# PROSODIC REALIZATION OF FOCUS IN SECOND LANGUAGE SPEECH:

# EFFECTS OF LANGUAGE EXPERIENCE

by

YING CHEN

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

September 2014

DISSERTATION APPROVAL PAGE

Student: Ying Chen

Title: Prosodic Realization of Focus in Second Language Speech: Effects of Language Experience

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

| | |
|---|---|
| Melissa Redford | Chairperson |
| Melissa Baese-Berk | Core Member |
| Vsevolod Kapatsinski | Core Member |
| Kaori Idemaru | Institutional Representative |

and

| | |
|---|---|
| J. Andrew Berglund | Dean of the Graduate School |

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded September 2014

DISSERTATION ABSTRACT

Ying Chen

Doctor of Philosophy

Department of Linguistics

September 2014

Title: Prosodic Realization of Focus in Second Language Speech: Effects of Language Experience

Prosodic focus is phonetically realized by increasing duration, F0 and intensity on the focused constituents in a sentence. In some languages, there is a concomitant compression of F0 and intensity after the focused item, referred to as post-focus compression (PFC). Southern Min is a tone language that does not have PFC, while Beijing Mandarin is a tone language that does. Like Mandarin, American English has PFC; unlike Mandarin, American English has lexical stress rather than lexical tone. The current dissertation investigated the phonetic realization of focus in second language Mandarin by Southern Min and English learners and its realization in English by Mandarin learners. Second language experience was also manipulated in each of the investigations. The findings were that younger Southern Min speakers, who used more L2 Mandarin than the mid-age and older speakers, produced substantial PFC in Mandarin. Chinese-heritage American learners, who were exposed to Mandarin earlier than non-Chinese-heritage learners, produced some PFC in Mandarin while non-Chinese-heritage learners did not produce any. Finally, Chinese students in college with longer residencies in the United States produced more PFC in English than those with shorter residencies. American English speakers were also found to have more difficulties producing contour tones compared to the high-level tone on target focused items in L2

Mandarin while Mandarin Chinese speakers had more difficulties in producing unstressed syllables compared to stressed syllables on target focused items in L2 English. Overall, the results support the Speech Learning Model prediction that similarities in L1 and L2 sound system result in difficulty acquiring L2 sounds. This may be especially true for prosody because there are interactions between word- and sentence-level patterns. The results also confirm that age of learning is especially important for native-like acquisition of an L2; however, for early learners, the amount of L2 use and the length of residence in the L2-speaking environment also clearly impact the acquisition of L2 prosody. Finally, the results suggest that production of PFC in a language that requires it provides a good index of second language speech proficiency.

CURRICULUM VITAE

NAME OF AUTHOR: Ying Chen

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Southwest University, Chongqing, China
Fujian Normal University, Fuzhou, China
Quanzhou Normal University, Quanzhou, China

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2014, University of Oregon
Master of Arts, Modern English and Linguistics, 2002, Southwest University
Bachelor of Arts, English Language and Literature, 1999, Fujian Normal University
Associate of Arts, English Education, 1994, Quanzhou Normal University

AREAS OF SPECIAL INTEREST:

Phonetics and Phonology
Second Language Acquisition
Cognitive Linguistics

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, Department of Linguistics, University of Oregon, 2008-2014

Research Assistant and Teaching Assistant, Center of Applied Second Language Studies, University of Oregon, 2007-2008

Visiting Scholar, Department of Linguistics, University of Rochester, 2005-2007

Lecturer, Department of English, Quanzhou Normal University, 2002-2005

Assistant Instructor, Department of English, Quanzhou Normal University, 1994-2002

GRANTS, AWARDS, AND HONORS:

Jacqueline Schachter Outstanding Conference Presentation Award, Prosodic realization of focus in Mandarin by advanced American learners of Chinese, oral presentation at the 163rd meeting of the Acoustical Society of America, Department of Linguistics, University of Oregon, 2013

Gary E. Smith Summer Professional Development Award, enrollment in the Summer Institute of Linguistics Society of American at University of California Berkeley, Graduate School, University of Oregon, 2009

Graduate Teaching Fellowship, Department of Linguistics, University of Oregon, 2008-2014

Graduate Teaching Fellowship, Center for Applied Second Language Studies, University of Oregon, 2007-2008

University Research Grant, A contrastive study of sound system in British English and Quanzhou Southern Min, Quanzhou Normal University, 2003-2006

University Teaching Innovation Grant, An innovative teaching method: tutorial training on English pronunciation for Southern Min speakers, Quanzhou Normal University, 2003-2005

First Prize of Teaching Achievement, Tutorial training on English pronunciation, Quanzhou Normal University, 2004

Annual Excellent Article Award, A contrastive analysis of rhythm in English and Chinese, Foreign Language and Literature Association of Fujian Province, 2002

Second Prize of University Teaching Contest, English Pronunciation, Quanzhou Normal University, 2001

Excellent Graduate Award, Quanzhou Normal University, 1994

First Prize of University Scholarship, Quanzhou Normal University, 1993

PUBLICATIONS:

Chen, Y., & Jing-Schmidt Z. (2014). The Mandarin LVS construction: verb lexical semantics and grammatical aspect. *Cognitive Linguistics*, *25 (1)*, 1-27.

Chen, Y. (2013). Theoretical models and pedagogical implications of second language speech perception (In Chinese). *Journal of Foreign Languages*, *37 (3)*, 68-76.

Chen, Y., Kapatsinski, V., & Guion-Anderson, S. (2012). Acoustic cues of vowel quality to coda nasal perception in Southern Min. *Proceedings of the 13th Annual Conference of the International Speech Communication Association*, 1412-1415.

Chen, Y., Guion-Anderson, S., & Xu, Y. (2012). Post-focus compression in second language Mandarin. *Proceedings of the 6th International Conference of Speech Prosody*, 410-413.

Chen, Y. (2012). Prosodic variations of focus on Mandarin monosyllabic words. In X. Wen & C. Liu (Eds.), *Perspectives on Language* (pp. 107-130), Beijing: Foreign Language Teaching and Research Press.

Chen, Y., & Guion-Anderson, S. (2011). Perceptual confusability of word-final nasals in Southern Min and Mandarin: Implications of coda nasal mergers in Chinese. *Proceedings of the 17th International Congress of Phonetic Sciences*, 464-467.

Chen, Y. (2007). A preliminary contrastive analysis of speech sounds in Southern Min Chinese and British English by Praat (In Chinese). *Journal of Quanzhou Normal University*, *25 (5)*, 64-69.

Chen, Y. (2006). Effects of suprasegmentals on ambiguity in English and Chinese (In Chinese). *Journal of Jimei University, 3: 9 (2)*, 67-73.

Chen, Y. (2004). A contrastive analysis of rhythm in English and Chinese (In Chinese). *Journal of Xi'an International Studies University, 34: 12 (2)*, 1-4.

Chen, Y. (2004). A functional comparison between English stress and Chinese neutral tone (In Chinese). *Journal of Quanzhou Normal University*, *22 (3)*, 117-124.

Chen, Y. (2001). A pragmatic analysis on nucleus choice in English intonation-units (In Chinese). *Journal of Quanzhou Normal University*, *19 (3)*, 142-145.

ACKNOWLEDGMENTS

This dissertation would not have been finished without the encouragement, support and help of many of my professors, colleagues and friends, especially after I lost my former advisor, academic model, life inspiration and best friend, Professor Susan Guion Anderson, in the fifth year of my PhD program.

First of all, I would like to express my sincere appreciation to my current advisor, Professor Melissa Redford, for her taking over advising me with no hesitation after Susan left us and supporting me to continue working on the dissertation topic and experiments that Susan had directed me to. I am always impressed by Lisa's profound knowledge and insight in phonetics and psycholinguistics and inspired by her enthusiasm and experience in research. Whenever I encountered difficulties in dissertating, Lisa was able to point out my problems promptly and give me valuable suggestions. I cannot imagine how I could have completed this dissertation without Lisa's guidance, advice and assistance.

My hearty gratitude also goes to my dissertation committee: Professor Melissa Baese-Berk, Professor Vsevolod Kapatsinski and Professor Kaori Idemaru. I thank Melissa for her expertise in acquisition of second language phonology that filled in the gap Susan left for my dissertation. I admire Volya's talents and accomplishments in linguistics and statistics and appreciate his patience and responses to all of my questions. I remember Kaori was the professor who first taught me to use Pratt scripts and consistently provided me with useful references and suggestions for my projects. The precious comments of my committee enabled the completion and a better version of this dissertation.

I am also deeply indebted to my mentors Professor Zhuo Jing-Schmidt and Professor Eric Pederson and other professors of mine— Professor Scott DeLancey, Professor Spike Gildea, Professor Doris Payne and Professor Cynthia Vakareliyska, for their instruction and encouragement in both my study and life these seven years.

My special appreciation also goes to the professors in the Department of Linguistics at the University of Rochester— Professor Greg Carlson, Professor Christine Gunlogson, Professor Joyce McDonough, Professor Scott Paauw and Professor Jeff Runner, for their direction to theoretical linguistics when I was a visiting scholar there for two years.

I would like to thank my academic sister Julia Trippe, for her encouragement and help out of academia, and other academic siblings— Danielle Barth, Brian Butler, Wook Kyung Choe, Chris Doty, Kyoungho Kang, Linda Konnerth, Grace Oh, Paul Olejarczuk, Anna Pucilowski, Shahar Shirtz, Irina Shport, Hema Sirsa, Hideko Teruya and Prakaiwan Vajrabhaya, for their care of me and sharing my ups and downs on the way to the PhD.

Last but not least, without the love and support of my family, my dream of PhD would not come true. I thank my brother Wei Chen for taking care of my parents when I am not at home. I thank my mother Cuihua Zeng for not only giving the life to me but also looking after my father day and night since he got sick sixteen years ago. I also want to thank my father, Yonghuang Chen, for his initial and solid instruction of the knowledge in Chinese languages to me. My utmost gratitude goes to my husband Wei Wang. Despite being in a different field, Wei can always guide me to the right path in both academia and everyday life whenever I am confused or lost. He is always the one that first enjoys any progress and achievements of mine in this long and challenging journey. I would not have completed my PhD study and even cannot survive without his endless love and support, his generosity and strength that no one else can give me, and his faith in me in the past fifteen years when we were either together or long-distance.

This dissertation is dedicated to the memory of

Susan Guion Anderson (1966-2011)

who should have lived to see it completed.

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF TABLES

# CHAPTER I

# INTRODUCTION

Focus, according to Bolinger (1972), is used to signal newness, contrast, or other special informativeness in an utterance. From a pragmatic point of view, focus is on the unpredictable element in an utterance; the complement of the presupposition, which transforms the utterance into an assertion (Lambrecht, 1994: 207). In a conversational setting, focal information refers to the most important information that the speaker gives the hearer for integration with discourse pragmatics (Dik, 1997: 326). Recent studies have found that not only does a focused constituent itself get prosodically highlighted, with the in-focus item undergoing suprasegmental increases in fundamental frequency (F0, the acoustic correlate of pitch), intensity (the acoustic correlate of loudness), and duration. Elements after the focused constituent are prosodically reduced relative to those that occur before the focused item. The reduction is called post-focus compression (PFC). PFC has been found in many non-tone languages, such as English (Eady & Cooper, 1986; Xu & Xu, 2005), Finnish (Vainio & Järvikivi, 2007), Dutch (Hanssen, Peters & Gussenhoven, 2008), and in some tone languages, like Beijing Mandarin (Jin, 1996; Xu, 1999). In non-tone languages, PFC is marked with a decrease in F0 and intensity. In tone languages, it is marked with a narrower F0 range and by a decrease in intensity.

This dissertation will investigate the second language acquisition (SLA) of prosodic focus and its concomitant, PFC, in Mandarin and English. It will also examine whether increased language experience with the second language (L2) plays a positive role in improving the production of all suprasegmentals in a language or whether some aspects are advantaged over others. Several specific factors will be examined: the influence of a first language (L1) sound system on the acquisition of L2 phonology, age of learning (AOL) L2, amount of L1/L2 use in daily life, and length of residence (LOR) in the L2 speaking environment. The results are interpreted within the framework provided by the Speech Learning Model (Flege, 1995, 2007). These specific choices are justified in the following review of work on SLA of phonology and of the major theoretical frameworks that have been proposed to account for the empirical results.

### 1.1. Second language acquisition of phonology

Research on the acquisition of second language (L2) sound systems has bloomed in recent decades along with the development of theoretical models and research technologies. This research, focused mainly on the acquisition of segments, has been motivated by two crucial questions. First, whether and how does an L1 sound system influence the acquisition of L2 phonology? Second, whether and how does L2 experience facilitate learning L2 phonology?

With regard to the first question, several decades of research indicate that the L1 phonemic inventory influences the acquisition of an L2 inventory in both perception and production. Although most of the studies have focused on the acquisition of consonants (Flege, McCutcheon & Smith, 1987; Flege, 1991; Kang & Guion, 2006) and vowels (Flege, Bohn & Jang, 1997; McAllister, Flege & Piske, 2002; Baker & Trofimovich, 2005), there are some that have investigated the acquisition of lexical tone in Mandarin and other languages. These studies are of interest here given that the acquisition of Mandarin is investigated in the present dissertation. Lexical tone is like a phoneme in that different tones give rise to different lexical meanings. Consider for example the Mandarin syllable /mi/: with Tone 1 it is, mī, ‘nap’; with Tone 2, m í, ‘riddle’; Tone 3, mǐ, ‘rice’;

and, Tone 4, mì 'honey'. Since Chao (1933) tones have been described on a 1-5 scale of relative pitch, with 1 as the lowest relative value and 5 as the highest. The shape of the contour is conveyed by sequencing the numbers. In Mandarin, Tone 1 is a high level tone with a stable pitch described as 55; Tone 2 is a rising tone with a change in relative pitch described by the values 35; Tone 3 is a low dipping tone with the relative pitch values 214; Tone 4 is a falling tone with the relative pitch values 51. Figure 1.1 shows the relative pitch and the contours of the four Mandarin tones.

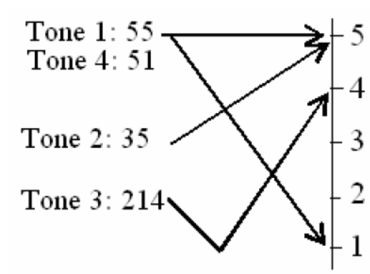


**Figure 1.1.** Pitch values of the four Mandarin tones (based on Chao, 1933).

Tone is conveyed by modulating F0, the acoustic correlate of pitch. In non-tone languages, F0 modulation is used only to convey intonation. Studies on the acquisition of lexical tone suggest that speakers of an intonation-only language focus on different aspects of F0 than speakers of a tone language. For example, Leather (1987) tested English and Dutch speakers' ability to discriminate Mandarin Tone 1 (high-level) from Tone 2 (rising). Listeners were presented with a synthetic tonal continuum with Tone 1 at one end and Tone 2 at the other. The discrimination curve was markedly less sigmoidal for native Dutch- and English-speaking listeners compared to native Mandarin speaking listeners. The location of the category boundary between Tone 1 and Tone 2 also varied more among the Dutch and English listeners compared to Mandarin listeners, suggesting an unstable or linguistically inappropriate weighting of the F0 contour, related to the absence of a tonal contrast in the native language. Gandour (1983, 1984) found that native English listeners focused more on F0 height than Mandarin speakers, who focus on both F0 height and F0 contour. Guion and Pederson (2007) among others (e.g., Massaro, Cohen & Tseng, 1985) replicated this finding and suggested that it reflects the fact that English is an intonation language rather than a tone language. Intonation contours have been described as a sequence of target H and L tones with F0 interpolated between the tones (e.g., Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986; Ladd, 1996). It is for this reason that speakers of an intonational language might attend more to absolute pitch rather than to the falling and rising contours that distinguish between lexical tones in a language like Mandarin.

Other studies show that L2 lexical tone is more easily acquired when it is present in the L1 inventory. For example, Wayland and Guion (2004) trained English speakers and Mandarin speakers to learn Thai tones and found a significant improvement in identifying the mid tone versus the low tone from the pretest to the posttest in the Mandarin group but not in the English group. Similar results were found in Wayland and Li (2008) for the same groups, and by Francis and colleagues (2008) for English- and Mandarin-speaking learners of Cantonese tones.

With regard to the question of L2 experience, decades of research has shown that age of learning (AOL) is the most important factor for second language acquisition, starting with Lenneberg (1967) who proposed the Critical Period Hypothesis (CPH), which posits a decline in language learning ability from ages 6 or 7 to puberty. Although originally proposed for first language acquisition, CPH was extended early on to account for the effect of AOL in the second language acquisition of grammar (Patkowski, 1980; Johnson & Newport, 1989; Newport, 1990; DeKeyser, 2000). For example, Johnson and Newport (1989) found that Chinese and Korean immigrants showed a very strong negative correlation between AOL and the accuracy of English grammaticality judgments prior to age 15, but no correlation after age 15. The CHP has also been proposed to account for the effect of AOL on learners' foreign accent (Scovel, 1969, 1988; Patkowski, 1990). For example, Scovel (1969, 1988) reported the inability of late language learners to achieve native-like pronunciation in L2 after age 12. According to Patkowski (1980, 1990), the critical period for learning both L2 speech and morphosyntax ends at about age 15.

A problem with many of the studies on AOL is that these have focused on immigrant learners of an L2 (e.g., Patkowski, 1980; Scovel, 1988; Johnson & Newport, 1989; Newport, 1990; Patkowski, 1990). For immigrant L2 learners, the length of residence (LOR) in the L2-speaking environment is often confounded with AOL. For example, Flege, Munro and MacKay (1995) and Flege, MacKay and Meador (1999) found a significant inverse correlation between age of arrival (AOA) and LOR in Canada. The earlier the subjects began learning English, which should contribute to accurate L2 English pronunciation, the longer they had spoken English, which might also contribute to accurate English pronunciation. Whereas Flege, MacKay and Meador showed that LOR predicted the acquisition of English vowels by Italian immigrants, Flege, Munro and MacKay found that LOR contributed little to the degree of foreign accent in English sentences produced by Italian immigrants. The latter result is consistent with the majority of studies that report no significant influence of LOR on L2 pronunciation when examining degree of L2 foreign accents (Oyama, 1976; Tahta et al., 1981; Flege, 1988; Tompson, 1991; Moyer, 1999; Piske et al., 2001). Piske et. al (2001) and Piske (2007) suggested that the contradictory results on the effect of LOR on L2 acquisition might be explained if additional years of L2 experience leads to a decrease of L2 foreign accent in the early phases of L2 learning but asymptotes in proficient L2 speakers.

Flege and Liu (2001), bothered by the conflicting results on the effect of LOR, reexamined its role in L2 acquisition by testing different groups of Chinese learners of English in three L2 tasks: identification of word-final consonants, grammaticality judgments, and listening comprehension. The Chinese learners of English, who were students and had a long LOR, performed significantly better than those with a short LOR; however, LOR did not affect performance in the Chinese learners of English who were not students. Flege and Liu suggested that the difference between the groups was due to the different quality of L2 input that the learners had received. Students likely received a substantial amount of high quality L2 input (e.g., from native speakers); nonstudents much less so since they spent more time in the Chinese immigrant community and less time with native speakers of English. Flege and Lui concluded that a long LOR does not lead directly to progress in L2 learning, but it can provide a good index of learning for those who come into regular contact with native speakers. Relatedly, Flege and Fletcher (1992) found that "years of formal education in English" correlated with AOL and LOR and had a significant influence on the degree of foreign accent in L2 English among Spanish learners of English.

Related to the quality of language input, there is language use. Some studies show that foreign accent persists even among early bilinguals who rarely use the L2 they have acquired. Flege, Frieda and Nozawa (1997) examined foreign accents of English produced by Italian immigrants who started learning Canadian English in early childhood. Even some very early bilinguals with L1 Italian and L2 English were found to speak English with a slight but detectable foreign accent while others had no detectable accent. The two groups of learners had similar AOL (5.6 vs. 5.9 years) and had spoken English for 34 years on average. It was their self-reported use of Italian that varied (3% vs. 36%). English production by subjects who used more Italian was rated as more foreign accented than English produced by those who seldom spoke Italian. Piske et al. (2001) also found an effect of use on foreign accent in two groups of late Italian-English bilinguals with similar AOL (both 20 years) and different L1 use (10% vs. 53%). Like Flege et al. (1997), they found that L1 use also affected production accuracy in late bilinguals. Guion, Flege and Loftin (2000) reported similar results for early Quichua-Spanish bilinguals. Subjects with more L1 Quichua use demonstrated stronger foreign accent in L2 Spanish than those with less Quichua use; however, the degree of foreign accent in L1 Quichua was not significantly different among the monolingual Quichua group and the two bilingual groups. Altogether, these studies reveal that amount of L1 use affects the degree of foreign accent in L2 speech.

In summary, more L2 experience results in better acquisition of the L2 sound system. Whereas AOL may be the most important factor affecting L2 pronunciation accuracy, other factors, such as LOR in an L2 speaking environment, the quality of L2 input, and the amount of L1/L2 use also impact the acquisition of L2 phonology.

## 1.2. Theoretical models of second language phonology

Different theoretical models have been proposed to account for the effects of L1 on the acquisition of L2 phonology and the role of experience in the acquisition process. Here we compare and contrast three models: the structuralist Contrastive Analysis Hypothesis (CAH) (Lado, 1957), and the phonetically-based Speech Learning Model (SLM) (Flege, 1995) and Perceptual Assimilation Model (PAM) (Best, 1995). We conclude that the most appropriate theoretical framework for the current dissertation on the L2 acquisition of prosodic focus is SLM.

### *1.2.1. The structuralist hypothesis: CAH*

The Contrastive Analysis Hypothesis (CAH) was proposed by Lado in his landmark work, *Linguistics Across Cultures* (Lado, 1957), and provided an important early framework for studies of SLA, including L2 phonology. CAH predicts specific difficulties in learning L2 based on a comparison of the L1 and L2 linguistic features and the assumption that learners transfer features of their L1 to L2 such that when the forms of the two languages are similar, positive transfer occurs, and when the forms are different, negative transfer occurs. When there is no correlation between forms in the two languages, zero transfer occurs. Positive transfer facilitates acquisition, negative transfer inhibits it, and zero transfer has no effect. For example, the neutral lexical tone in Mandarin is phonetically like the F0 associated with unstressed syllables in English, so English learners of Mandarin are expected to easily acquire the neutral tone. However, Mandarin learners of English may have more difficulties producing unstressed syllables in English because the neutral tone is realized on a full syllable in Mandarin—a syllable-timed language (Chen, 2000), so Mandarin learners have trouble rendering the shorter and more centralized vowel quality of unstressed English syllables due to negative transfer. With regards to zero transfer, because there are no English (intonational) correlates with the Mandarin dipping lexical tone, no transfer is expected and so English learners of Mandarin will acquire the dipping tone more easily than if negative transfer were to apply, but they are expected to have more difficulties with its acquisition than in the case of positive transfer.

Soon after CAH became popular, its explanatory power was criticized because transfer could not account for many of the errors that L2 learners made. For example, Oller and Ziahosseiny (1970) found that foreign students whose L1 employed a Roman alphabet produced more spelling errors in English than those whose L1 writing system was non-Roman. To explain the negative effect of what should be a positive transfer they proposed the notion of interference, which is that when the L2 has similar structures to the L1, these will be more difficult to acquire than when the L2 has wholly different structures from the L1.

Whereas CAH is still useful in that it provides teachers with a basic and easy-to-grasp understanding of their students' language acquisition problems (River & Temperley, 1978: 152), empirical studies in the past several decades have shown that the simple structuralist comparisons of L1 and L2 phoneme inventories to predict transfer or interference effects oversimplifies the problem of L2 acquisition (Bohn, 1995, 2007). In particular, transfer and interference cannot explain the details of what is or is not attended to in learning and it cannot account for the difference in the success of the acquisition of certain L2 sounds among learners from similar L1 language background. Consider, for example, Bohn's (1995) investigation of Spanish-, German-, and Mandarin-speaking learners' acquisition of English vowels, which contrast in both quality and duration. The native inventories of Spanish, German, and Mandarin are different in specific ways from English vowels. In Spanish, vowels contrast only in quality; duration is not used to differentiate vowel categories. In German, vowels contrast in quality and duration. In Mandarin, vowels contrast in quality as well, but there are also duration differences associated with the four lexical tones, and every vowel is produced with a tone. Surprisingly, though, these attributes of the native vowel systems did not predict learners' perceptual discrimination of English /i/ vs. /ɪ/ and /ɛ/ vs. /æ/. Spanish learners used both spectral (= quality) and durational cues to differentiate the vowels, German learners used durational cues more than spectral ones, and so did Mandarin learners. In fact, the magnitude of the effect was much larger for Mandarin learners than for German learners. These results indicated that the structure of L1 phonology does not directly transfer to or interfere with the acquisition of L2 phonology.

It turns out instead that when transfer is observed, it is better explained by speech sound acoustics and contextual factors than by abstract features that are the mainstay of structuralist descriptions of language. For instance, Major (2008) gives the example of how a Japanese learner of English might produce *spy* as /supai/ in a citation speech because Japanese phonotactics does not allow for consonant clusters, but that same learner is likely to produce the word as /spai/ in running speech because vowel devoicing and deletion between voiceless obstruents occurs in fluent Japanese. Major considered the production of these two different forms, one nonnative and the other native-like, as both resulting from transfer, but only if transference is understood in terms of surface speech and language behavior. Next, we introduce two influential models of L2 phonological acquisition that are based on surface behavior, that is, on phonetics rather than on phonology.

*1.2.2. Phonetically-based models: SLM and PAM*

Flege's (1995, 2007) Speech Learning Model (SLM) formalizes the hypothesis that perception precedes production in L2 acquisition. Specifically, the hypothesis is that accurate perceptual targets are critical to guide the sensorimotor learning of L2 sounds. Accordingly, L2 phonological acquisition is hypothesized to be influenced by sound similarities across languages, albeit in the opposite direction suggested in CAH. Learners are hypothesized to more easily acquire L2 sounds that are more different from L1 sounds in acoustic-phonetic space. This is because, in SLM, L1 and L2 phonetic categories are established within a common acoustic space. Further, the mechanisms and processes of L1 sound learning are assumed to remain intact over the learners' lifespan, allowing these to be applied to L2 sound learning. Learners are argued to perceive L2 sounds by relating them to L1 sounds at a position-sensitive allophonic level; that is, at a phonetic rather than phonemic level.

By focusing on the acoustic-phonetic level of speech sound acquisition, SLM provides a way to understand the interesting variability that is observed in L2 phonological acquisition. For example, Japanese speakers have difficulty perceiving and producing English /1/ and /ɹ/ because there is only one liquid in Japanese. Nonetheless, Japanese learners of English perceive and produce English liquid more accurately in word-final than world-initial position. Research within an SLM framework seeks an explanation for this behavior in the acoustics. The explanation for the asymmetry of /1/

and /ɹ/ acquisition by word position is that the acoustic difference between English /l/ and /ɹ/ is more robust in word-final than in word-initial position (Strange, 1992).

Although learners are assumed to establish L1 and L2 sound categories in the same acoustic-phonetic space, the hypothesis in SLM is that proficient bilinguals do their best to maintain contrasts between L1 and L2 phonetic categories. Separate categories are possible when the L2 sound is phonetically distinct from the L1 sound. When this is not the case, an L2 category will be perceptually assimilated to the L1 category. A case study by Mack (1990) of a French-English bilingual 10-year-old boy illustrates how L1 and L2 categories influence one another. The boy produced /b, d, g/ with short-lag VOT in both languages and /p, t, k/ with VOTs too long for both French and English compared to native monolingual speakers of these two languages. This result shows that although the boy managed to maintain the voicing contrast in both languages with a short/long lag distinction, voiceless VOT values were non-native in both languages due, presumably, to a merged voiceless category. Similar intermediate patterns of vowel duration, closure duration and VOT have been described by Wang and Behne (2007) for Mandarin learners of English.

SLM has also been used to understand how phonological representations change with L2 experience. The hypothesis is of an age-related decrease in the discrimination of L1 and L2 sounds and between L2 sounds that are non-contrastive; that is, SLM predicts an effect of AOL based on perception, i.e., the discriminability of phonetic differences between L1 and L2 sounds or between two L2 sounds that are not contrastive in L1. In this way the prediction from SLM differs from that of the Critical Period Hypothesis (Scovel, 1969, 1988; Patkowski, 1990), which posits a genetically-linked decrement in language learning. According to SLM, even adults can acquire L2 sounds in a native-like manner given sufficient L2 experience. SLM encourages researcher to seek other factors in addition to AOL to explain success in L2 learning. These factors, which were reviewed above, include L1/L2 use, LOR, quantity and quality of L2 input. Other research in this framework suggests a positive role for special training of perception and production in the acquisition of L2 sounds (Piske, 2007).

Another well-known framework in L2 speech learning is the Perceptual Assimilation Model (PAM) proposed by Best (1995). PAM assumes that listeners' nonnative speech sound perception is based on the similarities and differences in how nonnative and native sounds are produced. Like SLM, it provides a way to understand non-native sound assimilation to native categories. Unlike SLM, perception and perceptual categories are understood as gesture-based, rather than as acoustically-based. Like SLM, PAM predicts that discrimination of non-native sound contrasts increases as similarity to the native speech sounds decrease. For example, although Zulu click consonants are non-speech sounds in English, American English speakers are very accurate in discriminating these sounds (Best, McRoberts & Sithole, 1988). By contrast, discrimination of a velar voiceless aspirated vs. ejective stop is significantly lower, presumably due to assimilation to the English /k/ category (Best, 1995).

PAM has been widely applied in perceptual studies of nonnative sounds by naïve listeners. By contrast, SLM has focused on experienced learners. Best and Tyler (2007) extended PAM to account for similarities and differences between inexperienced and experienced listeners (PAM-L2). Best and Tyler argued that "perceptual learning abilities are available to adults learning an L2 as to children learning an L1 or L2", which is similar to the theme of SLM. Also like SLM, PAM-L2 predicts that L2 phonetic categories are formed in the phonological space shared by the L1 and L2.

Despite the many similarities between SLM and PAM-L2, the current dissertation adopts SLM to discuss the acquisition of L2 prosody for two principle reasons. First, there is no clear gestural vocabulary for the kinds of prosodic phenomena that are investigated here, but there are well defined acoustic correlates of these phenomena. Second, unlike PAM, SLM has already been extended from L2 acquisition of segments to L2 acquisition of prosody. For example, McGory (1997) used SLM to explain the English production of lexical stress by Korean and Mandarin speakers. That said, extensions of SLM to L2 prosody at the phrase-level remain rare.

**1.3. Prosodic focus in second language speech**

There have been only a few studies related to the second language acquisition of prosody. In her dissertation, McGory (1997) investigated the production of intonational prominence contrasts in American English by Mandarin and Korean speakers differing in amounts of L2 English experience. She examined word pairs differing in the location of stress (e.g. “memorize” vs. “memorial”) and produced in statements and questions, and in in-focus, pre-focus and post-focus contexts. F0 timing, F0 change and duration were measured in the target words. Results indicated that, unlike native English speakers, Mandarin and Korean learners of English did not vary F0 values of stressed syllables according to the intonational context. They also produced a higher F0 in stressed than in unstressed syllables in both focused and non-focus contexts, unlike native English speakers. The degree of this tendency varied with the learners’ first language and L2 English experience. The results also showed the influence of L1 on the duration pattern in

L2 English. Specifically, Korean learners had more difficulty acquiring English-like reduction patterns than Mandarin speakers; Mandarin speakers overshot the English target, producing unstressed syllables that were shorter in duration than native English speakers. Neither learner group had difficulty producing the stressed syllables of English.

He, Hanssen, van Heuven and Gussenhoven (2011) examined prosodic focus in Dutch produced by northern Mandarin speakers, whose Mandarin has a similar pattern of prosodic focus including PFC as in Dutch. They found that, compared to native Dutch speakers, Mandarin learners of Dutch did not show a regular pattern of prosodic features in broad, narrow or corrective focus. Intriguingly, the Mandarin speakers' productions did not vary with their proficiency in Dutch. The authors concluded that the L1 positive transfer (as reviewed in Section 1.2.1) did not work in the realization of prosodic focus in L2 Dutch by L1 Mandarin speakers.

Wu and Chung (2011) examined the production of prosodic focus in both English and Cantonese by simultaneous bilinguals of these two languages. They found that the English-Cantonese bilinguals produced prosodic focus in English in the same way as monolingual English speakers: eight out of ten of the bilinguals had PFC in their English. Their Cantonese focus production was slightly different from that of monolingual Cantonese speakers; however, they were native-like in their production of the most salient focus marker--lengthening of focused words. None of the English-Cantonese bilinguals produced PFC in Cantonese, which does not occur in Cantonese in any case. The authors' conclusion was similar to He et al. (2011), namely, that PFC does not easily transfer from language to language even in  simultaneous bilingualism, the most intimate form of language contact.

According to SLM (Flege 1995:239), “the mechanisms and processes used in learning the L1 sound system, including category formation, remain intact over the life span, and can be applied to L2 learning;” and “phonetic categories established in childhood for L1 sounds evolve over the life span to reflect the properties of all L1 or L2 phones identified as a realization of each category.” From these two postulates we might infer that any L1 sound category can be expanded to incorporate a related L2 category. For example, in the common phonological space where L1 and L2 prosodic focus is represented, the acoustic-phonetic patterns that vary with sentence location relative to focus (pre-focus, in-focus, and post-focus) will be represented. The same mechanisms and processes that allow these to be realized in the L1 should transfer to their realization in the L2. So, for example, Beijing Mandarin learners of English should be able to transfer the realization of PFC from their L1 to the L2, since both languages include this acoustic-phonetic pattern as a feature of the category “prosodic focus.” However, He et al.’s (2011) results suggest that PFC does not transfer from one language to another even among high-proficiency learners. This relevant study raises several research questions regarding the acquisition of the acoustic-phonetic patterns associated with the realization of prosodic focus that the current dissertation will explore.

This dissertation presents an investigation of the acquisition of prosodic focus, including PFC, by learners whose L1 shares some features and differ in others. The questions addressed are as follows: How do speakers of a tone language without PFC (L1 Southern Min) behave when learning another tone language with PFC (L2 Beijing Mandarin)? How do speakers of a non-tone language with PFC (L1 American English) behave when learning another PFC language with tone (L2 Beijing Mandarin)? Finally, how do speakers of a tone language with PFC (L1 Beijing Mandarin) behave when learning another PFC language with no tone (L2 American English)? In exploring these questions, the dissertation explores the interaction between the acquisition of phonological categories at both the word (lexical tone, lexical stress) and sentence (prosodic focus) level. This is because sentence-level prosody is realized via changes to word-level suprasegmentals. Three factors of L2 experience: age of learning (AOL) L2, amount of L1/L2 use in daily life, and length of residence (LOR) in the L2 speaking environment, are investigated for insight into their effect on the acquisition of L2 prosodic focus.

# CHAPTER II

# TONES AND INTONATION IN MANDARIN AND ENGLISH

Prosodic focus has been defined as the means by which a part of a sentence is highlighted against the rest of its constituents (Bolinger, 1972; Gussenhoven, 1983; Lambrecht, 1994; Ladd, 1996; Selkirk, 2006). The focused constituent is typically acoustically marked with increasing F0, intensity and duration (Pierrehumbert, 1980, 1993; Cooper, Eady and Mueller, 1985; Eady and Cooper, 1986). These phrase-level suprasegmental patterns necessarily interact with the realization of lexical tone in Chinese and lexical stress in English. Thus, in studying the second language acquisition of prosodic focus, a phrase-level phenomenon, we cannot help but to also investigate the second language acquisition of lexical tone and/or lexical stress, which are both word-level phenomena. In this chapter, we will introduce these interactions in native speakers' Mandarin and English and seek an appropriate acoustically-based model with which to understand the interactions in both languages. We also introduce the different ways in which prosodic focus is realized across languages through changes in F0, intensity and duration. Finally, at the end of the chapter, the organization and contents of the following chapters is briefly introduced.

