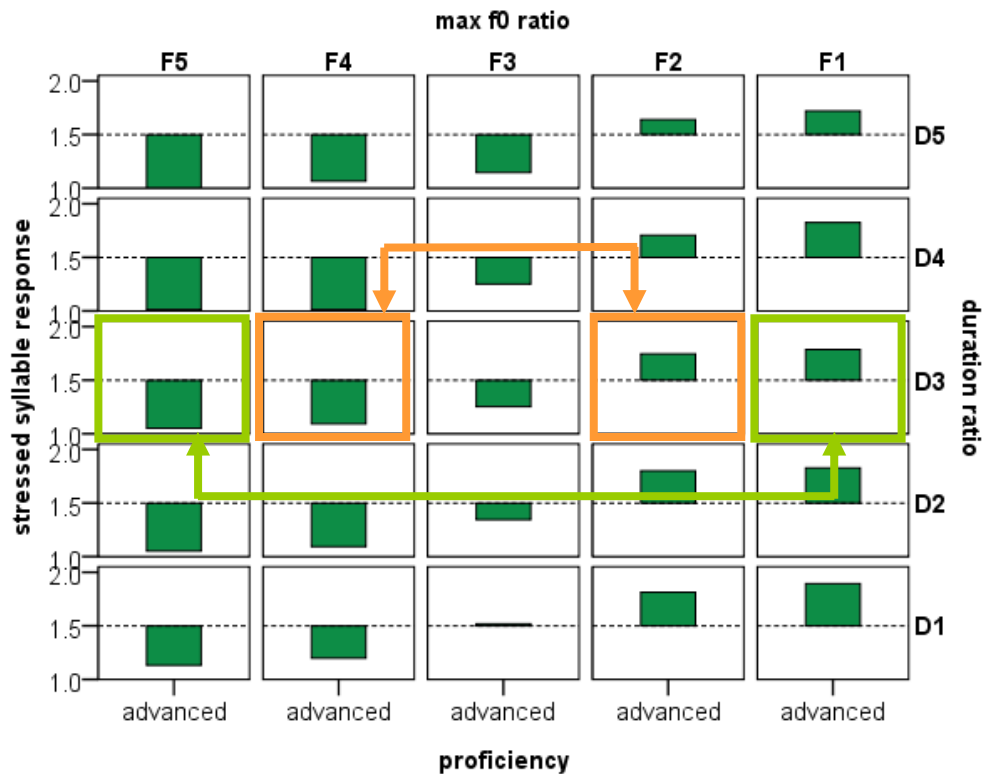


Figure 4. 11. Responses from advanced learners for all cue combinations.

I propose these results are possibly caused by three factors. Firstly, the max F0 magnitude differences between nouns and verbs in advanced learners' productions are not as large as in those of native speakers. Hence the congruence phenomenon may not be shown as clearly. Secondly, advanced learners were found to have the strongest noun-biased response pattern (see Figure 3.7). The noun-biased pattern may enhance the noun response in F5 and F4, generating more noun responses.

Finally and most importantly, I argue that the phonemic feature of F0 in Mandarin might affect the advanced learners' perception.

4.2.3 Prosodic transfer from Mandarin L1

For Mandarin L2 learners, no production asymmetry was observed in the learners' production data hence no parallel perceptual asymmetry can be claimed. Secondly, L2 learners' prosodic acquisition is believed to be affected by their first language hence the prosodic features of their L1 need to be taken into consideration in addition to the parallel pattern between productions and perception.

It has been shown that the perceptual weight of F0 height and F0 contour is correlated with the listeners' linguistic experience and that tone language users have been shown to be more sensitive to F0 contours instead of F0 height (Gandour, 1983). Based on this finding, I argue that beginning Mandarin L2 learners are not sensitive to the max F0 register changes due to their perceptual weight in Mandarin L1. They instead focus more on duration. Furthermore, advanced learners, once aware of F0 as a relevant perceptual cue in English, are capable of utilizing this cue to a great extent and base their stress perception on F0 cues, which can be seen from the categorical perception across max F0 ratios (cf. Figure 4.6).

Mandarin L2 learners' performance is greatly influenced by the phonemic role of F0 in Mandarin. In beginning learners' production, they are found to consistently use the same magnitude of mean F0 and max F0 across noun and verb readings (cf. Figure 4.3 as shown by the similar stressed-to-unstressed vowel ratios). I argue that the phonemic feature of F0 in the learners' L1 forces them to use this feature in a very consistent manner. In terms of perception, I propose that beginning learners carry over their insensitivity to F0 height (Gandour, 1983) from the perceptual habit in Mandarin L1 and hence are not sensitive to the max F0 changes. They then rely solely on duration in identifying stress.

With more English exposure, learners may become aware that F0 is one of the correlates utilized in stress production and perception and begin to use it differently

from the beginning learners. Results from the production study suggested that advanced learners, although utilizing F0 cues in verb readings, can successfully control the magnitude of these cues to match that of native speakers' (cf. Figure 4.2 advanced learners' mean F0 and max F0 ratios are between the beginning learners and native speakers).

4.2.4 The present study and current theoretical framework

The three perceptual models mentioned in section 1.1.1 - the Speech Learning Model, Perceptual Assimilation Model, and Native Language Magnet Model - all base their arguments on L2 perception and from this base make inferences about the production of second language learners. The production and perception results uncovered by the present study demonstrate a discrepancy between these two aspects - beginning learners are not sensitive to max F0 cues in perception, but are nevertheless capable of utilizing this cue in production. It is therefore proposed that second language production and perception may not be as well-aligned as in first language acquisition. Such a discrepancy raises serious questions about the reliability of using perception to predict the production of second language learners, and warrants a thorough reexamination of the topic.

In addition, the three models under discussion are primarily directed at accounting for learning at the segmental level, and may not be appropriately applied to suprasegmental learning. It is argued that prosody is utilized in linguistic and non-linguistic domains (emotion, music among others) of nearly all languages, and that as a result neither the production nor the perception of prosody is entirely novel to second language learners. For instance, the 'new' condition proposed in the Speech Learning Model and the 'Nonassimilable type' proposed by the Perceptual Assimilation Model may never occur in the learning process.

Furthermore, while stress (and pitch accent in a loose sense) is implemented by the relative magnitude differences between F0, intensity, and duration between stressed and unstressed syllables, tone is realized in the syllable domain. None of the three perceptual models mentioned above can account for learning across different domains. Hence, it is difficult to define the relationship between the target L2 attribute (in the present study, acquiring relative magnitude differences between stressed and unstressed syllables) and that of learners' L1 (tone implemented in each syllable).

In sum, L2 prosodic learning seems to present a complex of challenges for the existing perception models, some or all of which may be insurmountable. It is suggested that level-specific (segmental or suprasegmental) models should be proposed for these two types of learning.

4.2.5 An alternative view

An alternative view for lexical stress acquisition is to consider stressed and unstressed vowels as allophones of the same vowel. Learners need to learn that when located in an unstressed position the vowel needs to be implemented with less duration and intensity, lower F0, and with reduced vowel quality. It is proposed that once L2 learners recognize the existence of lexical stress, they can utilize duration, F0, and intensity to signal the difference between stressed and unstressed vowels. This phenomenon may have been overlooked by previous studies due to the fact that data analyses were performed with data merged across contexts.

