

A STUDY ON MANDARIN FOCUS PRODUCED BY ENGLISH L2 LEARNERS

Siyun Zhu

A thesis submitted to the faculty at the University of North Carolina at Chapel Hill in
partial fulfillment of the requirements for the degree of Master of Arts
in the Department of Linguistics.

Chapel Hill
2014

Approved by:

Elliott Moreton

Jennifer L. Smith

David Mora-Marín

©2014
Siyun Zhu
ALL RIGHTS RESERVED

ABSTRACT

Siyun Zhu: A Study on Mandarin Focus Produced by English L2 Learners
(Under the direction of Elliott Moreton)

This thesis investigates how Mandarin focus intonation is realized in the production of English L2 learners by examining relevant measurements of its acoustic carrier, fundamental frequency (f_0), and comparing the observed results with those of the native speakers. The experiment and comparative results show that the English L2 learners are able to realize focus by manipulating duration and pitch range of the syllable accordingly. At the same time, most of the subjects are able to realize lexical tones, though not equally proficient among the four tones, instead of substituting them with a pitch accent. Result of an individual subject resolves a controversy over English focus and confirms it as a pitch accent. This thesis not only fills in a gap in the study of SLA of Mandarin intonation, but also provides pedagogical implications for L2 instruction of Mandarin.

ACKNOWLEDGEMENTS

First of all, I'd like to thank my thesis advisor, Elliott Moreton, who encouraged me when I had doubt on my ability, inspired me when I faced difficulties and supported me whenever I was in need. He taught me how to think, learn and solve problems independently. Words cannot describe how grateful I am for having him as my advisor throughout the past year.

Advice and support from the other two committee members, Dr. Jennifer Smith and Dr. David Mora Marin are invaluable as well. Their detailed comments are extremely insightful.

I'd also like to thank Chris Wiesen for his statistic support of the thesis. His expertise and patience verifies and standardizes the critical analysis and results of this study.

I'm grateful to have all the members in the department, especially the weekly p-side, for their kind feedback and abundant encouragement.

Last but not least, I would have never been able to enjoy my graduate study in U.S. without the unconditional support of my parent, both financially and spiritually. I dedicate this thesis to my dearest mother and father.

TABLE OF CONTENTS

LIST OF TABLES	viii
LIST OF FIGURES	ix
CHAPTER 1 INTRODUCTION	1
1.1 Problem and hypothesis	1
1.2 Lexical tones	4
1.2.1 Lexical tones in Mandarin Chinese.....	4
1.2.2 Second language acquisition of Mandarin lexical tones.....	6
1.3 Focus	7
1.3.1 Focus	7
1.3.2 Focus in English.....	9
1.3.3 Focus in Mandarin	10
1.3.4 Second language acquisition of focus in Mandarin	11
1.4 Studies of Xu.....	12
1.4.1 Xu (2005) on English focus	12
1.4.2 Xu (1999) on Mandarin focus and lexical tone	13
1.5 Summary and overview	13
CHAPTER 2 METHODOLOGY	14
2.1 Subjects	14

2.2 Materials	15
2.3 Recording.....	17
2.4 Measurements	18
CHAPTER 3 RESULTS	20
3.1 Realization of Focus	20
3.1.1 F0	21
3.1.2 Pitch Range	27
3.1.3 Duration	28
3.1.4 Summary	29
3.2 Realization of lexical tones	29
3.2.1 F0	31
3.2.2 Slope	37
3.2.3 Summary	38
3.3 Comparison between L2 subjects and native speakers	38
3.3.1 L2 vs Native: Realization of focus.....	39
3.3.2 L2 vs Native: Realization of focus over four lexical tones.....	43
3.3.3 L2 vs Native: Realization of lexical tones	48
3.3.4 Summary	54
3.4 Comparison between subjects.....	55
3.4.1 Individual subject vs. Native speaker: Realization of focus	55
3.4.2 Individual subject vs. Native speaker: Slope	58
3.5 Summary	62
CHAPTER 4 DISCUSSIONS.....	64