## 2.1 Relationship between lexical tone and intonation in Mandarin

Recall that in Mandarin (see Figure 1.1), Tone 1 is a high level tone described as having a pitch value of 55; Tone 2 is a rising tone, described as moving from a mid-range pitch to a high pitch, or 35; Tone 3 is a low dipping tone, described by a 214 sequence of pitch values; Tone 4 is a falling tone, described as moving from the highest pitch value to the lowest one, or 51.

The relative pitch values and contours in the four Mandarin tones are phonologically ideal. Their phonetic realization does not exactly match the idealized description. Figure 2.1 is adapted from Xu (1997), which illustrates the mean F0 contour of the four tones in the monosyllable /ma/, produced by eight male speakers with six times of each tone. We can see that although the onset of Tone 1 and Tone 4 are the same in Chao's paradigm, the onset of Tone 1 is actually lower than that of Tone 4 in its phonetic realization. The phonetic offset of Tone 2 is also phonetically lower than the phonological onset of Tone 4.

According to Chao (1933), four forms are combined with tones to create speech melodies in Chinese "(a) generalized raised level of pitch, (b) generalized lowered level of pitch, (c) widening of range, (d) narrowing of range". These four forms indicate that the intonation in Chinese is conveyed by not only mean pitch level but also by pitch range. Tones at the word level are the algebraic sums or resultants of phrase-level intonation and original lexical tones (Chao, 1933). Chao (1932) described the relationship between tones and intonation in Chinese as "ripples riding on the top of waves." Chao only analyzed the "tonal addition" between tone and intonation in sentence-final position, allowing for many decades of continued work and interpretation of the tone-intonation relationship by Chinese phoneticians (Wu, 1982, 1988, 1997; Shen, 1985, 1992, 1994; Shen, 1990; Shih, 2000, 2004; Cao, 2002). Several important findings have resulted from this work.

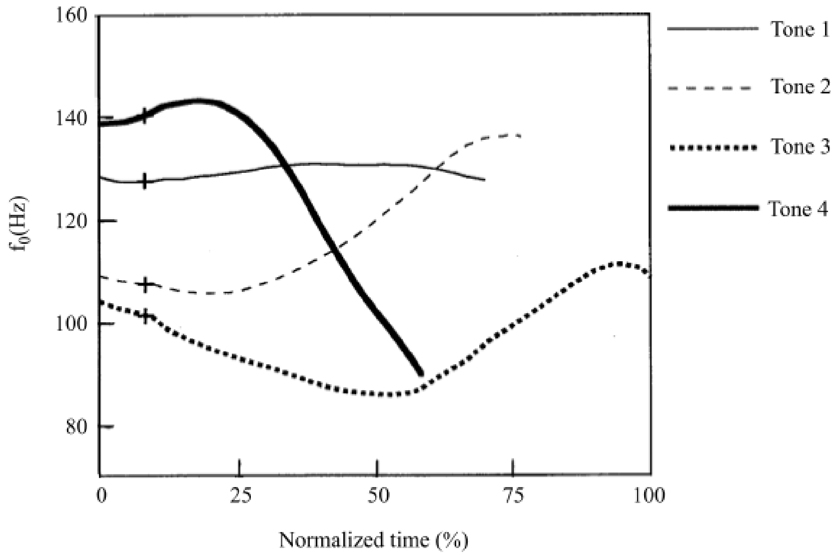


**Figure 2.1**. F0 contours of the four Mandarin tones of the syllable /ma/ produced in isolation (adapted from Xu 1997).

Wu (1982) noted that monosyllabic tones and sandhi tones on disyllabic words (including the neutral tone) are the basic units of intonation. According to Wu, intonation has some effects on these basic units, but these effects are on the pitch range rather than on the pitch contour. Wu (1988) also noted that pitch range is affected by other factors, including speech rate. Similarly, Shen (1985, 1992, 1994) argued that pitch range is the basis of Chinese intonation, but this is not the simple differences in pitch level across a sentence. Instead, the adjustment of pitch range occurs within the syllable domain. Intonation relies on a two-way adjustment: the maximum height and the minimum height of the tonal pitch range, which can be varied within a sentence.

Wu (1997) found that the algebraic sum did not describe the F0 contour as well as it described the overall or mean F0 within at least one syllable or within a higher-level constituent within the sentence (see also Cao, 2002). The F0 contour in Chinese intonation is a combination of lexical tones and coupled with changes in mean pitch. For example, Cao (2002) analyzed mean pitch across declarative and interrogative sentences in a broadcasting corpus of Mandarin. She found that there was a strong tendency for decreases in mean F0 across declarative sentences, i.e., declination, but the F0 contours associated with individual lexical tones remained intact. As for interrogative sentences, the mean F0 rose or remained relatively flat across the sentence. Again, the F0 contours associated with individual lexical tones remained intact. The relationship between tone and intonation goes to the metaphor of ripples and waves that Chao (1932) used. It also applies to the realization of prosodic focus, as will be evident below.

### 2.2. Relationship between lexical stress and intonation in English

In English, lexical stress can convey meaning at the word level. Consider pairs such as PERmit (noun) and perMIT (verb). A stressed syllable is perceived as more prominent than an unstressed syllable. Prominence is realized acoustically with increased intensity and duration (Ladefoged, 2006: 110-114). Unlike Chinese, F0 change within a syllable in an English word does not change its meaning. Instead, changes in the global F0 contour and mean F0 convey linguistic meaning at the phrase level. English is therefore considered as an intonation language. The traditional notion of "sentence stress" refers to the most prominent word in the sentence. The word that receives sentence stress is also the one that is in focus. It is "pitch accented." Pitch accented words are acoustically realized not only with increased intensity and duration but also with F0 raising or lowering, i.e., with the expansion of pitch (Ladd, 1996; Cruttenden, 1997; Ladefoged, 2006).

The autosegmental-metrical (AM) framework describes intonation in American English with reference to two basic tone levels: H and L (Pierrehumbert, 1980; Beckman

& Pierrehumbert, 1986; Ladd, 1996; Pierrehumbert, 2000). An H is identified as corresponding to an F0 peak while an L to an F0 valley. Pitch accents consist of a single tone or two successive tones that are "phonologically located on metrically prominent syllables" (Pierrehumbert, 2000: 20). The analysis and transcription conventions associated with the AM model are known as ToBI (Tone and Break Index; Beckman & Ayer, 1994). The AM theory proposes six pitch accents, which are labeled according to the ToBI conventions as follows: H*, L*, L+H*, L*+H, H+L*, H*+L (Pierrehumbert, 1980; Beckman & Pierrehumbert, 1986). The star (*) indicates which tone is aligned with the stressed syllable, especially important for bi-tonal pitch accents. The AM theory also proposes two levels of intonational phrasing: the intermediate phrase and the full intonation phrase. Each of the phrases has a boundary tone, either L or H. An intermediate phrase is marked by a phrase accent at its right or left edge (H- or L-) and an intonational phrase by a boundary tone at its right edge (H% or L%). The overall structure of English intonation is shown in Figure 2.2.

In the AM model, tonal sequences are mapped onto an F0 target by locally context sensitive implementation rules (Pierrehumbert, 2000), but the entire space of F0 targets depends on the speaker's choice of pitch range. In addition, similar to tone languages, there is a downstep rule in English. H tones successively lower after the nuclear pitch accent to the end of the phrase. The downstep rule in English is considered to be triggered by two-tone accents. American English also has an upstep rule that only applies at an intonation phrase boundary after an H pitch accent. F0 contours with an H pitch accent realize subsequent H tones with either the same pitch level or a rising level towards the end of the phrase.

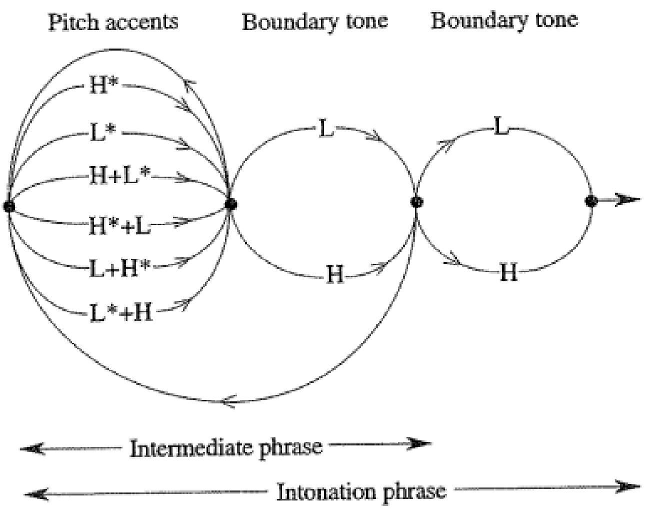


**Figure 2.2.** English intonation patterns according to the autosegmental-metrical (AM) model (adapted from Pierrehumbert 2000).

Pierrehumbert and Hirschberg (1990) described the linguistic functions of different F0 contours in English intonation based on the relationship of each contour to the mutual beliefs of the interlocutors during a discourse. They argued that the H* accent is used to mark focused information that adds to the mutual beliefs of the speaker-listener pair. Conversely, they argued that the L* accent is used to mark information that is salient to the speaker, but not additional to the discourse. The L*+H and L+H* accents mark information which is selected from a small domain of alternatives, with the L+H* marking new information and the L*+H marking already presented information. Boundary tones differ from pitch accents in that they mark the information status of the phrase as a whole rather than the status of individual elements within the phrase.

To summarize, according to AM theory, starred tones align with stressed syllables in an English sentence. These focus-marking tones are not necessarily realized with a high pitch. Instead, the F0 contour and range varies with the position of these words in a sentence and with respect to their information status in the discourse. The global F0 contour of a sentence adjusts the location and type of pitch accent assigned to individual words on the basis of the general information that interlocutors want to convey in the discourse.

### 2.3. The Target Approximation Model of tone, stress and intonation

Based on the review above, we can see some differences in the relationship between lexical tone and intonation in Mandarin compared to the relationship between lexical stress and intonation in English. In Mandarin, the local F0 contour of lexical tones remains the same whether the word is uttered in isolation or in a phrase. Intonation is conveyed based on the global mean of F0 and F0 range. In English, stressed syllables are the locus of nuclear pitch accents, but the local contour is determined by the type of pitch accent chosen to convey a particular discourse meaning. Intonation is defined by the pitch accents and boundary tones.

Interestingly, Bolinger (1972:19) used a metaphor to describe the pitch variation in English intonation similar to the one that Chao (1932) used to describe the relationship between tone and intonation in Mandarin. According to Bolinger, pitch variation is like "ripples on waves on swells on tides. In speech…the ripples are the accidental changes in pitch, the irrelevant quavers. The waves are the peaks and valleys that we call accent. The swells are the separations of our discourse into its larger segments. The tides are the tides of emotion." How can these metaphors be understood more concretely and with respect to the acoustic-phonetic realization of prosody? According to SLM, L2 phonological acquisition can only be understood with reference to perception, which is based on speech acoustics. Xu and colleagues (Xu, 1999; Xu & Wang, 2001; Liu & Xu, 2005) have provided us with a model based on a whole series of quantitative studies on F0 variations in tone and intonation in Chinese that provides a concrete characterization of tone-intonation interactions in Mandarin and other tone language. The model has also been extended to account for the relationship between accent and intonation in English, which is also of value to the present dissertation (Xu, 2005; Xu & Xu, 2005; Liu, 2009; Prom-on, Xu & Thipakorn, 2009).

Similar to the descriptive analyses provided by many Chinese phoneticians (e.g., Wu, 1981, 1988, 1997; Shen, 1985, 1994, 1995; Cao, 2002), Xu (1999) argues that the local F0 contour of a syllable is determined by its lexical tone while focus modulates the global F0 curve and affects the height and shape of the local F0 contour on the syllable. However, the arguments of Chinese phoneticians before Xu were that the global contour could be expressed as an algebraic sum of tonal contour and mean F0 for every syllable in a sentence, which ignores the effects of contextual tonal coarticulation (Xu, 1994; Xu 1997). Xu found that the tones of adjacent syllables also affect the height and shape of the lexical tone's F0 contour and the preceding tone effects the following tone more than vice versa (Xu, 1999; Xu & Wang, 2001). To accommodate this more nuanced view of tone-intonation interactions in Mandarin, Xu and Wang (2001) proposed the Target Approximation (TA) model. The model assumes that a tone has only one pitch target, defined as the smallest phonetically operable unit associated with a linguistic function and comparable to a segmental phone. The surface F0 contour is implemented based on coarticulatory constraints. Due to those constraints, the surface F0 form reflects partially its underlying pitch targets, which are as follows for Mandarin: Tone 1, Tone 3 and a neutral tone have static targets (Tone 1 = high, Tone 3 = low, neutral = mid); Tone 2 and Tone 4 have dynamic targets (rise and fall, respectively). No matter what types of pitch targets are next to each other, if the offset of the first pitch target is different from the onset of the second pitch target, the second one will be assimilated or partially assimilated to the first one due to the coarticulatory constraints in the model. See Figure 2.3, which illustrates these effects.

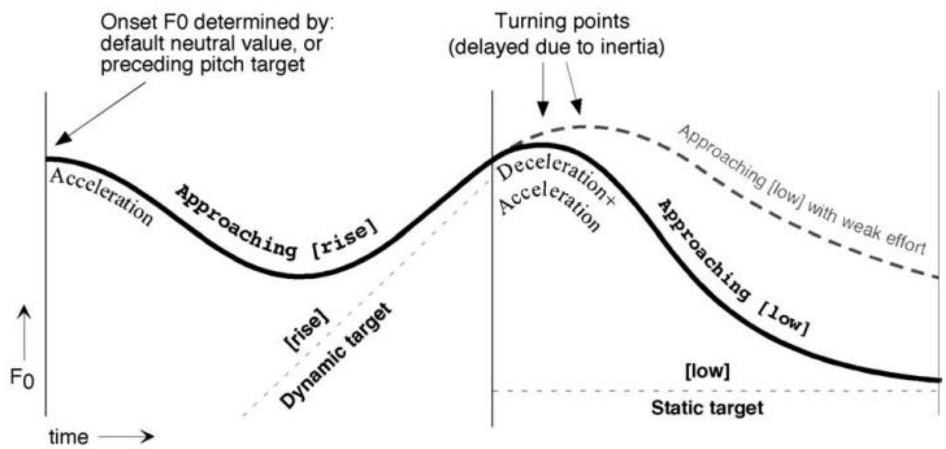


**Figure 2.3.** The Target Approximation (TA) model (adapted from Xu 2005). The vertical lines indicate syllable boundary. The straight dashed lines represent local pitch targets. The solid curve depicts the F0 contour resulting from asymptotic approximation of the pitch targets.

As shown in Figure 2.4, during target approximation, the difference between the present articulatory state and the desired state for the target determines the direction and speed of further F0 movement. However, the implementation of the previous target carries over an influence to the initial articulatory state for the implementation of the current target. This carryover influence diminishes over time as the current target is being approached. The approximation of the next target starts as soon as the current syllable is over, but not any time sooner (Xu, 2005).

Besides the local pitch targets that are the lexical tones, the surface F0 contour in TA is also affected by non-local factors, such as focus and topic initiation. The non-local factors may specify the pitch range over which local pitch targets are implemented. Focus results in an expansion of tonal contours for the item in focus and compression after the focused items. Tonal contours before the focus do not deviate much from the neutral-focus condition. Topic initiation results in a decline of mean F0 over a declarative sentence, in accordance with observations of declination in Mandarin. Liu and Xu (2005) further confirmed that focus generated the same pitch range modification in interrogative sentences as in declarative sentences. This dissertation adopts the assumptions that Xu and colleagues have formalized in the TA model. Specifically, we assume static targets for Tone 1 and Tone 3, dynamic targets for Tone 2 and 4, that coarticulatory constraints will be evident in the shape of local (tonal) contours, and that the local contours will be effected by changes in pitch range due to focus and by mean F0 due to position within a sentence.

In addition to providing a good account of the phonetic realization of tone and intonation in Mandarin, the TA model is an attractive framework within which to study L2 acquisition of prosodic focus because it also generalizes to English. Xu and Xu (2005) pointed out that the AM model reviewed earlier in this chapter and other modeling efforts (e.g., the Fujisaki Model, 1983) do not provide a mechanism for alignment of F0 peaks and valleys with the stressed syllable to which a starred tone is assigned. Xu and Xu further observe that the AM model does not account for the F0 shape of pitch accents or the shapes of the F0 contour in nonaccented syllables and words. This is because the AM model assumes a simple interpolation between pitch targets. The interpolation can be either linear or nonlinear and "sagging," but either way coarticulatory constraints are not accounted for and neither are global changes to mean F0 or to F0 range. For these reasons, Xu (2005) and Xu and Xu (2005) extended the TA model proposed in Xu and Wang (2001) to characterize the interaction between local F0 contours associated with stressed syllables in English and the global intonation contour. The extended model for English is called the parallel encoding and target approximation (PENTA) model.

Xu and Xu (2005) proposed that every syllable in English is associated with a local pitch target, but that unstressed syllables are represented by a static mid-tonal target while stressed syllables are associated with static or dynamic targets depending on a number of lexical and postlexical factors. Focus and other discourse-level meanings are modeled as a parallel process that results in the manipulation of four melodic primitives: local pitch targets, pitch range, articulatory strength and duration. The particular manipulation depends on the specific communicative function. When these assumptions are made, there is consistent alignment of an F0 valley with the onset of a stressed syllable and consistent alignment of F0 peak with the offset of a stressed syllable in words that are the targets of prosodic focus. Xu and Xu also showed that F0 peaks occur well before the syllable offset in word-final stressed syllables when these are in focus or in sentence-final position. The PENTA model is illustrated in Figure 2.4.

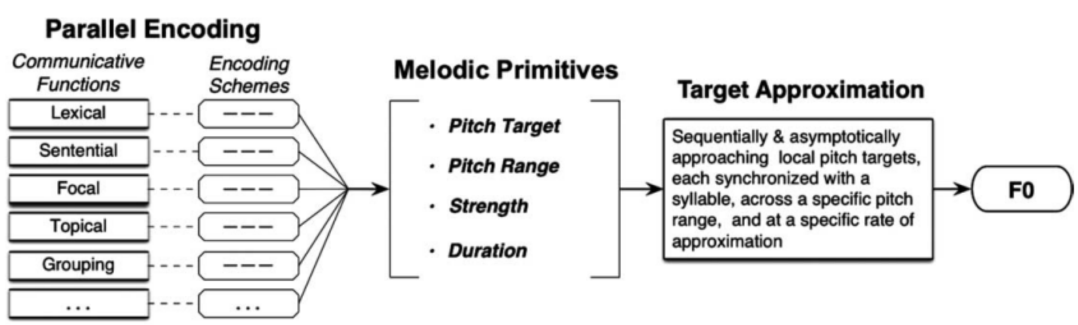


**Figure 2.4.** A brief sketch of the Parallel Encoding and Target Approximation (PENTA) model (adapted from Xu & Xu 2005). The unnamed block at the bottom left indicates communicative functions yet to be identified.

Prom-on, Xu and Thipakorn (2009) developed a quantitative target approximation (qTA) model based on TA and PENTA in order to simulate the production of tone and intonation in Mandarin and English. The simulations were evaluated both numerically and perceptually. Prom-on et al. found that the quality of the simulations was high and concluded that the TA and PENTA models provide effective tools for research on tone and intonation and potentially effective frameworks for the automatic synthesis of tone and intonation.

The assumptions of the TA model (and its extended versions) are adopted in the present dissertation because they provide a basis for the careful phonetic measurement and computation of tone, stress and intonation. This is important because we assume, following SLM, that the second language acquisition of phonology is acoustically-based, and the acquisition of prosodic focus should be no different.

## 2.4. The realization of prosodic focus and PFC across languages

The PENTA model formalizes the point that communicative functions at lexical and sentential levels are mediated by a set of parameters of which pitch is just one. In addition to pitch, "strengthening" and duration are invoked. Studies on the realization of prosodic focus in different languages have made this point as well, noting expansion of pitch, intensity and duration of the focused item. In some language, there is a concomitant compression of pitch, intensity, and duration after the focused item (PFC). PFC has been found in many non-tone languages, such as English (Cooper, Eady & Mueller, 1985; Eady & Cooper, 1986; Xu & Xu, 2005), Finnish (Vainio & Järvikivi, 2007), Dutch (Hanssen, Peters & Gussenhoven, 2008), Hindi (Patil et al. 2008), Japanese (Kubozono, 2009), Korean (Lee & Xu, 2010), etc., and in some tone languages, such as

Mandarin (Jin, 1996; Xu, 1999) and Nanchang dialect (Wang, Wang & Kadir, 2011). Figures 2.5 and 2.6 illustrate both the expansion and compression of pitch range in Mandarin and English, respectively.

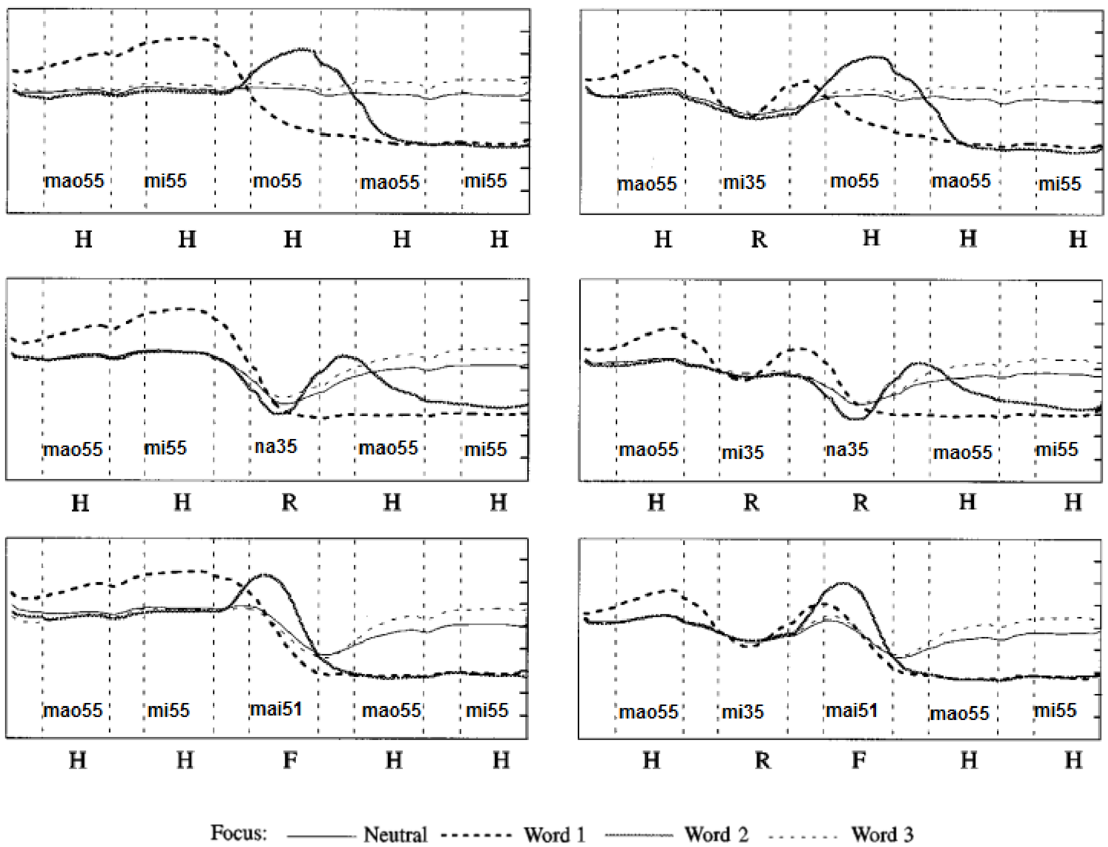


**Figure 2.5.** Effects of focus on F0 contours in Beijing Mandarin (adapted from Xu 1999).

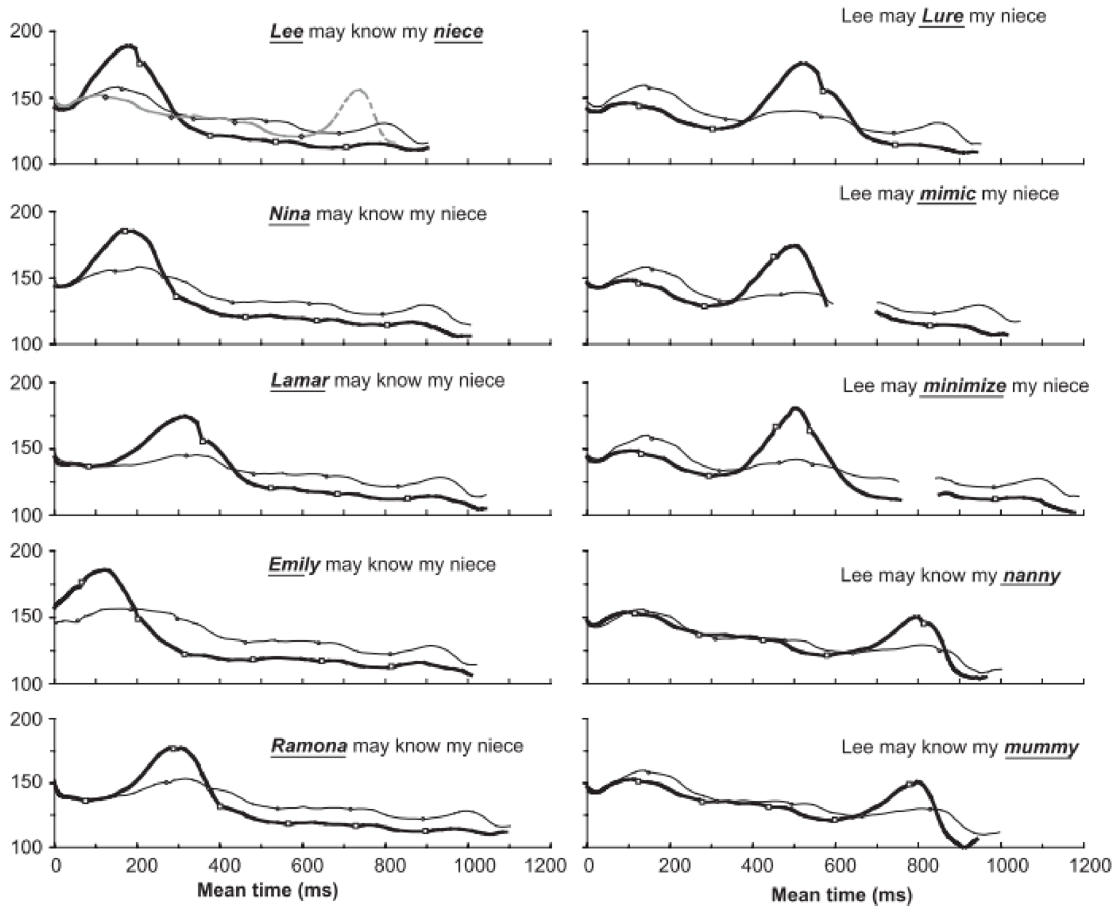


**Figure 2.6.** Effects of focus on F0 contour in American English (adapted from Xu and Xu 2005).

The figures illustrate that the in-focus expansion of pitch range results in the increase of mean F0 while the post-focus compression of pitch range results in the decrease of mean F0. Furthermore, as mentioned in the previous section, the local pitch contour of a syllable is associated with the local pitch target of the syllable, which is determined by lexical tone in Mandarin and by lexical stress in English. Meanwhile, focus extensively modulates the global shape of the F0 curve and in turn affects the height and even the shape of the local F0 contour. The observation that is of key interest to the present dissertation is that native speakers of Mandarin and English realize focus with similar prosodic patterns even though Mandarin is a tone language and English is an intonation language. The similarity is especially in the presence of PFC across the two languages. Although not illustrated here, in-focus expansion and PFC also applies to intensity in Mandarin and English. There is no PFC of duration, but the in-focused items are also lengthened in both languages.

Of additional interest to the present dissertation, PFC is not a universal feature of prosodic focus. In particular, many tone languages do not show the same kind of pitch range compression after the focused item as is observed in Mandarin. For example, PFC is absent in Cantonese (Wu & Xu, 1999) and Yi (Wang et al., 2011) and in Southern Min (Pan, 2007), one of the target languages in this dissertation. Xu, Chen and Wang (2012) compared the acoustic realization of focus in Mandarin and Southern Min. Four groups of speakers were recruited from Taiwan and Beijing: monolingual speakers of Taiwan Southern Min (Taiwanese), monolingual speakers of Taiwan Mandarin, bilingual speakers of Taiwan Southern Min and Taiwan Mandarin, and monolingual speakers of Beijing Mandarin. Although all four groups produced similar changes in F0, intensity and duration for in-focused words, only monolingual Beijing Mandarin speakers produced PFC of F0 and intensity. PFC was absent in Taiwan Southern Min produced by both Taiwan Southern Min monolinguals and Taiwan Southern Min-Mandarin bilinguals, and even in Taiwan Mandarin produced by Taiwan Mandarin monolinguals and Taiwan Southern Min-Mandarin bilinguals (See Figure 2.7). This result is consistent with the studies reviewed in Chapter I, which also showed that PFC is not easily transferred from one language to another. The results also suggest that PFC can be lost, assuming that, before extended contact with Taiwan Southern Min, Taiwan Mandarin had PFC.

The preceding review establishes that (a) PFC is present in languages that are otherwise typologically quite different from one another (e.g., Beijing Mandarin and English); (b) PFC differs across languages/dialects even when these are closely related (e.g., Beijing Mandarin and Taiwan Mandarin); and, as indicated in Chapter I and reiterated here via the review of Xu et al. (2012) PFC may be difficult to acquire or transfer from one language to another. The Xu et al. (2012) study also suggests that PFC may be lost through language contact.

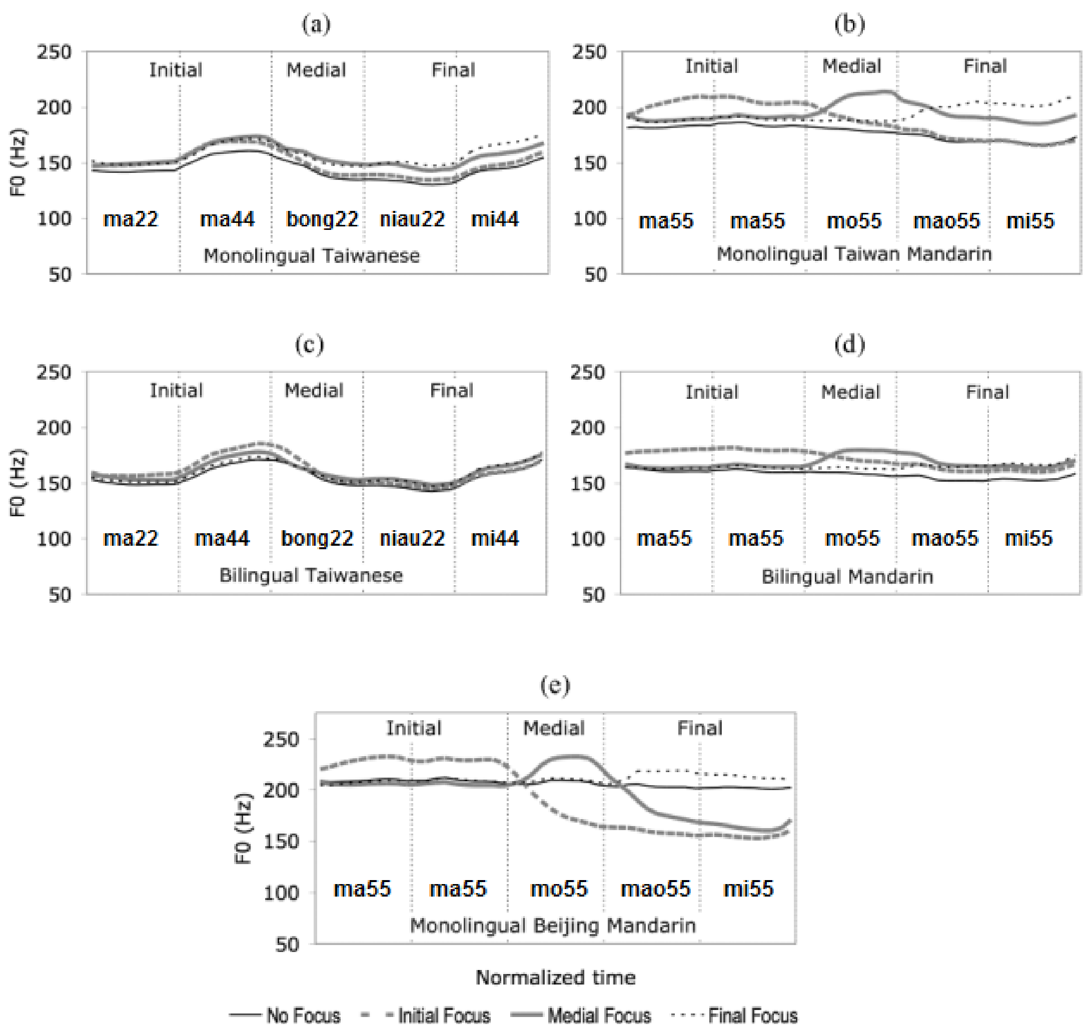


**Figure 2.7.** Time-normalized mean F0 contours in Taiwanese, Taiwan Mandarin and Beijing Mandarin (adapted from Xu et al. 2012).

## 2.5. Current dissertation

To review, a great number of studies have been conducted in second language phonology, especially at the segmental level. The results from these studies can be understood with reference to the assumptions of the Speech Learning Model (SLM), which takes a phonetically-based approach to the second language acquisition of phonology, and assumes that perception drives production and is mediated by acoustics. Finally, studies on second language prosody have typically been carried out either at word level or at phrase level, but very few have examined the acquisition of prosodic focus, which involves changes at both the word and phrase level.

The present dissertation investigates the acquisition of prosodic focus. The experiments reported were motivated by the assumption that the acquisition of prosodic focus, a phonological category, can be understood and measured in terms of the acoustic phonetics. Of particular interest is the question of PFC, a concomitant of prosodic focus in some languages and not others. The hypothesis is that learners may apply the mechanism and process of category formation of prosodic focus in L1 to establishing a similar category in L2 based on the same phonetic features, such as PFC, if both L1 and L2 are PFC languages; learners may also apply the mechanism and processes associated with learning L1 phonetic patterns to the acquisition of patterns such as PFC when their L1 lacks PFC. These hypotheses are investigated in three languages: (1) Southern Min – a tone language with no PFC, (2) Beijing Mandarin – a tone language with PFC, and (3) English – a non-tone language with PFC. In addition to the structure of the L1, the effects of L2 experience are investigated; specifically, the effects of language use, age of learning (AOL), and length of residence (LOR) in the L2 speaking environment. Three experiments were conducted:

Experiment 1 (Chapter III) examined the realization of prosodic focus in both L1 Southern Min and L2 Mandarin by societal bilinguals who differ in chronological age and the amount of L2 Mandarin they use in daily life. The prediction was that L1 Southern Min speakers would easily acquire L2 Mandarin lexical tones since Southern Min is tonal, but that they may not be able to acquire a native-like realization of prosodic focus in Mandarin, since Southern Min does not have PFC but Mandarin does. L1 Southern Min speakers with more L2 Mandarin use were predicted to have acquired more Beijing-like prosodic focus than Southern Min speakers who use Mandarin less.

Experiment 2 (Chapter IV) examined the realization of prosodic focus in L2 Mandarin by L1 American English learners grouped by Chinese heritage status and thus age of learning (AOL) L2 Mandarin. The prediction was that L1 English speakers would have trouble acquiring lexical tones since English is not a tone language, but they would be able to accurately realize prosodic focus in Mandarin because its realization is similar to the English realization of prosodic focus. The Chinese-heritage learners, who started learning L2 Mandarin in early childhood, were predicted to produce more native-like prosodic focus than the non-Chinese-heritage learners, who started learning Mandarin as teenagers.

Experiment 3 (Chapter V) examined the realization of prosodic focus in L2 English by L1 Beijing Mandarin speakers, who were students in an American university at the time of test and grouped by the length of residence (LOR) in the US. The Beijing Mandarin-speaking learners of English were predicted to easily acquire prosodic focus in English for the same reason that English learners of Mandarin were expected to acquire prosodic focus in Mandarin: because the two languages realize prosodic focus in a similar way. Just as English learners of Mandarin were expected to have trouble with the acquisition of lexical tone, Mandarin learners of English were expected to have trouble with the acquisition of lexical stress. Beijing Mandarin speakers with long LOR in the US were predicted to produce more native-like prosodic focus in English than those with short LOR due to greater exposure of high-quality L2 input.