It is argued that the *magnitude* of F0, duration, and intensity utilized by L2 learners is greatly affected by the prosodic attributes of learners' L1. The phonemic feature of F0 in Mandarin gives rise to a consistent use of F0 magnitude differences across different contexts when Mandarin speakers produce English lexical stress. It is further predicted that under the same condition (lack of overt lexical stress in L1), L2

learners with duration as a phonemic attribute in their L1 will utilize the duration correlate with a more fixed magnitude across contexts.

The difficulty in modifying vowel quality is argued to be based on three different levels. First, the learners need to acquire the concept of reducing unstressed vowels. Second, they need to correctly identify the unstressed vowel and produce it with reduction. Finally, upon reducing the vowels, the prosodic context (in the current study, initial or final position) may cause different levels of difficulty for the learners (more reduction for unstressed vowels in final position). Furthermore, due to the strong correlation between vowel duration and reduced vowels, it is predicted that learners who have acquired near-native duration differences between stressed and unstressed vowels should perform near-native vowel quality changes.

Taken together, it is argued that existing models, which were established for L2 segmental acquisition, cannot sufficiently account for L2 suprasegmental learning, and that an encompassing model for the production and perception of suprasegmental learning is necessary. This model should incorporate the phonemic features of F0, duration, and intensity as well as syllable structure and stress metricality in learners' L1, as well as L2. In addition, it needs to be able to account for the possibility of re-attunement as learner proficiency improves, as shown in the present study.

4.3. Conclusion

The acoustic correlates and perceptual cues for English stress have been widely discussed but the relationship between acoustic correlates and perceptual cues has not been thoroughly studied. The current study discovered an asymmetry in correlates realization of noun and verb pairs (e.g. *contract* and *contract*) which has been overlooked in previous studies. Results from a production and perception study revealed a congruence pattern between the utilization of acoustic correlates and the sensitivity of the perceptual system to the given cue. More specifically, the magnitude of a correlate required to obtain stressed syllable perception is parallel to the magnitude of its realization in production. Also, this pattern appears to be position-specific.

The relationship between production and perception in Mandarin L2 learners' performance are not affected by the general congruence pattern for the noun and verb due to the lack of clear asymmetry in their stress realization. However, the phonemic role of F0 in Mandarin as well as the learners' English proficiencies both play a role in stress perception.

The current study also provides some suggestions for pedagogical application and further L2 studies of prosodic transfer. Firstly, I propose that it is necessary to define the term "native-like" more specifically. Nearly all the second language studies reviewed claim that their participants' use of a certain correlate "native-like" when it has a greater magnitude in stressed than in unstressed vowels. Although in some cases, the magnitude ranges don't quite match those of the native speakers closely. Furthermore, I suggest that further studies should aim to explore the implementation of stress in various contextual positions but analyze them with caution. As shown in this dissertation, native English speakers demonstrate different implementation patterns in noun (iambic) and verb (trochaic) readings. Neglecting the contextual differences in native speakers may result in setting the incorrect norm to be compared

with the second language learners' productions. Finally, based on an understanding of the contextual differences in native norms, second language teaching should reveal this diverse pattern in English production and provide training for stressed domains in various prosodic positions.

References

- Al-Jarrah, Rasheed Saleem. 2002. An Optimality-Theoretic Analysis of Stress in the English of Native Arabic Speakers [Dec]. *Dissertation Abstracts International, A: The Humanities and Social Sciences* 63:2220-A.
- Altmann, Heidi. 2006. The perception and production of second language stress: A cross-linguistic experimental study, University of Delaware: Ph.D.
- Anderson-Hsieh, Janet, Johnson, Ruth, and Koehler, Kenneth. 1992. The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentals, prosody, and syllable Structure. *Language Learning* 42:529-555.
- Anderson, P. J. 1993. The Interstress Interval as an Indicator of Perceived Intelligibility among Non-native Speakers of English, Wichita State University, Kansas.
- Archibald, John. 1997. The acquisition of English stress by speakers of nonaccentual languages: Lexical storage versus computation of stress. *Linguistics* 35:167-181.
- Archibald, John. 1998. Second language phonology, phonetics, and typology. *Studies in Second Language Acquisition* 20:189-211.
- Beckman, M. E., and Edwards, J. . 1994. Articulatory Evidence for Differentiating Stress Categories. In *Phonological Structure and Phonetic Form: Papers in Laboratory Phonology III*, ed. P.A. Keating. Cambridge: Cambridge University Press.
- Beckman, Mary. 1986. *Stress and non-stress accent*. Foris, Holland: Dordrecht.
- Bohn, Ocke-Schwen, and Flege, James Emil. 1992. The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14:131-158.
- Bohn, Ocke-Schwen, and Flege, James Emil. 1993. Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics* 21:267-290.
- Bolinger, Dwight. 1958. A theory of pitch accent in English. *Word* 14:109-149.
- Chen, Yang, Robb, Michael P., Gilbert, Harvey R., and Lerman, Jay W. 2001. A study of sentence stress production in Mandarin speakers of American English. *The Journal of the Acoustical Society of America* 109:1681-1690.
- Cooper, Nicole, Cutler, Anne, and Wales, Roger. 2002. *Constraints of Lexical Stress on Lexical Access in English: Evidence from Native and Non-Native Listeners*.
- Coster, D. C., and Kratochvil, P. 1984. Tone and stress discrimination in normal Peking dialect speech. In *New papers in Chinese linguistics*, ed. B. Hong, 119-132. Canberra: Australian National University Press.
- Eckman, Fred R. 1977. Markedness and the contrastive analysis hypothesis'. *Language Learning* 27:315-330.
- Eckman, Fred R. 2004. From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition* 26:513-549.
- Ellis, R. 1994. *The Study of Second Language Acquisition*. New York: Oxford University Press.