4.1 Answers to research questions	64
4.1.1 Research question 1	64
4.1.2 Research question 2	65
4.1.3 Research question 3	71
4.2 The nature of English focus	73
4.3 Pedagogical implication on L2 instruction	74
CHAPTER 5 CONCLUSIONS AND FUTURE DIRECTIONS	76
5.1 Overall conclusions.....	76
5.2 Limitations of the current study and directions for future study	77
REFERENCES	80

LIST OF TABLES

Table 1. Lexical tones of Mandarin Chinese in various representations	5
Table 2-1. Words with varying lexical tones used as recording materials	15
Table 2-2. Example of precursor questions	16
Table 3-1. Realization of Focus	25
Table 3-2. Pitch ranges in different focus conditions	27
Table 3-3. Durations in different focus conditions	28
Table 3-4. Realization of lexical tones reflected in maximum and minimum f0 of each syllable	34
Table 3-5. Slope of lexical tones in four focus conditions	38
Table 3-6. L2 vs Native: Realization of focus and lexical tones	54
Table 3-7. Average slope of S3	60

LIST OF FIGURES

Figure 1. An example of segmentation of the second and third nucleus in Praat	19
Figure 2. Realization of focus produced by L2 shown in f0 curves	22
Figure 3. Realization of lexical tones of word 1 (the second syllable)	31
Figure 4. Realization of lexical tones of word 2 (the third syllable)	32
Figure 5. L2 versus native speakers: Extent of variation in duration under various focus conditions	40
Figure 6. L2 versus native speakers: Extent of variation in pitch range under various focus conditions	41
Figure 7. L2 versus native speakers: Extent of variation in maximum f0 value under various focus conditions	42
Figure 8. L2 versus native speakers: Extent of variation in minimum f0 value between various focus conditions	42
Figure 9. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under focus condition compared to neutral condition	43
Figure 10. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under pre-focus condition compared to neutral condition	45
Figure 11. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under post-focus condition compared to neutral condition	45
Figure 12. Difference in the extent of variation between native speakers and L2	47
Figure 13. L2 vs Native: F0 curves of the second syllable in the neutral condition	49
Figure 14. L2 vs Native: F0 curves of the third syllable in the neutral condition	50

Figure 15. L2 vs Native: Mean f0-max and f0-min of four lexical tones in neutral condition	52
Figure 16. L2 vs Native: Mean f0-max and f0-min of four lexical tones under focus	53
Figure 17. L2 vs Native: Mean f0-max and f0-min of four lexical tones in pre-focus condition .	53
Figure 18. L2 vs Native: Mean f0-max and f0-min of four lexical tones in post-focus condition	54
Figure 19. Comparison between subjects: Variation in duration under varying focus condition..	57
Figure 20. Comparison between subjects: Variation in pitch range under varying focus conditions	57
Figure 21. Comparison between subjects: Variation in f0-max under varying focus conditions..	58
Figure 22. Comparison between subjects: Variation in f0-min under varying focus conditions..	58
Figure 23. Slope of word 1 (the second syllable) produced by S3	60
Figure 24. Slope of word 2 (the third syllable) produced by S3	61
Figure 25. Realization of lexical tones of the word 2 by S3	62

CHAPTER 1

INTRODUCTION

1.1 Problem and hypothesis

Focus intonation is a communicative function used to emphasize semantic prominence of a sentence (Ladd, 1980; Xu, 2005; etc.). It is acoustically manifested through variations of fundamental frequency (f_0) (Johnson, 2012). Since f_0 carries not only focus, but also lexical tones, it has been an interesting topic to explore how focus and lexical tones are realized simultaneously in tonal languages such as Mandarin Chinese (MC). Numerous theoretical models (Chao, 1968; Garding, 1987; Xu, 2001; etc.) have been proposed from various perspectives with different specific predictions. Two general consensuses they have reached are: expansion of pitch range and increase of duration mainly characterize the effect of focus; the identity of the lexical tone under focus should be preserved.

As interest in Mandarin intonation, such as focus, has grown rapidly in recent years, few studies have shed light on the second language (L2) acquisition of the intonation of MC. Considering the importance of suprasegmental features in the learning and use of L2, the present study aims to describe and analyze how L2 learners of MC realize one type of intonation, focus, based on their production. Not only will it potentially fill in the gap in this area of L2 study, but also provide linguistic reference for L2 instruction of focus intonation.