Measurement and data analysis in all three experiments assumed the Target Approximation (TA) model introduced in this chapter. The results are discussed with reference to SLM. Chapter VI closes the dissertation with general conclusions and future directions.

# CHAPTER III

# BILINGUAL PRODUCTION OF PROSODIC FOCUS IN SOUTHERN MIN AND MANDARIN

## 3.1. Introduction

Since the National Language Movement in the mid 1940s, Mandarin has been promoted and spread as a national language in Taiwan (Li & Lee, 2006). As a consequence, Taiwan has since developed a large speech community of Taiwan Southern Min-Mandarin bilinguals. Due to the political separation of Taiwan and mainland China since 1949, the standard Mandarin in Taiwan has deviated from Beijing Mandarin. Taiwan Mandarin has also changed due to different degrees of contact with Taiwan Southern Min (Liao, 2008). Taiwan Southern Min-Mandarin bilinguals typically acquire Southern Min earlier than Mandarin. They learn and use Southern Min as L1 at home and Mandarin as L2 at school (Huang & Fon, 2007). Therefore, their Mandarin phonological system is influenced by their Southern Min phonological system (Kubler, 1985). Xu et al. (2012) finding of PFC in Beijing Mandarin but not in Taiwan Mandarin is consistent with the idea that Southern Min has influenced the prosodic aspect of Mandarin in the bilingual Taiwan Southern Min-Mandarin speakers.

Diglossia in Quanzhou, a city where Southern Min has been spoken for over 1,500 years in mainland China and one of the cities from which most of the population of Taiwan is derived, has a similar diglossic situation to Taiwan. The residents there speak not only Quanzhou Southern Min, but also Beijing Mandarin (Putonghua) as required by the government policy of National Popularization of Putonghua. Since the implementation of the policy in the 1950s, local residents in Quanzhou have been immersed and educated in Beijing Mandarin and have become Southern Min-Mandarin bilinguals. People around 60 years of age are considered to be the first generation of Southern Min-Mandarin bilinguals in Quanzhou. However, their daily use of Mandarin is quite limited. After several decades of popularization of Putonghua, and with a continuous increase in language contact, the younger generation uses more Mandarin than older generations. Additionally, most of the younger speakers, unlike the older ones, receive preschool education in Mandarin; so their age of learning (AOL) tends to be slightly earlier than the older speakers. Language experience therefore varies by age in Quanzhou Southern Min-Mandarin bilinguals. This provides an opportunity to determine whether the lack of PFC in Mandarin spoken by Southern Min-Mandarin speakers in Taiwan is due to contact with a non-PFC variety, and whether more extensive contact with Beijing Mandarin and greater use of this language enables the acquisition of PFC. The experiment reported in this chapter makes use of this opportunity.

Previous studies of the effects of language experience on bilingual speech production have mostly focused on immigrant bilinguals. The factors found to be relevant to immigrant bilinguals have also been confirmed in studies of societal bilinguals (Peng, 1993; Guion, Flege & Loftin, 2000; Guion, 2003). This experiment will take the amount of L2 Mandarin use as the primary index of language experience to examine the bilingual production of prosodic focus in Southern Min and Mandarin by Quanzhou speakers. The following specific research questions will be investigated: (1) Are there expansions of F0, intensity and duration of focused words in both Southern Min and Mandarin by Quanzhou bilingual speakers? (2) Does PFC of F0 and intensity occur in Quanzhou bilinguals' production of L1 Southern Min and L2 Mandarin? (3) Do different age groups produce different patterns of prosodic focus in both languages? Is there an intermediate pattern of prosodic focus in the bilingual production between the older and the younger generations? (4) Is there any reverse influence of L2 Mandarin on the prosodic patterns of focus in L1 Southern Min?

Based on what is already known about prosodic focus in the two languages, the expectation is that speakers will produce in focus items with an expanded F0 range, increased intensity and duration in both Mandarin and Southern Min, but PFC of F0 and intensity only in Mandarin. The prediction from SLM regarding the acquisition of focus patterns is that all the three age groups will produce some PFC in their L2 Mandarin production because they have all been exposed to and immersed in Beijng Mandarin since early childhood. That said, the younger generation, with greater L2 Mandarin experience and who use Mandarin more on a daily basis, are expected to produce a more Beijing-like pattern of prosodic focus in Mandarin, especially more PFC, than the older generation. Finally, based on reports of mutual influence between L1 and L2 in studies of early bilinguals (Peng, 1993; Guion, 2003) and the possibility that Taiwan Mandarin lost PFC due to contact with Taiwan Southern Min (Xu et al., 2012), we might expect some reverse influence from L2 Mandarin to L1 Southern Min in the younger speakers' Southern Min productions.

## 3.2. Methods

### *3.2.1. Participants*

Three age groups of Quanzhou Southern Min-Mandarin bilinguals participated in this study: younger, mid-age and older. The younger speakers were between 18 and 21 years of age, the mid-age speakers between 35 and 43 years, and the older speakers between 55 and 64 years. There were four males and four females in each group. All participants were born and raised in Quanzhou City, Fujian Province, mainland China. Participants in the younger group had always lived in Quanzhou and were students in Quanzhou Normal University at the time of testing. Participants in the older and mid-age groups had never lived longer than three months out of Quanzhou. All three groups learned Southern Min first, followed by Mandarin in childhood. However, the amount of Mandarin use and the age that speakers learned Mandarin varied among the age groups. All the participants reported having normal hearing and not speaking Chinese languages other than Southern Min and Mandarin.

The language experience of L2 Mandarin was determined by participants' responses to a language background questionnaire (LBQ). The LBQ requested participants to report their chronological age, the age at which they learned Southern Min and Mandarin, when and where they received their education, any places where they had traveled to for more than three months, scores on the National Test of Oral Putonghua Proficiency (if applicable), other languages they had learned, and self-estimates of Southern Min and Mandarin proficiency (1-10 scale) (See Appendix A). As for the amount of Mandarin use, the participants were requested to report the percentage of use with their grandparents, parents, children, other family members and relatives, teachers and classmates, coworkers, friends, use at home and outside home, and to estimate their overall use of Mandarin (See Appendix B). The overall information of LBQ by age group is reported in Table 3.1.

**Table 3.1.** Language background of the three age groups of Quanzhou Southern Min (SM)-Mandarin (MD) bilinguals. Mean and standard deviation (in parentheses) are shown for the age of test, years of education, age of learning Mandarin, amount of Mandarin use, self-estimated Southern Min proficiency and self-estimated Mandarin proficiency.

| Age Group | Age of test | Years of education | MD AOL | MD use | SM self estimate | MDself estimate |
|---|---|---|---|---|---|---|
| Younger | 19.9 | 14 | 4.0 | 63% | 7.6 | 8.5 |
| | (1.1) | (0.9) | (0.8) | (10%) | (0.9) | (0.8) |
| Mid-age | 39.6 | 16.9 | 6.3 | 43% | 9.1 | 8.1 |
| | (3.4) | (2.5) | (1.0) | (13%) | (0.8) | (1.1) |
| Older | 58.6 | 11.5 | 7.8 | 25% | 9.8 | 7.5 |
| | (3.2) | (2.9) | (0.5) | (15%) | (0.5) | (1.3) |

Table 3.1 indicates that the younger group learned Mandarin earlier and used it more than the mid-age group and the mid-age group learned Mandarin earlier and used it more than the older group. Both the older and mid-age groups estimated their proficiency of Southern Min to be higher than their proficiency in Mandarin; however, the younger group reported higher proficiency in Mandarin than in Southern Min. All the groups reported speaking relatively more Southern Min at home and relatively more Mandarin outside home. Nevertheless, younger speakers reported more Mandarin use with parents, children, relatives, teachers, classmates, coworkers, and friends. The mid-age group reported speaking more Mandarin to their children than the older group, which could foreshow more Mandarin use in the younger group.

*3.2.2. Stimuli*

Materials were adapted from Xu et al. (2012). Participants were requested to produce a target sentence in both Mandarin and Southern Min. In both languages, the target sentence had three words and five syllables. The surface form of the sentences is shown in Table 3.2 along with the widely used tone values based on the 5-point scale for both languages (Chao, 1933).

**Table 3.2.** Target sentences in Mandarin and Quanzhou Southern Min in Experiment 1.

| | Word 1 | Word 2 | Word 3 |
|---|---|---|---|
| Character | 妈妈 | 摸 | 猫咪 |
| Gloss | 'mom' | 'pet' | 'kitty' |
| Mandarin | ma55 ma3 | mo55 | mau55 mi55 |
| QZ. SM | ma22 ma24 | mɔ33 | niau33 mi24 |

All syllables of the target sentence in Mandarin had the high-level tone 55 underlyingly, but the second syllable /ma/ was realized with a neutral tone due to a rule of reduplication, so that the surface tone value was 3 (Chen & Xu, 2006). In Quanzhou Southern Min, /ma/ had the rising tone 24 underlyingly, but the first syllable /ma/ was realized with the surface low-level tone 22 due to a tone sandhi rule (Lin, 1993: 60). Both /mɔ/ and /niau/ had the mid-level tone 33 and /mi/ had the rising tone 24 in both underlying and surface forms.

A picture illustrating the target sentence ('Mom is petting the kitty') was shown to the participants in order to set up a focus-eliciting situation. Four prompt questions were used to elicit four types of focus: no focus, initial focus (on Word 1), medial focus (on Word 2) and final focus (on Word 3). The prompt questions are shown in Table 3.3.

**Table 3.3.** Prompt questions for eliciting foci in both Southern Min and Mandarin.

| Focus | Prompt questions | English translation |
|---|---|---|
| None | 图中你看到什么? | What do you see in the picture? |
| Initial | 谁摸猫咪? | Who is petting the kitty? |
| Medial | 妈妈对猫咪做什么? | What is Mom doing to the kitty? |
| Final | 妈妈摸什么? | What is Mom petting? |

*3.2.3. Recording*

The prompt questions were asked by the experimenter, who is bilingual in Southern Min and Mandarin, in the relevant target language (Southern Min or Mandarin) and each question was repeated five times in a random order. The experimenter requested the subjects to use the target sentence to answer the prompt questions as naturally as possible. The inter-trial interval (interval between adjacent question-answer pairs) was about three seconds. Participants answered the questions with the target sentence with appropriate focus. Each participant produced 40 sentences (4 foci × 5 repetitions × 2 languages). Experiment instructions were given in both Mandarin and Southern Min. The Mandarin production was recorded prior to Southern Min, except for two speakers in the older group, who preferred using Southern Min first. The recording in each language lasted about five minutes and there was a two-minute break between the two recordings. Recording was conducted in a quiet room with a Marantz professional solid state recorder

PMD660 and a Shure professional unidirectional head-worn dynamic microphone. Target sentences were directly recorded into a computer SD card with a sampling rate of 44,100Hz.

*3.2.4. Analyses*

Data were analyzed in Praat version 5.1.32. ProsodyPro, a custom-written script (Xu, 2005-2014) was used to sequentially open the sound files and generate measurements as well as continuous F0 contours from each file. The script allows users to mark syllable boundaries and rectify vocal pulse markings initially generated by Praat. Measurements used in Experiment 1 included maximum F0, minimum F0, mean F0, intensity, duration, and time-normalized F0 with 10 points in each syllable interval. The time-normalized F0 contours were used only in the graphical analysis, and all the other F0 measurements were taken from the original non-time-normalized F0 contours. To assess the effect of focus, differences of mean F0, intensity and duration between in-focus, pre-focus and post-focus words and their no-focus counterparts were calculated. In-focus change was calculated as the mean of measured values of the focused syllable minus that of their no-focus counterparts. Pre-focus change and post-focus change were calculated as the mean of the differences between the pre-focus or post-focus syllables and their no-focus counterparts. For instance, post-focus change in Mandarin was the mean of the values of /mo55/, /mau55/ and /mi55/ after focused /ma55 ma3/ minus the values of the three syllables in the no-focus sentence and the values of /mau55/ and /mi55/ after focused /mo55/ minus the values of the two syllables in the no-focus sentence (see Table 3.2). These measurements therefore enabled comparisons of F0, intensity and duration patterns across the sentences in the no-focus, initial-focus, medial-focus and final-focus conditions for the five repetitions of each sentence by each speaker.

## 3.3. Results

### *3.3.1. The overall F0 contours*

Before making statistical comparisons, time-normalized F0 contours of the stimulus sentences were first examined for an assessment of the overall differences across the experimental conditions. Figures 3.1 and 3.2 show the mean time-normalized contours associated with different focus conditions for the different age groups and languages.

Figure 3.1 shows that almost all the F0 curves overlap in each panel, which indicates that none of the three age groups produced either noticeable F0 expansion on in-focus constituents or F0 compression on post-focus constituents under any of the focus conditions in Southern Min.

Figure 3.2 indicates that the in-focus words produced by the younger and mid-age groups had higher F0 than their no-focus counterparts in Mandarin. In the younger group, post-focus F0 contours in sentences with both initial focus and medial focus were lower than their counterparts in the no-focus sentences. In the mid-age group, post-focus F0 was lower than in no-focus F0 in sentences with initial focus but not in sentences with medial focus. The older group did not show clear post-focus lowering of F0 or any other noticeable F0change relative to the no-focus sentences.

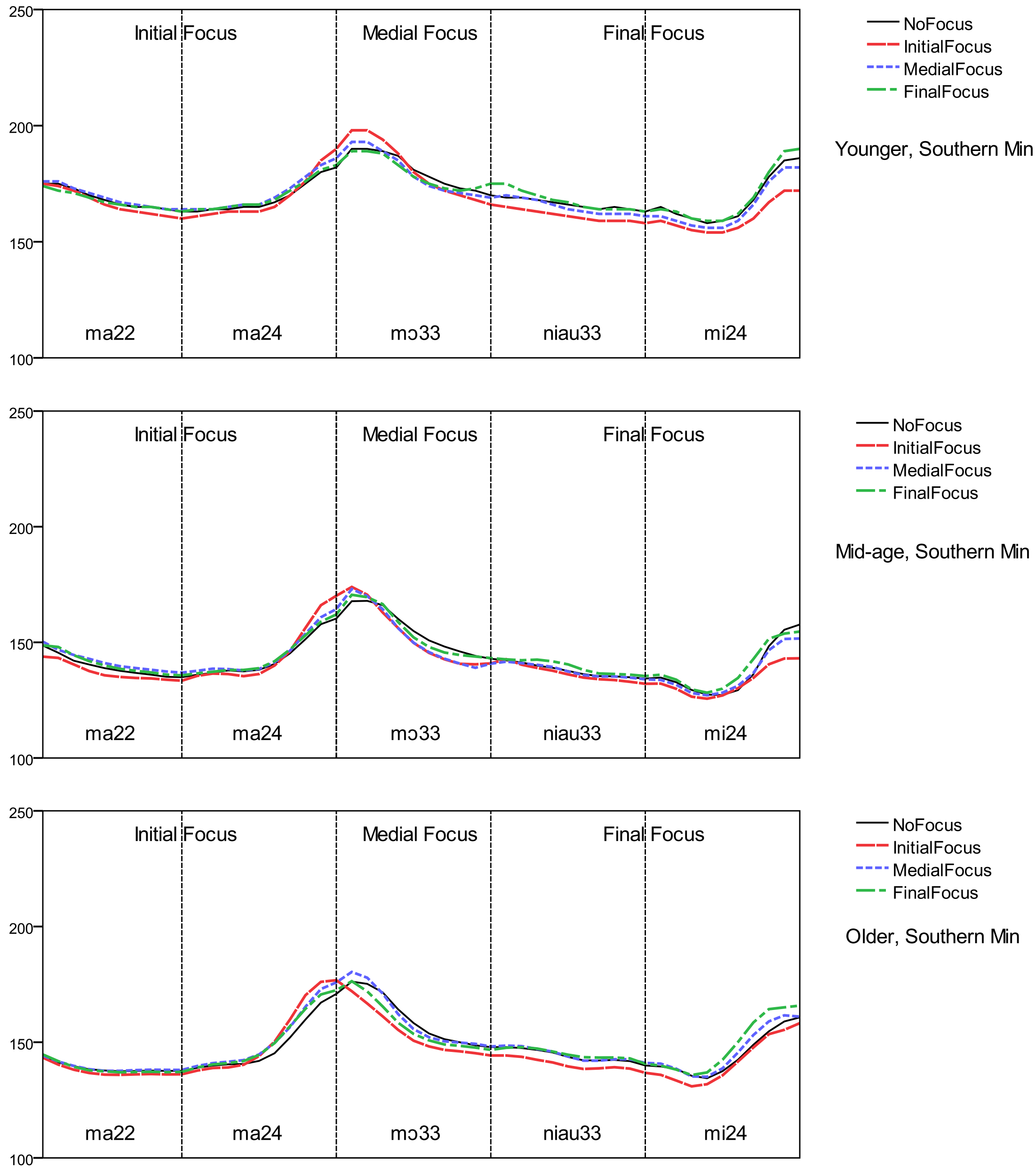


**Figure 3.1.** Time-normalized F0 contours (Hz) in Southern Min by three age groups. Each curve represents an average of the five repetitions by eight speakers under the same focus condition. Syllable boundaries are marked with vertical dashed lines.

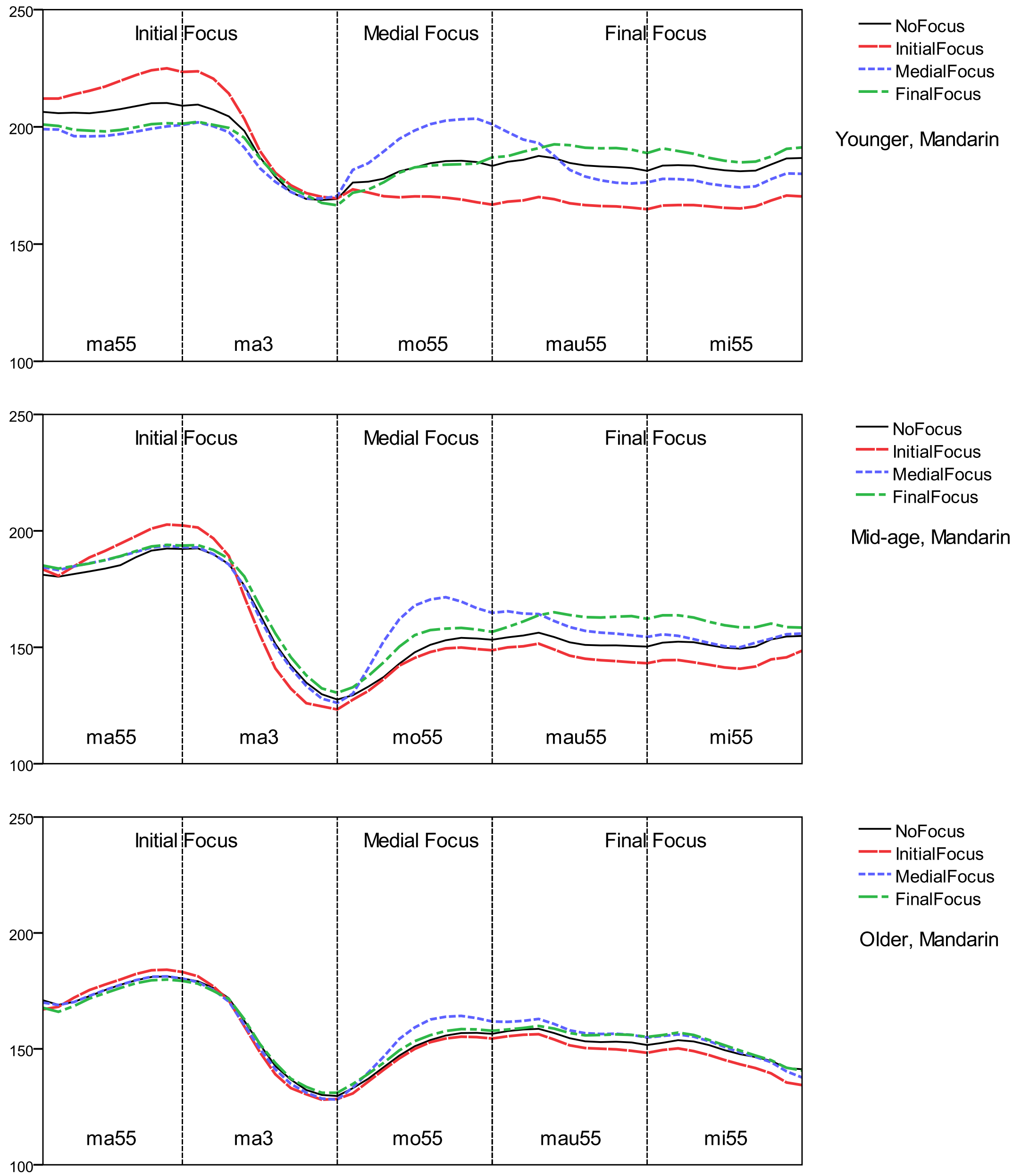


**Figure 3.2.** Time-normalized F0 contours (Hz) in Mandarin by the three age groups. Each curve represents an average of the five repetitions by the eight speakers under the same focus condition. Syllable boundaries are marked with vertical dashed lines.

Note that all words in the Mandarin stimulus sentence are rendered with Tone 1 /55/ except for the second syllable *ma* of the first word *mā ma*, which is normally produced in a neutral tone due to the reduplication of the previous syllable in addressing.

According to Chen and Xu (2006), the default surface tone value of a neutral tone after a high level tone is /3/, a short and weak mid tone— M. According to the Target Approximation (TA) model (Xu & Wang, 2001), both Tone 1 and neutral tones represent static targets. However, because the offset of the first pitch target /55/ in *mā* is different from the onset of the second pitch target /3/ in *ma*, the F0 surface form of the second *ma* is assimilated to the first *mā*. For this reason, the F0 peak of the word *mā ma* is located around the syllable boundary in all the three panels in Figure 3.2. Also, a similar pattern of assimilation occurs between the second syllable *ma* and the third syllable *mō*, and we can see that the F0 valley is located around the syllable boundary of these two syllables in Figure 3.2.

As for the effect of focus, the expected pattern for Bejing Mandarin is the expansion of F0 contours associated with tones when there is non-final focus and a compression of the F0 contours associated with tones (i.e., PFC) after a focused item. Nevertheless, the older group of Southern Min-Mandarin bilinguals showed no expansion or compression of F0 in their L2 Mandarin in the focus condition compared to the no-focus condition (bottom panel of Figure 3.2). By contrast, the younger group produced both in-focus expansion and post-focus compression (PFC) while maintaining the tonal F0 contours in the initial and medial focus condition (top panel of Figure 3.2). The mid-age group of speakers showed a similar pattern to that of the younger age group (middle panel of Figure 3.2), albeit one that was somewhat less dramatic.

*3.3.2. The change of mean F0, intensity and duration by sentence location relative to focus*

To test whether the differences in F0 change across the four focus conditions (no focus, initial focus, medial focus and final focus) and three age groups was statistically significant, F0 difference values (in focus – no focus) were converted from Hz to semitone. A repeated measures ANOVAs with three factors—language (Southern Min, Mandarin), age group (younger, mid-age, older) and sentence location relative to focus item (pre-focus, in-focus, post-focus) was then conducted on these values. The results showed a three-way interaction between the factors for mean F0 change ($F(4,42) = 4.436$, $p = 0.004$) and two-way interactions between language and location ($F(2,42) = 17.233$, $p < 0.001$) and between location and age ($F(4,42) = 2.694$, $p = 0.044$). Figure 3.3 displays the means and standard errors according to sentence location relative to focus and age groups in the two languages.

Given the significant interactions with language, the data were split by language to further examine the effects of primary interest, age group and sentence location, on F0 difference values. As expected based on Figure 3.1, there was no interaction between age group and focus condition and no significant main effect of age group on F0 change in the Southern Min data. However, the main effect of focus condition on F0 change was significant ($F(2,21) = 15.679$, $p < 0.001$). The source of the effect of location seems to come from some slight variations of pre-focus and post-focus F0 change compared to small in-focus F0 changes by all the three groups (top panel in Figure 3.3 below). However, follow up testing of the magnitude of these differences by sentence location was compared by subtracting pre-focus values from in-focus values and post-focus values from in-focus values. The comparison showed no significant difference of magnitude of F0 change for any sentence location between any two groups of speakers, consistent with the initial observation that F0 change is not used to code focus in Southern Min.

In contrast to Southern Min, the Mandarin data showed a significant interaction between age group and sentence location relative to focus on F0 change ($F(4,42) = 4.518$, $p = 0.004$) and significant main effects of both focus condition ($F(2, 21) = 34.685$, $p <$ 0.001) and age group ($F(2, 21) = 3.625$, $p = 0.044$). Post-hoc independent samples t-tests showed significant differences[1] in pre-focus F0 change for younger and mid-age groups ($t(14) = -3.097$, $p = 0.015$) and for younger and older groups ($t(14) = -2.798$, $p = 0.014$). A significant difference in post-focus F0 change was also found between younger and older groups ($t(14) = -2.719$, $p = 0.022$). The magnitude of F0 change from in-focus to pre-focus was significantly different between the younger and mid-age groups ($t(14) = -2.250$, $p = 0.041$) and between the younger and older groups ($t(14) = -2.971$, $p = 0.010$). The magnitude of F0 change from in-focus to post-focus was significantly different between the mid-age and older groups ($t(14) = 2.188$, $p = 0.046$) and between the younger and older groups ($t(14) = -3.351$, $p = 0.005$). These results are evident in the bottom panel of Figure 3.3.

To understand the relationship between F0 change across the sentence (pre-focus, in-focus, post-focus) in Mandarin, Pearson correlations were examined for each speaker group. None of the speaker groups showed a significant correlation between in-focus and pre-focus F0 change or between in-focus and post-focus F0 change in sentences with

[1] Note that not all the differences in the post hoc t-tests would be considered to be significant if the *p* values were adjusted to 0.017 for the three comparisons (0.05/3), but it has been argued that this kind of adjustment may be overly conservative (Perneger, 1998).

initial or medial focus. However, the mid-age and older groups showed a significant correlation in F0 change between in-focus and pre-focus for sentences with final focus (mid-age, $r(8) = 0.823$, $p = 0.012$; older group, $r(8) = 0.8$, $p = 0.018$). Figure 3.3 indicates that the mid-age and older groups increased mean F0 increased in pre-focus position in sentence where the final item was in focus.

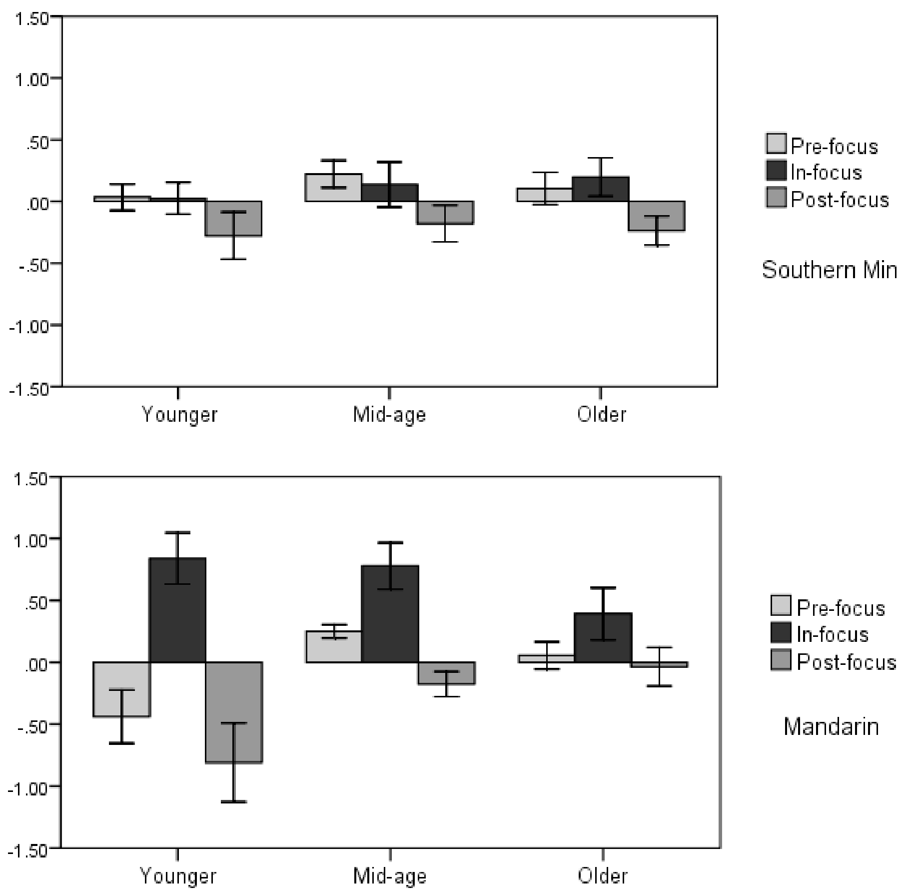


**Figure 3.3.** Mean F0 change (semitone) by sentence location relative to focus item and age group in Southern Min and Mandarin.

Figure 3.3 shows that the younger group produced both pre-focus and post-focus compression of F0; the mid-age group did not significantly reduce post-focus F0 and F0 was higher in pre-focus condition; F0 did not vary systematically with sentence location in the older group. Older speaker also did not expand F0 as much as the younger and mid-age speakers. Thus, overall, the younger group produced an F0 pattern in Mandarin that was different from that produced by the mid-age and older groups of Southern Min-Mandarin bilinguals.

Repeated measures ANOVAs similar to those on F0 change were also conducted on mean intensity change. No three-way interaction was found among the three factors: language, age group and sentence location relative to focus. There were interactions between language and age group ($F(2,21) = 5.909$, $p = 0.009$) and between age group and location ($F(4,42) = 5.080$, $p = 0.002$). Figure 3.4 displays the pattern of intensity change for the two languages as a function of location and age group. Within language analysis of intensity change indicated a two-way interaction between age group and location in Southern Min ($F(4,42) = 3.466$, $p = 0.016$) and Mandarin ($F(4,42) = 3.995$, $p = 0.008$); significant main effects of location in Southern Min ($F(2, 21) = 44.507$, $p < 0.001$) and Mandarin ($F(2, 21) = 43.292$, $p < 0.001$); and an effect of age group in Southern Min ($F(2, 21) = 3.710$, $p = 0.042$) and Mandarin ($F(2, 21) = 5.437$, $p = 0.013$). These results are shown in Figure 3.4 below.

Post-hoc independent samples t-tests on the Southern Min data showed significant differences in pre-focus intensity change for younger and mid-age groups ($t(14) = -2.200$, $p = 0.045$) and for mid-age and older groups ($t(14) = -2.251$, $p = 0.041$). Significant differences of in-focus intensity change were also found between younger and older groups ($t(14) = 5.502$, $p < 0.001$) and between mid-age and older-group ($t(14) = 2.403$, $p = 0.031$). The magnitude of intensity change from in-focus to pre-focus was significantly different between the younger and mid-age groups ($t(14) = -2.573$, $p = 0.022$). The magnitude of intensity change from in-focus to post-focus was significantly different between the younger and older groups ($t(14) = -4.707$, $p < 0.001$). The upper panel of Figure 3.4 indicates that the younger group increased intensity more for the in-focus position than mid-age and older groups, and that the mid-age group increased intensity more for the pre-focus position than younger and older groups. None of the groups showed post-focus compression of intensity in Southern Min.

Similar to Southern Min, the post-hoc independent samples t-tests on the Mandarin data showed significant differences in pre-focus intensity change for younger and mid-age groups ($t(14) = -3.308$, $p = 0.005$) and post-focus intensity change between younger and older groups ($t(14) = -3.728$, $p < 0.001$) and between mid-age and older-group ($t(14) = -2.874$, $p = 0.012$). The magnitude of intensity change between post-focus and in-focus positions was significantly different between the younger and older groups ($t(14) = -4.222$, $p = 0.001$) and between the mid-age and older groups ($t(14) = -3.583$, $p = 0.003$). The lower panel in Figure 3.4 indicates that all the three groups expanded intensity on in-focus position; however, only the younger and mid-age groups produced post-focus compression of intensity in Mandarin. The younger group also produced pre-focus compression of intensity in Mandarin.

Similar Pearson correlation coefficients were conducted in the intensity changes between in-focus and pre-focus and between in-focus and post-focus in the Mandarin production for each speaker group. None of the speaker groups showed any significant correlation of intensity change between in-focus and pre-focus or between in-focus and post-focus at any focus location.

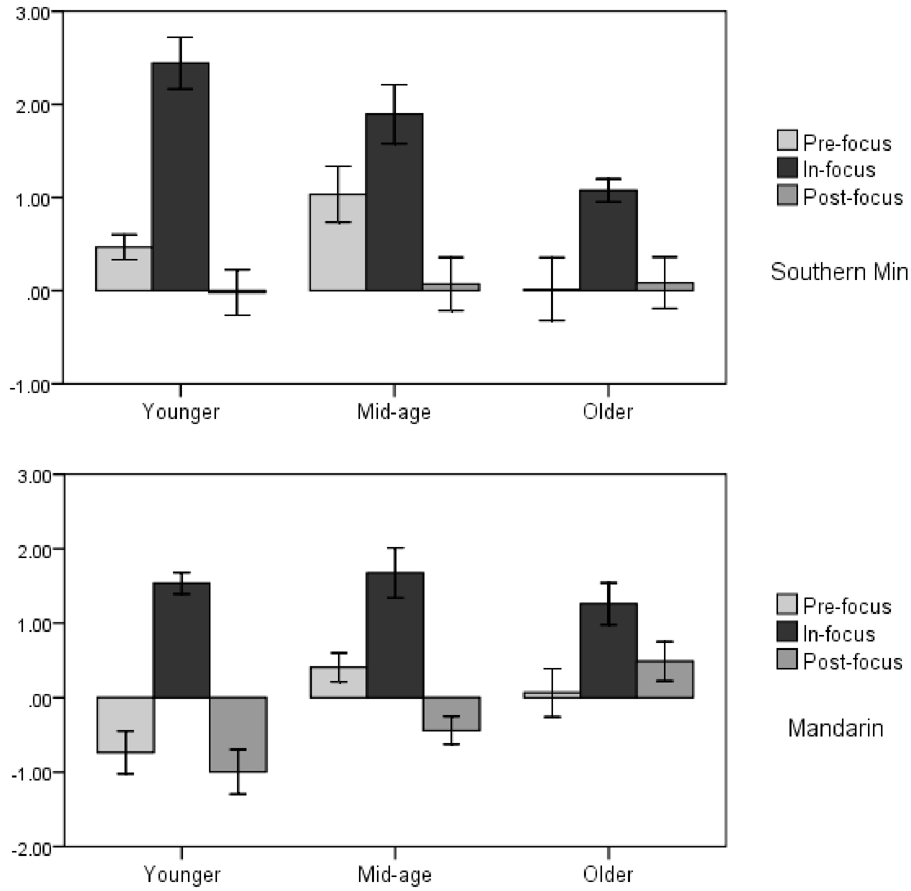


**Figure 3.4.** Mean intensity change (dB) by sentence location relative to focus item and age group in Southern Min and Mandarin.

Finally, a repeated measures ANOVA on duration change showed no three-way interaction and no two-way interactions between any of the factors. The main effects of age group on change in duration were not significant either; however, there were significant main effects of language ($F(1,21) = 4.507$, $p = 0.046$) and sentence location ($F(2,42) = 47.896$, $p < 0.001$). These results are shown in Figure 3.5.

When split by language, there was still no interaction between age group and location and no significant main effect of age group on change in duration. The Southern

Min data showed. However, the main effect of location on duration change was significant ($F(2,21) = 46.984$, $p < 0.001$). The results did not vary at all by speakers' age.

Similar to Southern Min, the analysis of duration change in Mandarin showed no effect of age group or any interaction between age group and sentence location relative to focus. Like Southern Min, there was a significant main effect of location ($F(2,21) = 35.166$, $p < 0.001$). Just as with Southern Min, the magnitude of this effect did not vary with speakers' age.