- Flege, James Emil. 1987. The instrumental study of L2 speech production: Some methodological considerations. *Language Learning* 37:285-296.
- Flege, James Emil. 1995. Second-language speech learning: Theory, findings, and problems. In *Speech perception and linguistic experience: Theoretical and methodological issues*, ed. Winifred Strange, 229-273. Baltimore: York Press.
- Fry, D. B. 1955. Duration and intensity as physical correlates of linguistic stress. *Journal of the acoustical society of America* 27:765-768.
- Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1:126-152.
- Fry, D. B. 1972. French Listeners and the Tonic Accent. In *Papers in Linguistics and Phonetics to the Memory of Pierre Delattre*, ed. A. Valdman. The Hague: Mouton.
- Gandour, Jack. 1983. Tone perception in far eastern languages. *Journal of Phonetics* 11:149-175.
- Gay, T. 1978. Physiological and acoustic correlates of perceived stress. *Language and Speech* 21:347-353.
- Howie, John Marshall. 1976. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.
- Ioup, Georgette. 1984. Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. *Language Learning* 34:1-17.
- Jongman, Allard, Wang, Yue, and Sereno, Joan. 2006. Perception and Production of Mandarin Tone. In *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)*, ed. P. Li, Tan, L.H., Bates, E., and Tzeng, O.J.L. Cambridge: Cambridge University Press.
- Koo, H. -S. 1997. Umhyang chukjungkwa jikak pandane uihan hankukin yeongeoui unyul yeongu [A study using acoustic measurements and perceptual judgment to identify prosodic characteristics of English as spoken by Koreans] *Speech Sciences* 2:95-108.
- Ladefoged, Peter. 1982. *A Course in Phonetics*. San Diego: Harcourt Brace Jovanovich Publishers.
- Lado, R. 1957. *Linguistics Across Cultures*. Ann Arbor: University of Michigan Press.
- Lea, W. A. 1977. Acoustic correlates of stress and juncture. In *Studies in stress and accent*, ed. L. M. Hyman, 83-119. Los Angeles: Department of Linguistics, University of Southern.
- Lee, Borim, Guion, Susan G., and Harada, Tetsuo. 2006. Acoustic analysis of the production of unstressed English vowels by early and late Korean and Japanese bilinguals. *Studies in Second Language Acquisition* 28:487-513.
- Lehiste, Ilse 1970. *Suprasegmentals*. Cambridge, MA: MIT.
- Lieberman, Phillip. 1960. Some acoustic correlates of word stress in American English. *Journal of Acoustical Society of America* 22:451-454.
- Lin, M. C. 1988. Putong Hua shen diao de sheng xue te xing he zhi jue zheng zhao (Standard Mandarin tone characteristics and perceptual cue). *Zhongguo Yuyan* 3:182-193.

- Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of Acoustical Society of America* 35:1773-1781.
- McAllister, Robert, Flege, James Emil, and Piske, Thorsten. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30:229-258.
- McGory, Julia Tevis. 1998. Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers [Apr]. *Dissertation Abstracts International, A: The Humanities and Social Sciences* 58:3906-A.
- Moore, C. B., and Jongman, A. 1997. Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America* 102:1864-1877.
- Morton, J., and Jassem, W. 1965. Acoustic correlates of stress. *Language and Speech* 8:159-181.
- Munro, Murray J., and Derwing, Tracey M. 1998. The effects of speaking rate on listener evaluations of native and foreign-accented speech. *Language Learning* 48:159-182.
- Nguyen, Thu T. A., and Ingram, John. 2005. Vietnamese acquisition of English Word Stress. *TESOL Quarterly* 39:309-319.
- Nordenhake, M. , and Svantesson, J. O. . 1983. Duration of standard Chinese word tones in different sentence environments. *Lund University Working papers in Linguistics* 25.
- Odlin, T. 1989. *Language Transfer: Crosslinguistic Influence in Language Learning*. Cambridge: Cambridge University Press.
- Okobi, Anthony O. 2006. Acoustic correlates of word stress in American English, Massachusetts Institute of Technology: Ph.D.
- Peterson, Gordon E., and Barney, Harold. 1952. Control methods used in a study of the vowels. *Journal of Acoustical Society of America* 24:175-184.
- Ramus, Franck, Nespors, Marina, and Mehler, Jacques. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73:265-292.
- Shen, Xiaonan Susan. 1993. Relative Duration as a Perceptual Cue to Stress in Mandarin. *Language and Speech* 36:415-433.
- Sluijter, Agaath M. C., and Van Heuven, Vincent J. 1996. Spectral balance as an acoustic correlate of linguistic stress. *The Journal of the Acoustical Society of America* 100:2471-2485.
- Sluijter, Agaath M. C., Van Heuven, Vincent J., and Pacilly, Jos J. A. 1997. Spectral balance as a cue in the perception of linguistic stress. *The Journal of the Acoustical Society of America* 101:503-513.
- Stragray, J. R., and Downs, D. 1993. Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics* 21:143-163.
- Studdert-Kennedy, M., and Shankweiler, D. 1970. Hemispheric specialization for speech perception. *Journal of Acoustical Society of America* 48:579-594.

- Trofimovich, Pavel, and Baker, Wendy. 2006. Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech [Mar]. *Studies in Second Language Acquisition* 28:1-30.
- Tseng, E. 2001. *The Use of Stress and Vowel Reduction by Chinese Learners of English*, The University of Texas at Austin.
- Ueyama, Motoko. 2001. Prosodic Transfer: An Acoustic Study of L2 English vs. L2 Japanese [May]. *Dissertation Abstracts International, A: The Humanities and Social Sciences* 61:4366-A.
- van Santen, Jan PH , and Shih, Chilin 2000. Suprasegmental and segmental timing models in Mandarin Chinese and American English. *Journal of Acoustical Society of America* 107:1012-1026.
- Weinreich, U. 1953. *Language in Contact*. The Hague: Mouton.
- Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. *Journal of Phonetics* 27:55-105.
- Yang, B. 1996. A comparative study of American English and Korean vowels produced by male and female speakers. *Journal of Phonetics* 24 245-261.
- Zhang, Yanhong, Nissen, Shawn L., and Francis, Alexander L. 2008. Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *Journal of the acoustical society of America* 123:4498-4513.
- Zuraig, Wael. 2005. *The Production of Lexical Stress by Native Speakers of Arabic and English and by Arab Learners of English*, University of Kansas.

Appendices

Appendix A - Target words in context sentences.

Experiment 1

	WORD	σ^1	NOUN	VERB
1	conflict	8	There is a conflict between them.	This does not conflict with her plan.
2	contract	8	The new contract is much better.	Steel will contract when it is cooled.
3	desert	7	This desert is very hot.	It is bad to desert pets.
4	digest	8	Reader's Digest is popular.	It is hard to digest this book.
5	import	10	Bananas are an import for Japan.	We import apples from America.
6	insult	8	That is the worst insult I know.	It is mean to insult people.
7	object	9	The object of this game is to win.	I object this proposal strongly.
8	permit	9	I got my learner's permit this year.	I will not permit you to go there.
9	present	9	Jack gave me a present for Christmas.	We will present our project today.
10	progress	9	I am making progress on my book.	Children learn as they progress in school.
11	project	10	Mary started her project yesterday.	He likes to project photos on big screens.
12	rebel	10	He is a rebel in his own country.	It is unwise to rebel against him.
13	record	8	It is a record of your grades.	We record ten native speakers.
14	subject	11	It is an easy subject for him to learn.	It makes no sense to subject yourself to this.

Table A- 1. The stimuli and the carrier sentences for experiment 1.

Experiment 2

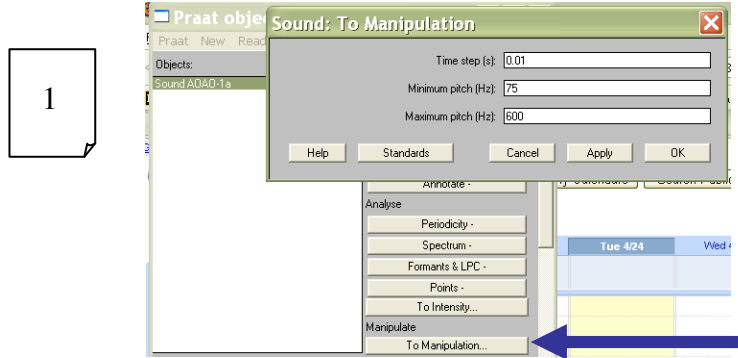
SYLLABLE	STRESSED	UNSTRESSED
[di]	The word <i>dog</i> starts with the letter "D".	That man is a wonderful daddy.
[ki]	Uncle Tom lost his car key.	The zoo has a white monkey.
[bi]	Sue caught a bumble bee.	She is a quiet baby.
[ni]	My brother hurt his knee.	I gave her some money.