Inevitably, this study includes examination of lexical tones produced by the L2 learners. On the one hand, as mentioned earlier, lexical tones of MC are supposed to closely interplay with focus intonation: they manipulate f_0 movement simultaneously. Therefore, a comprehensive analysis of focus needs to take a look at realization of lexical tones. Specifically, it is crucial to check whether tone identities are preserved under the influence of focus. On the other hand, given the fact that lexical tones of MC have been a big challenge (Ke, 2012; Zhang, 2013) to the English L2 learners of MC, who come from a non-tonal L1 background, examination on realization of four lexical tones under the influence of focus might shed light on L2 acquisition of lexical tones.

After describing how focus is realized in L2 production, together with lexical tones, this study compares the current result with that of the native speakers collected by Xu (Xu, 1999), which explicitly shows how the production of L2 learners deviates from the production of the native speakers. Therefore, the study not only shows how the L2 learners produced focus, more importantly, it shows whether they were doing it right. The observed deviations from the native speakers reveal challenges in learning prosody of MC for L2 learners.

L1 transfer of English focus intonation is expected to affect realization of Mandarin focus produced by English L2 learners. This potential negative transfer provides a second theoretical angle to this thesis. Traditionally, English focus is believed to be a pitch accent (Pierrehumbert, 1980; Ladd, 1996; etc.), which works differently from Mandarin focus intonation. However, a controversy over the nature of English focus has been raised by Xu (Xu, 2005), where he suggests that focus in English is not a simple pitch accent but a pitch range manipulation that works fundamentally the same as the Mandarin focus over the lexical tones: expanding the pitch range, adjusting maximum and minimum pitch values, etc. The second purpose of this study,

therefore, is to serve as a possible solution to the controversy based on performance of the English L2 learners of MC: If any typical English pitch accent in the focus condition could be found in the interlanguage, which is clearly a sign of L1 transfer of English prosody, then Xu's conclusion is wrong. Because based on his idea, English L2 learners of Mandarin should produce focus just as the native speakers do without any negative transfer effect.

In sum, the purpose of this thesis is twofold: By looking at the interlanguage of English L2 learners when producing Mandarin focus, one could: (1) find out how English L2 learners of MC realize Mandarin focus; (2) to reveal whether the nature of English focus is a pitch accent or not. The first is a standard comparative analysis between L2 and the target language (Mandarin Chinese). The second is an innovative analytic method that using interlanguage data to examine properties of L1 (English) of the L2 speakers.

In order to find out the answers, specific research questions are raised as follows:

1) Will the English L2 learners of Mandarin be able to produce focus (not necessarily native-like) under the influence of lexical tones? How will they do that?

2) How might the production of the L2 learners deviate from that of native speakers? Are the deviations due to different ways of realizing focus?

3) How will each lexical tone be realized with the effect of focus? Will different lexical tones bring different levels of difficulty to the L2 learners in the production of focus?

The respective hypotheses are:

- 1) Yes, they will be able to realize focus by producing an English pitch accent, which has wider pitch range (especially higher pitch peak) and longer duration. They will also shorten the duration and compress the pitch range of the syllable after the focus.
- 2) First and most importantly, they wouldn't preserve the tone identities while realizing

focus. More specifically, they will substitute lexical tones of MC with a pitch accent, which is transferred from intonation of L1. Secondly, their production of lexical tones in the neutral condition may not be perfectly native-like.

- 3) Most substitution errors will be found in the High and Rising tones, because they both end in a high pitch level, unlike the final low pitch level of a pitch accent. The Falling tone is the easiest to produce under focus, because it is similar to the pitch accent in terms of f0 contour. The Low tone, with a low dipping contour, brings a certain difficulty but less than the High and Rising tones.

The main findings of this thesis are: on average, the L2 subjects managed to realize focus without a pitch accent substitution over lexical tones. However, one particular subject failed to preserve tone identities by producing a pitch accent under focus, which proves Xu is wrong in terms of the nature of English focus intonation. Excluding the substitution error found in this one subject, based on averaged results, the pitch range, f0-max and f0-min in the production of the L2 learners varied in a much less extent compared to those of the native speakers between varying focus conditions. The L2 learners had most difficulty in producing the L tone, which is almost the same as the R tone in their production regardless of the focus condition.