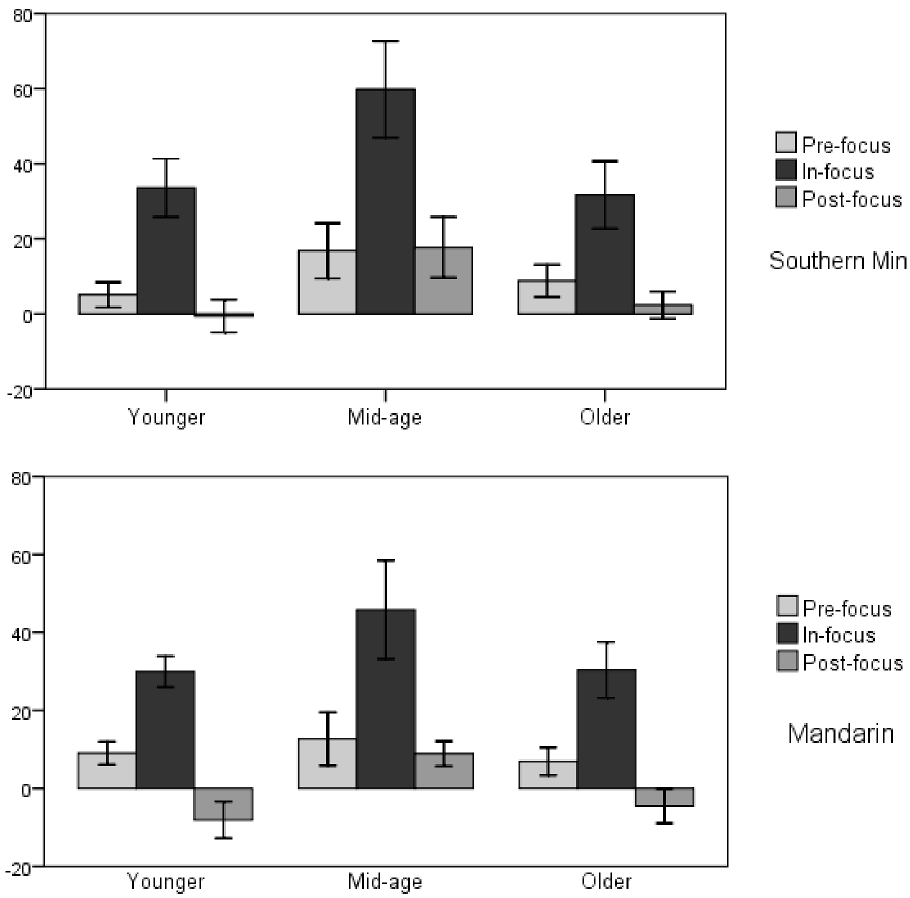


**Figure 3.5**. Mean duration change (ms) by sentence location relative to focus item and age group in Southern Min and Mandarin.

*3.3.3. Correlations between L2 experience and PFC of F0 and intensity*

As expected more experience in L2, in particular, greater amount of L2 Mandarin use resulted in the production of PFC in Mandarin: older speakers produced no PFC in Mandarin, the mid-age speakers produced some and the younger speakers produced the most. To explore the relationship between use and the production of PFC more thoroughly, a final set of analyses was conducted. The amount of use reported by individual speakers was used to predict the degree of PFC in Mandarin using linear regression. Since age of learning (AOL) also varied somewhat between the groups, it was entered as a control variable. Here PFC was calculated by subtracting the no-focus mean F0 and intensity from the post-focus mean. The analysis showed that AOL was not a significant predictor of either feature. L2 use significantly predicted PFC of F0 ($\beta$ = -0.657, $p$ = 0.007) and also explained a significant proportion of variance in PFC of F0 ($R^2$ = 0.482, $F$(2, 21) = 9.827, $p$ = 0.001). L2 use was also a significant predictor of PFC of intensity ($\beta$ = -0.746, $p$ = 0.001) and explained a significant proportion of variance in PFC of intensity ($R^2$ = 0.576, $F$(2, 21) = 14.289, $p$ < 0.001). The regression plots are shown in Figures 3.6 and 3.7. Both figures show that the amount of L2 use varied across the three age groups, although there are also clear overlaps. In both Figures 3.6 and 3.7, there seem to be a strong relationship between use and PFC in the younger speakers. This result might suggest that only when L2 proficiency exceeds some kind of threshold does PFC start to emerge.

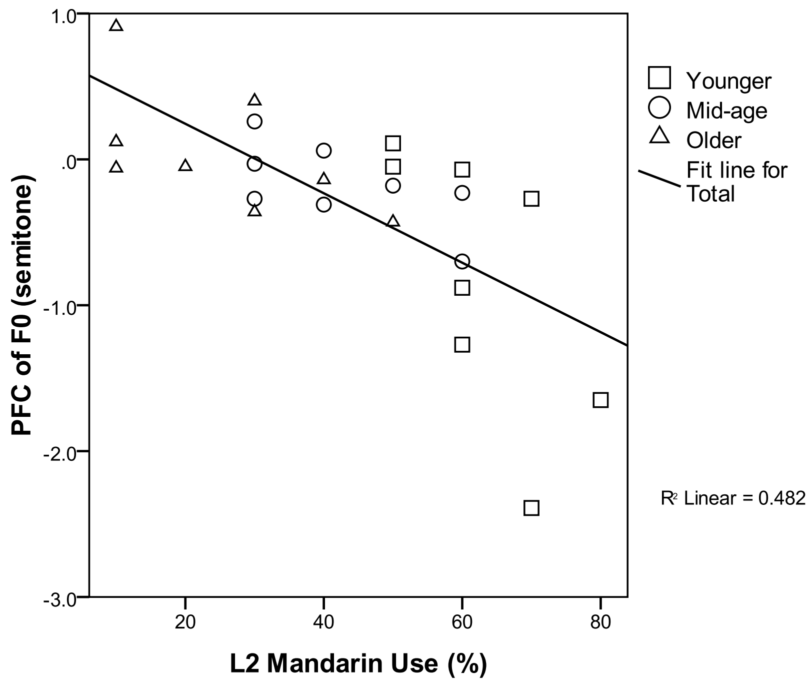


**Figure 3.6.** Regression plots of PFC of F0 in Mandarin as a function of L2 Mandarin use.

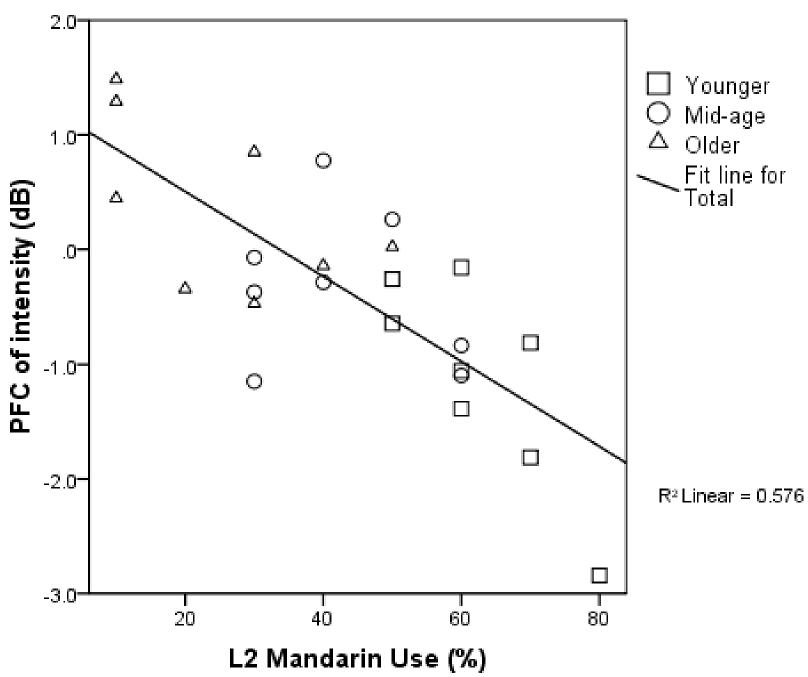


**Figure 3.7.** Regression plots of PFC of intensity in Mandarin as a function of L2 Mandarin use.

Based on the findings in Xu et al. (2012) and the results in this experiment, PFC of F0 and PFC of intensity do not seem to be independent of one another. This observation is confirmed by a highly significant correlation of post-focus change between F0 and intensity within subject ($r(24) = 0.756$, $p < 0.001$).

**3.4. Discussion**

The results just presented have provided answers to the four questions investigated in this study. First, in answer to the question of in-focus change in Southern Min and Mandarin, all age groups expanded duration and intensity on the focused words in both languages. This is consistent with the findings of Pan (2007) and Xu et al. (2012) in Taiwan Southern Min and those of Jin (1996) and Xu (1999) in Mandarin. However, unlike in Mandarin and Taiwan Southern Min, we found that none of the Quanzhou bilinguals expanded F0 for in-focus items in Southern Min, which again confirmed the findings in Pan (2007). Since we asked our subjects to speak as naturally as possible, this result suggests that only intensity and duration are used to convey focus in Southern Min.

Second, in answer to the question about the presence of PFC in the two languages, none of the Quanzhou speakers in the current study produced PFC in Southern Min. No group used F0 to code focus in Southern Min. Although all Quanzhou bilingual speakers expanded duration, intensity and F0 on focused words in Mandarin, only the younger group produced significant PFC of F0 and intensity in Mandarin. This finding was unexpected given the early exposure that all speakers had to Beijing Mandarin and the importance of AOL to second language acquisition. These results are, however, consistent with our prediction that younger speakers would be more native-like in their production of focus in Mandarin due the fact that they use Mandarin more than the other age groups. Younger Quanzhou bilinguals tended to compress F0 and intensity in the pre-focus constituents. This pre-focus compression goes beyond the typical Mandarin norms and may reflect an extra effort in this group to realize focus as clearly as possible.

In addition, the Pearson correlation results indicate that the pre- and post-focus compressions were independent of in-focus expansion in Mandarin. This finding could be attributed to the tones of the stimuli—except for the neutral tone in the second syllable, other constituents in the stimulus sentences were all Tone 1 and thus lack of local pitch range change. Nonetheless, a more plausible reason is that speakers tended to maintain the local pitch contour of the in-focus words and did not extend the mean F0 change pattern to the post-focus constituents. Finally, the top panel in Figure 3.2 also shows that the neutral tone (with a relatively low pitch value 3) did not block PFC in the Mandarin production of younger speakers. This is consistent with findings from previous studies, which indicate that PFC occurs regardless of whether the neighboring syllable has a high or low tone[2] (Xu, 1999; Wu and Xu, 2010; Wu and Chung, 2011; Xu et al., 2012).

Other group differences were also as expected. Mid-age group speakers realized PFC in their Mandarin speech, but the effect was smaller than that realized by the younger group. The mid-age group may therefore represent a transitional or intermediate stage of realizing prosodic focus that lies somewhere between the norms for Beijing Mandarin and Southern Min and between the productions of younger and older generations of Quanzhou Southern Min-Mandarin bilingual speakers.

As noted several times already, the differences observed between the age groups are likely attributable to differences in language experience, particularly, the amount of

[2] The lack of PFC in Southern Min as shown in Figure 3.1 is therefore unlikely due to the relatively low tonal values (22 or 33) of the Southern Min stimulus sentence.

L2 use (see Figures 3.6 and 3.7). This explanation is even more likely when we consider that speakers were otherwise quite similar. All were from similar societal communities, some even from the same families; all had similar education and economic backgrounds; were exposed to similar, political environments and mass media; and all had similar contact with broader circles of places and people. Data from the language background questionnaire (LBQ) indicated that age of learning Mandarin differed by only about two years between adjacent age groups and all the speakers started learning Mandarin in early childhood. All speakers could therefore be considered "early learners" (Flege, MacKay & Meador, 1999; Guion, 2003; Hojen & Flege, 2006). The results in Section 3.3.3 then confirmed that AOL was not a predictor of the presence of PFC in Mandarin. By contrast, data from the LBQ indicated that the amount of L1/L2 use differed by about 20% between the age groups – the younger the speakers, the more the Mandarin they used (see Appendixes A and B). This difference in language use appears to have impacted speakers' self-assessments of language proficiency. Younger speakers indicated higher proficiency in Mandarin than mid-age and older speakers. Younger speakers also thought their Mandarin proficiency was higher than their Southern Min proficiency. Finally, the finding that self-estimated L2 use predicted a significant proportion of the variance in PFC of F0 and intensity in Mandarin supports the hypothesis that L2 use is the primary factor responsible for the acquisition of Mandarin prosodic focus by native speakers of Southern Min.

The effects of L1/L2 use on L2 speech production has previously been examined for foreign accent with native speaker rating of global pronunciation (Flege, Frieda & Nozawa, 1997; Guion, Flege & Loftin, 2000; Piske, MacKay & Flege, 2001). Foreign accent rating studies suggest that L2 mastery at the segmental level does not imply mastery at the suprasegmental level (Birdsong, 2007). The contribution of the present experiment is that it examined the effect of L1/L2 use on suprasegmental patterns using detailed acoustic measurements instead of subjective ratings of global pronunciations. In another study of societal bilingualism, Guion et al. (2000) found that that the group with the most L1 Quichua use had the strongest accent in L2 Spanish, but that the amount of L1 Quichua use did not result in different degrees of Spanish accents in Quichua. The present results are similar to those obtained by Guion et al. in that more L2 Mandarin was correlated with the realization of more Beijing-like prosodic focus in Mandarin but had no effect on L1 Quanzhou Southern Min productions.

Mutual effects of L1 and L2 sound systems in societal bilinguals have been found for the production of vowels (Guion, 2003) and consonants (Peng, 1993). Unlike these studies, the present results indicate only an effect of L1 Quanzhou Southern Min prosody on L2 Mandarin prosody in the production of the older group, and no effect of L2 Mandarin prosody on the production of L1 Southern Min prosody production in any of the three age groups. This finding is consistent with the suggestion that post-focus compression is not easily transferred from one language to another (Wu & Chung, 2011; Xu, 2011). It is, however, at odds with the suggestion we have made that Taiwan Mandarin lost PFC due to contact with Taiwan Southern Min.

Another important factor we must consider when examining L2 speech production is L2 input. Flege and Liu (2001) found that even adults' L2 performance could be improved if they received a substantial amount of native speakers' input. However, the amount of formal instruction in L2 did not positively influence L2 pronunciation (Piske, MacKay & Flege, 2001; Piske, 2007). Flege's findings suggest that a large amount of L2 input is not sufficient to achieve a native-like accent if the speakers are exposed to non-native L2 speech. Younger speakers in the present study were more immersed in Mandarin than mid-age or older speakers. They also received more and higher-quality Beijing Mandarin input in school and in the ambient environment due to a surge in access to Mandarin media during the younger speakers' life times. The older speakers, on the other hand, as the first generation of Southern Min-Mandarin bilinguals in Mainland China, may not have received high quality Beijing Mandarin input in school; as the first generation of Mandarin teachers may have had a lower Beijing Mandarin proficiency than the current generation of Mandarin teachers. It is certain that they received less Beijing Mandarin input outside of school, since at that point societal bilingualism was just emerging and the (centrally programmed) media were not as pervasive as they are today. In addition to receiving better input, most of the speakers in the younger age group and some speakers in the mid-age group were required to take the National Putonghua Proficiency Test for their future or current profession. These speakers also received intensive Mandarin training in class before they took the oral test. Intensive training at both segmental and suprasegmental levels has been found to be effective in decreasing foreign accent (Moyer, 1999; Missaglia, 1999; Piske et al., 2001; Piske, 2007). The effect of special training may have supplemented the effect of higher quality input to ensure that younger speakers were Beijing-like in their L2 productions of focus in Mandarin.

According to the Speech Learning Model (SLM) (Flege, 1995), sounds that are noncontrastive in L1 but phonetically similar between L1 and L2 are difficult to discern for L2 learners and this kind of difficulty is reflected in the learners' L2 production. It has been also suggested that L2 learners have difficulty perceiving, and so acquiring, L2 phonetic patterns that are not used to signal phonological contrast in L1 (McAllister et al., 2002). In the context of the present findings, note that the in-focus increase of F0 and post-focus reduction of F0 and intensity are both noncontrastive in L1 Southern Min. Southern Min learners of Mandarin may therefore have had difficulty perceiving these features and applying them in their L2 Mandarin production.

SLM further suggests that bilinguals' category representations are based on different features or feature weightings of acoustic-phonetic features than monolinguals' representations. If this is the case, then we can imagine that the phonological category established for L2 production of focus by bilingual speakers of Southern Min and Mandarin may differ from that of monolingual Mandarin speakers. Older Quanzhou bilingual speakers produced prosodic focus in L2 Mandarin using only in-focus expansion of duration and intensity and no overall F0 change, which is in contrast to native Beijing Mandarin monolinguals, who produce prosodic focus with in-focus expansion of duration, F0 and intensity, and post-focus compression of $F_0$ and intensity (Xu, 1999; Xu et al., 2012). Thus, older Quanzhou bilinguals seem to have not gone much beyond what they do in Southern Min for prosodic focus. Younger Quanzhou bilinguals, however, seem to have established a native-like phonological category of focus in both languages, including the relevant features of in-focus expansion and post-focus compression in their L2 Mandarin production.

SLM also suggests that bilinguals try to maintain a phonetic contrast between phonological categories in the common acoustic-phonetic L1/L2 space by defining new L2 categories that are "deflected" away from existing L1 categories. This could account for why younger Quanzhou bilinguals engaged in *pre-focus compression* of F0 and intensity in their L2 Mandarin production (Figures 3.3 and 3.4), which is rarely done by native Beijing Mandarin speakers (Xu, 1999; Xu et al., 2012). This new feature may have arisen to create additional contrast between the phonological representation of focus in the L1 Southern Min and L2 Mandarin productions by the younger bilingual speakers.

In summary, Experiment 1 suggests language experience affects the phonetic realization of prosodic focus in L2 Mandarin. PFC can be learned anew by speakers if given sufficient language experience, i.e., early exposure to high-quality L2 input and extensive use of L2. The findings in Experiment 1 can be interpreted within the framework provided by the Speech Learning Model (Flege, 1995).

# CHAPTER IV

# PRODUCTION OF PROSODIC FOCUS IN MANDARIN BY AMERICAN ENGLISH LEARNERS

## 4.1. Introduction

The findings from Experiment 1 suggest that an increase in L2 experience, driven by factors such as greater L2 use, results in more native-like production of L2 prosodic focus. Although older bilingual Southern Min-Mandarin speakers who used Mandarin less on a daily basis realized in-focus items in L2 Mandarin similarly to younger speakers who used the language more on a daily basis, only younger speakers showed a significant decrement in F0 and intensity following the in-focus item. The absence of PFC was attributed to the structure of Southern Min, which has lexical tone and conveys prosodic focus through changes in duration and intensity on the in-focus item, but does not have post-focus compression (PFC) of F0 and intensity. Experiment 1 sheds light on the positive role of increased L2 experience in the acquisition of L2 prosodic focus, especially in the acquisition of the associated phonetic feature: PFC.

In this chapter, we investigate the L2 Mandarin production of prosodic focus by L1 speakers of American English. In the United States, many students have been interested in learning Mandarin Chinese in recent years. The Chinese Language Flagship Program, sponsored by the US government, aims to exponentially increase the number of American students graduating from college with fluent and professional Mandarin Chinese. For example, at the University of Oregon (UO), the Chinese Flagship Program is a comprehensive four-year program for undergraduate students who have at least an intermediate level proficiency in Mandarin when admitted to the program. Once admitted to the Flagship program, students pursue advanced Mandarin language courses and take other content courses from across academic disciplines in Mandarin as well. The professors who teach the language and content classes are all native Mandarin speakers from mainland China; they all speak a variety of Mandarin with PFC. The American students in the Flagship Program also spend a semester or a year at an academic institution in Nanjing, Qingdao or Beijing, where the variety of Mandarin spoken has PFC. The present study investigates how well these advanced learners are able to convey prosodic focus in Mandarin.

In English, prosodic focus is realized by increased F0, intensity, and duration of the in-focus item and by a significant decrement of F0 and intensity following the item. In this way, American English is similar to Mandarin, which also realizes focus through an expansion of in-focused items and compression of post-focus items. In other ways, the languages differ substantially. The difference that is of particular relevance to the second language acquisition of prosodic focus is that Mandarin is a tone language and English is not. According to the Target Approximation (TA) model (Xu & Wang, 2001; Xu & Xu, 2005), prosodic focus in the two languages can nonetheless be modeled in a similar way—each syllable specified with a tone. The difference is that in Mandarin tonal targets are both static and dynamic. In English, the tonal targets are mostly static. Given the interaction between tone and intonation presented in Chapter II, the dearth of dynamic targets in English may mean that American English learners of Mandarin will have difficulty implementing prosodic focus in a native-like fashion.

Four research questions are addressed in this experiment: (1) Can advanced American learners of Chinese produce prosodic focus in L2 Mandarin with in-focus expansion of duration, intensity and F0? (2) Do they also produce post-focus compression of F0 and intensity similarly to native Bejing Mandarin speakers? (3) Do they have any difficulties producing the four tones of Mandarin, and does their ability to produce Mandarin tones vary with the realization of prosodic focus? (4) Finally, does age of learning influence the realization of L2 prosodic focus in speakers with similar levels of L2 proficiency? With respect to the last question, the students in the Flagship program at the UO include both Chinese heritage (CH) and non-Chinese heritage (NCH) American students. The CH students are typically exposed to Mandarin in early childhood and were sent to local Chinese language schools on the weekend to learn Mandarin. The NCH students typically started learning Mandarin in high school. Therefore, the CH learners and NCH learners of Mandarin differed from one another in age of learning (AOL), one of the most important factors affecting the acquisition of L2 phonology (Piske, 2007).

Based on the prosodic patterns of English and Mandarin, the predictions from SLM are that: (1) American learners will produce PFC in Mandarin because PFC occurs in English; (2) assuming the TA model of tone and intonation, American learners are expected to produce in-focus expansions in Mandarin using F0, intensity, and duration because English also uses these parameters to implement prosodic focus; (3) American learners of Mandarin are expected to have more difficulty producing the dynamic tones of Mandarin (Tones 2 and 4) than the static tones because most tonal targets in English are static with the exception of certain pitch accents; and (4) CH learners will produce more native-like prosodic focus and lexical tones than NCH learners due to earlier exposure to Mandarin.

## 4.2. Methods

### *4.2.1. Participants*

Two groups of American English-speaking learners of Mandarin, Chinese-heritage (CH) and non-Chinese-heritage (NCH), participated in the experiment. A control group of native Beijing (BJ) Mandarin speakers was also included. There were five male and five female speakers in each group. All participants were advanced learners of Mandarin and undergraduate or first-year graduate students at the University of Oregon. A language background questionnaire (LBQ) was completed by each learner in the two learner groups. The overall profile of the students in each group is shown in Table 4.1.

**Table 4.1.** Language background of the two groups of American learners of Mandarin. Mean and standard deviation shown for the age of test, age of learning Mandarin, length of residence by year in China, self-reported amount of Mandarin use, and self-estimated Mandarin proficiency in listening and speaking.

| Learner Groups | | Age of Test | Age of Learning | Length of Residence | Mandarin Use | Self-est. Listening | Self-est. Speaking |
|---|---|---|---|---|---|---|---|
| Chinese-Heritage | Mean | 20.9 | 4.3 | 0.9 | 37% | 7.8 | 7.5 |
| | *SD* | *1.0* | *3.1* | *0.6* | *18%* | *0.6* | *0.7* |
| Non-Chin-Heritage | Mean | 22.1 | 17.3 | 1.1 | 27% | 6.5 | 6.4 |
| | *SD* | *3.5* | *3.2* | *0.6* | *13%* | *1.1* | *1.1* |

CH learners were first exposed to Mandarin between birth to age 9. By contrast, NCH learners were first exposed to Mandarin in high school. Some of the CH learners' parents were Cantonese or Taiwanese speakers, and some were exposed to a mixture of Beijing Mandarin and Taiwanese Mandarin in their local Chinese language schools. At the time of test, all CH learners and eight NCH learners were studying in or just finished the Chinese Flagship Program at the University of Oregon. The other two NCH learners had just finished a year study-abroad program in Beijing. All the NCH learners and eight CH learners had six months to two years study-abroad experience in China. The other two CH learners were about to go to China for a year study when tested.

*4.2.2. Stimuli*

The present experiment investigates focus-induced changes as a function of position within the sentence. Tone 1 was used for both the pre-focus and post-focus constituents. Tone 1 was chosen because it is the only true level tone in Mandarin. In order to examine learners' production of lexical tones as a function of focus, tonal targets for in-focus words were varied. Otherwise, the target sentences in Experiment 2 were similar to those used in Experiment 1. As in Experiment 1, prompt questions were used to elicit focus in the desired position (See Table 4.2). Because pre-focus and post-focus items are controlled in Tone 1, the no-focus answer /u55ma55 mo55 ni55la55/ was used three times respectively paired with the initial focus on /u55ma55/, medial focus on /mo55/ and final focus on /ni55l55/. Therefore, there were twenty-two target sentences in total.

**Table 4.2.** Prompt questions and answers of the Mandarin stimuli in Experiment 2. The in-focus words are underscored.

| | | |
|---|---|---|
| No focus | Question | 你说什么?<br>'What did you say?' |
| | Answer | *See initial, medial, and final focus sentences below.* |
| Initial focus | Question | 谁摸妮拉?<br>'Who patted Nila?' |
| | Answer | 邬妈/ 刘妈/ 李妈/ 魏妈 摸 妮拉。<br>u55ma55/ liou35ma55/ li214ma55/ uei51ma55/ mo55 ni55la55<br>'Wuma/ Liuma/ Lima/ Weima patted Nila.' |
| Medial focus | Question | 邬妈对妮拉做什么?<br>'What did Wuma do to Nila?' |
| | Answer | 邬妈 摸/ 挠 /搂 /骂 妮 拉。<br>u55ma55 mo55/ nau35/ lou214/ ma51 ni55la55<br>'Wuma patted/ scratched/ hugged/ cursed Nila.' |
| Final focus | Question | 邬妈摸谁?<br>'Who did Wuma pat?' |
| | Answer | 邬妈 摸 妮拉/ 妮兰/ 妮美/ 妮娜。<br>u55m55 mo55 ni55la55/ ni55lan35/ ni55mei214/ ni55na51<br>'Wuma patted Nila/ Nilan/ Nimei/ Nina'. |

### *4.2.3. Recording*

Recording took place in a sound-attenuated booth in the Phonetics Laboratory at the University of Oregon. Target sentences were presented in Pinyin, simplified characters, and traditional characters using PowerPoint. Participants clicked through the slides to play the prompt questions and to answer with the target sentences. Each target sentence was produced to answer the paired non-focus vs. focus questions in three pre-determined pseudorandom orders. A Marantz professional solid state recorder PMD670 and a Shure professional unidirectional head-worn dynamic microphone were used. Target sentences were directly recorded into a computer SD card with a sampling rate of 44,100Hz.

### *4.2.4. Analyses*

According to the convention of L2 speech production analysis, acoustic measures were made on the second repetition of the sentence unless this was damaged, in which case the third repetition was used. Data were analyzed using Praat version 5375 and ProsodyPro version 5.5.2. Similar to Experiment 1, time-normalized F0 was collected at 10 points in each syllable across the entire stimulus sentence.

Prosodic changes due to focus were defined as the difference in mean F0, intensity and duration between sentences with and without prosodic focus (focus condition minus no-focus condition). Because tone was controlled in the pre-focus and post-focus constituents, analysis of prosodic changes by location relative to focus (in focus, pre-focus, and post-focus) was examined for Tone 1 only. In theses analyses, values were  collapsed across the three focus locations (initial, medial and final) in the target sentences such that in-focus change was computed by taking the syllable means for the three focused words with Tone 1 /u55 ma55, mo55, ni55 la55/ in the different sentence positions; pre-focus change was computed based on the words /u55 ma55, mo55/ in the final focus condition and the words /u55 ma55/ in the medial focus condition; and, post-focus change was computed based on the words /mo55, ni55 la55/ in the initial focus condition, and the words /ni55 la55/ in the medial focus condition.

To examine tone production as a function of focus, mean F0, F0 range, intensity and duration were measured for /u55, liu35, li214, uei51/ in initial focus vs. no focus, /mo55, nao35, lou214, ma51/ in medial focus vs. no focus, and /ni55, lan35, mei214, na51/ in final focus vs. no focus. Changes in mean F0, F0 range, intensity and duration were collapsed across the three focus locations for each tone type. In order to specifically examine the production of static versus dynamic tones, specifically the F0 slope, F0 velocity (i.e. F0 change rate over time) was also measured at the 30 ms before the offset of the syllable for each tone produced in focus. According to the Target Approximation Model (Xu & Wang, 2001; Xu, 2005; Xu & Xu, 2005) and the studies of Xu and colleagues (Xu & Liu, 2006, 2007; Gauthier, Shi & Xu, 2007), the F0 value and the F0 velocity at the 30 ms before the syllable offset represent the target height and the target slope for the underlying pitch target for a tone at phrase level. F0 velocity was computed as $F0' = (F0_{i+1} - F0_{i-1}) / (t_{i+1} - t_{i-1})$ based on sampled F0 contour with the sampling rate 100 points per second.

## 4.3 Results

### *4.3.1. The overall F0 contours*

Similar to Experiment 1, time-normalized F0 contours of the target sentences were first plotted by focus location and subject group. In Figures 4.1-4.3, each curve represents an average of the ten speakers' production of the same tone type. Solid curves represent the non-focus condition and dashed curves represent the focus condition. Syllable boundaries are marked with vertical dashed lines.

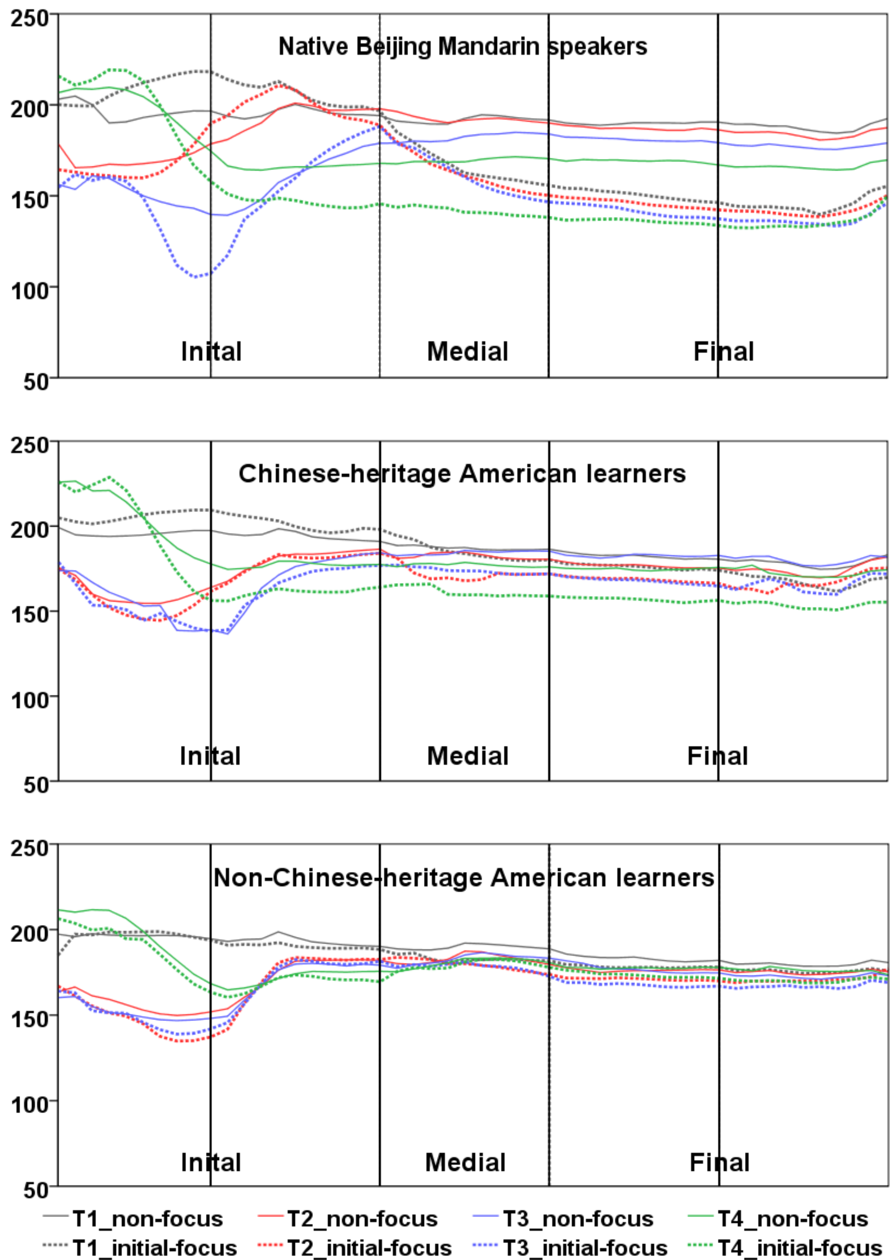


**Figure 4.1.** Time-normalized F0 contours (Hz) with initial focus by tone type and speaker group in Mandarin.

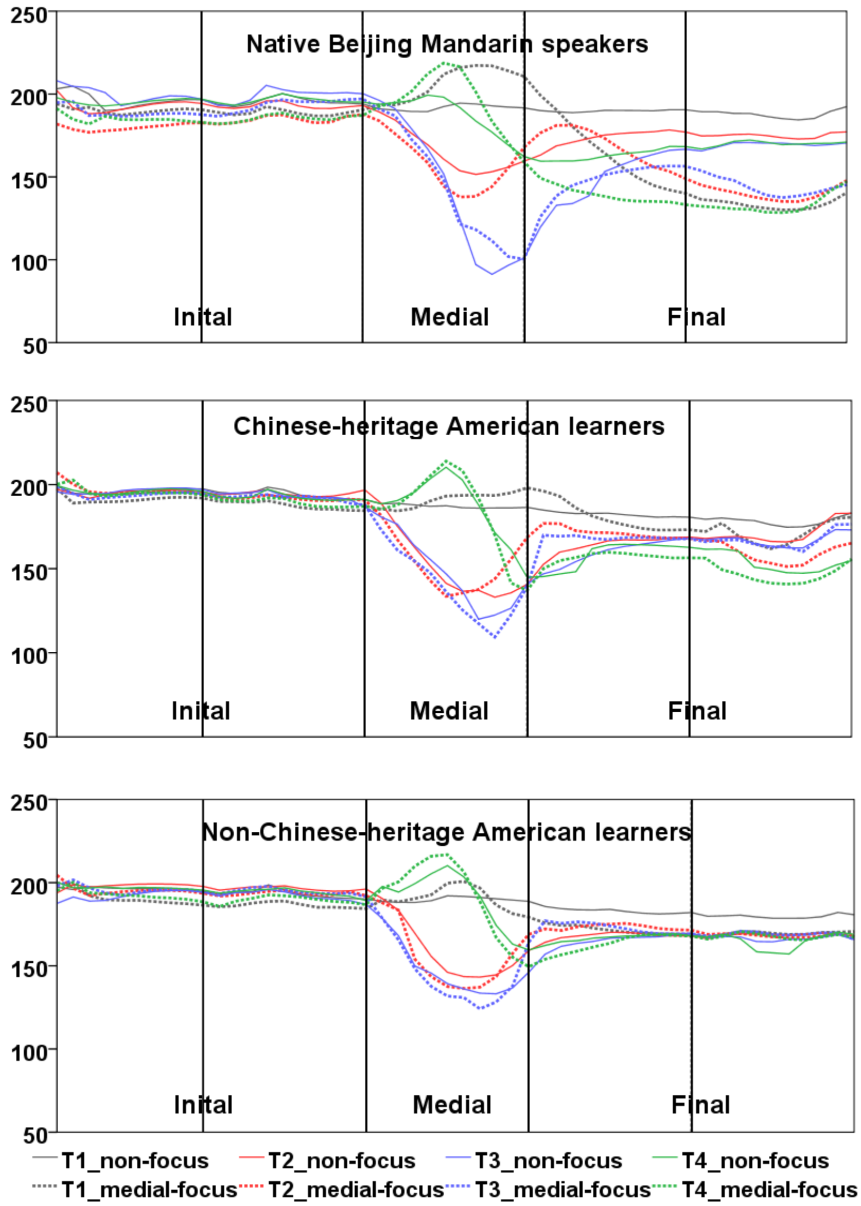


**Figure 4.2.** Time-normalized F0 contours (Hz) with medial focus by tone type and speaker group in Mandarin.

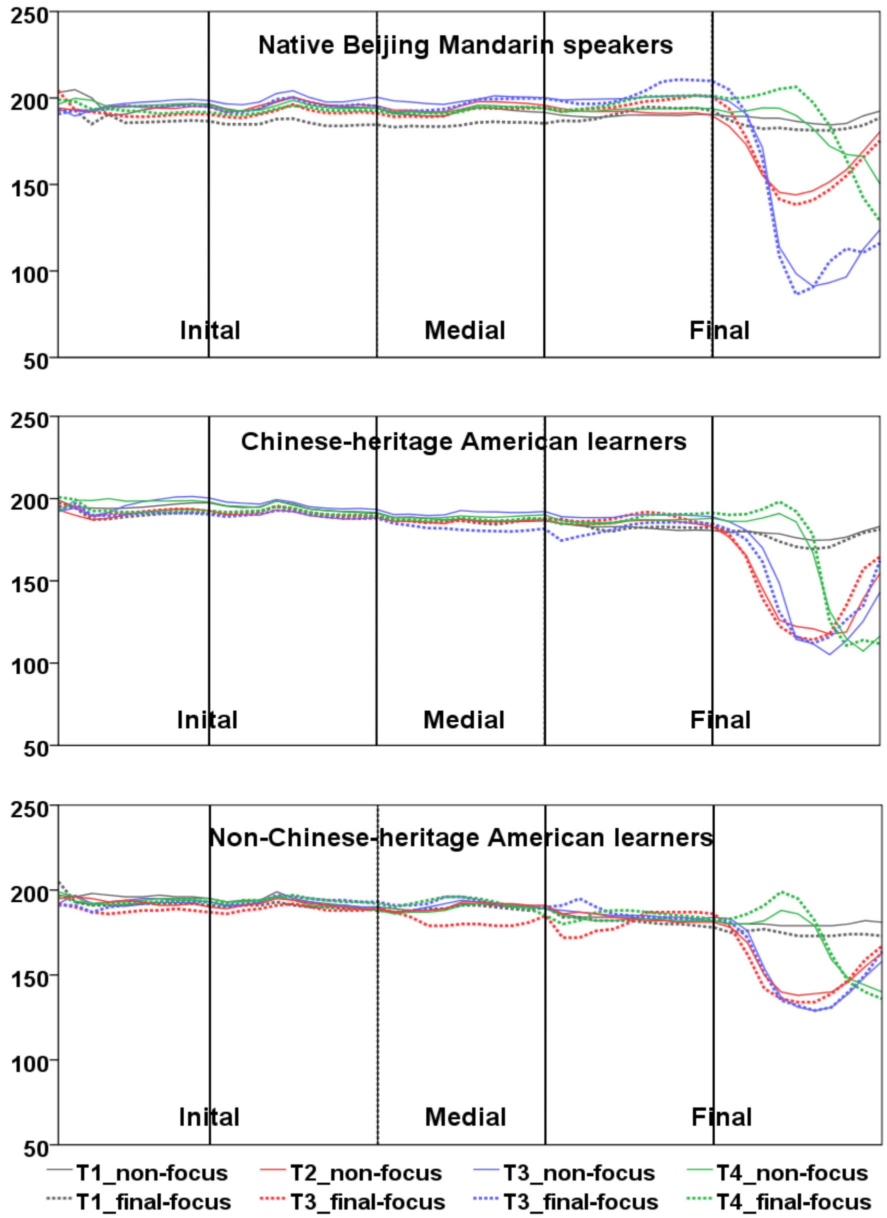


**Figure 4.3.** Time-normalized F0 contours (Hz) with final focus by tone type and speaker group in Mandarin.

Figure 4.1 indicates that native Beijing Mandarin speakers used an expanded F0 range to produce the lexical tones associated with in-focus items in sentence-initial position. It also shows a distinct pattern of post-focus compression of F0 for these speakers. By contrast, Chinese-heritage American learners' speech exhibited less in-focus expansion and less post-focus compression. Non-Chinese-heritage American learners' speech exhibited almost no in-focus expansion and no post-focus compression.