Table A- 2. The stimuli and the carrier sentences for experiment 2.

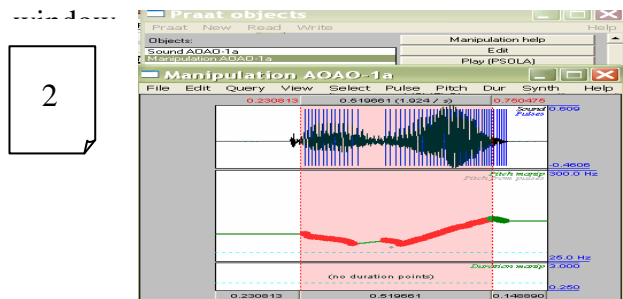
¹ Number of syllable in the sentence

Appendix B - Duration resynthesis

select the source sound and click Manipulation



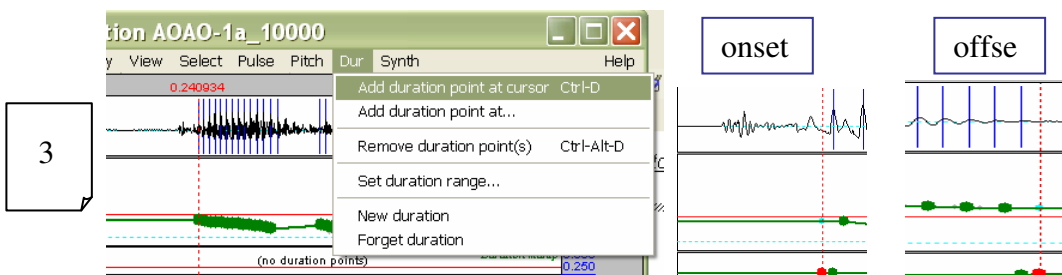
The manipulation



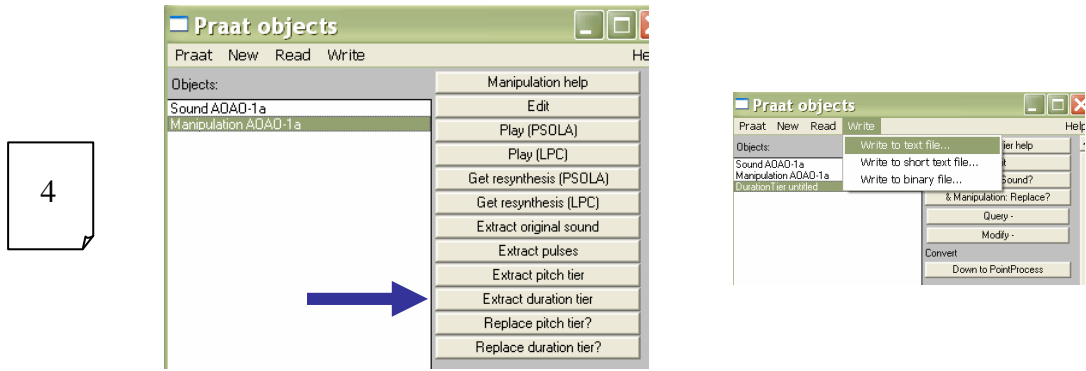
Set duration cursors at the onset and offset of the segment intend to modify
Set one cursor at the onset of the segment and another cursor at the offset.

Set one extra cursor a fraction of a second to the **left** of the **onset**
Set one extra cursor a fraction of a second to the **right** of the **offset**

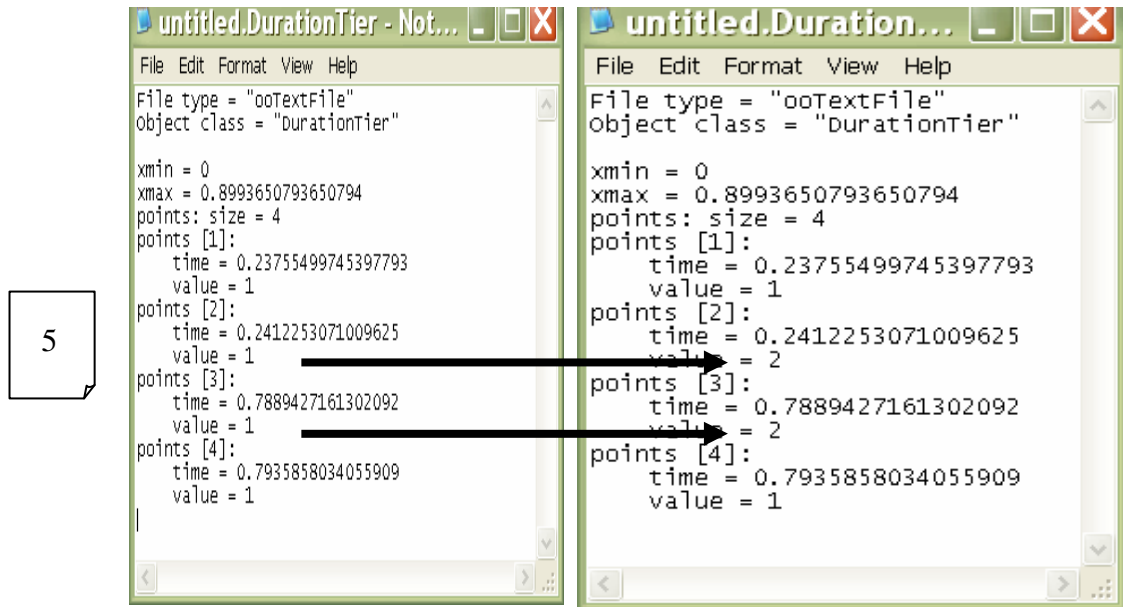
cf <http://www.ling.rochester.edu/people/cross/intro-to-praat.pdf> for further details



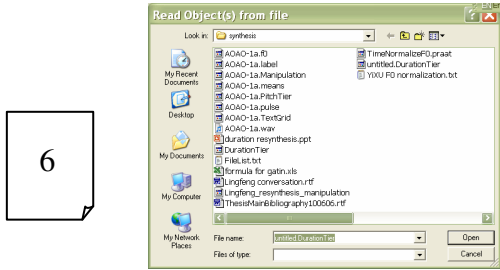
Extract the duration tier and write the duration tier to a text file
 you may change rename the duration tier in Praat and then save it.
 Rename the file in the “write to” process will make the connect
 between the duration tier with the source manipulation



Open the text file in Notepad. You may specify the ratio here.
 Example shown here will increase the vowel duration from 547 ms to 1094 ms
 If you move the duration cursor up or down during setting them and results in slight
 value change in irrelevant points, you may adjust their values here back to “1”