Figure 4.2 shows that native Beijing Mandarin speakers used an expanded pitch range to produce the different tones associated with in-focus items in sentence-medial position. It also shows a distinct pattern of post-focus compression of F0 for these speakers, but no clear pre-focus F0 change. Similar to the initial focus condition, Chinese-heritage American learners' speech exhibited much less in-focus expansion and post-focus compression than native Beijing Mandarin speech; non-Chinese-heritage American learners' speech exhibited almost no in-focus expansion and post-focus compression. Pre-focus F0 change was not observed in the speech of either learner group.

Figure 4.3 shows no salient in-focus or pre-focus change of F0 in the production of any speaker group in the final focus condition. However, the learner groups' production of individual tones differed from that of the native Beijing speakers, particularly that of Tone 3. Differences between native speakers and the learners' production of Tone 3 can also be seen in Figures 4.1 and 4.2. Also, the F0 contours of Tone 2 and Tone 3 looked very similar in the learners' production, especially in the NCH group.

*4.3.2. The change of mean F0, intensity and duration by sentence location relative to focus*

To test whether the focus-related F0 changes displayed in Figures 4.1-4.3 truly differed across the speaker groups, F0 difference values were calculated for words with Tone 1 (focus condition – no focus condition), converted from Hz to semitone, and subjected to a repeated measures ANOVA with two factors—sentence location relative to focus (pre-focus, in-focus, post-focus), a within-subjects factor, and speaker group (BJ, CH, NCH), a between-subjects factor. The results showed a two-way interaction between location and speaker group ($F(4,54) = 17.453$, $p < 0.001$). The main effects of location ($F(2,54) = 51.174$, $p < 0.001$) and speaker group ($F(2,27) = 25.345$, $p < 0.001$) were also highly significant. Figure 4.4 displays the means and standard errors of F0 change as a function of sentence location relative to focus and speaker group.

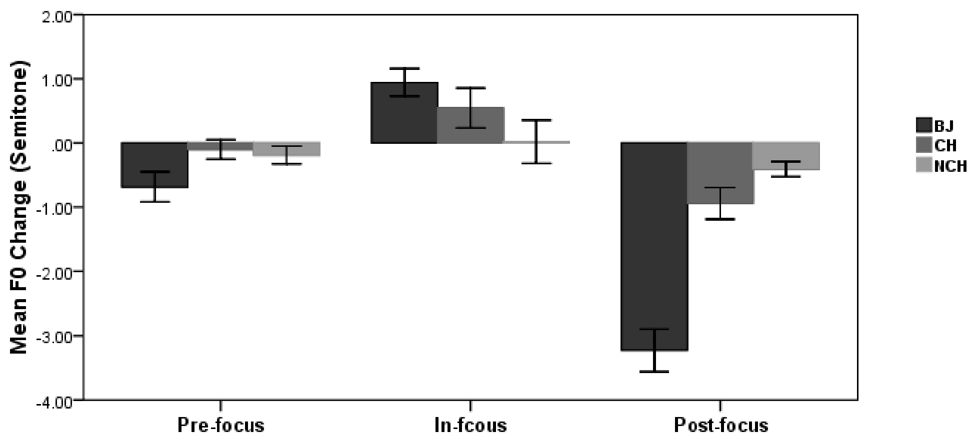


**Figure 4.4.** Mean F0 change (semitone) by sentence location relative to focus item and speaker group in Mandarin.

Follow up post-hoc independent samples t-tests were used to assess differences by speaker group. These tests showed a significant difference between the BJ and CH groups ($t(18) = -2.097$, $p = 0.05$) and a nearly significant difference between the BJ and NCH groups ($t(18) = -1.805$, $p = 0.088$) in F0 change for items in the pre-focus location. A significant difference between the BJ and NCH groups ($t(18) = 2.296$, $p = 0.034$) was also found for items in the in-focus location. Both learner groups differed from native speakers in F0 change for items in the post-focus location: BJ vs. CH ($t(18) = -5.548$, $p <$

0.001) and BJ vs. NCH ($t(18) = -8.034$, $p < 0.001$). The difference between the learner groups (CN vs. NCH) was also nearly significant for items in the post-focus location ($t(18) = -1.94$, $p = 0.068$).

In a separate analysis, the difference in the magnitude of F0 change from in-focus to pre-focus (in-focus minus pre-focus) and from in-focus to post-focus (in-focus minus post-focus) was examined. The results showed a two-way interaction between location and speaker group ($F(2,27) = 16.312$, $p < 0.001$). The main effects of location ($F(1,27) = 48.637$, $p < 0.001$) and group ($F(2,27) = 18.24$, $p < 0.001$) were also significant. Follow up post-hoc tests showed that both learner group differed significantly from native speakers in the magnitude of F0 change from in-focus to pre-focus: BJ vs CH ($t(18) = 2.758$, $p = 0.013$) and BJ vs NCH ($t(18) = 3.26$, $p = 0.004$). All groups differed from one another in the magnitude of F0 change from in-focus to post-focus: BJ vs CH ($t(18) = 4.61$, $p < 0.001$), BJ vs NCH ($t(18) = 6.343$, $p < 0.001$), and NC vs. NCH ($t(18) = 2.113$, $p = 0.049$).

To understand the relationship between F0 change across the sentence, Pearson correlation coefficients were computed for items with Tone 1 for each speaker group. None of the speaker groups showed a significant correlation between in-focus and post-focus changes in sentences with either at initial or medial focus. The CH group showed significant correlation between in-focus and pre-focus changes in sentences with medial ($r(10) = 0.66$, $p = 0.038$) and focus ($r(10) = 0.844$, $p = 0.002$).

Repeated measures ANOVAs similar to those on mean F0 change were also conducted for mean intensity change (focus condition – no focus condition). The results showed a two-way interaction between sentence location relative to focus and speaker group ($F$(4,54) = 7.997, $p < 0.001$). The main effects of location ($F$(2,54) = 101.402, $p < 0.001$) and group ($F$(2,27) = 15.444, $p < 0.001$) were both highly significant. Figure 4.5 displays the means and standard errors of intensity change by sentence location and speaker group. Post-hoc independent samples t-tests showed significant differences between all groups, albeit only in the post-focus location: BJ vs. CH ($t$(18) = -4.192, $p$ = 0.001), BJ vs. NCH ($t$(18) = -7.973, $p < 0.001$), and CH vs. NCH ($t$(18) = -4.43, $p < 0.001$).

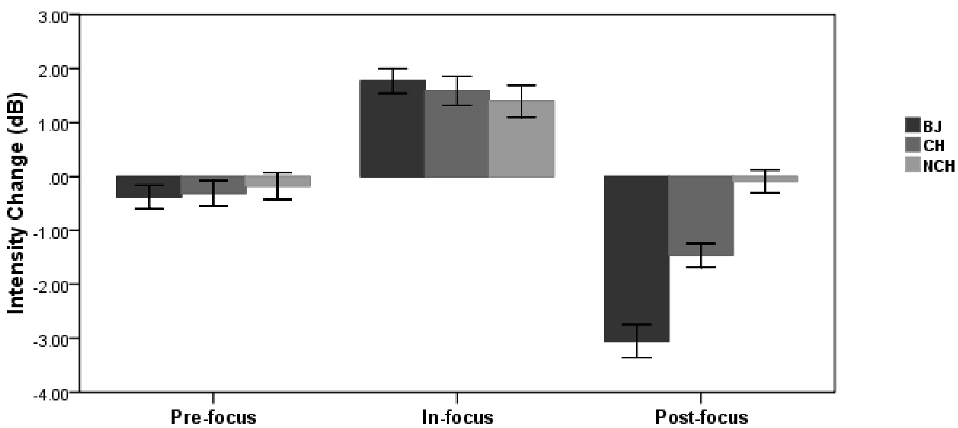


**Figure 4.5.** Mean intensity change (dB) by sentence location relative to focus item and speaker group in Mandarin.

With respect to the magnitude of change in intensity from in-focus to pre-focus and in-focus to post-focus, the ANOVA results showed a two-way interaction between location and group ($F$(2,27) = 21.62, $p < 0.001$). The main effects of location ($F$(1,27) = 52.659, $p < 0.001$) and group ($F$(2,27) = 6.33, $p$ = 0.006) were also both significant. Post hoc tests indicated that the magnitude of intensity change from in-focus to post-focus was significantly different between all groups: BJ vs. CH ($t$(18) = 3.52, $p$ = 0.002), BJ vs. NCH ($t$(18) = 5.914, $p < 0.001$), and CH vs. NCH ($t$(18) = 2.758, $p$ = 0.013). There were no significant group differences in mean intensity change or in the magnitude of change from in-focus to pre-focus locations.

Similar Pearson correlation coefficients were conducted in the intensity changes between in-focus and pre-focus and between in-focus and post-focus at each focus location for each speaker group. Again, none of the speaker groups showed a significant correlation of intensity change between in-focus and post-focus. A significant correlation of F0 change between in-focus and pre-focus was found at final focus in CH group ($r(10) = 0.705$, $p = 0.023$) and at medial focus in NCH group ($r(10) = -0.751$, $p = 0.012$). However, Figure 4.5 shows that both groups had very little pre-focus change of intensity.

A final set of ANOVA analyses similar to those on F0 and intensity change were conducted for duration changes (focus condition - no focus condition). An overall analysis of change showed a two-way interaction between sentence location relative to focused item and speaker group ($F(4,54) = 8.43$, $p < 0.001$). The main effects of location ($F(2,54) = 116.283$, $p < 0.001$) and group ($F(2,27) = 5.616$, $p = 0.009$) were both highly significant. Figure 4.6 displays the means and standard errors of duration change by sentence location and speaker groups. Post-hoc independent samples t-tests showed significant differences of duration change between BJ and CH groups ($t(18) = -5.25$, $p < 0.001$) and between BJ and NCH groups ($t(18) = -2.487$, $p = 0.029$).

An analysis of the magnitude of duration change from in-focus to pre-focus and in-focus to post-focus showed no main effect of location, but the main effect of group was significant ($F(2,27) = 11.167$, $p < 0.001$). The magnitude of duration change from in-focus to pre-focus was significantly different between the BJ and CH groups ($t(18) = -4.996$, $p < 0.001$) and between the CH and NCH groups ($t(18) = 2.223$, $p = 0.039$); the

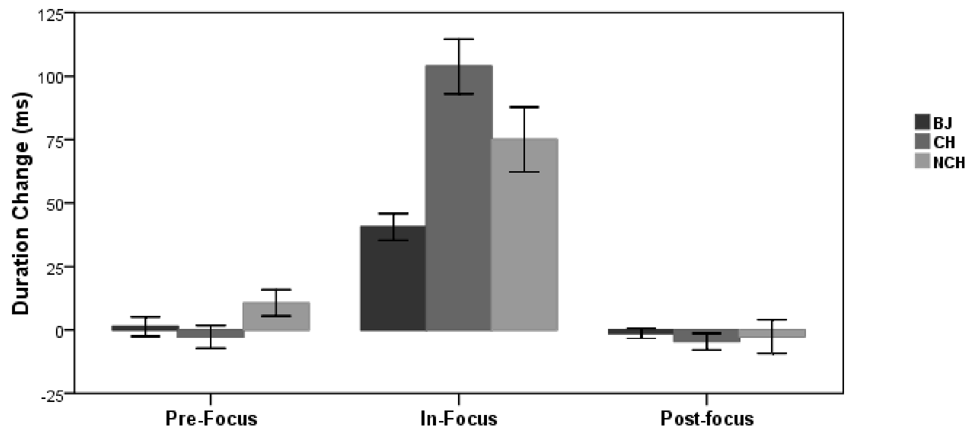


**Figure 4.6.** Mean duration change (ms) by sentence location relative to focused item and speaker group in Mandarin.

magnitude of change from in-focus to post-focus was significantly different between the BJ and CH groups ($t(18) = -5.842$, $p < 0.001$), the BJ and NCH groups ($t(18) = -2.79$, $p = 0.012$), and nearly significant between the CH and NCH groups ($t(18) = 2.024$, $p = 0.058$).

*4.3.3. In-focus change of F0, intensity and duration by tone type*

The next set of analyses investigated the effect of focus on the production of different tone types. Change in mean F0, F0 range, and mean intensity and duration were the dependent variables. Repeated measures ANOVAs with two factors— tone type (Tones 1-4) and speaker group (BJ, CH and NCH)— showed no main effect of group or any interaction between tone type and group, but there was a significant main effect of tone type on mean F0 change ($F(3,81) = 4.559$, $p = 0.005$); F0 range ($F(3,81) = 4.957$, $p = 0.003$); intensity change ($F(3,81) = 5.478$, $p = 0.002$); and duration change ($F(3,81) = 3.903$, $p = 0.012$).

Since the absence of a group effect on the difference values does not necessarily imply an absence of group difference in the default or in focus production of items, another set of repeated measures ANOVAs were conducted on the values obtained for each condition.

The ANOVA results indicated significant interactions between focus condition and speaker group ($F(2,27) = 4.022$, $p = 0.03$) and between focus condition and tone type ($F(3,81) = 4.876$, $p = 0.004$). The main effects of focus condition ($F(1,27) = 4.433$, $p = 0.045$) and tone ($F(3,81) = 114.899$, $p < 0.001$) were also significant, but there was no main effect of group. These results are shown in Figure 4.7 below.

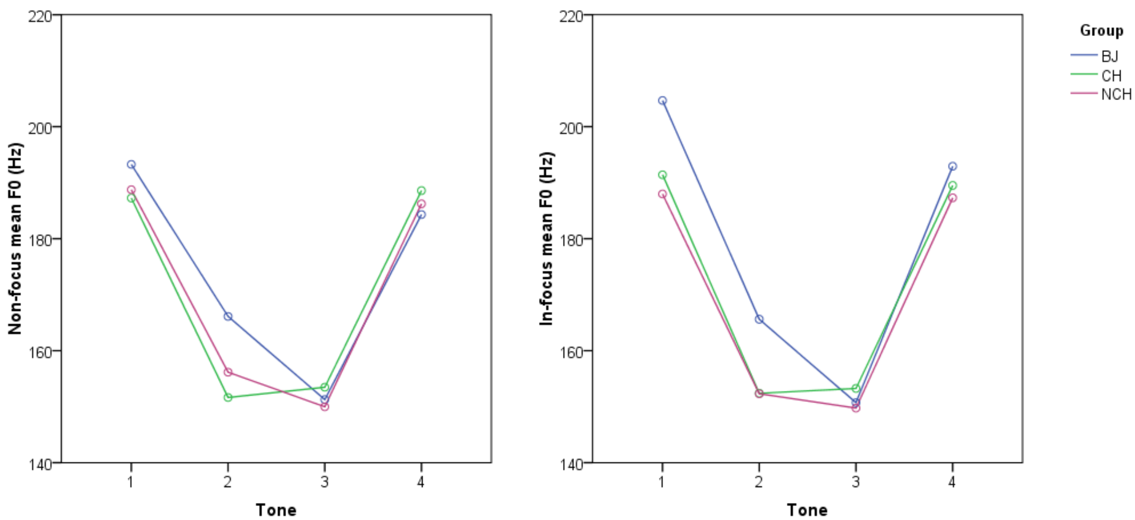


**Figure 4.**7. Mean F0 in the no-focus and in-focus conditions as a function of tone type and speaker group in Mandarin.

Post-hoc independent samples t-tests showed no significant difference between BJ and CH groups or between BJ and NCH groups for mean F0 in the default (no focus) and focus conditions. There was a significant difference in mean F0 change (in focus – no focus) between BJ and CH groups for Tone 4 ($t(18) = 2.134$, $p = 0.047$) and between BJ and NCH groups for Tone 1 ($t(18) = 2.395$, $p = 0.028$).

There were significant interactions between focus condition and tone for F0 range ($F(3,81) = 4.957$, $p = 0.003$) and between group and tone ($F(2,27) = 3.15$, $p = 0.008$). Also significant were the main effects of focus condition ($F(1,27) = 193.74$, $p < 0.001$) and tone ($F(3,81) = 48.562$, $p < 0.001$). These results are shown in Figure 4.8 below.

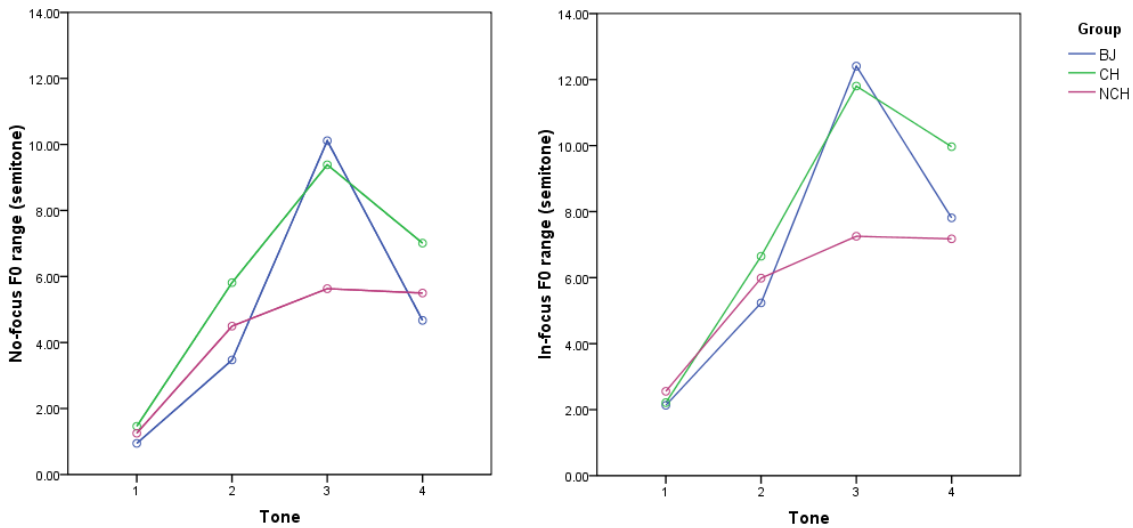


**Figure 4.8.** F0 range in the no-focus and in-focus conditions as a function of tone type and speaker group in Mandarin.

Post-hoc independent samples t-tests showed a significant difference between the BJ and CH groups in F0 range for Tone 1 ($t(18) = -3.446$, $p = 0.003$) and Tone 2 ($t(18) = -2.942$, $p = 0.009$) in the default no focus condition. Significant differences were also found between BJ and NCH groups for Tone 3 in the default no focus condition ($t(18) = -3.036$, $p = 0.007$) and the in-focus condition ($t(18) = 3.318$, $p = 0.004$). The change in F0 range from no focus condition to the in-focus condition was significantly different between the BJ and NCH groups for Tone 4 ($t(18) = 2.264$, $p = 0.036$).

Similar to F0, intensity changes varied systematically with focus condition and tone ($F(3,81) = 5.478$, $p = 0.002$) and with group and tone ($F(6,27) = 2.578$, $p = 0.025$). There were also main effects of focus condition ($F(1,27) = 99.845$, $p < 0.001$) and tone ($F(3,81) = 65.837$, $p < 0.001$) on intensity change, but no significant effect of group. These results are shown in Figure 4.9 below.

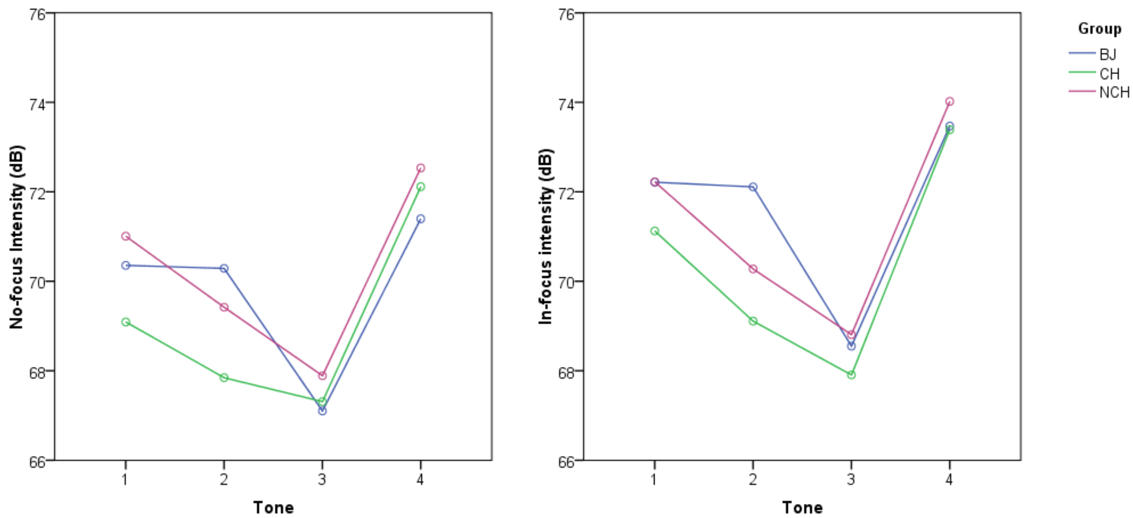


**Figure 4.9.** Mean intensity in the no-focus and in-focus conditions as a function of tone type and speaker group in Mandarin.

Post-hoc independent samples t-tests showed a significant difference of intensity change between BJ and CH groups in Tone 3 ($t(18) = 2.377$, $p = 0.029$). The BJ and NCH groups differed significantly in their production of Tone 2 ($t(18) = 2.712$, $p = 0.014$).

Finally, like F0 and intensity, duration varied systematically with focus condition and tone ($F(3,81) = 4.519$, $p = 0.006$) and with group and tone ($F(6,27) = 4.421$, $p =$

0.001). The main effects of focus ($F(1,27) = 254.506$, $p < 0.001$), group ($F(2,27) = 28.813$, $p < 0.001$), and tone ($F(3,81) = 84.995$, $p < 0.001$) were also all highly significant. These results are shown in Figure 4.10 below.

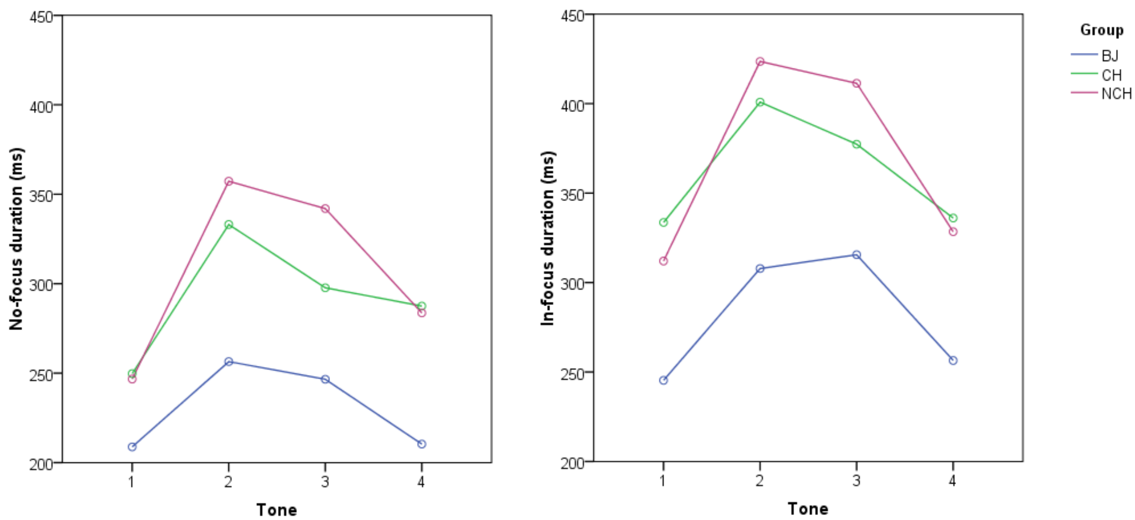


**Figure 4.10.** Mean duration in the no-focus and in-focus conditions as a function of tone type and speaker group in Mandarin.

Post-hoc independent samples t-tests showed that the BJ and CH groups and the BJ and NCH groups produced items with significantly different duration in the default no focus. These results are shown in Table 4.3.

Post-hoc independent samples t-tests also showed a significant difference of duration change from non-focus condition to in-focus condition in Tone 1 between BJ and CH groups ($t(18) = 11.416$, $p = 0.005$) and a near significant difference between BJ and NCH groups ($t(18) = -1.856$, $p = 0.08$).

**Table 4.3.** The results of t-tests of duration by tone type in no-focus and in-focus conditions between groups in Experiment 2 (*df* = 18).

| | No-focus | | | | In-focus | | | |
|---|---|---|---|---|---|---|---|---|
| Tone | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| CH vs. BJ | $t$= 3.09<br>$p$= .006 | $t$= 4.21<br>$p$= .001 | $t$= 3.36<br>$p$= .003 | $t$= 5.19<br>$p$< .001 | $t$= 5.64<br>$p$< .001 | $t$= 4.62<br>$p$< .001 | $t$= 3.69<br>$p$= .002 | $t$= 4.96<br>$p$< .001 |
| NCH vs. BJ | $t$= 3.35<br>$p$= .004 | $t$= 6.76<br>$p$< .001 | $t$= 6.45<br>$p$< .001 | $t$= 5.40<br>$p$< .001 | $t$= 4.17<br>$p$= .001 | $t$= 5.75<br>$p$< .001 | $t$= 6.45<br>$p$< .001 | $t$= 4.13<br>$p$= .001 |
| NCH vs. CH | --<br>-- | --<br>-- | $t$= 2.55<br>$p$= .020 | --<br>-- | --<br>-- | --<br>-- | $t$= 2.11<br>$p$= .049 | --<br>-- |

*4.3.4. Tone production as a function of focus condition and sentence position*

A final set of analyses were conducted on F0 velocity in order to assess whether American learners of Mandarin produced the static tones (Tone 1 and 3) in a more native-like fashion than the dynamic tones (Tones 2 and 4) as predicted. The F0 velocity at the 30 ms before the offset of the syllable was measured for each tone type by syllable position in the sentence and focus condition. The reason of taking sentence position as a variable is that different syllable location in a sentence results in different F0 slope due to the tonal coarticulation with the adjacent tones. In this set of stimuli, each tone type is followed by a high level tone at sentence-initial position, preceded and followed by high level tones at sentence-medial position, and preceded by a high level tone at sentence-final position. Therefore, a repeated measures ANOVA was conducted with four factors— focus condition (two levels: no-focus, in-focus), sentence position (three levels: initial, medial, final), tone type (four levels: Tones 1-4), and group (three levels: BJ, CH, NCH). There was no four-way interaction and neither three-way interactions. There were two-way interactions between tone and group ($F(6,81) = 3.011$, $p = 0.01$) and between tone and position ($F(6,162) = 7.183$, $p < 0.001$). The main effects of tone ($F(3,81) = 30.358$, $p < 0.001$), position ($F(2,54) = 2.709$, $p = 0.001$) and group ($F(2,27) = 7.294$, $p = 0.031$) were significant. There was no main effect of focus condition.

The data were then split into two sets of ANOVAs with three factors-- tone, position and group, separately in no-focus condition and in in-focus condition. The results of no-focus condition showed no three-way interaction. The interaction between tone and position ($F(6, 162) = 2.925$, $p = 0.01$) was significant. The main effects of tone ($F(3,81) = 24.265$, $p < 0.001$) and position ($F(2,54) = 7.383$, $p = 0.001$) were significant. There was no main effect of group on the tone production in no-focus condition. The results of in-focus condition showed no three-way interaction. The interactions between tone and position ($F(6, 162) = 3.201$, $p = 0.005$) and between tone and group ($F(6,81) = 3.909$, $p = 0.002$) was significant. The main effects of tone ($F(3,81) = 17.032$, $p < 0.001$) and group ($F(2,27) = 5.973$, $p = 0.007$) were significant. The main effect of position was not significant.

Post-hoc independent samples t-tests showed no significant difference of F0 velocity at the final position between any two of the three groups in either the non-focus or in-focus condition. The results of significant or near significant ($p < 0.1$) at the initial and medial positions between the control and the learner groups are reported in Table 4.4.

**Table 4.4.** The results of t-tests of F0 velocity in no-focus and in-focus conditions at initial (I) and medial positions (M) between groups in Experiment 2 (*df* = 18).

| | | Non-focus condition | | | | In-focus condition | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Location | Tone 1 | Tone 2 | Tone 3 | Tone 4 | Tone 1 | Tone 2 | Tone 3 | Tone 4 |
| CH vs. BJ | I | -- | -- | -- | -- | -- | *t*=-1.93 | -- | *t*=2.01 |
| | | -- | -- | -- | -- | -- | *p*=.069 | -- | *p*=.059 |
| | M | -- | -- | -- | *t*=1.88 | *t*=3.3 | -- | -- | *t*=2.74 |
| | | -- | -- | -- | *p*=.077 | *p*=.004 | -- | -- | *p*=.014 |
| NCH vs. BJ | I | -- | *t*=-2.14 | *t*=3.07 | -- | -- | *t*=-4.05 | -- | *t*=1.79 |
| | | -- | *p*=.046 | *p*=.007 | -- | -- | *p*=.001 | -- | *p*=.091 |
| | M | *t*=-2.44 | | *t*=2.42 | *t*=2.07 | -- | *t*=-1.93 | -- | *t*=3.23 |
| | | *p*=.026 | | *p*=.026 | *p*=.053 | -- | *p*=.069 | -- | *p*=.005 |

## 4.4. Discussion

The results from Experiment 2 indicate that PFC in Mandarin is difficult for American learners of Chinese to acquire even though there is PFC in English. This result suggests that PFC does not transfer easily from one language to another. The results also indicate that AOL influences learners' ability to phonetically realize focus in a native-like manner. Chinese heritage students, who were exposed to Mandarin earlier than non-Chinese heritage peers, were more likely to realize PFC than non-Chinese heritage students, suggesting that language-specific patterns of PFC can be learned given sufficient L2 experience. Both CH and NCH groups produced in-focus expansion of duration and intensity but only the CH group increased in-focus F0. Similar to the results in Experiment 1, the Pearson correlations indicate that post-focus compression was independent of in-focus expansion. Again, this finding is likely attributable to speakers' efforts on maintaining the local pitch contour of in-focus words and also the adoption of Tone 1 in the analysis.

As for interactions with tone production, American-English speaking learners of Mandarin produced all four tones with longer syllable durations than native Beijing Mandarin speakers. This result suggests a slower speech rate, which is consistent with other studies in the literature on L2 speech (Guion et al., 2000; Trofimovich & Baker, 2006; Aoyama & Guion, 2007; Huang & Jun, 2011). Learners also had more difficulties with F0 range than with mean F0, especially Tone 2 for CH learners and Tone 3 for NCH learners. Difficulties in Tone 2 and Tone 3 production were also observed in the results on F0 velocity and intensity. Learners produced Tone 2 and Tone 3 with similar F0 contours in contrast to the native Beijing Mandarin speaking participants who produced different F0 contours for each tone. This result is consistent with previous findings in the literature on the acquisition of Mandarin: Tone 2 and Tone 3 are frequently reported to be the most confusable pair of tones (Sun, 1998; Wang et al., 1999). Further, American-English speaking learners only produced native-like F0 velocity profiles for tones in sentence-final words. In sentence-initial and -medial position, learners had more difficulties. This result suggests that learners had more difficulties with the anticipatory coarticulation of tones than with perseveratory coarticulation of tones. Learners' particular difficulties producing Tone 3 and Tone 4 in sentence-initial and sentence-medial positions may be attributed to the target height of these two tones— Tone 3 is realized with a 21 pitch value, a low tone in non-final position, and the offset value for Tone 4 is also low. Coarticulation from a low offset to a high onset represents a conflicting tonal context (Xu, 1994) that may be especially difficult for L1 English learners (Yang, 2011). Finally, learners had more difficulties producing F0 velocity in contour tones than in the high level tone. This result is consistent with the literature, suggesting that English speakers focus more on F0 height than F0 contours (Gandour, 1983, 1984; Massaro et al., 1985; Guion & Pederson, 2007). It is also possible; however, that this results is due to the design of the stimuli in that neighboring tones were all Tone 1.