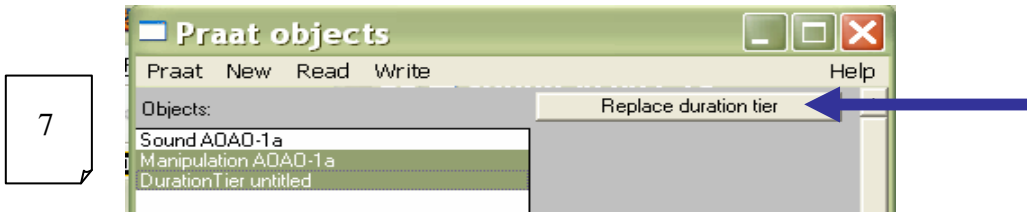


Read the modified duration tier file to Praat object window

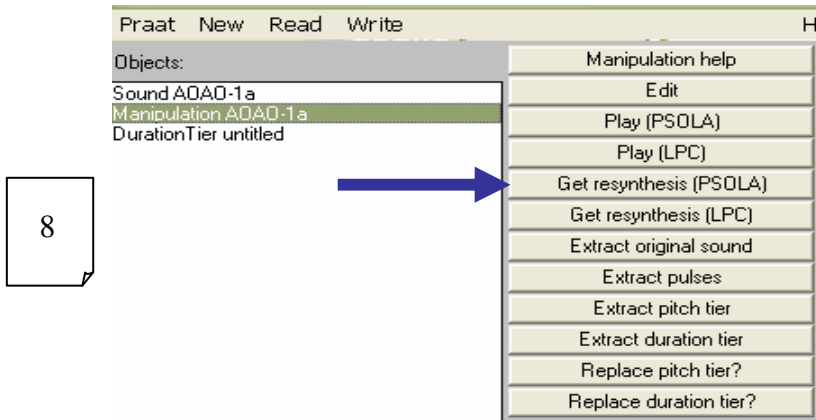


Highlight both the manipulation and the duration tier

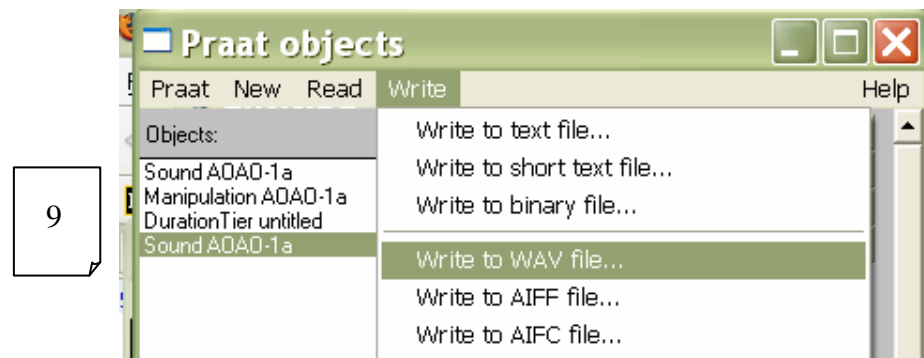
Click Replace duration tier



Highlight the manipulation and click Get resynthesis (PSOLA)



A new sound will be generated. Save the new sound file.



Appendix C - Questionnaires

1: Questionnaire for native English speakers

Questionnaire for personal information and language information

Date: _____ Email: _____ Name: _____

Gender: Female Male

1. Year of birth: _____ Where were you born?
_____(city)_____(country)
2. Where have you lived most of your life? _____(city)
_____(country) For how long? _____
3. What is your native language(s): _____
4. What is the language you feel most comfortable using? _____
5. What languages can you speak? _____
6. What foreign speaking countries have you lived/traveled in? For how long?

8. Do you have a lot of contact with people who do not speak English as their native languages? _____ if Yes, what languages do they use? _____
How much time do you spend with them per week? _____hours

2: Questionnaire for Mandarin L2 learners of English (in Mandarin)

所有您提供的背景資料都將被妥善保密，如果您需要更多空間書寫資料，可以寫在紙的背面。

1. 姓名-_____ (eg. Christina Lin):email:
2. 年齡: 性別: 男 女 出生地(城市/國家):
3. 您是學生嗎? 是 否
如果您是學生請註明您的年級:
(例如: 高中一年級; 大學三年級; 研究所一年級)
4. 你的母語是 國語 台語 台國雙語 偏台語 台國雙語 偏國語
5. 你與人溝通最自然的語言是 國語 台語 台國雙語 偏台語 台國雙語 偏國語
6. 你與父母溝通 是使用 國語 台語 台國雙語 偏台語 台國雙語 偏國語
7. 您的母親使用哪種(些)語言? 您的父親呢?.....
(請依使用該語言的比例依序列出)
8. 您幾歲開始學英文?
9. 您學習英文多久了?
10. 你曾經跟隨以英語為母語的老師學習英文嗎? 是 否
如果是，他(她)來自哪個國家 在你幾歲時教你 教你多久
11. 您覺得您的英文程度如何?
 很流利 還不錯 我學英文一段時間了， 很不好
可是說的不好
12. 您覺得您的英文聽說的程度如何_____ (請以 1-10 評分 1 為不好 10 為最好)
13. 您平日如何學習英文?
 看書
 看電影
 聽音樂
 其他.....
14. 如果您有在學校或補習班上英文課你在英文課外會說英文嗎? 是 否
如果是，您一週大概說英文幾個小時?
15. 您跟誰說英文?
 母語是英文的外國朋友
 母語不是英文的外國朋友
 公司的老闆或同事
 其他
16. 你的工作需要每天使用英文嗎? 是 否
如果是，您的工作是
17. 您會說其他的語言嗎? 是 否
如果是，您會說的其他語言是.....
18. 您覺得這個語言的程度如何?
 很流利 還不錯 我學一段時間了，可 很不好
是說的不好
19. 您曾經在台灣以外的地區居住過嗎(超過一個月)?
 否
 是 請簡短的說明您什麼時候(幾歲至幾歲)在哪個國家居住了多長的時間:
謝謝您的合作!

3: Questionnaire for Mandarin L2 learners of English (English translation)

All personal information you will provide is confidential. Feel free to use the back of the sheet if you need more room.

1. Name _____ (e.g. Christina Lin) Email: _____
2. Age: _____ Sex: male female City/Country of birth: _____
3. Are you a student currently? yes no
If yes, please indicate your current level of education:
(For example: high school- first year; college-third year)
4. What's your native language? Mandarin Taiwanese
Taiwanese/Mandarin bilingual Mandarin/Taiwanese bilingual
5. What's the language you feel most comfortable using to communicate with others?
 Mandarin Taiwanese Taiwanese and Mandarin with dominant Taiwanese
 Mandarin and Taiwanese with Mandarin dominance
6. What language do you use to talk with your parents? Mandarin Taiwanese
Taiwanese and Mandarin with dominant Taiwanese Mandarin and Taiwanese
with Mandarin dominance
7. What language(s) does your mother speak? _____ your father?
8. How old were you when you started to learn English?
9. How many years have you studied English?
10. Have you ever study English from a teacher who speaks English as his/her native
language? yes no
11. If yes, which country is he/she from? _____ How old were you when this
happened? ____ How long did s/he teach you? _____
12. How do you think of your speaking and listening ability in English?
 very fluent somewhat good so somewhat poor very poor
13. How would you rate your overall speaking and listening ability in English on a
scale of 1 to 10, 1 being very poor and 10 being very fluent _____.
14. Which of the below activities do you engage in at least once a week?