The results from the present experiment confirmed once again the importance of AOL to second language acquisition in that CH learners' realization of prosodic focus in Mandarin was much more native-like than NCH learners'. That said, it is surprising that American-English speaking learners of Mandarin were unable to realize prosodic focus in the manner of native Mandarin speakers. According to the Speech Learning Model, difficulties of acquiring an L2 category increase as the similarities between the L2 category and an L2 category increase, and the ability of discerning the differences between similar categories in L1 and L2 decreases as AOL increases. However, SLM also proposes that learners will apply the mechanisms and processes for establishing L1 categories to L2 category learning, and this is more easily done when there are shared acoustic-phonetic features in the two languages. SLM does suggest that native-like L2 categories are less likely to be acquired once the L1 system is fully mature because late learners are less capable than early learners to discern the phonetic differences between L1 and L2 sounds and between L2 sounds. This may explain in part NCH learners' failure to acquire Mandarin prosodic focus. It appears that these learners may have weighted the phonetic features associated with prosodic focus differently than native Mandarin speakers in that they produced the in-focus expansion of duration and intensity, but not F0. It is also likely that the NCH learners were distracted from learning the global phonetic patterns associated with prosodic focus because they were focused on the acquisition of lexical tones, a more local suprasegmental pattern. This possibility is supported by the finding that the tonal patterns were similar across focus conditions, suggesting that learners made more effort to maintain tonal contrasts than to fully convey prosodic focus, even though their production of lexical tones was also not native-like in all phonetic aspects. Although the CH learners produced some PFC of F0 and intensity, they also realized with more in-focus expansion of duration than the native Beijing Mandarin speakers did. It is possible that this too represents a differential weighting of the phonetic parameters (melodic primitives) associated with prosodic focus in American English and Beijing Mandarin.

# CHAPTER V

# PRODUCTION OF PROSODIC FOCUS IN ENGLISH BY BEIJING MANDARIN LEARNERS

## 5.1. Introduction

The findings from Experiment 2 were consistent with the findings from Experiment 1 in that even highly proficient second language learners of Mandarin did not produce prosodic focus in a native-like manner, and were especially unlikely to produce PFC. Experiment 2 also suggested that the presence of PFC in a learner's L1 does not facilitate its acquisition in L2, suggesting that PFC is not easily transferred from one language to another. In discussing the results of Experiment 2, it was suggested that transference of PFC from English to Mandarin may be impeded by English-speaking learners' focus on correct production of lexical tone. Experiment 3 tests whether transfer occurs if a learner is able to focus just on the phrase-level patterns in the L2, not having to worry so much at least about F0 contours at the level of the syllable. Specifically, Experiment 3 investigates the acquisition of prosodic focus in L2 English, a non-tone language with PFC, by L1speakers of Beijing Mandarin, a tone language with PFC.

Since the late 1990s, most elementary school students in China have been taught English starting from first to third grade for about 3 class hours a week. In the early 2000s, the number of class hours devoted to English was raised from 3 to 5. In addition, some children are also given additional afterschool English practice. There are 6 regular class hours and 3 practice class hours in middle school, and this continues through high school. In contrast to the traditional grammar-translation method of teaching foreign languages, modern EFL (English as a foreign language) teaching in China tries to balance the four language skills—listening, speaking, reading and writing. However, the assessments still normally include reading, writing and listening and not speaking. Also, the Chinese teachers of English frequently themselves have foreign-accented English. For this reason, students are rarely as balanced in listening, speaking, reading and writing as the teachers would like. In fact, it is rare for Chinese students to learn English from native English speakers before they go abroad to study in an English-speaking country. Because one cannot expect students to acquire native-like pronunciation of a second language without exposure to native speakers of the language, the present experiment investigated the acquisition of prosodic focus in English by Chinese learners who were undergraduate students in an American university. These students all had similar AOL, AOA, and quality of L2 English input, but they had different LOR in the United States.

Four research questions are explored in Experiment 3: (1) Can Beijing Mandarin learners of English produce prosodic focus in L2 English with a native-like pattern of in-focus expansion of duration, F0 and intensity? (2) Can they also produce post-focus compression (PFC) of F0 and intensity in L2 English? (3) In light of interactions between lexical stress and phrase-level prosody, we also wanted to know whether Mandarin learners have difficulties producing English word stress for items that are produced with prosodic focus. (4) Finally, what is the effect of LOR on the L2 acquisition of prosodic focus in English?

With respect to question (1), the expectation is that Beijing Mandarin learners will expand duration, intensity and F0 for in-focused items in English since doing so involves a simple extension of the phonetic parameters used to realize focus in Mandarin to its realization in English. Because prosodic focus in Mandarin and English are both accompanied by PFC, it is also expected that Mandarin learners will produce PFC in English. With respect to question (3), Mandarin learners are expected to realize stressed syllables better than unstressed syllables in English because mainstream Mandarin does not have a lexical stress contrast and the rhythm is accordingly described as syllable-timed (Duanmu, 1994; Chen, 2000). Finally, LOR is expected to positively influence L2 acquisition of English prosodic focus, assuming that learners with longer LORs have received more exposure to high quality English than learners with shorter LORs.

## 5.2. Methods

### *5.2.1. Participants*

Two groups of Beijing Mandarin learners of English, who were respectively freshmen and seniors at the University of Oregon, participated in the experiment. A control group of native American-English speakers was also included. There were five male and five female speakers in each group. Participants in the learner groups were all from north China and spoke Beijing Mandarin as their L1. They were all experienced learners of English, having passed the TOEFL test required to be admitted to the university. Since one group was freshmen students and the other senior students at the time of test, the learners' LOR differed. The freshman group had an LOR from 3 to 7 months, having studied at the UO for one to two terms. The senior group had an LOR from 3.5 to 4.5 years. The two students with the longest LORs in the senior group had completed their last year of high school in the US. The overall language background of the learner groups is reported in Table 5.1.

**Table 5.1.** Language background of the two groups of Beijing learners of English. Mean and standard deviations are shown for the age of test, age of learning English, length of residence by year in the US, self-reported amount of English use, and self-estimated (Self-est.) English proficiency in listening and speaking.

| Learner Groups | | Age of Test | Age of Learning | Length of Residence | English use | Self-est. Listening | Self-est. Speaking |
|---|---|---|---|---|---|---|---|
| Senior Students | Mean | 22.7 | 9.0 | 3.9 | 56% | 6.8 | 6.7 |
| | *SD* | *1.0* | *1.8* | *0.4* | *16%* | *0.6* | *0.8* |
| Freshman Students | Mean | 19.5 | 8.5 | 0.4 | 44% | 5.0 | 4.9 |
| | *SD* | *1.0* | *2.2* | *0.1* | *16%* | *0.9* | *1.2* |

### *5.2.2. Stimuli*

The stimuli are listed in Table 5.2. The non-focus prompt question was the same for all the three focus locations—initial, medial and final—and used to elicit a default no-focus production of each sentence. Sentences with contrastive focus in the tree different focus locations were elicited using different prompt questions. The in-focus words also varied in five types of lexical stress. There were thirty target sentences in total (15 sentences each in the no focus and in-focus conditions).

### *5.2.3. Recording*

Recording took place in a sound-attenuated booth in the Phonetics Laboratory at the University of Oregon. Target sentences were presented in PowerPoint. Participants clicked through the slides to play the prompt questions and answered the questions with the target sentences. Each target sentence was also produced in a default manner, without contrastive focus (i.e., the no-focus condition). The stimuli were presented in three pre-determined pseudorandom orders. A Marantz professional solid state recorder PMD670 and a Shure professional unidirectional head-worn dynamic microphone were used. Target sentences were directly recorded into a computer SD card with a sampling rate of 44,100Hz.

**Table 5.2.** Prompt questions and answers of the English stimuli in Experiment 3. The syllables with lexical stress are underscored.

| | | |
|---|---|---|
| Non-focus | Question | What's the news? |
| | Answer | *See initial, medial, and final focus sentences below.* |
| Initial focus | Question | Who may marry Ray? |
| | Answer | Leigh / Nina / Melanie / Marie / Ramona may marry Ray. |
| Medial focus | Question | What may Leigh do to Norman? |
| | Answer | Leigh may leave / marry / nominate / remind / remember Norman. |
| Final focus | Question | Who may Ray marry? |
| | Answer | Ray may marry Leigh / Nina / Melanie / Marie / Ramona. |

*5.2.4 Analyses*

Acoustic measures were again made on the second repetition of a sentence unless this was disrupted, in which case the third repetition was used. Data were analyzed by Praat version 5375 and ProsodyPro version 5.5.2. Similar to Experiments 1 and 2, time-normalized F0 was collected at 10 points in each syllable.

To examine the prosodic change associated with in-focus productions, difference values were used. F0, intensity and duration of each stressed syllable produced in the default no focus condition was subtracted from that of each stressed syllable produced in the contrastive focus conditions. Thus, in-focus change was based on stressed syllable difference values in the three focus locations (initial, medial, and final). Pre-focus change was based on the overall difference values for “Leigh may” in the medial focus condition and “Ray may marry” in the final focus condition. Post-focus change was based on the overall difference values of “may marry Ray” in the initial focus condition and “Norman” in the medial focus condition.

To examine the production of English lexical stress as a function of focus, mean F0, F0 range, intensity and duration were measured for both stressed syllables and unstressed in the focused words at three sentence locations (initial, medial and final) in both the default no focus and in-focus conditions. Once again, change was calculated as the in-focus value minus the no-focus value. Prosodic change was assessed by collapsing across all stressed and unstressed syllables in all in-focus target words. Unstressed syllables in the target words were treated differently depending on whether these occurred before or after the target stressed syllable. The motivation for this is that, according to the AM theory (Pierrehumbert, 1980; Beckman & Pierrehambert, 1986; Ladd, 1996; Pierrehambert, 2000), pitch accents (i.e., F0 changes associated with focus) align with the stressed syllable in a target word. Here, unstressed syllables that occur before the stressed syllable are labeled “pre-stressed” syllables, and those that occur after the target stressed syllable are labeled “post-stressed” syllables. In order to examine lexical stress production as a function of, say, initial focus location, prosodic change for the stressed syllable was captured as the mean difference value of the stressed syllable in “Leigh, Nina, Melanie, Marie, Ramona;” pre-stressed syllable change by the mean difference value of the unstressed syllable in “Marie” and the first unstressed syllable in “Ramona;” and post-stressed syllable change by the mean difference value of the unstressed syllable in “Nina”, the second and third syllables of “Melanie” and the third syllable of “Ramona”. The motivation of adding one more measurement, F0 range over syllables, is that Mandarin speakers may realize English sentence-level word stress in a Mandarin-like tonal way, which means they may vary more on F0 range than the mean F0 to code focus over word stress.

## 5.3. Results

### *5.3.1. The overall F0 contours*

Similar to Experiments 1 and 2, time-normalized F0 contours of the stimulus sentences were first plotted by focus location and subject group. In Figures 5.1-5.3, each curve represents an average of the ten speakers’ production. Solid curves represent the no focus condition and dashed curves represent the focus condition. Syllable boundaries are marked with vertical dashed lines. Due to the different number of syllables in the in-focus words and same number of syllables in the pre-focus and post-focus words, the F0 contours for initial focus and medial focus were right aligned and left aligned for final focus.

Figure 5.1 indicates that native American-English (henceforth AE) speakers produced in-focus expansion of F0 range on stressed syllables in sentence-initial position and distinct post-focus compression of F0. If there was an unstressed syllable in the in-focus word linked to the post-focus constituents, PFC started on that syllable; for example, the ‘na’ or “NIna”, the ‘la’ of “MElanie”, and the ‘na’ of “RaMOna”. Compared to native American-English speakers, the Senior Chinese (henceforth SC) learners of English produced less in-focus expansion and post-focus compression, but like American-English speakers they started PFC from the last unstressed syllable in the in-focus words. The Freshman Chinese (henceforth FC) learners of English produced in-focus expansion but little post-focus compression, and PFC did not start until the first syllable of the post-focus constituents.

Figure 5.2 indicates that AE speakers produced in-focus expansion of F0 range on stressed syllables in sentence-medial position, clear post-focus compression, and some pre-focus compression. The SC learners again produced less in-focus expansion and post-focus compression than native American-English speakers, and almost no pre-focus F0 change. The FC learners produced clear in-focus expansion but little post-focus compression and pre-focus compression.

Figure 5.3 indicates that none of the groups produced F0 change in pre-focus position when contrastive focus was on the final word in the phrase; however, they all produced in-focus expansion of F0 on the in-focus words. The FC group appeared to have more in-focus expansion than the SC group, which can be also seen in Figures 5.1 and 5.2.

*5.3.2. The change of mean F0, intensity and duration by sentence location relative to focus*

The pre-focus and post-focus constituents were the same items in the same sentence locations relative to the focus item for each of the different in-focus words. The in-focus words varied in length and lexical stress pattern, the in-focus variable was calculated on the stressed syllable to examine the overall prosodic change from no-focus to focus conditions. To statistically verify the F0 change over three focus locations across

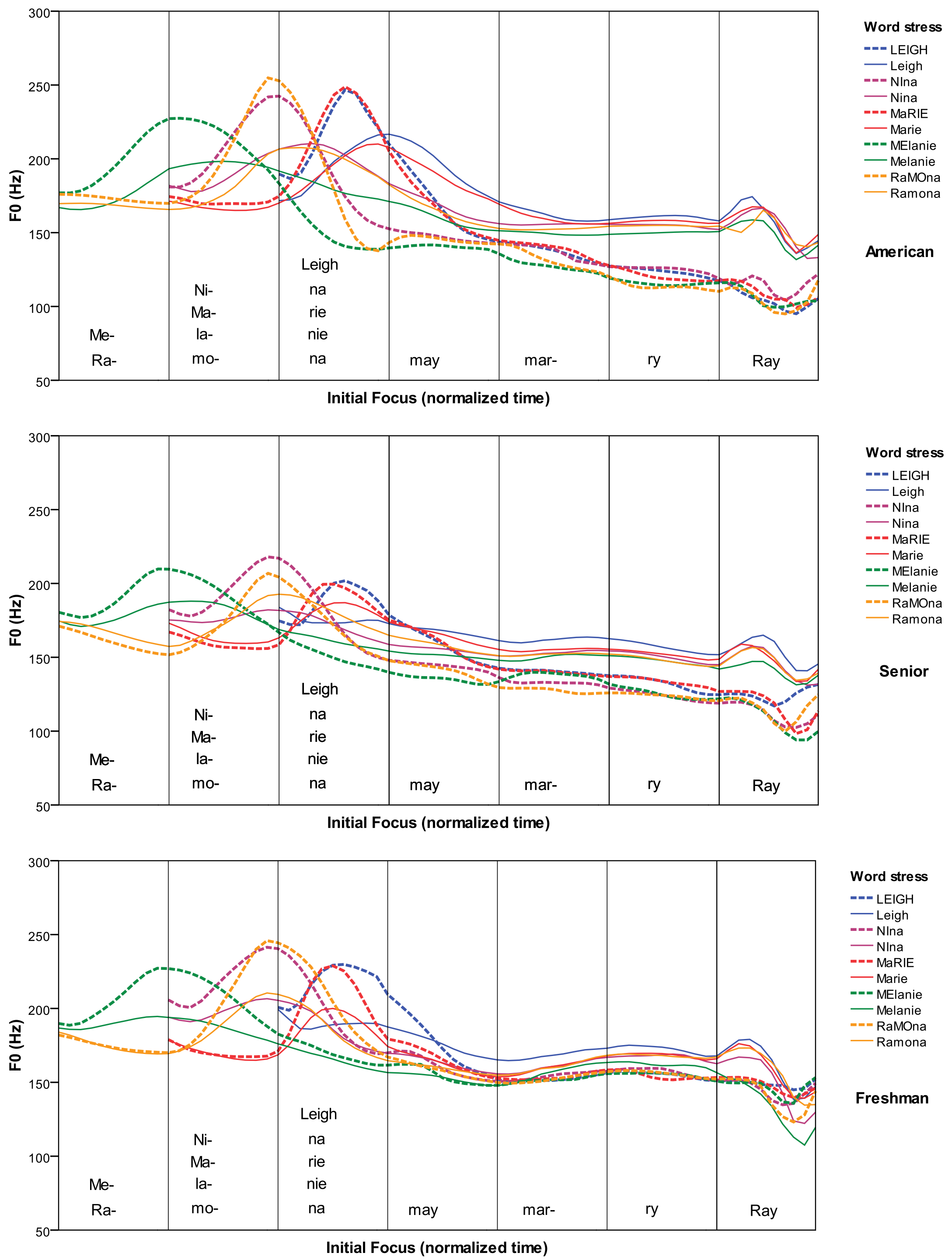


**Figure 5.1.** Time-normalized F0 contours (Hz) with initial focus by word stress and speaker group in English.

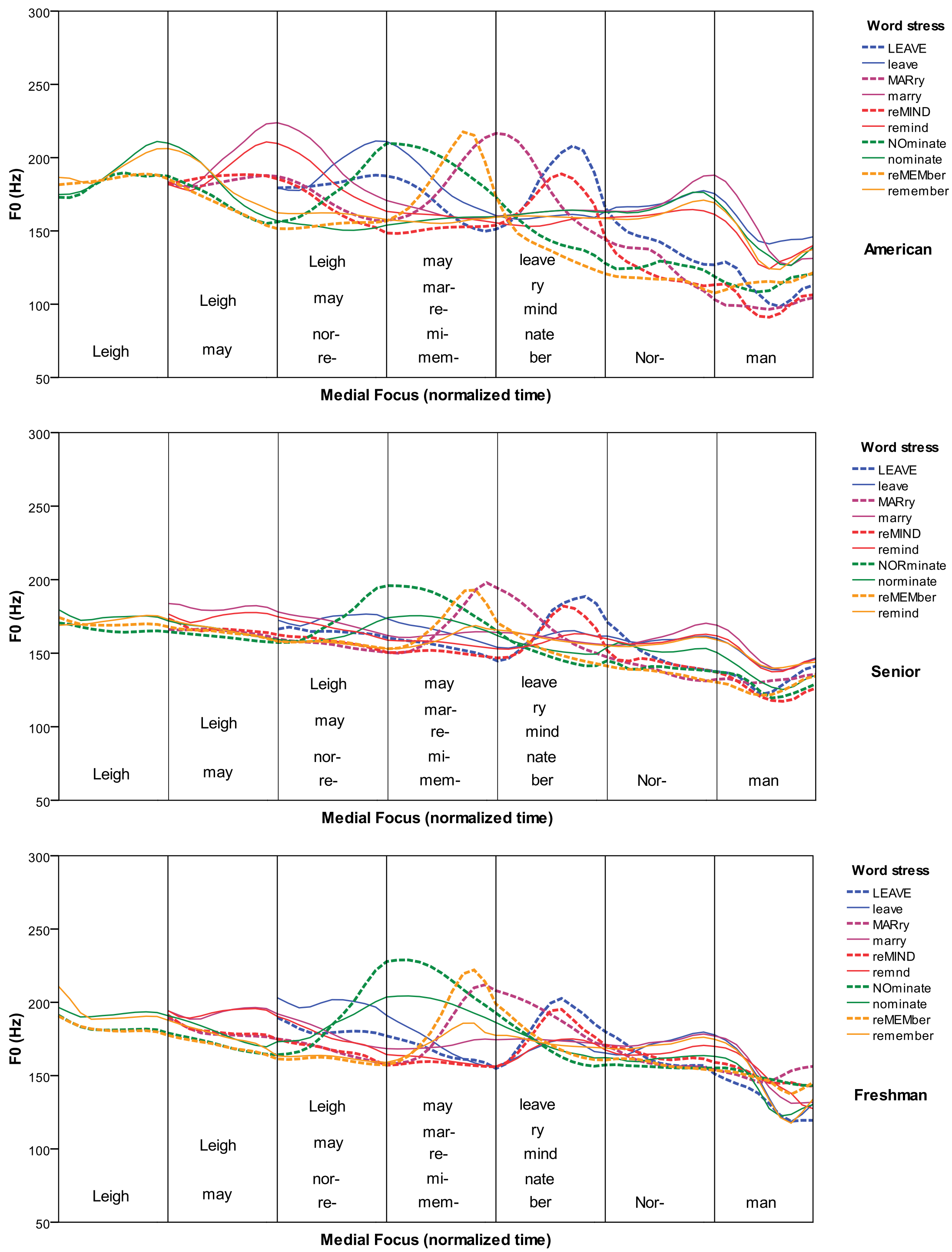


**Figure 5.2.** Time-normalized F0 contours (Hz) with medial focus by word stress and speaker group in English.

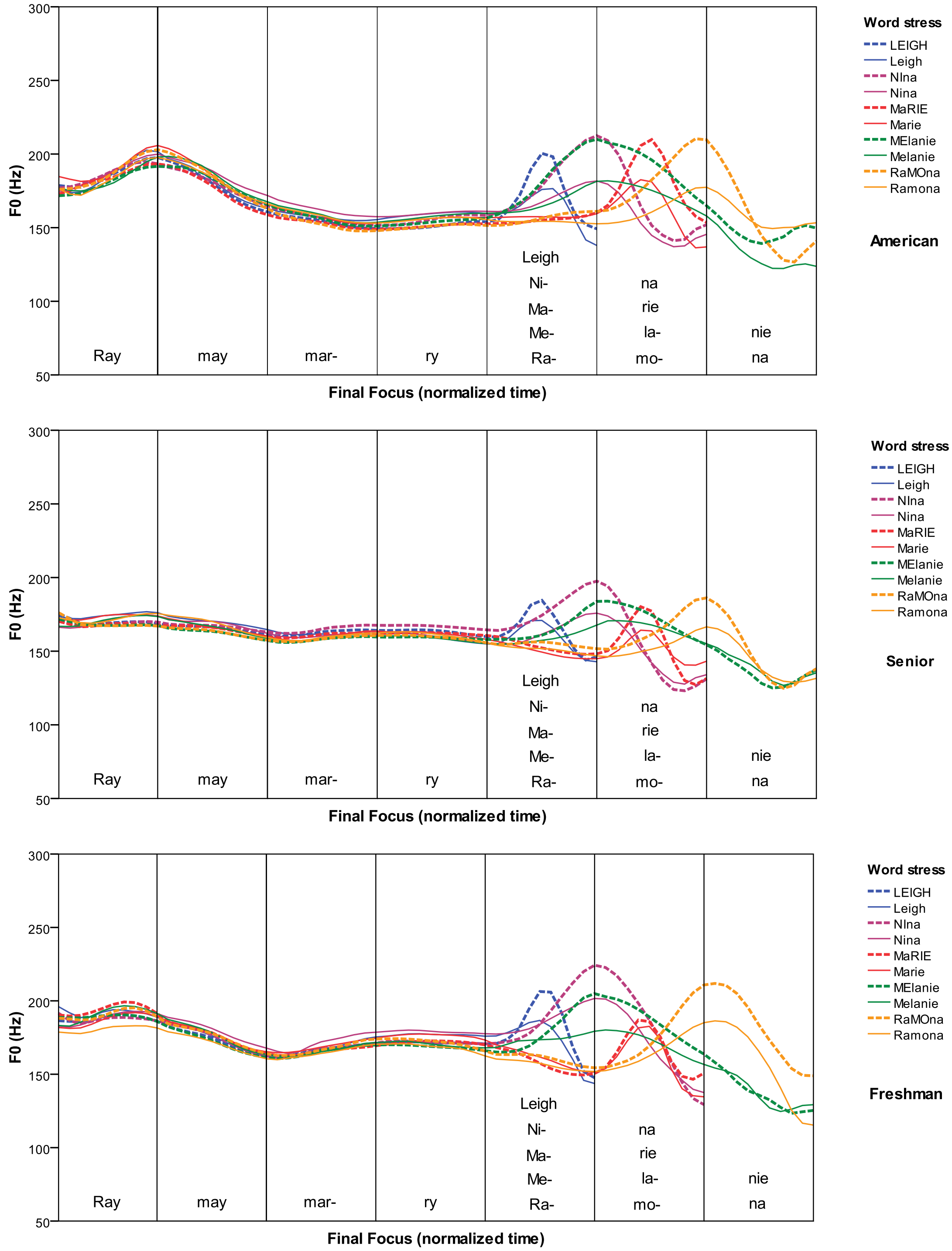


**Figure 5.3.** Time-normalized F0 contours (Hz) with final focus by word stress and speaker group in English.

the three groups as seen in Figures 5.1-5.3, mean F0 differences were converted from Hz to semitone and compared in a repeated measures ANOVA with two factors— sentence location relative to focus item (pre-focus, in-focus, post-focus) as a within-subjects factor and group (native AE speakers, SC learners of English, FC learners of English) as a between-subjects factor. The results showed a two-way interaction between location and group ($F(4,54) = 9.726$, $p < 0.001$). The main effects of location ($F(2,54) = 79.378$, $p < 0.001$) and group ($F(2,27) = 5.92$, $p = 0.007$) were both highly significant. Figure 5.4 displays the means and standard errors of F0 change according to sentence locations and speaker groups.

The magnitude of F0 change from in-focus to pre-focus locations and from in-focus to post-focus locations was examined in another ANOVA. The results showed a two-way interaction between location and group ($F(2,27) = 7.309$, $p = 0.003$). The main effects of location ($F(1,27) = 26.001$, $p < 0.001$) and group ($F(2,27) = 13.361$, $p < 0.001$) were also both significant.

Post-hoc independent samples t-tests showed significant differences of mean F0 for in-focus change between the AE and the SC groups ($t(18) = 2.378$, $p = 0.036$) and a nearly significant difference between the AE and the FC groups ($t(18) = 1.954$, $p = 0.066$). Significant differences for post-focus change were also found between the AE and SC groups ($t(18) = -2.324$, $p = 0.032$), between the AE and FC groups ($t(18) = -4.401$, $p < 0.001$), and between the SC and FC groups ($t(18) = -3.526$, $p = 0.003$). The magnitude of mean F0 change from in-focus to pre-focus locations was significantly different between the AE and SC groups ($t(18) = 4.069$, $p = 0.001$) and between the AE and FC groups ($t(18) = 3.038$, $p = 0.007$). The magnitude of F0 change from in-focus to

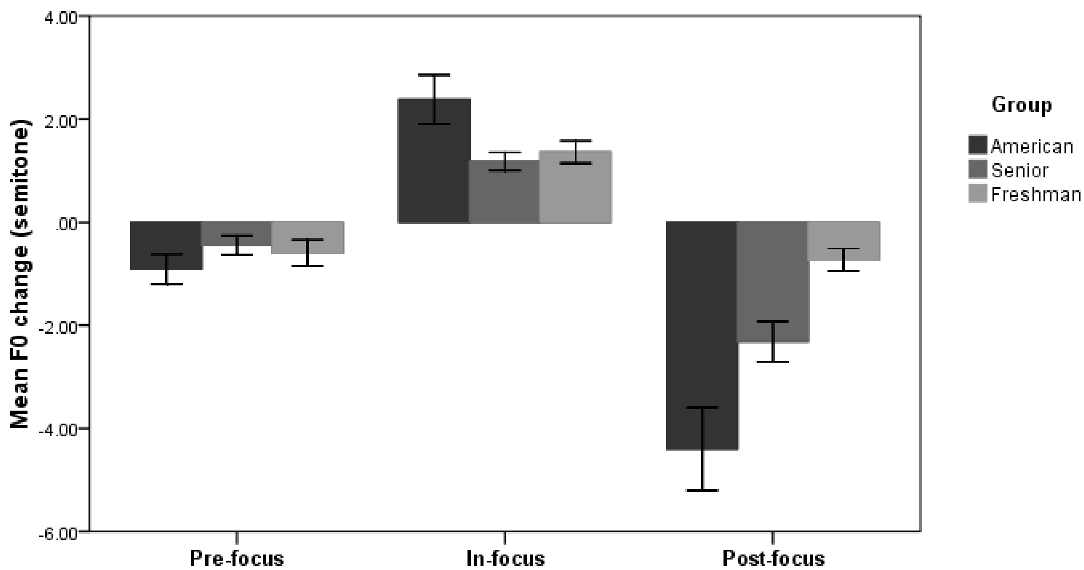


**Figure 5.4.** Mean F0 change (semitone) by sentence location relative to focus item and speaker group in English.

post-focus locations was also significantly different between the AE and SC groups ($t(18) = 2.78$, $p = 0.012$), between the AE and FC groups ($t(18) = 4.017$, $p = 0.001$), and between the SC and the FC groups ($t(18) = 2.733$, $p = 0.014$).

A repeated measures ANOVA on mean intensity change showed a two-way interaction between sentence location relative to focus item and speaker group ($F(4,54) = 10.687$, $p < 0.001$). The main effects of location ($F(2,54) = 187.211$, $p < 0.001$) and group ($F(2,27) = 21.668$, $p < 0.001$) were also both highly significant. Figure 5.5 displays the means and standard errors of intensity change according to sentence locations and speaker groups.

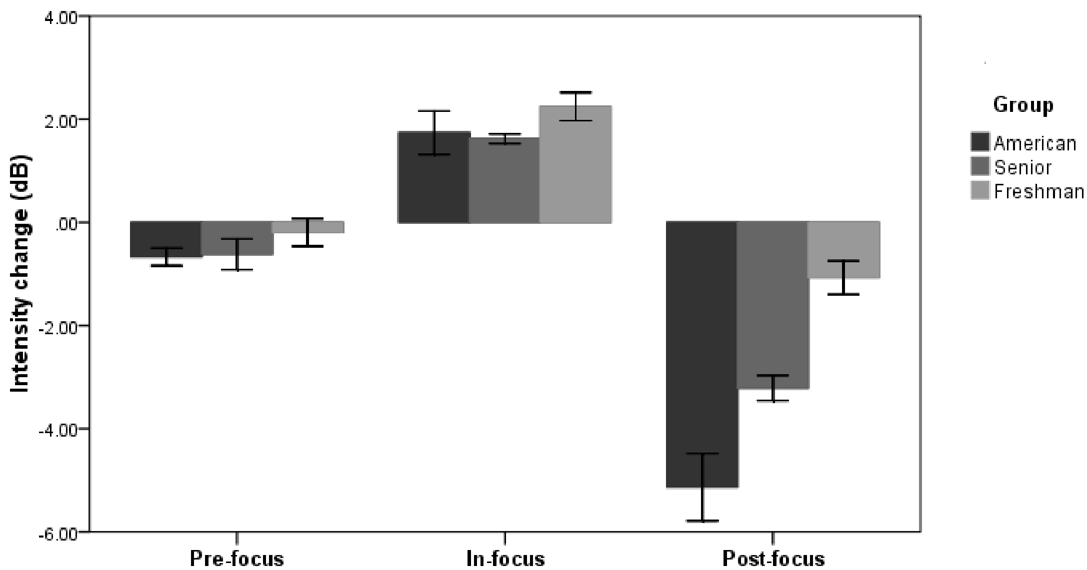


**Figure 5.5.** Mean intensity change (dB) by sentence location relative to focus item and speaker group in English.

The magnitude of intensity change from in-focus to pre-focus locations and from in-focus to post-focus locations was also examined. The ANOVA results showed a two-way interaction between location and group (*F(*2,27) = 18.384, *p* < 0.001). The main effects of location (*F(*1,27) = 120.068, *p* < 0.001) and group (*F(*2,27) = 4.651, *p* = 0.018) were also both significant.

Post-hoc independent samples t-tests showed a significant difference of mean intensity in the post-focus change between the AE and SC groups (*t*(18) = -2.378, *p* = 0.018), between the AE and FC groups (*t*(18) = -5.576, *p* < 0.001), and between the SC and FC groups (*t*(18) = -5.279, *p* < 0.001). The magnitude of intensity change from in-focus to post-focus locations was also significantly different between the AE and SC groups (*t*(18) = 2.483, *p* = 0.031), between the AE and FC groups (*t*(18) = 3.949, *p* = 0.001), and between the SC and FC groups (*t*(18) = 2.981, *p* = 0.008).

A final repeated measures ANOVA on duration change showed no interaction between sentence location relative to focus item and speaker group and also no main effect of group. The main effect of sentence location was highly significant ($F(2,54) = 124.735$, $p < 0.001$). Figure 5.6 displays the means and standard errors of duration change according to sentence location and subject groups.

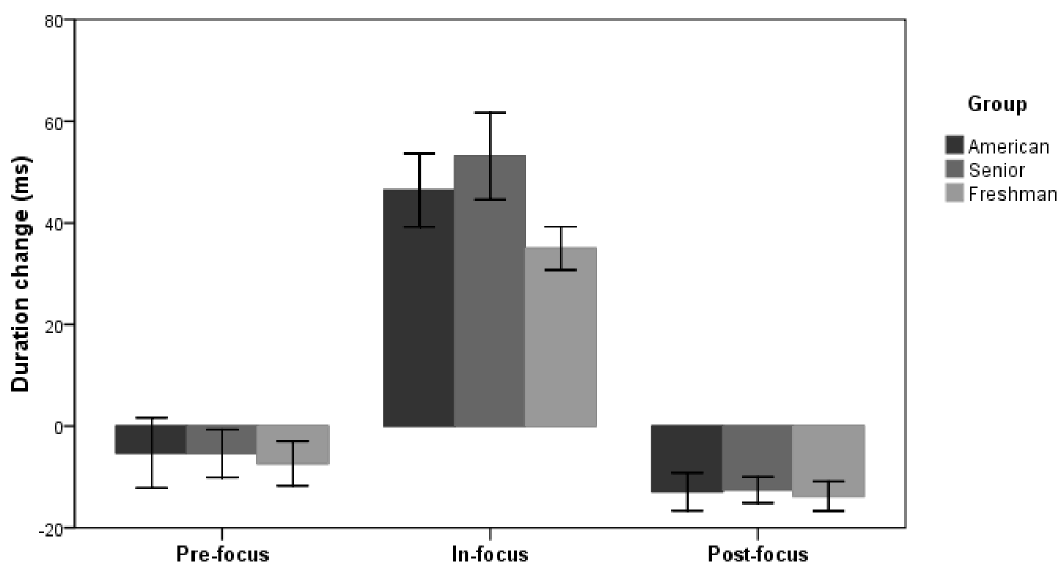


**Figure 5.6.** Mean duration change (dB) by sentence location relative to focus item and speaker group in English.

The magnitude of duration change from in-focus to pre-focus locations and from post-focus to in-focus locations showed a nearly significant interaction between location and group ($F(2,27) = 3.202$, $p = 0.057$). There was a main effect of location ($F(1,27) = 123.455$, $p < 0.001$) but no main effect of group. No group difference in duration change was found for any sentence location relative to focus item or for the magnitude of duration changes between any two locations.

*5.3.3. In-focus change of F0, intensity and duration in stressed and unstressed syllables*

To investigate the production of lexical stress in words targeted to be produced with prosodic focus, prosodic change was assessed in repeated measures ANOVAs with stress (pre-stressed, stressed, and post-stressed) and focus location (initial, medial, final) as a within-subjects factors and speaker group (American, Senior, Freshman) as a between-subjects factor.

Prosodic changes in mean F0 were examined first. There was no three-way interaction among stress, focus location and speaker group, but there was a two-way interaction between stress and location (F(4, 108) = 2.485, $p$ = 0.048). The main effects of stress (F(2, 54) = 18.808, $p$ < 0.001), location (F(2, 54) = 13.45, $p$ < 0.001) and group (F(2, 27) = 4.219, $p$ = 0.025) were also all significant. Figure 5.7 shows the prosodically-driven F0 change of stressed and unstressed syllables from no focus production to in-focus productions at three focus locations in the sentences.

Post-hoc independent samples t-tests showed a significant difference between the AE and SC groups ($t$(18) = -2.197, $p$ = 0.041) for stressed syllables when contrastive focus was in sentence final. The differences between the AE and FC groups were significant for stressed syllables in sentence-initial words under prosodic focus ($t$(18) = 2.47, $p$ = 0.024) and in sentence-medial words under prosodic focus ($t$(18) = 2.398, $p$ = 0.028), and for post-stressed syllables in sentence-initial words ($t$(18) = 3.15, $p$ = 0.006) and sentence-final words ($t$(18) = 2.578, $p$ = 0.019). The difference between AE and FC groups was also nearly significant in post-stress unstressed syllables in sentence-medial words under prosodic focus ($t$(18) = 1.971, $p$ = 0.064).