- read magazine/book in English
 - watch television/movie in English
 - listen to music in English
 - other
15. Do you speak English outside of English classes? yes no
If yes, how many hours do you speak English in a week? _____
16. Who do you speak English with?
- friends who are native speakers of English
 - friends who are not native speakers of English
 - boss or other people at work
 - other
17. Do you currently have a job that requires you to use English on a daily basis?
 yes no If yes, what job do you have?
18. Do you know any additional language? yes no
If yes, what language is it? Please rate your proficiency in that language: very fluent somewhat good so-so somewhat poor very poor
19. Have you spent any time longer than a month living in an environment where a language other than English or Japanese is the majority language?
 No. Yes. Describe briefly where, when, and for how long: _____

Thanks for your cooperation! Please take a moment now to make sure that you have filled in all the blanks.

Appendix D - Sample mean test

Statistical results of the one-sample t-tests comparing the first-to-second ratios of mean F0, max F0, intensity, and duration with 1.0

Native speakers

One-Sample Test

	Test Value = 1					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
f0_n	10.636	114	.000	.1749	.142	.207
f0_max_n	13.578	130	.000	.22694	.1939	.2600
inte_n	28.114	139	.000	.1458	.136	.156
dur_n	5.280	139	.000	.2414	.151	.332
f0_v	.398	138	.691	.0034	-.013	.020
f0_max_v	-.137	139	.891	-.00146	-.0224	.0195
inte_v	-2.916	139	.004	-.0127	-.021	-.004
dur_v	-28.997	139	.000	-.5554	-.593	-.518

Advanced learners of English

One-Sample Test

	Test Value = 1					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
f0_n	14.505	122	.000	.1706	.147	.194
f0_max_n	13.232	125	.000	.14287	.1215	.1642
inte_n	18.841	125	.000	.0733	.066	.081
dur_n	5.354	125	.000	.3059	.193	.419
f0_v	-3.115	125	.002	-.0152	-.025	-.006
f0_max_v	-4.267	125	.000	-.02495	-.0365	-.0134
inte_v	-3.152	125	.002	-.0098	-.016	-.004
dur_v	-14.413	125	.000	-.3989	-.454	-.344

Beginning learners of English

One-Sample Test

	Test Value = 1					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
f0_n	12.153	107	.000	.1827	.153	.213
f0_max_n	12.349	122	.000	.12059	.1013	.1399
inte_n	11.822	125	.000	.0677	.056	.079
dur_n	2.764	125	.007	.1335	.038	.229
f0_v	-7.363	123	.000	-.0687	-.087	-.050
f0_max_v	-9.903	125	.000	-.07729	-.0927	-.0618
inte_v	-5.637	125	.000	-.0223	-.030	-.014
dur_v	-11.375	125	.000	-.2903	-.341	-.240

Appendix E - Results overall analysis (four-way ANOVA)

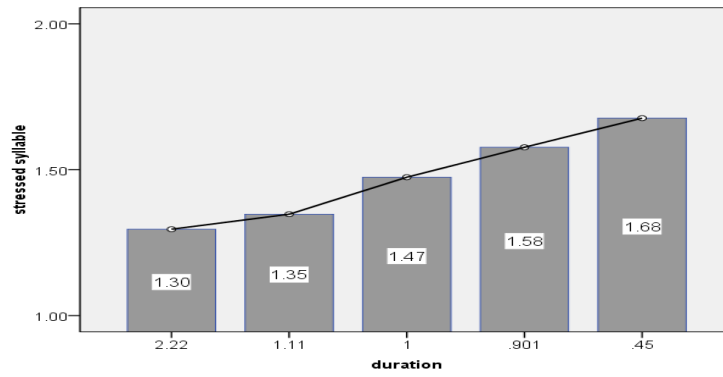


Figure A- 1. Responses for first-to-second duration ratios.

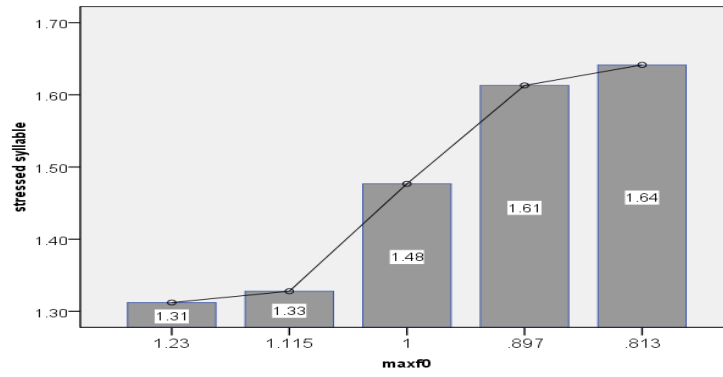


Figure A- 2. Responses for first-to-second max F0 ratios.

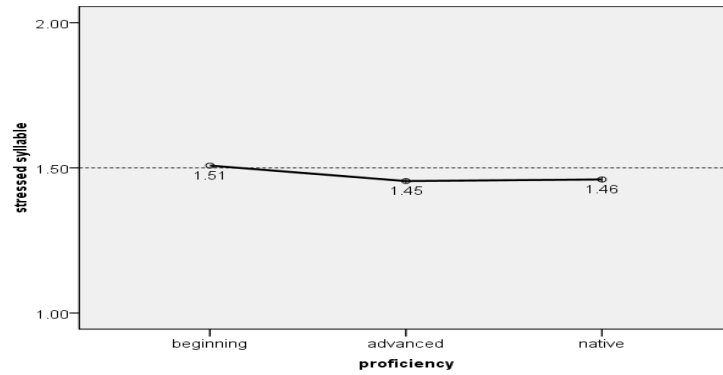


Figure A- 3. Responses across proficiency groups.

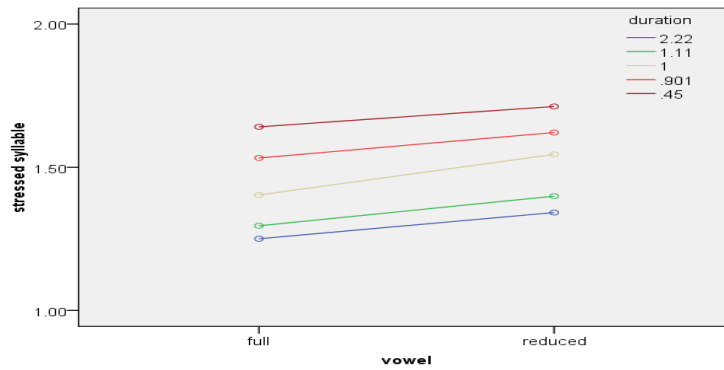


Figure A- 4. Responses for full and reduced vowel across five duration ratios.

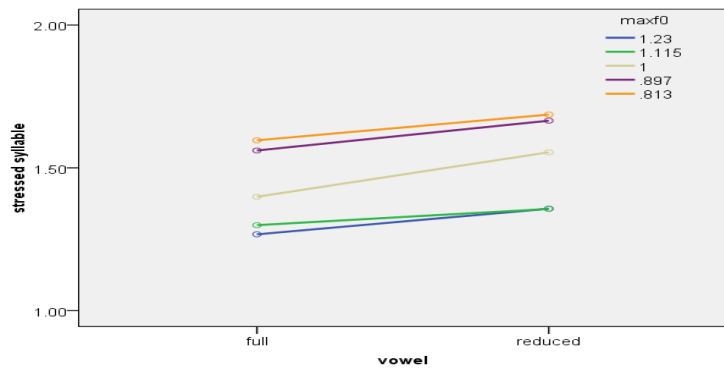


Figure A- 5. Responses for full and reduced vowel across five max F0 ratios.

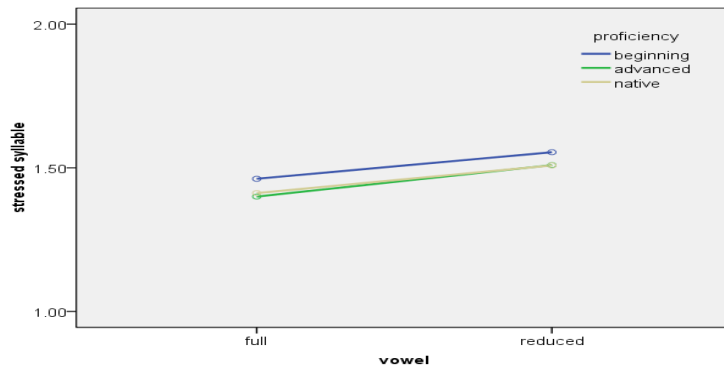


Figure A- 6. Responses for full and reduced vowel across three proficiencies.

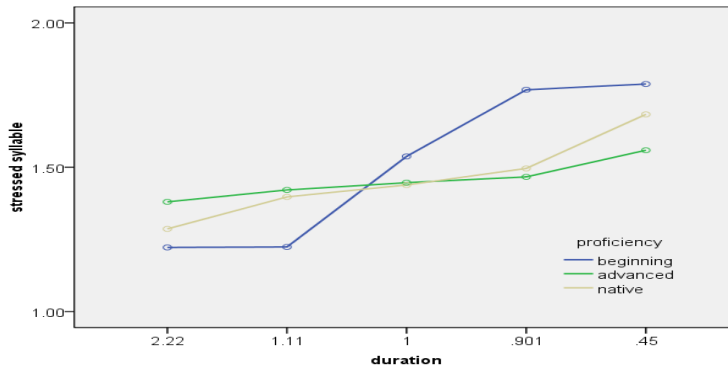


Figure A- 7. Responses for five duration ratios from three proficiencies

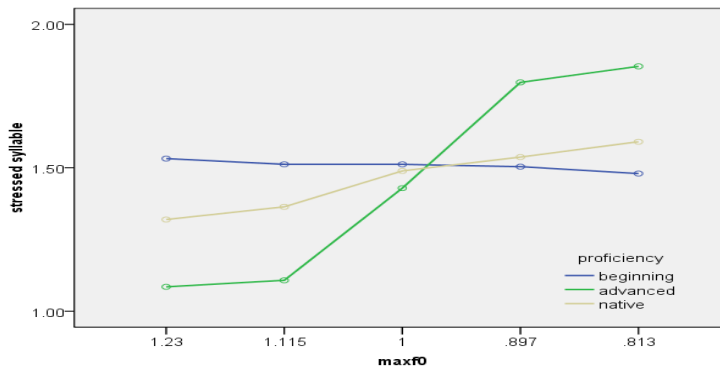


Figure A- 8. Responses for five max F0 ratios from three proficiencies

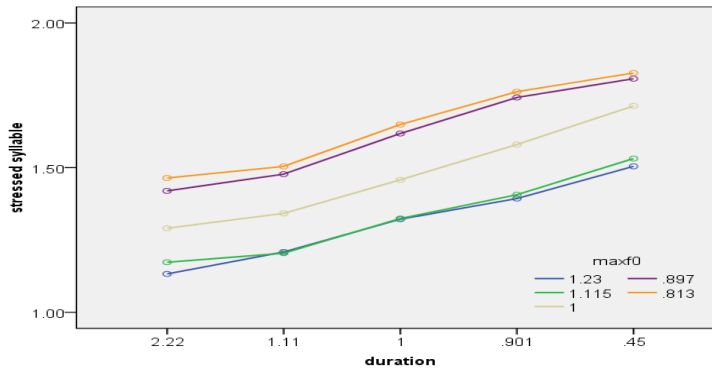


Figure A- 9. Responses with duration ratios along the x-axis and max F0 ratios along the x-axis (bottom).

Appendix F - F0 contour from each individual speaker

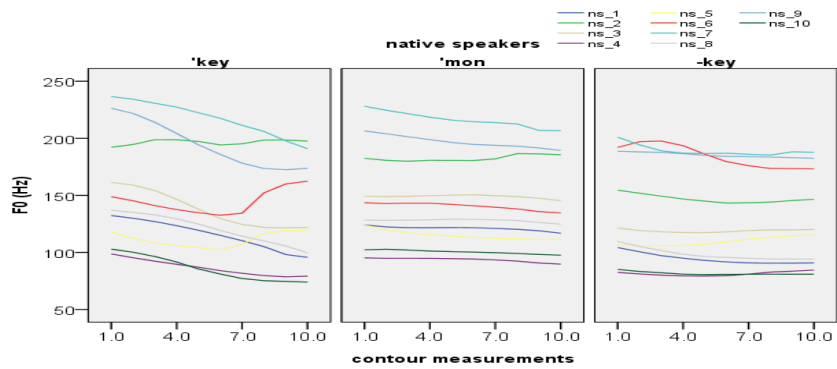


Figure A- 10. F0 contours from each native speaker across three syllable types.

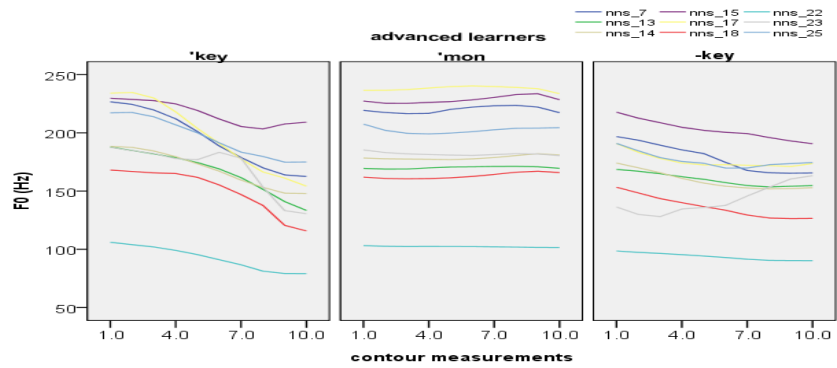


Figure A- 11. F0 contours from each advanced learners across three syllable types.