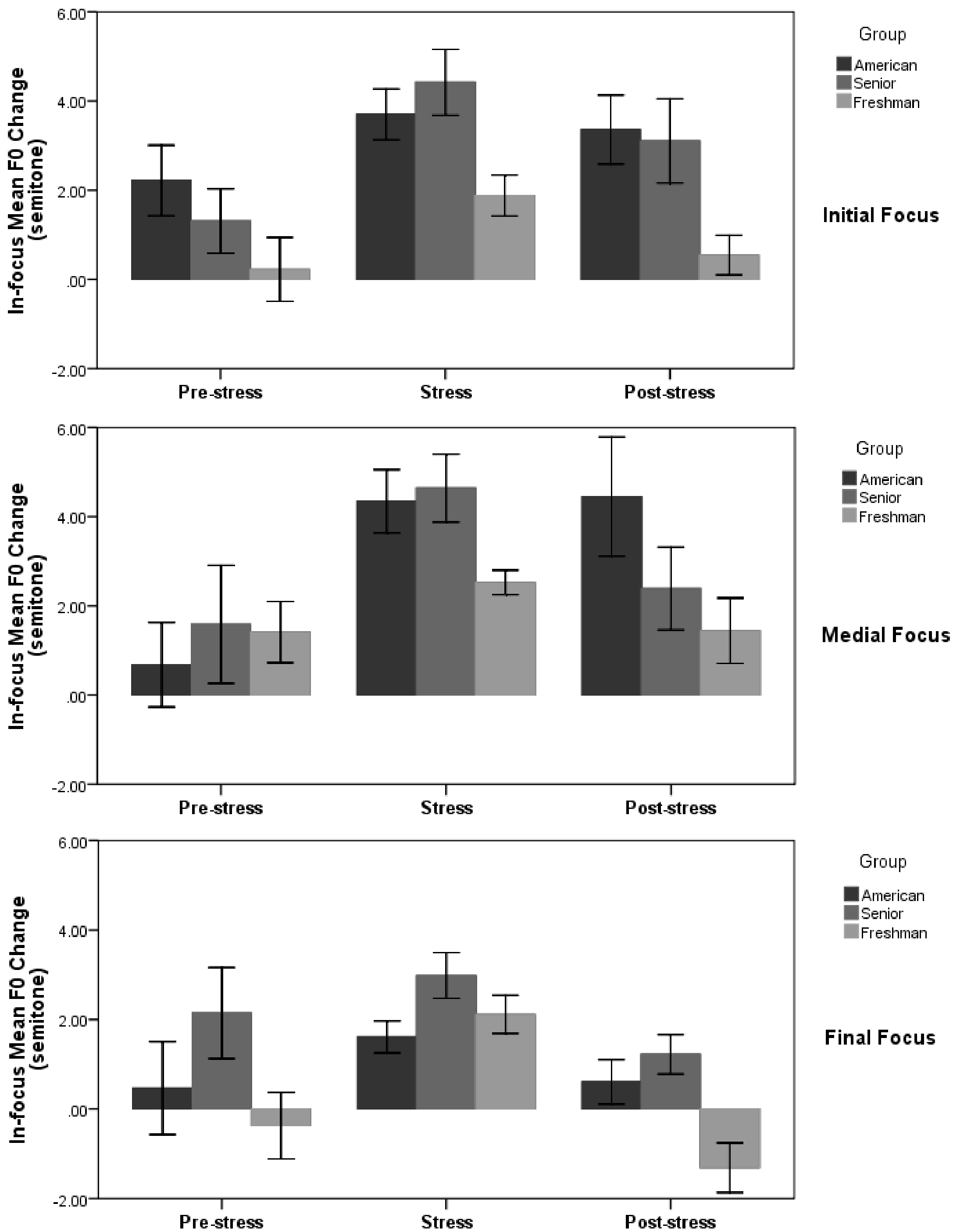


**Figure 5.7.** In-focus change of mean F0 in pre-stressed, stressed and post-stressed syllables by focus location and speaker group in English.

The analysis of prosodic change in for F0 range showed a three-way interaction among stress type, focus location and speaker groups (F(8, 108) = 2.796, $p = 0.007$) and a two-way interaction between stress and location (F(4, 108) = 6.237, $p < 0.001$). The main effects of stress (F(2, 54) = 53.073, $p < 0.001$), location (F(2, 54) = 11.59, $p < 0.001$) and group (F(2, 27) = 5.885, $p = 0.008$) were also all significant. Figure 5.8 shows the change in F0 range for stressed and unstressed syllables for target words in the three sentence positions.

Post-hoc independent samples t-tests showed that differences in the change in F0 range was significant for the AE and FC groups ($t(18) = 2.534$, $p = 0.021$) and nearly significant for the AE and SC groups ($t(18) = 2.025$, $p = 0.058$) in pre-stress unstressed syllables in sentence-final target words. Change in post-stress unstressed syllables in sentence-initial target words sentence initial position was significantly different for the AE and SC groups ($t(18) = 2.741$, $p = 0.018$) and the AE and FC groups ($t(18) = 3.244$, $p = 0.006$). F0 range changes in post-stress unstressed syllables in sentence-medial target words was also significantly different between the AE and SC groups ($t(18) = 2.981$, $p = 0.012$) and the AE and FC groups ($t(18) = 2.599$, $p = 0.018$).

Change in intensity also varied systematically with stress and focus location (F(4, 108) = 8.341, $p < 0.001$). The main effects of stress (F(2, 54) = 22.126, $p < 0.001$) and group (F(2, 27) = 4.24, $p = 0.025$) were also significant. Figure 5.9 shows the intensity change of stressed and unstressed syllables in target words across the three sentence positions.

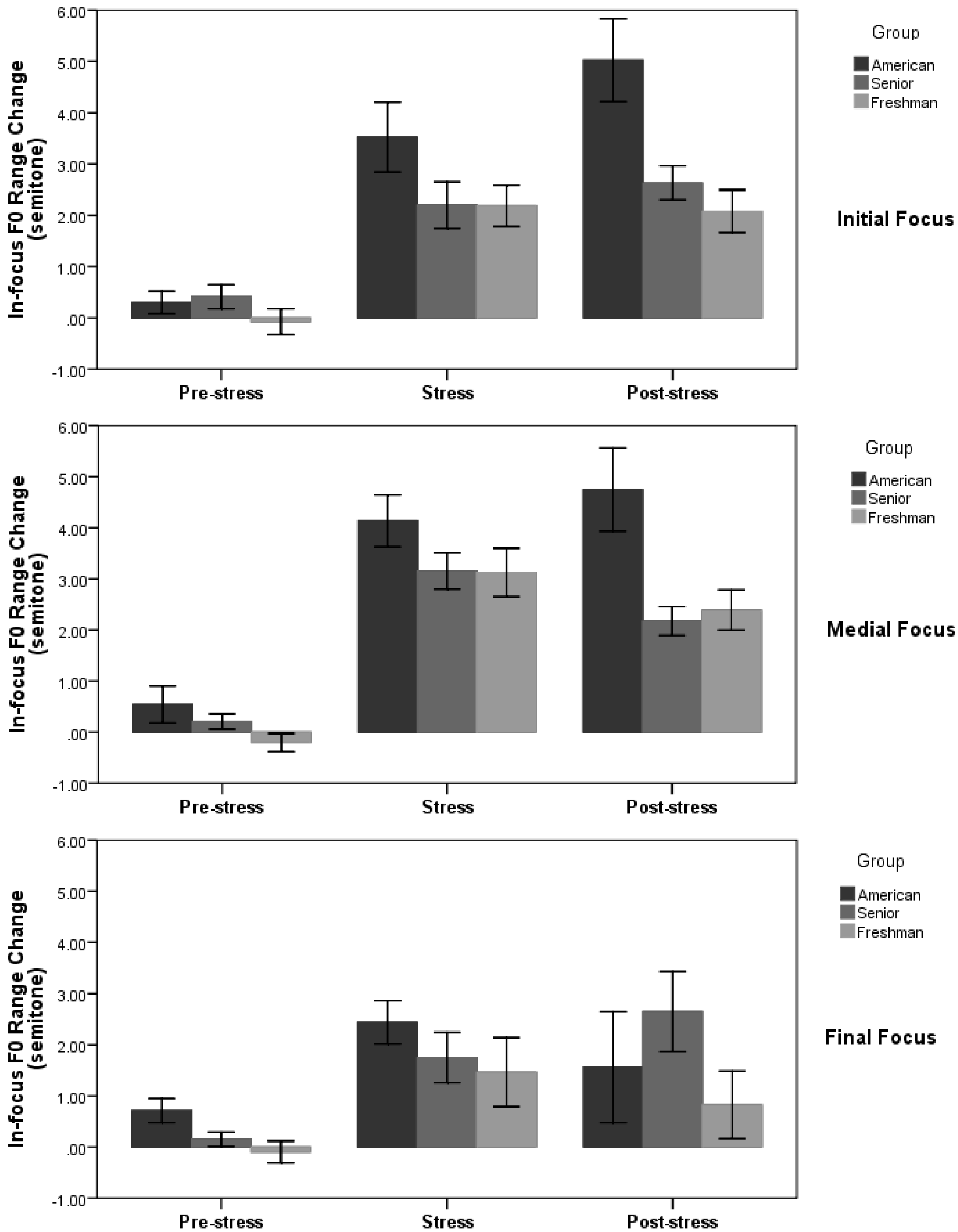


**Figure 5.8.** In-focus change of F0 range in pre-stressed, stressed and post-stressed syllables by focus location and speaker group in English.

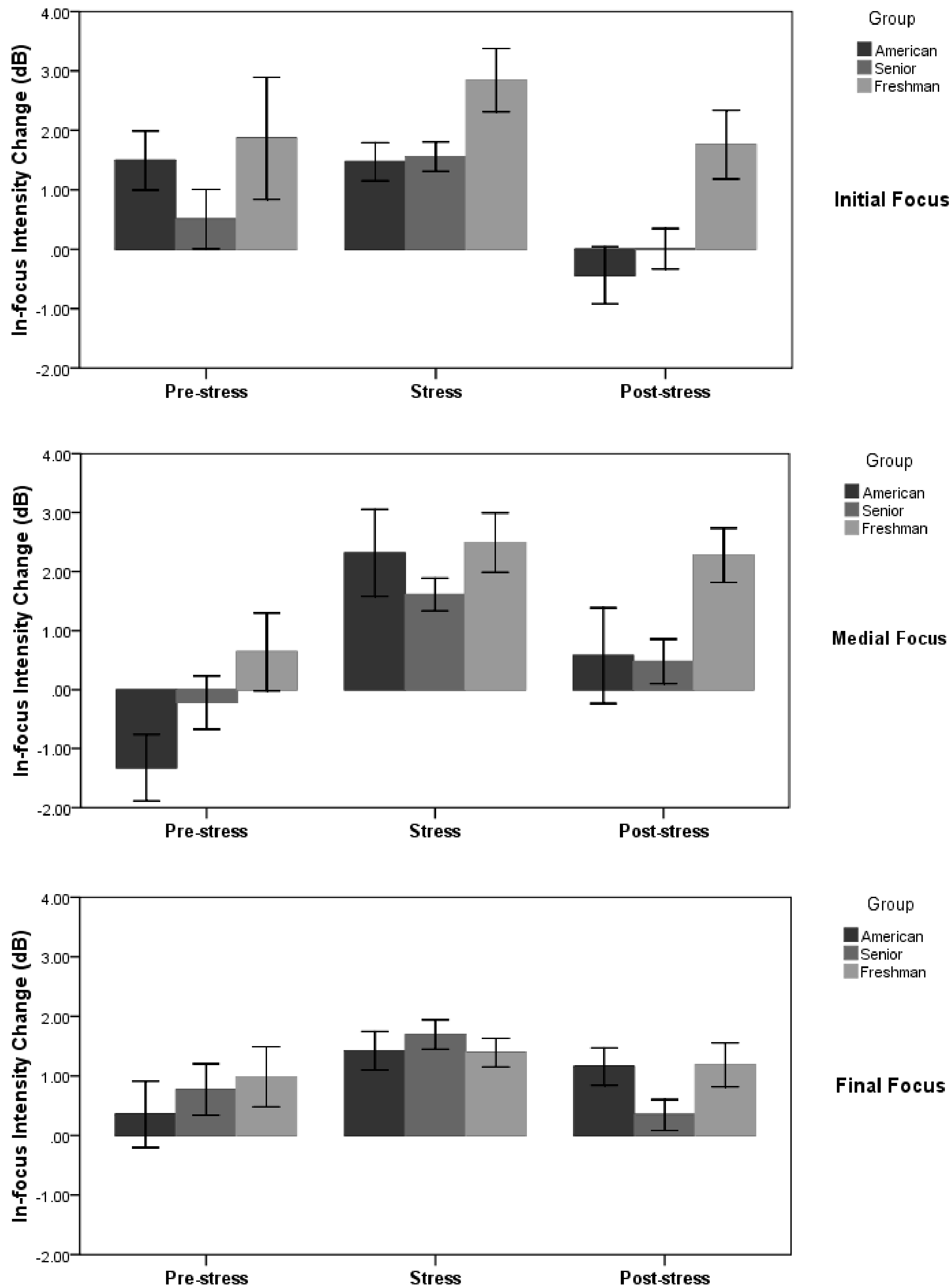


**Figure 5.9.** In-focus change of intensity in pre-stressed, stressed and post-stressed syllables by focus location and speaker group in English.

Post-hoc independent samples t-tests showed significant differences between the AE and FC groups for intensity change for stressed syllables in sentence-initial target words ($t(18) = -2.216$, $p = 0.04$) as well as for pre-stress unstressed syllables in sentence-medial target words ($t(18) = -2.265$, $p = 0.036$). The difference of in-focus intensity change on post-stress syllables was significant at sentence initial position between the AE and the FC groups ($t(18) = -2.911$, $p = 0.009$) and near significant between the AE and SC groups ($t(18) = 2.004$, $p = 0.06$).

A final set of analyses investigated prosodically-driven changes in duration as a function of stress and target word position within the sentence. Once again, there was a significant two-way interaction between stress and focus location ($F(4, 108) = 4.577$, $p = 0.002$) and a nearly significant interaction between focus location and group ($F(4, 108) = 2.372$, $p = 0.064$). The main effect of location was significant ($F(2, 54) = 10.308$, $p < 0.001$) and that of group was nearly significant ($F(2, 27) = 2.987$, $p = 0.067$). Figure 5.10 shows the duration change in stressed and unstressed syllables for target words in the three sentence positions.

Post-hoc independent samples t-tests showed significant differences in duration change between the AE and the FC groups for stressed syllables in sentence-initial target words ($t(18) = 2.113$, $p = 0.049$), and between the AE and SC groups for stressed syllables in sentence-final target words ($t(18) = -2.891$, $p = 0.01$). Significant differences between the AE and FC groups were found for duration changes in post-stress unstressed syllables at sentence-initial target words ($t(18) = 2.551$, $p = 0.02$) and in sentence-final target words ($t(18) = 2.197$, $p = 0.41$).

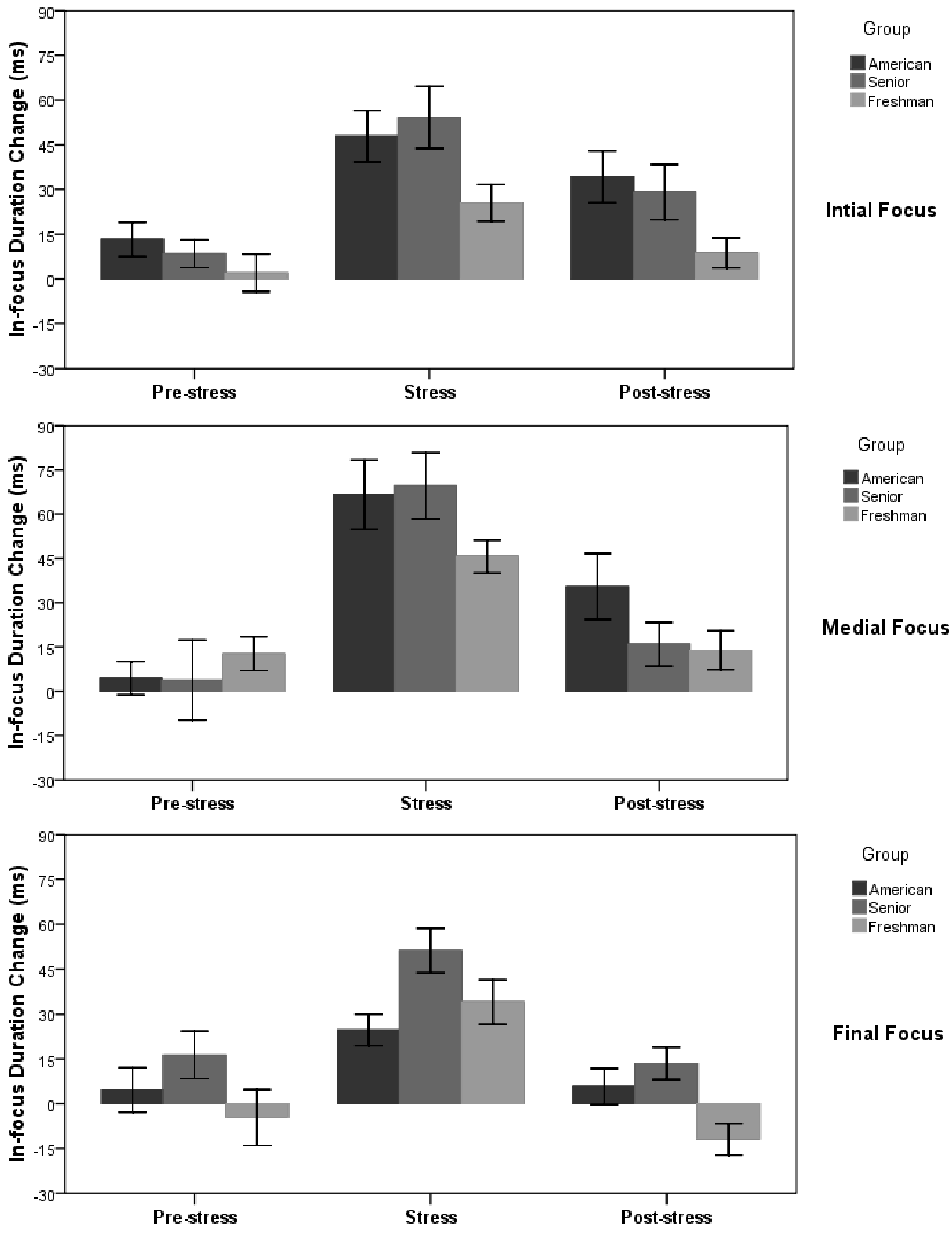


**Figure 5.10.** In-focus change of duration in pre-stressed, stressed and post-stressed syllables by focus location and speaker group in English.

*5.3.4. Correlations in mean F0 and intensity change between the in-focus syllable and other constituents in the sentence*

To better understand the relationship between changes in the focused stressed syllable and other changes in the sentence (pre-focus, pre-stress, post-stress, post-focus). Pearson correlation coefficients were conducted. Tables 5.3 and 5.4 show the results for F0 and intensity changes, respectively.

**Table 5.3.** Correlations of mean F0 change of in-focus syllables relative to other constituents by focus location and speaker group in Experiment 3 (*n*=10, $p < 0.1$).

| | Initial-focus vs. | | | Medial-focus vs. | | | | Final-focus vs. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre-stress | Post-stress | Post-focus | Pre-focus | Pre-stress | Post-stress | Post-focus | Pre-focus | Pre-stress | Post-stress |
| AE | -- | *r*=.628 | -- | -- | *r*=.817 | *r*=.696 | -- | -- | *r*=.782 | -- |
| | -- | *p*=.052 | -- | -- | *p*=.004 | *p*=.025 | -- | -- | *p*=.007 | -- |
| SC | -- | *r*=.588 | -- | -- | *r*=.576 | *r*=.72 | *r*=.681 | *r*=.834 | *r*=.927 | *r*=.699 |
| | -- | *p*=.074 | -- | -- | *p*=.081 | *p*=.019 | *p*=.03 | *p*=.003 | *p*<.001 | *p*=.025 |
| FC | *r*=.642 | -- | -- | -- | *r*=.633 | -- | -- | -- | -- | *r*=.868 |
| | *p*=.046 | -- | -- | -- | *p*=.049 | -- | -- | -- | -- | *p*=.001 |

Table 5.3 indicates that the F0 changes on the pre-stress and post-stress syllables are closely correlated to in-focus F0 change on the stressed syllables in the production of native AE speakers. The CH learners also produced correlated changes, but unlike the AE speakers, the correlations went beyond the focused words. Table 5.4 indicates that the in-focus change of intensity was not as closely correlated to the intensity change on the other constituents as the F0 change showed.

**Table 5.4.** Correlations of mean intensity change of in-focus syllables relative to other constituents by focus location and speaker group in Experiment 3 ($n$=10, $p$ < 0.1).

| | Initial-focus vs. | | | Medial-focus vs. | | | | Final-focus vs. | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | Pre-stress | Post-stress | Post-focus | Pre-focus | Pre-stress | Post-stress | Post-focus | Pre-focus | Pre-stress | Post-stress |
| AE | --<br>-- | --<br>-- | --<br>-- | --<br>-- | --<br>-- | *r*=.697<br>*p*=.025 | --<br>-- | --<br>-- | --<br>-- | --<br>-- |
| SC | --<br>-- | *r*=.787<br>*p*=.007 | --<br>-- | --<br>-- | --<br>-- | --<br>-- | --<br>-- | --<br>-- | --<br>-- | --<br>-- |
| FC | *r*=.669<br>*p*=.034 | *r*=.563<br>*p*=.09 | --<br>-- | --<br>-- | --<br>-- | *r*=.712<br>*p*=.021 | --<br>-- | --<br>-- | *r*=.601<br>*p*=.066 | *r*=.718<br>*p*=.019 |

## 5.4. Discussion

Experiment 3 reveals that that PFC in English is not easy for Chinese learners to acquire even though there is PFC in Beijing Mandarin. This result confirms once again that PFC does not easily transfer from one language to another. The results also show that LOR affects the phonetic realization of focus in L2 English by experienced Beijing Mandarin speaking learners. The college seniors, who had been residing in the US for several years, were more likely to realize native-like PFC than the college freshmen, suggesting that the full phonetic realization of prosodic focus can be acquired given sufficient L2 experience.

The results indicate that both learner groups increased in-focus F0, but not as much as the native English speakers. The FC group produced little PFC of mean F0 whereas the SC group produced significantly more PFC. The results of PFC of intensity were similar; however, both learner groups produced native-like in-focus expansion of intensity. Like the AE speakers, the SC group produced more post-focus change of F0 and intensity than in-focus change; in contrast, the FC produced less post-focus change of F0 and intensity than in-focus change (See Figures 5.4 and 5.5). The results in Section 5.3.4 indicate that F0 change on the post-focus constituents was not correlated to the in-focus change on the focused stressed syllables; however, as shown in Figures 5.1 and 5.2, native AE speakers and the SC group started PFC on the post-stress unstressed syllables, whose F0 change from non-focus to focused conditions was correlated to the in-focus change on the focused stressed syllables. These results revealed that PFC in English is not independent of in-focus expansion at the word level for native speakers. All three groups produced little in the way of a pre-focus decrease in mean F0 (See Figure 5.4 and compare Figure 5.3 with Figure 5.2) and also little pre-focus intensity change. As for duration, both learner groups produced native-like in-focus, pre-focus and post-focus change.

Together these results address the 1st and 2nd research questions regarding in-focus prosodic change and PFC. In particular, the results provide partial support the expectation that Mandarin learners would not have difficulties in producing both in-focus expansion and post-focus compression in English since these patterns also exist in Mandarin. The results also address the question about the effect of LOR on focus production. Although the SC group did not produce native-like PFC of mean F0 and intensity, their patterns were significantly more native-like than those produced by the FC group. In other words, long LOR learners show an intermediate pattern of PFC between the native English speakers and the short LOR learners.

The analyses of in-focus change of mean F0 on stressed syllables showed that SC learners had particular difficulty producing a native-like pattern in sentence-final target words, whereas FC learners had difficulty with the pattern in sentence-initial and –medial target words. However, neither learner group had a difficulty producing a native-like change in F0 range on stressed syllables, suggesting that Mandarin learners of English are especially apt to realize prosodic focus using an expanded F0 range. The FC learners also had difficulties producing native-like F0 mean patterns in post-stress unstressed syllables in target words across all three sentence positions. The two learner groups had similar difficulties realizing native-like patterns of change in F0 range for pre-stress unstressed syllables in sentence-final target words and for post-stress unstressed syllables in sentence-initial and -medial target words. Together, these results indicate that Mandarin learners of English have more difficulties producing native-like F0 patterns in unstressed syllables than in stressed syllables, as expected.

Learners' difficulties producing native-like changes in F0 in post-stress unstressed syllables may be related to the realization of PFC at the global sentential level. Xu & Xu (2005) found that native American English speakers reduce F0 immediately after the stressed syllable peak, resulting in large F0 range, basically measured from the in-focus F0 peak to an F0 valley located at the offset of the post-stress unstressed syllable. The findings in this experiment were similar, as can be seen in the F0 contours for AE productions in Figures 5.1 and 5.2. As we pointed out in Section 5.3.1, the SC learners started PFC on the post-stress unstressed syllables in a native-like manner while the FC learners did not, but because the magnitude of in-focus expansion and PFC of F0 was not as large in the SC learners' production as in the AE speakers' production, the SC learners' F0 range change on post-stress unstressed syllables in sentence-initial and -medial target words still deviated from those of the AE speakers' productions. The FC learners produced little PFC; therefore, their F0 production on post-stress unstressed syllables in sentence-initial and -medial target words deviated from the AE speakers not only in F0 range but also in mean F0.

Like the results from Experiment 2, the results from Experiment 3 suggest that F0 and intensity are not independent of one another. The difficulties that FC learners had in producing native-like PFC of F0 were also observed in the highly non-native-like intensity changes in post-stress unstressed syllables in sentence-initial target words. Of course, FC learners' difficulties extended beyond post-stress unstressed syllables and beyond changes in F0 and intensity: FC learners had difficulties producing native-like changes of intensity and duration for stressed syllables in sentence-initial target words, and their production of in-focus duration change on post-stress syllables in these words was also non-native-like.

In sum, Mandarin learners of English had more difficulties producing native-like prosodic changes in unstressed syllables than in stressed syllables, and they had particular difficulty with changes in post-stress unstressed syllables in sentence-initial and -medial target words, which suggests an interaction with the difficulty that learners had realizing native-like PFC. Learner difficulties were more evident for F0 than for intensity and duration. FC learners had more difficulty than SC learners, suggesting an effect of LOR on the production of word stress as a function of focus.

According to the Target Approximation (TA) model, unstressed syllables in English are represented by static pitch targets while pitch accents associated with stressed syllable are sometimes represented by dynamic pitch targets (Xu, 2005; Xu & Xu, 2005). The results for F0 change as a function of lexical stress in English suggest that Mandarin speakers had more difficulties achieving a static pitch target than a dynamic one. This result could be due to the tonal system in L1 Mandarin. Mandarin tones are contour tones. They contrast not only along the high-low pitch dimension, but also in F0 contour shape. F0 contour varies more dramatically in Mandarin than in English at sentential level due to tonal coarticulation. Therefore, it is perhaps unsurprising that Chinese learners do better in the prosodic variations on stressed syllables than on unstressed syllables.

To consider all the findings in this experiment under the SLM framework, the similarities in phonetically realizing focus in Beijing Mandarin and English may result in difficulties of discerning and thus acquiring prosodic focus in English by Beijing learners; however, the SC Mandarin learners of English apparently established an L2 category for prosodic focus based on acoustic-phonetic input from English though this category was not native-like. This result is consistent with that in Experiment 2 and does not support our prediction that PFC would transfer from Mandarin to English or vice versa. The fact that the L1 category is not simply expanded to accommodate the realization of the L2 category, challenges the SLM postulate of acoustic-phonetic transference across languages.

One explanation for the surprisingly poor acquisition of prosodic focus across the two languages was offered in Chapter IV and is reiterated here. Learners are more focused on the acquisition of local patterns, and this disrupts their ability to acquire global ones. Here, it may be that Mandarin learners of English focused unduly on the acquisition of L2 lexical stress, which blocked the successful acquisition of prosodic focus, including the important feature of PFC. Note, though, that both learner groups produced more native-like in-focus change of F0 in stressed syllables than in unstressed syllables. This result is more consistent with SLM because the pitch accented syllables of English are more akin to all syllables in Mandarin with specific—and often dynamic—tonal targets (Xu, 2005; Xu & Xu, 2005).

# CHAPTER VI

# CONCLUSIONS AND FUTURE DIRECTIONS

The current dissertation investigated the prosodic realization of focus in second language speech as a function of first language sound patterns and second language experience. With regard to sound patterns, the study investigated: (1) L1 Southern Min, a tone language that realizes prosodic focus without PFC, and L2 Beijing Mandarin, a tone language that realizes prosodic focus with PFC; (2) L1 American English, a non-tone language that realizes prosodic focus with PFC and L2 Beijing Mandarin; (3) L1 Beijing Mandarin and L2 American English. With regard to L2 experience, the study investigated: (1) the amount of L2 use; (2) AOL; and (3) LOR. This chapter will summarize the main findings from the three experiments and discuss these with reference to the Speech Learning Model (SLM) framework; especially, with reference to the nature of the L1 sound system and L2 category formation. After this, the chapter will discuss directions for potential future research based on the limitations of and illuminations from the current experiments.

## 6.1. The effect of L1 prosody system on L2 prosodic focus

In Experiment 1, old Southern Min speakers increased duration and intensity on the in-focus items, but did not produce post-focus compression of F0 or intensity or even in-focus expansion of F0 in their Mandarin production of prosodic focus, suggesting that F0 change is not obligatory for realizing focus even in a tone language. This shows that the phonetic realization of focus varies across tone languages, an observation that was also made with reference to PFC, which some tone languages have and others do not (Jin 1996; Xu, 1999; Pan, 2007; Wu & Xu, 2010; Wang et al., 2011; Xu, 2011). Variation in the realization of prosodic focus is also observed even within the same language: in one variety of Mandarin, Beijing Mandarin, there is PFC; but in another, Taiwan Mandarin, there is not (Xu et al., 2012). This variation in the realization of prosodic focus suggests that tone and intonation are independent, and so that the acquisition of new tones in an L2 tone language need not interfere with the acquisition of L2 intonation, if L1 is also a tone language.

However, when the L1 is a tone language and L2 is not or if the L1 is not a tone language and the L2 is, then lexical tone can interfere with the acquisition of L2 prosodic focus even when prosody is defined by the same phonetic parameters in the two languages and can be modeled as an independent process. The Target Approximation (TA) model provides an explanation for the interference of lexical suprasegmentals in L1 with the acquisition of phrase-level patterns in L2. In the TA model, Mandarin Tone 1 and Tone 3 and unstressed syllables in English are represented with static pitch targets while Mandarin Tone 2 and Tone 4 and pitch accented stressed syllables in English are represented with dynamic pitch targets (Xu & Wang, 2001; Xu, 2005; Xu & Xu, 2005). In spite of the fact that both languages have both static and dynamic pitch targets, the results from Experiment 2 and 3 revealed that American learners had more difficulties to achieve dynamic pitch targets in L2 Mandarin than static pitch targets, but the reverse was true for Chinese learners of L2 English. The explanation for these findings lies in the other assumption of the TA model. In Mandarin, most syllables are realized as if the targets were dynamic because tone-to-tone coarticulation in Mandarin means that even static tonal targets are realized with rapid changes in F0, whereas in English there is only one nuclear pitch accent per intonation unit, thus F0 changes are much slower.

Although we may explain the absence of in-focus expansion and post-focus compression of F0 in the L2 Mandarin produced by older L1 speakers of Southern Min to the absence in their L1; it is more challenging to explain the difficulty with the acquisition of prosodic focus in L2 Mandarin and English by L1 speakers of either English or Mandarin. In particular, the correlation results suggest that PFC is an independent feature of prosodic focus in both languages. In Experiments 1 and 2 there was some correlation between pre-focus changes and in-focus changes, but none between in-focus changes and post-focus changes. This result is consistent with Cao's (2002) description of Mandarin intonation as the combination of lexical tones and local changes in pitch range. Although the results from Experiment 3 suggested that PFC is correlated with in-focus F0 expansion in English within the target word, in-focus changes were not correlated with pre- and post-focus changes across the entire sentence. Thus, as in Mandarin, PFC at the sentence-level is probably a distinct feature of prosodic focus. Together these findings justify the assumption that prosodic focus in Mandarin and English is implemented very similarly. This similarity, according to the SLM framework, on one hand, may not be easy to detect for L2 learners, and on the other hand, should make for easy transference from one language to the other. However the findings indicate that the transfer did not happen. Non-Chinese-heritage learners in Experiment 2 were not able to produce F0 expansion on the in-focus items even though F0 expansion is a feature of prosodic focus in their L1 English. They were also unable to produce PFC of F0 and intensity in Mandarin even though there is PFC in their L1 English. Similar difficulties were observed in the acquisition of L2 English by L1 Mandarin speakers: Freshman Chinese learners of English were not able to produce PFC in English even though PFC exists in their L1 Mandarin. If we treat prosodic focus as a category, as required in SLM, then these findings indicate that the phonetic aspects of L1 categories do not necessarily form the basis of an L2 phonetic category even when commonalities are assumed. Thus, contra SLM, we must assume that the mechanism and processes used in L1 category formation do not automatically apply to L2 learning and that category assimilation which occurs frequently at the segmental level does not occur in prosody.

The failure to transfer a category from one language to another may occur when categories interact within a language. In Chapter II, we described interactions between tone and intonation in Mandarin and between lexical stress and intonation in English. Note that even though we might describe the phonetic realization of prosodic focus similarly in Mandarin and English, the suprasegmentals in the two languages are different at the lexical level. If learners have not acquired the novel lexical tone or stress categories in the L2, but these interact with a common phrase-level category, then the learners may not be able to identify the commonalities between the languages because they are so focused on what is new. In particular, if a learner pays more attention to those categories or phonetic features that are novel in the L2, they may make more efforts to realize them, blocking the acquisition of the *familiar* L2 features. In this way, we might explain the absence of PFC in L2 Mandarin by NCH learners and in L2 English by FC learners: the NCH learners attached more importance to lexical tone category formation and the FC learners to lexical stress category formation.

It is worth noting that SLM also claims that the blockage of category formation does not prevent phonetic learning from taking place as long as learners remain sensitive to subcategorical cross-language phonetic differences (Flege, 2007). And, indeed, with more extensive experience, both Mandarin and English learners were better able to produce native-like phrase-level patterns in the L2.

### 6.2. The effect of L2 experience on L2 production of prosodic focus

Xu et al. (2012) argued that the loss of PFC in Taiwan Mandarin might be caused by the long-term contact with Southern Min. Wu and Chung (2011) found that none of the simultaneous bilinguals of Cantonese and English produced PFC in Cantonese and two of ten subjects did not produce PFC in English. Experiment 1 in the current dissertation found that none of the age groups produced PFC in Southern Min and the older speaker did not produce PFC in Mandarin. These findings suggest that language contact may result in the loss of PFC. Interestingly, long-term contact and bilingualism does not seem to result in non-PFC varieties gaining PFC even though, with sufficient L2 experience, PFC can be learned and acquired.

Experiment 1 indicated that the greater amount of L2 Mandarin use resulted in more PFC in Mandarin by Southern Min learners. Although the AOL was similar among the three speaker groups, the younger speakers used more Mandarin than the older and mid-age speakers. However, the younger speakers likely had also had more, and higher quality, Beijing Mandarin input since childhood. In particular, in the early years of the National Popularization of Putonghua Southern Min speakers were less immersed in Beijing Mandarin than in later years when most of the population in Quanzhou had become bilingual. Because diglossia became more and more well-established over the years, younger Quanzhou residents would have had more Beijing Mandarin input across more settings and at younger AOL than older speakers. This generational difference is confounded with use, and so it may not only be use that has resulted in the younger

Quanzhou bilinguals having more native-like L2 Mandarin prosody than the mid-age and older bilinguals.