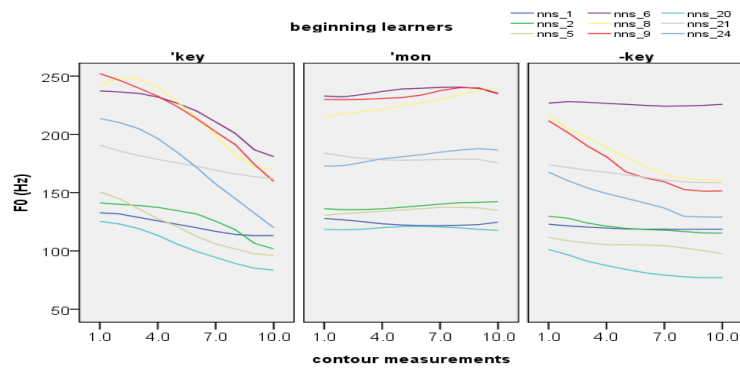
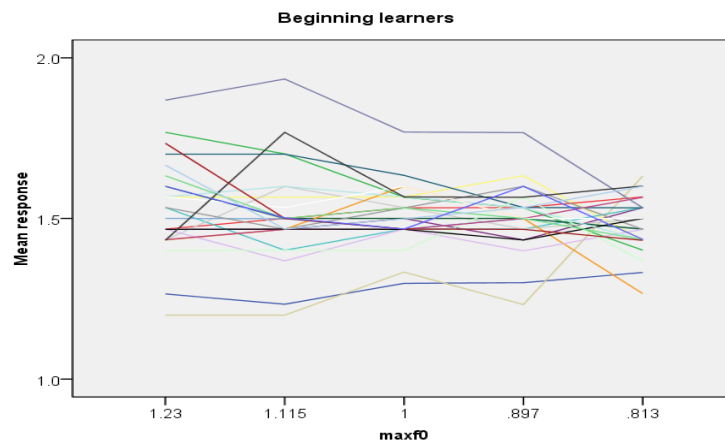
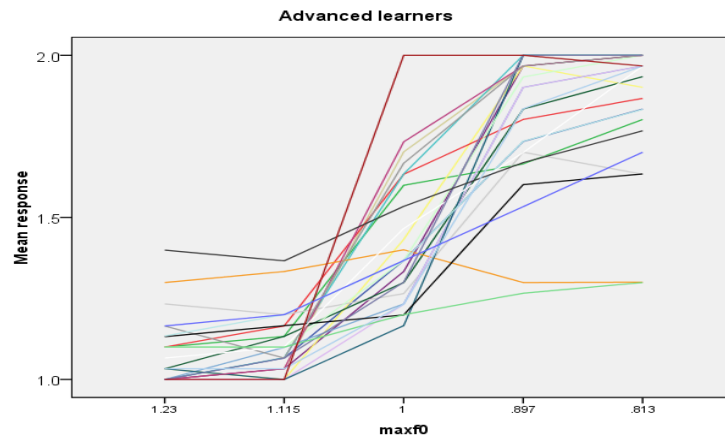
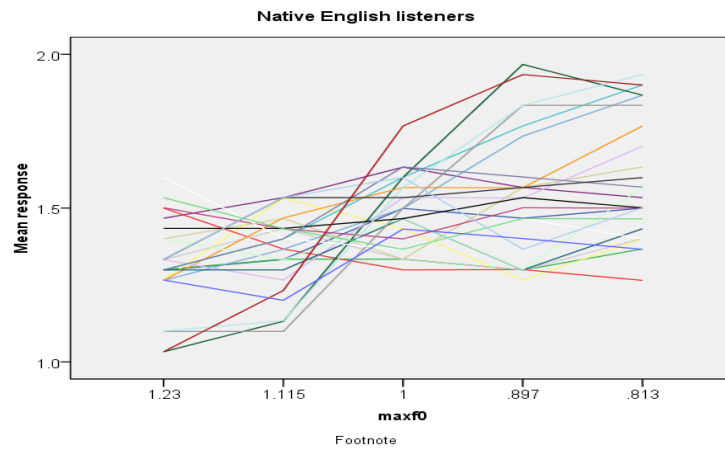
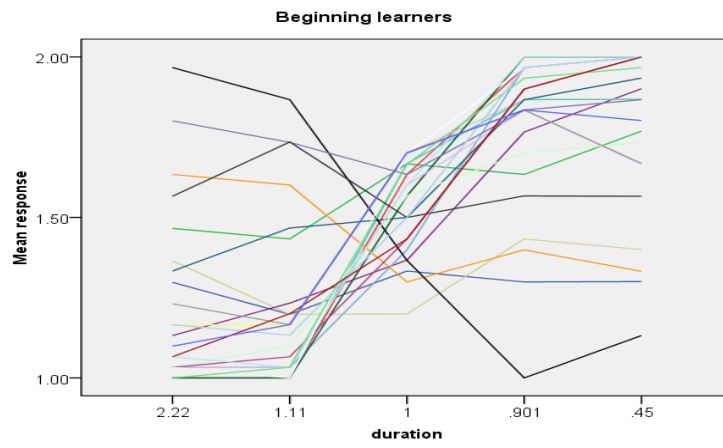
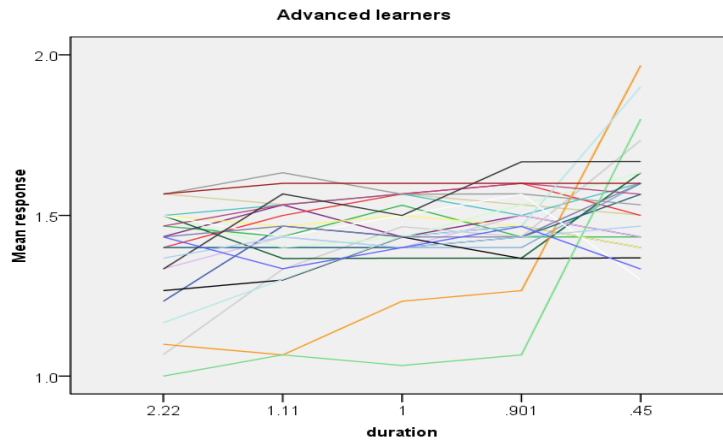
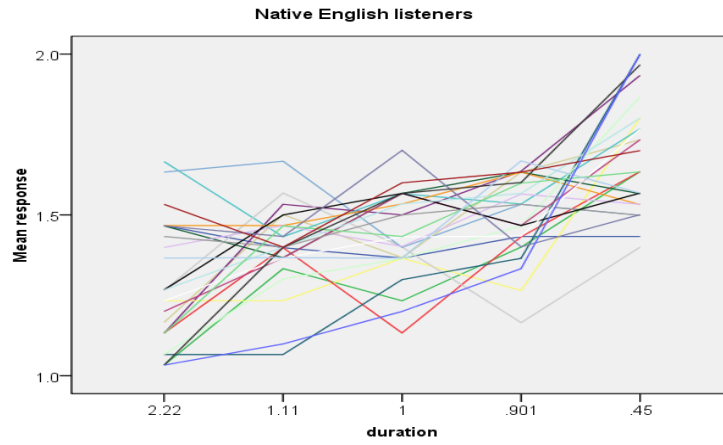


Figure A- 12. F0 contours from each beginning learners across three syllable types.

Appendix G - Results of perception study from individual participant





Appendix H - Perception results in all cue combinations across three proficiencies