This possibility seems even more likely when we consider the results from Experiment 2, which indicated that earlier AOL resulted in more native-like L2 Mandarin by American learners. In particular, Chinese-heritage American learners of Mandarin were exposed to the language in childhood, substantially earlier than the non-Chinese-heritage American learners of Mandarin. The Chinese-heritage learners produced Mandarin prosodic focus with F0 expansion on the in-focus item, more PFC, and with a better pattern of lexical tones. Of course, AOL is also confounded with use: the earlier one acquires a language, the more life-time use that person will have with the language.

In Experiment 3, learners varied in LOR. This factor correlates with L2 input, but also with L2 use. The Senior Chinese learners of English produced more native-like PFC and in-focus change of lexical stress in English than the Freshman Chinese learners of English. The Senior Chinese learners had also been immersed in American English for longer in a university setting, and obligated to use more L2 English in that setting, than the Freshman Chinese learners.

In all, careful characterization of the populations studied in these three experiments demonstrates that L2 experience is a multi-factorial concept, where AOL is not easily segregated from language use or from quality input and so on. However, the results from all experiments would appear to underscore the importance of AOL for native-like performance in a second language, albeit only if the input received early on is of high enough quality. This caveat is especially well demonstrated in the current dissertation. The older Southern Min-Mandarin speakers in Experiment 1 and the

Freshman Chinese learners of English in Experiment 3 had early AOLs, but neither group could produce native-like prosodic focus in the target L2 because the speech input they received early on was not of sufficiently high quality for them to have acquired the L2 patterns. Thus, the current study reconfirms the importance of the quality and quantity of L2 input in acquisition of L2 phonology.

### 6.3. Future directions

The current dissertation assumed an SLM framework for understanding the influences of the L1 on the acquisition of the L2 and for understanding the effects of AOL, use, and input (Flege, 1995, 2007). The central hypothesis of SLM is that learners' production is driven by their perception. But we argued that the absence of PFC in L2 Mandarin and English by L1 English and Mandarin speakers might be due to the allocation of attention to the novel word-level categories that interact with the realization of prosodic focus, resulting in accented productions of prosodic focus in the L2, surprising because phrase-level prosody in Mandarin and English is similar. Future work should investigate L2 perception of prosodic focus in these languages to determine the extent to which it tracks production, and so whether the attention explanation is apt. For example, we could examine whether non-Chinese-heritage American learners perceive PFC in native Beijing Mandarin by manipulate the extent to which it occurs and their attention to word- or phrase-level patterns, perhaps by asking some to identify lexical tones and others the overall gist of the sentence.

Other ideas for future work are suggested by certain limitations of the current experiments. For example, we controlled the lexical tone associated with pre- and post-focus items in Experiments 1 and 2, using Tone 1 because its high level target made it easy to assess PFC in learners' production. Future work should investigate PFC using other tone types to see whether the patterns observed here are general or specific to the type of tone being produced. Similarly, future work in L2 English might investigate the interaction between lexical stress and prosodic focus more completely if multiple different items with the same stress pattern were produced with prosodic focus, but also in pre- and post-focus positions. Other work could also use more repetitions of different patterns of lexical stress to examine which types of stress patterns are easier for Mandarin learners to acquire and which are more difficult. Finally, it would be interesting to investigate L2 English prosodic focus by Taiwanese learners, who have no access to a language with PFC to determine if or when they might acquire PFC in a language with lexical stress. This setting would allow the possibility to explore how learners establish phonetic categories in L2 prosody when these are wholly new at the word- and phrase-level.

Lastly, we note that the finding that learners' production of prosodic focus becomes more native-like with increased L2 experience has pedagogical implications for L2 Mandarin or L2 English teaching. Early exposure to an L2 is obviously important, but it is also clearly important to increase the quality and quantity of L2 input and the amount of L2 use to achieve higher L2 speech accuracy. Birdsong (2007) found that performance at the global level predicts performance at the segmental level but not vice versa, suggesting the importance of phonetic training at the suprasegmental level. Indeed, training at the segmental and suprasegmental levels has been shown to more effective in decreasing foreign accent than training only at the segmental level (Moyer, 1999). "Prosody-centered" phonetic training has also been shown to be more effective than “segment-centered” phonetic training (Missaglia, 1999). Relatedly, Viger (2007) attributed the failure of intonation acquisition in L2 Mandarin by L1English learners to the sacrifice of global prosody for the sake of local prosody in teaching and training.

Entrenched L2 training on word-level prosody in the classroom may have impacted learners’ production of phrase-level prosody in the current experiments. In most L2 Mandarin classroom teaching, teachers attach more importance on lexical tones than other phonetic categories, and they typically train the students on lexical tones with isolated syllables rather than at the phrase level. Similarly, L2 English teachers train students specifically about lexical stress, but again this training is focused on the word-level rather than on the phrase-level. It is highly like that this kind of initial training has an impact on the categories that learners establish, making it more difficult for them to produce the distributed changes associated with prosodic focus in L2.

All together the current findings coupled with those already reported in the literature suggest that the ideal learning situation for achieving high L2 pronunciation accuracy is a high-quality language immersion environment with specific perception and production training on both L2 segments and prosody. The present findings also suggest that assessment of prosodic patterns at the phrase-level will provide a good indication of L2 speech proficiency. With respect to L2 speech proficiency, balancing pronunciation accuracy and speech intelligibility is difficult for both learners and teachers (Bent, Bradlow & Smith, 2007; Munro, 2008). The lack of PFC in L2 production may not immediately result in the failure of perceiving focus if there are in-focus change and duration cues; however, it may result in a foreign accent. Insofar as phrase-level patterns relate both to pronunciation accuracy and speech intelligibility, future work could assess the extent to which secondary patterns, like PFC, contribute to overall L2 higher accuracy and intelligibility. Instead of paired no-focus vs. focus questions and answers, this kind of examination can be set in more spontaneous speech, such as story retelling or longer conversation. This may be especially useful in learners with intermediate or higher proficiency levels when segmental and word-level inaccuracies are perhaps less salient.

## APPENDIX A

## LANGUAGE BACKGROUND OF ALL BILINGUAL SPEAKERS OF SOUTHERN MIN (SM) AND MANDARIN (MD) IN EXPERIMENT 1

| | Age of Test | Year of Education | MD AOL | MD Use | SM self Estimate | MD self Estimate | PTH Test[3] |
|---|---|---|---|---|---|---|---|
| **Younger** | | | | | | | |
| Y-f-1 | 21 | 15 | 4 | 60% | 8 | 9 | 2A |
| Y-f-2 | 20 | 14 | 4 | 70% | 8 | 8 | 2A |
| Y-f-3 | 18 | 13 | 3 | 80% | 6 | 8 | 2B |
| Y-f-4 | 19 | 13 | 5 | 60% | 9 | 10 | |
| Y-m-1 | 20 | 14 | 4 | 70% | 7 | 8 | 2A |
| Y-m-2 | 19 | 13 | 3 | 60% | 7 | 8 | 2B |
| Y-m-3 | 21 | 15 | 4 | 50% | 8 | 8 | 2A |
| Y-m-4 | 21 | 15 | 5 | 50% | 8 | 9 | |
| **Mean** | **19.9** | **14** | **4.0** | **63%** | **7.6** | **8.5** | |
| **SD** | **1.1** | **0.9** | **0.8** | **10%** | **0.9** | **0.8** | |
| **Mid-age** | | | | | | | |
| M-f-1 | 35 | 19 | 6 | 50% | 9 | 10 | 2A |
| M-f-2 | 37 | 14 | 7 | 60% | 9 | 8 | |
| M-f-3 | 42 | 16 | 6 | 30% | 10 | 8 | |
| M-f-4 | 43 | 16 | 7 | 60% | 9 | 8 | |
| M-m-1 | 37 | 22 | 4 | 40% | 8 | 8 | 2B |
| M-m-2 | 37 | 16 | 6 | 30% | 10 | 8 | |
| M-m-3 | 42 | 16 | 7 | 30% | 10 | 9 | 2B |
| M-m-4 | 43 | 16 | 7 | 40% | 8 | 6 | |
| **Mean** | **39.5** | **16.9** | **6.3** | **43%** | **9.1** | **8.1** | |
| **SD** | **3.3** | **2.5** | **1.0** | **13%** | **0.8** | **1.1** | |
| **Older** | | | | | | | |
| O-f-1 | 55 | 9 | 8 | 20% | 10 | 8 | |
| O-f-2 | 57 | 9 | 8 | 50% | 9 | 6 | |
| O-f-3 | 56 | 12 | 7 | 10% | 10 | 9 | |
| O-f-4 | 62 | 9 | 8 | 10% | 9 | 5 | |
| O-m-1 | 60 | 14 | 7 | 30% | 10 | 8 | |
| O-m-2 | 64 | 14 | 8 | 40% | 10 | 8 | |
| O-m-3 | 59 | 9 | 8 | 10% | 10 | 8 | |
| O-m-4 | 56 | 16 | 8 | 30% | 10 | 8 | |
| **Mean** | **58.6** | **11.5** | **7.8** | **25%** | **9.8** | **7.5** | |
| **SD** | **3.2** | **2.9** | **0.5** | **15%** | **0.5** | **1.3** | |

[3] 2B is regarded as an average grade in National Putonghua Proficiency Test.

## APPENDIX B

## PERCENT USE OF MANDARIN IN DIFFERENT SITUATIONS BY THE SOUTHERN MIN-MANDARIN BILINGUAL SUBJECTS IN EXPERIMENT 1

| | With grand-parents | With parents | With children | With relatives | At home | At school | With co-workers | With friends | Other occasions |
|---|---|---|---|---|---|---|---|---|---|
| **Younger** | | | | | | | | | |
| Y-f-1 | 0 | 20 | - | 40 | 30 | 90 | - | 70 | 80 |
| Y-f-2 | 0 | 30 | - | 30 | 30 | 90 | 90 | 90 | 60 |
| Y-f-3 | 0 | 10 | - | 20 | 20 | 90 | 90 | 90 | 90 |
| Y-f-4 | 0 | 50 | - | 20 | 50 | 90 | - | 60 | 70 |
| Y-m-1 | 0 | 20 | - | 20 | 20 | 90 | - | 40 | 90 |
| Y-m-2 | 0 | 10 | - | 40 | 20 | 60 | 70 | 60 | 70 |
| Y-m-3 | 0 | 10 | - | 40 | 10 | 60 | 70 | 40 | 50 |
| Y-m-4 | 0 | 10 | - | 30 | 10 | 80 | - | 50 | 90 |
| **Mean** | **0** | **21.3** | **-** | **30.0** | **23.8** | **81.3** | **80.0** | **62.5** | **75.0** |
| **SD** | **0** | **12.7** | **-** | **8.7** | **12.2** | **12.7** | **10** | **18.5** | **14.1** |
| **Mid-age** | | | | | | | | | |
| M-f-1 | 0 | 0 | 80 | 10 | 50 | 80 | 80 | 50 | 80 |
| M-f-2 | 0 | 10 | - | 50 | 60 | 90 | 90 | 80 | 80 |
| M-f-3 | 0 | 10 | 80 | 30 | 40 | 80 | 90 | 80 | 80 |
| M-f-4 | 0 | 0 | 80 | 20 | 40 | 50 | 20 | 20 | 40 |
| M-m-1 | 0 | 30 | - | 50 | 50 | 50 | 30 | 30 | 40 |
| M-m-2 | 0 | 0 | 70 | 10 | 10 | 40 | 10 | 10 | 40 |
| M-m-3 | 0 | 0 | - | 0 | 10 | 50 | 80 | 10 | 60 |
| M-m-4 | 0 | 0 | 30 | 0 | 20 | 50 | 50 | 20 | 50 |
| **Mean** | **0** | **6.3** | **68.0** | **21.3** | **35.0** | **61.3** | **56.3** | **37.5** | **58.8** |
| **SD** | **0** | **10.6** | **21.7** | **20.3** | **19.3** | **18.9** | **32.9** | **29.2** | **18.9** |
| **Older** | | | | | | | | | |
| O-f-1 | 0 | 0 | 40 | 0 | 20 | 70 | 20 | 20 | 50 |
| O-f-2 | 0 | 0 | 50 | 50 | 40 | 60 | 20 | 30 | 60 |
| O-f-3 | 0 | 0 | 10 | 10 | 10 | 20 | 0 | 0 | 10 |
| O-f-4 | 0 | 0 | 0 | 0 | 0 | 60 | 40 | 20 | 30 |
| O-m-1 | 0 | 0 | 30 | 20 | 20 | 50 | 40 | 40 | 50 |
| O-m-2 | 0 | 0 | 30 | 30 | 30 | 40 | 50 | 30 | 40 |
| O-m-3 | 0 | 0 | 10 | 0 | 10 | 60 | 10 | 20 | 40 |
| O-m-4 | 0 | 0 | 30 | 0 | 20 | 80 | 50 | 20 | 60 |
| **Mean** | **0** | **0** | **25.0** | **13.8** | **18.8** | **55.0** | **28.8** | **22.5** | **42.5** |
| **SD** | **0** | **0** | **16.9** | **18.5** | **12.5** | **18.5** | **18.9** | **11.6** | **16.7** |

## APPENDIX C

## LANGUAGE BACKGROUND OF ALL AMERICAN LEARNERS OF MANDARIN IN EXPERIMENT 2

| | Age of Test | Age of Learning | Length of Residence | Mandarin Use | Self-est. Listening | Self-est. Speaking |
|---|---|---|---|---|---|---|
| **Chinese-heritage learners** | | | | | | |
| CH_F1 | 23 | 8 | 0.5 | 40% | 8 | 8 |
| CH_F2 | 22 | 5 | 1.5 | 10% | 9 | 7.5 |
| CH_F3 | 21 | 1 | 0 | 20% | 8 | 6 |
| CH_F4 | 20 | 8 | 0.5 | 40% | 7 | 7 |
| CH_F5 | 20 | 1 | 0.5 | 10% | 8 | 8 |
| CH_M1 | 20 | 2 | 1 | 60% | 7 | 7 |
| CH_M2 | 21 | 2 | 1.5 | 40% | 7 | 7 |
| CH_M3 | 22 | 9 | 1.2 | 60% | 8 | 8 |
| CH_M4 | 20 | 1 | 2 | 30% | 8 | 8 |
| CH_M5 | 20 | 6 | 0 | 60% | 8 | 8 |
| **Mean** | **20.9** | **4.3** | **0.9** | **37%** | **7.8** | **7.5** |
| **SD** | **1.0** | **3.1** | **0.6** | **18%** | **0.6** | **0.7** |
| **Non-Chinese-heritage learners** | | | | | | |
| NCH_F1 | 21 | 15 | 0.5 | 20% | 7.5 | 7 |
| NCH_F2 | 32 | 26 | 1 | 10% | 5 | 5 |
| NCH_F3 | 19 | 14 | 0.5 | 30% | 7 | 7 |
| NCH_F4 | 22 | 16 | 2 | 60% | 7.5 | 7.5 |
| NCH_F5 | 23 | 18 | 1.5 | 30% | 6 | 7 |
| NCH_M1 | 20 | 16 | 1 | 30% | 6.5 | 6 |
| NCH_M2 | 22 | 18 | 0.5 | 20% | 5 | 6 |
| NCH_M3 | 21 | 16 | 2 | 30% | 7 | 7 |
| NCH_M4 | 20 | 16 | 1 | 10% | 8 | 7 |
| NCH_M5 | 21 | 18 | 0.5 | 30% | 5 | 4 |
| **Mean** | **22.1** | **17.3** | **1.1** | **27%** | **6.5** | **6.4** |
| **SD** | **3.5** | **3.2** | **0.6** | **13%** | **1.1** | **1.1** |

## APPENDIX D

## LANGUAGE BACKGROUND OF ALL BEIJING LEARNERS OF ENGLISH IN EXPERIMENT 3

| | Age of Test | Age of Learning | Length of Residence | Mandarin Use | Self-est. Listening | Self-est. Speaking |
|---|---|---|---|---|---|---|
| **Senior Chinese learners** | | | | | | |
| SC_F1 | 21 | 11 | 4 | 80% | 8 | 8 |
| SC_F2 | 24 | 12 | 4 | 50% | 6 | 7 |
| SC_F3 | 22 | 7 | 3.5 | 60% | 6 | 6 |
| SC_F4 | 24 | 9 | 4 | 90% | 7 | 7 |
| SC_F5 | 22 | 10 | 3.5 | 40% | 7 | 7 |
| SC_M1 | 22 | 9 | 4.5 | 60% | 7 | 7 |
| SC_M2 | 23 | 10 | 3.5 | 40% | 7 | 5 |
| SC_M3 | 22 | 6 | 3.5 | 50% | 6 | 6 |
| SC_M4 | 24 | 9 | 3.5 | 50% | 7 | 7 |
| SC_M5 | 23 | 7 | 4.5 | 40% | 7 | 7 |
| **Mean** | **22.7** | **9.0** | **3.9** | **56%** | **6.8** | **6.7** |
| **SD** | **1.0** | **1.8** | **0.4** | **16%** | **0.6** | **0.8** |
| **Freshman Chinese learners** | | | | | | |
| FC_F1 | 19 | 7 | 0.5 | 70% | 6 | 7 |
| FC_F2 | 20 | 12 | 0.3 | 50% | 4 | 5 |
| FC_F3 | 18 | 6 | 0.3 | 50% | 4 | 4 |
| FC_F4 | 18 | 6 | 0.3 | 30% | 5 | 4 |
| FC_F5 | 20 | 12 | 0.5 | 60% | 5 | 7 |
| FC_M1 | 20 | 7 | 0.25 | 60% | 7 | 6 |
| FC_M2 | 19 | 7 | 0.5 | 20% | 5 | 4 |
| FC_M3 | 21 | 10 | 0.6 | 30% | 4 | 4 |
| FC_M4 | 21 | 8 | 0.25 | 40% | 5 | 4 |
| FC_M5 | 19 | 10 | 0.25 | 30% | 5 | 4 |
| **Mean** | **19.5** | **8.5** | **0.4** | **44%** | **5.0** | **4.9** |
| **SD** | **1.0** | **2.2** | **0.1** | **16%** | **0.9** | **1.2** |

## REFERENCES CITED


Aoyama, K. & Guion, S. G. (2007). Prosody in second language acquisition: Acoustic analyses of duration and F0 range. In O.-S. Bohn & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 281-297). Amsterdam: John Benjamins.

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*, 1-27.

Beckman, M. & Pierrehumbert, J. (1986). Intonational Structure in Japanese and English, *Phonology Yearbook III*, 15-70.

Bent, T., Bradlow, A. R., & Smith, B. (2007). Phonemic errors in different word positions and their effects on intelligibility of non-native speech: All's well that begins well. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 331-348), Amsterdam: John Benjamins.

Best, C. T. (1995). A *direct realist* view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204), Timonium, MD: York Press.

Best, C. T., & Tyler, M. (2007). Nonnative and second-language speech perception. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 13-34), Amsterdam: John Benjamins.

Best, C. T., McRoberts, G. W., & Sithole, N. T. (1988). Examination of perceptual reorganization for non-native speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 345-360.

Birdsong, D. (2007). Nativelike pronunciation among late learners of French as a second language. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 99-116), Amsterdam: John Benjamins.

Bolinger, D. (1972). Accent is predictable (if you're a mind-reader). *Language*, *37*, 83-96.

Bohn, O.-S. (1995). Cross-language speech perception in adults: L1 transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279-304), Timonium, MD: York Press.

Cao, J. (2002). Hanyu shengdiao yu yudiao de guanxi (The relation of tone and intonation in Chinese). *Zhongguo Yuwen* (*Chinese Languages and Literature*), *2002 (3)*, 195-202.

Chao, Y. R. (1932). A preliminary study of English intonation (with American variants) and its Chinese equivalent. *Bulletin of the Institute of History and Philology, Suppl.1*, 105-156.

Chao, Y. R. (1933). Tone and intonation in Chinese. *Bulletin of the Institute of History and Philology*, *4*, 121-134.

Chen, M. Y. (2000). *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge: Cambridge University Press.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech -- Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, *63*, 47-75.

Cooper, W. E., Eady, S. J., & Mueller, P. R. (1985). Acoustical aspects of contrastive stress in question-answer contexts. *Journal of the Acoustical Society of America*, *77*, 2142-2156.

Cruttenden, A. (1997). *Intonation*. New York: Cambridge University Press.

DeKeyser, R. M. (2000). *The robustness* of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

Dik, S. (1997). *The Theory of Functional Grammar Part 1*. Berlin: Mouton de Gruyter.

Duanmu, S. (1994). Syllabic weight and syllabic duration: A correlation between phonology and phonetics. *Phonology*, *11* (*1*), 1-24.

Eady, S. J., & Cooper, W. E. (1986). Speech intonation and focus location in matched statements and questions. *Journal of the Acoustical Society of America*, *80*, 402-415.

Flege, J. E. (1988). Factors affecting degree of perceived foreign accent in English sentences. *Journal of the Acoustical Society of America*, *84*, 70-79.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395-411.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277), Timonium, MD: York Press.

Flege, J. E. (2007). Language contact in bilingualism: Phonetic system interactions. In J. Cole & J. Hualde (Eds.), *Laboratory Phonology 9* (pp. 353-382), Berlin: Mouton de Gruyter.

Flege, J. E., Bohn, O-S., & Jang, S. (1997). The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics, 25*, 437-470.

Flege, J. E., & Fletcher, K. (1992). Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America*, *91*, 370-389.

Flege, J. E., Frieda, E. M., & Nozawa, T. (1997). Amount of native language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169-186.

Flege, J. E. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527-552.

Flege, J. E., MacKay, I. R. A., & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of Acoustical Society of America*, *106*, 2973-2987.

Flege, J. E., McCutcheon, M., & Smith, S. (1987). The development of skill in producing word-final English stops. *Journal of the Acoustical Society of America*, *82*, 433-447.

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of Acoustical Society of America*, *97*, 3125-3134.

Francis, A.L., Kaganovich, N., & Driscoll-Huber, C.J. (2008). Cue-specific effects of categorization training on the relative weighting of acoustic cues to consonant voicing in English. *Journal of the Acoustical Society of America*, *124* (*2*), 1234-1251.

Fujisaki, H. (1983). Dynamic characteristics of voice fundamental frequency in speech and singing. In P. F. MacNeilage (Ed.), *The Production of Speech* (pp. 39-55). Heidelberg: Springer-Verlag.

Gandour, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, *11*, 149–175.

Gandour, J. T. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, *12*, 235–261.

Gauthier, B., Shi, R., & Xu, Y. (2009). Learning Prosodic Focus from Continuous Speech Input: A Neural Network Exploration. *Language Learning and Development*, *5*, 94-114.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals: An investigation into age of acquisition effects on the mutual influence of the first and second languages. *Phonetica*, *60*, 98-128.

Guion, S. G., Flege, J. E., & Loftin, J. (2000). The effect of L1 use on pronunciation in Quichua-Spanish bilinguals. *Journal of Phonetics*, *28*, 27-42.

Guion, S. G., Flege, J. E., Liu, S., & Yeni-Komshian, G. (2000). Age of learning effects on the duration of sentences produced in a second language. *Applied Psycholinguistics*, *21*, 205-228.

Guion, S. G. & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. Munro (Eds.) *Language Experience in Second Language Speech Learning* (pp. 57-77). Amsterdam: John Benjamins.

Gussenhoven, C. (1983). Focus, mode and the nucleus. *Journal of Linguistics, 19*, 377-417.

Hanssen, J., Peters, J., & Gussenhoven, C. (2008). Prosodic Effects of Focus in Dutch Declaratives. *ISCA Archive, Speech Prosody*, Campinas, Brazil, May 6-9, 2008.

He, X., Hanssen, J., van Heuven V. J., & Gussenhoven, C. (2011). Phonetic implementation must be learnt: native versus Chinese realization of focus accent in Dutch. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 843-846.

Hojen, A., & Flege, J. (2006). Early learners' discrimination of second-language (L2) vowels. *Journal of the Acoustical Society of America, 119*, 3072-3084.

Huang, B. H., & Jun, S-A. *(*2011*).* The *effect* of *age* on the acquisition of second language prosody. *Language and Speech*, *54* (*3*), 387–414.

Huang, Y-H., & Fon, J. (2007). The effect of acquisition order and word relatedness on code-switching costs in balanced bilingual speakers. *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1577-1580.

Jin, Shunde (1996). *An Acoustic Study of Sentence Stress in Mandarin Chinese*. Ph.D. dissertation. The Ohio State University.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, *21*, 60-99.

Kang, K.-H., & Guion, S. (2006). Phonological systems in bilinguals: Age of learning effects on the stop consonant systems of Korean-English bilinguals. *Journal of the Acoustical Society of America*, *119*, 1672-1683.

Kubler, C. C. (1985). The influence of Southern Min on the Mandarin of Taiwan. *Anthropological Linguistics*, *27*(*2*), 156-176.

Kubozono, H. (2007). Focus and Intonation in Japanese: Does Focus Trigger Pitch Reset? In Ishihara, S., Jannedy, S. & Schwarz, A. (Eds.) *Working Papers of the SFB632, Interdisciplinary Studies on Information Structure (ISIS)*, *9* (*1-27*). Potsdam: Universitätsverlag Potsdam.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge University Press.

Ladefoged, P. (2006). *A Course in Phonetics*. Boston: Thomson Wadsworth.

Lado, R. (1957). *Linguistics across cultures: Applied linguistics for language teachers*. Ann Arbor: University of Michigan Press.

Lambrecht, K. (1994). *Information Structure and Sentence Form*. Cambridge: Cambridge University Press.

Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In James, A. & Leather, J. (Eds.). *Sound Patterns in Second Language Acquisition* (pp. 59-81). Dordrecht: Foris.

Lee, Y-c., & Xu, Y. (2010). Phonetic Realization of Contrastive Focus in Korean. *Proceedings of Speech Prosody 2010*, Chicago.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York: Wiley.

Li, D. C. S., & Lee, S. (2006). Bilingualism in East Asia. In T. K. Bhatia, & W. C. Ritchia (eds.), *The Handbook of Bilingualism* (pp. 742-779). Cambridge, MA: Blackwell.

Liao, S. (2008). A Perceptual Dialect Study of Taiwan Mandarin:Language Attitudes in the Era of Political Battle. In M. K. M. Chan, & H. Kang (Eds.), *Proceedings of the 20th North American Conference on Chinese Linguistics (NACCL-20)*, *Vol. 1* (pp. 391-408), Columbus, OH: The Ohio State University.

Lin, L. (1993). *Quanzhoushi Fangyanzhi* (Dialect record of Quanzhou city). Beijing: Social Science Literature Press.

Liu, F., & Xu, Y. (2005). Parallel encoding of focus and interrogative meaning in Mandarin intonation. *Phonetica, 62*, 70-87.

Mack, M. (1990). Phonetic transfer in a French-English bilingual child. In P. H. Nelde (Ed.), *Language Attitudes and Language Conflict* (pp.107-124). Bonn, Germany: Dümmler.

Major, R. C. (2008). Transfer in second language phonology. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 63-84). Amsterdam: John Benjamins.

Massaro, D.W., Cohen, M.M.,& Tseng, C. Y. (1985). The Evaluation and Integration of Pitch Height and Pitch Contour in Lexical Tone Perception in Mandarin Chinese. *Journal of Chinese Linguistics*, *13*, 267-289.

McAllister, R., Flege, J., & Piske, T. (2002). The influence of the L1 on the acquisition of Swedish vowel quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, 30, 229-258.

McGory, J. T . (1997). *Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers*. Ph.D. dissertation. The Ohio State University.

Missaglia, F. (1999). Contrastive prosody in SLA: an empirical study with adult Italian learners of German. *Proceedings of the 14th International Congress of Phonetic Sciences*, *Vol. 1*, 551-554.

Moyer, A. (1999). Ultimate attainment in L2 phonology. *Studies in Second Language Acquisition*, *21*, 81-108.

Munro, M. J. (2008). Foreign accent and speech intelligibility. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition* (pp. 193-218). Amsterdam: John Benjamins.

Munro, M. J., & *Bohn, O.-S. (2007).* The study of second language speech learning: A brief overview. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 3-12), Amsterdam: John Benjamins.

Newport, E.L. (1990). Maturational constraints on language learning. *Cognitive Science*, *14*, 11-28.

Oller, J., & Ziahosseini, S. (1970). The contrastive analysis hypothesis and spelling errors. *Language Learning*, *20*, 183-189.

Oyama, S. (1976). A sensitive period for the acquisition of a non-native phonological system. *Journal of Psycholinguistic Research, 5 (3)*, 261-283.

Pan, H. (2007). Focus and Taiwanese unchecked tones. In C. Lee, M. Gordon, & D. Büring (Eds.) *Topic and Focus: Cross-linguistic Perspectives on Meaning and Intonation* (pp. 195-213), Dordrecht: Springer.

Patil, U., *Kentner, G., Gollrad, A., Kügler, F., Fery, C., & Vasishth, S.* (2008). Focus, word order and intonation in Hindi. *Journal of South Asian Linguistics*, 1, 1, 55-72.

Patkowski, M. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning*, *30,* 449-472.

Patkowski, M. (1990). Age and accent in a second language: A reply to James Emil Flege. *Applied Linguistics*, *11*, 73-89.

Peng, S-H. (1993). Cross-language influence on the production of Mandarin /f/ and /x/ and Taiwanese /h/ by native speakers of Taiwanese Amoy. *Phonetica*, *50* (*4*), 245-260.

Perneger, T. V. (1998). What's wrong with Bonferroni adjustments. *British Medical Journal*, 316, 1236-1238.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and Experiment* (pp. 11-26). Studies Presented to Gosta Bruce. Dordrecht: Kluwer.

Pierrehumbert, J., & Hirschberg J. (1990). In P. Cohen, J. Morgan, & M. Pollack (Eds.), Intentions *in Communication* (pp 271-311). Cambridge MA: MIT Press.

Piske, T. (2007). Implications of James E. Flege's research for the foreign language classroom. In O.-S. Bohn., & M. Munro (Eds.), *Language Experience in Second Language Speech Learning* (pp. 301-314). Amsterdam: John Benjamins.

Piske, T., MacKay, I.R.A. & Flege, J.E. (2001). Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, *29*, 191-215.

Prom-on, S., Xu, Y., & Thipakorn, B. (2009) Modeling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America*, *125*, 405-424.

Rivers, W. M., & Temperley, M. S. (*1978*). *A Practical Guide to the Teaching of English as a Second or Foreign Language*. New York: Oxford University Press.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19,* 245-253.

Scovel, T. (1988). A Time to Speak: Psycholinguistic Inquiry into the Critical Period for Human Speech. New York : Newbury House.

Selkirk, E. (1995). Sentence prosody: intonation, stress, and phrasing. In J. A. Goldsmith (Ed.), *The Handbook of Phonological Theory* (pp. 550-569). Oxford: Blackwell.

Selkirk, E. (2006). Bengali intonation revisited: An optimality theoretic analaysis in which FOCUS stress prominence drives FOCUS phrasing. In C. Lee, M. Gordon, & D. Büring (Eds.), *Topic and Focus: A Cross-Linguistic Perspective* (pp. 217-246). Dordrecht: Kluwer.

Shen, J. (1985). Beijinghua shengdiao de yinyu he yudiao (Pitch range of tone and intonation in Beijing dialect). In T. Lin & L. Wang (Eds.), *Beijing Yuyin Shiyanlu* (Working Papers in Experimental Phonetics) (pp. 73-130). Beijing: Peking University Press.

Shen, J. (1992). Hanyu yudiao moxing chuyi (Preliminary discussion of intonational models in Mandarin). *Yuyan Yanjiu* (Language Research), *1992* (*4*), 16-24.

Shen, J. (1994). Hanyu yudiao gouzao he yudiao leixing (Intonational structures and patterns in Mandarin). *Fangyan* (Dialect), *1994* (*3*), 221-228.

Shen, X.-N. S. (1990). The Prosody of Mandarin Chinese. Berkeley, CA: University of California Press.

Shih, C. (2000). A Declination Model of Mandarin Chinese. In A. Botinis (Ed.), *Intonation: Analysis, Modelling and Technology* (pp. 243-268). Dordrecht: Kluwer.

Shih, C. (2004). Tonal effects on intonation. *International Symposium on Tonal Aspects of Languages: Emphasis on Tone Languages*, (pp. 163-168). Beijing, China, March 28-30.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. Tohkura, E.Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech Perception, Production, and Linguistic Structure* (pp.197-219), Tokyo: Ohmsha.

Sun, S. H. (1998). *The Development of a Lexical Tone Phonology in American Adult Learners of Standard Mandarin Chinese.* Honolulu: University of Hawai'i Press.

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

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

Trofimovich, P., & Baker, W. (2006). Learning second-language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech. *Studies in Second Language Acquisition, 28,* 1-30.

Vainio, M., & Järvikivi, J. (2007). Focus in production: Tonal shape, intensity and word order. *Journal of the Acoustical Society of America*, *121*, *2*, EL55-61.

Viger, T. L. (2007). *Fundamental Frequency in Mandarin Chinese and English: Implications for Second Language Speakers*. Ph.D dissertation. The City University of New York.

Wang, Y. & Behne, D. (2007). Temporal remnants from Mandarin in nonnative English speech. In O.-S. Bohn & M. Munro (Eds.) *Language Experience in Second Language Speech Learning* (pp. 167-184). Amsterdam: John Benjamins.

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999) Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, *106*, 3649- 3658.

Wang, B., Wang, L., & Qadir, T. (2011). Prosodic encoding of focus in six languages in China. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 144-147.

Wayland, R. P., & Guion, S.G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, *54*, 681-712.

Wayland, R. P. & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, *36*, 250–267.

Wu, W. L., & Chung, L. (2011). Post-focus compression in English-Cantonese bilingual speakers. *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, 148-151.

Wu, W. L., & Xu, Y. (2010). Prosodic Focus in Hong Kong Cantonese without Post-focus Compression. *Proceedings of Speech Prosody 2010*, Chicago.

Wu, Z. (1982). Putonghua yuju zhong de shengdiao bianhua (Tonal variations in Mandarin sentences). *Zhongguo Yuwen* (Chinese Languages and Literature), *1982* (*6*), 439-450.

Wu, Z. (1988). Basic types of Mandarin Chinese intonation (Hanyu Putonghua yudiao de jiben diaoxing). In *Wang Li Xiansheng Linian Lunwenji* (A *collection of academic papersin Honor of Professor Wang Li),* reprinted in *Wu Zongji Yuyanxue Lunwenji* (A Collection of Linguistics Papers by Wu Zongji) (2004, pp. 281-300). Beijing: The Commercial Press.

Wu, Z. (1997). Cong shengdiao yu yuelu de guanxi tichu Putonghua yudiao chuli de xinfangfa (A new method of processing Mandarin intonation from the correlation of tone and prosody). In *Zhongguo Shenhui Kexueyuan Yuyan Yanjiusuo Jiansuo 45 Zhounian Jinian Wenji* (A Collection of Academic Papers for the 45th Anniversary of the Institute of Langauge Research at China Social Science Academy), reprinted in *Wu Zongji Yuyanxue Lunwenji* (A Collection of Linguistics Papers by Wu Zongji) (2004, pp. 320-358). Beijing: The Commercial Press.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America*, *95*, 2240-2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, *25*, 61-83.

Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*, 55-105.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech Communication*, *46*, 220-251.

Xu, Y. (2005-2014). "ProsodyPro.praat." Available from: <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>.

Xu, Y. (2011). Post-focus compression: Cross-linguistic distribution and historical origin. *Proceedings of The 17th International Congress of Phonetic Sciences*, Hong Kong: 152-155.

Xu, Y., Chen, S.-w., Wang, B. (2012). Prosodic focus with and without post-focus compression (PFC): A typological divide within the same language family? *The Linguistic Review*, 29, 131-147.

Xu, Y. and Liu, F. (2006). Tonal alignment, syllable structure and coarticulation: Toward an integrated model. *Italian Journal of Linguistics*, *18*, 125-159.

Xu, Y. and Liu, F. (2007). Determining the temporal interval of segments with the help of F0 contours. *Journal of Phonetics*, *35*, 398-420.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication*, *33*, 319-337.

Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative ntonation. *Journal of Phonetics*, *33*, 159-197.

Yang, C. (2011). The Acquisition of Mandarin Prosody by American Learners of Chinese as a Foreign Language (CFL). Ph.D. Dissertation. The Ohio State University.