1.2 Lexical tones

1.2.1 Lexical tones in Mandarin Chinese

Mandarin Chinese is a tonal language. Four lexical tones are employed to distinguish meanings of otherwise homophonous syllables. An additional neutral tone, which mostly occurs

in unstressed syllables and bears no pitch contour, is not discussed in this thesis and avoided in the experiment design. The acoustic carrier of lexical tones is f0.

Multiple systems of tonal representation of MC have been developed and adopted for various purposes. Some of them graphically display patterns of pitch contour, focusing on the movement of the tones; some others employ numerals to indicate the height of the tones. “This practice implies that tone consists of two independent orthogonal dimensions: register and pitch contour” (Chen, 2000). Tone diacritics of Pinyin, which are officially used in Mainland China and commonly employed in textbooks of second language acquisition (SLA) of MC, belong to the first category, as shown in the leftmost column of Table 1-1. A numeric system created by Chao (1930, 1968), has been widely used in Chinese linguistic study, where a scale of 1-5 represents the relevant pitch levels of a speaker, with 1 as the lowest and 5 the highest. In this thesis, the lexical tones are descriptively represented as the High, Rising, Low and Falling tones (abbreviated as the H, R, L and F tones), following the practice of Xu for ease of comparison.

Table 1. Lexical tones of Mandarin Chinese in various representations

Graphic Notation (Pinyin)	Numeric Notation (Chao)	Description (the present study)	Gloss
mā	55	High tone (H)	“mother”
má	35	Rising tone (R)	“hemp”
mǎ	214/21	Low tone (L)	“horse”
mà	51	Falling tone (F)	“scold”

When pronounced in isolated and stressed syllables, the H tone is a level tone with relatively high pitch level; the R tone starts from a middle pitch range and rises to the highest level, forming a rising contour shape; the L tone starts rather low, falls to the lowest pitch level

and then rises to a mid-high range, resulting in a “dipping” shape (In fact, the final raising is only produced in stressed, isolated and phrase-final positions. In most cases, the L tone is realized as a low tone, “21”); the F tone starts from the top and drastically declines to the bottom of the pitch range. Among the four lexical tones, the H, R and F tones start from a higher pitch range, hence in tonal categorization models that feature tone registers, they are characterized with a [+upper] feature (Yip 1980, 2001). As far as the target of the tonal movement, the H and R tones reach to a high pitch level, whereas the targets of the L and F tones are both relatively low.

Tone sandhi is an intensely discussed aspect of Mandarin tonal phonology (Chao, 1968; Chen, 2000; Wang, 2006; Zhang, 2103, among others). The experiment of this thesis was designed to exclude the effect of tone sandhi, in order to focus on the realization of focus intonation and simplify interactions between lexical tones themselves. Thus, detailed introduction of tone sandhi is omitted here.

1.2.2 Second language acquisition of Mandarin lexical tones

Studies have shown that the acquisition of MC lexical tones brings comparatively more challenge to the English L2 learners, who come from a non-tonal L1 background, than the segmental inventories (Miracle, 1989, Wang 2006, among others).

Previous studies mostly rely on tonal production or perception errors of L2 learners, mostly adults. As summarized by Sun (1998), some recurring themes include “(1) the establishment of an acquisition order among the four lexical tones; (2) the effect of a speaker’s pitch range on tone production; (3) the transfer of English suprasegmental features in tone perception and production; (4) the difference between native and non-native speakers of Mandarin in their categorical processing of tone; (5) the relationship between the perception and production of tone;

and (6) the ultimate phonological attainment by adult learners of tonal languages” (Sun, 1998, p. 6-17). This thesis is closely related to subject (3) but will shed light on subject (1) and (4) as well, so the results should be of theoretical and practical interest for the SLA studies of tones in MC.

Though differing in experiment design and subject selection, these studies have reached some similar findings, such as: (1) learners have most difficulty contrasting the R and L tones in the production for their intrinsic phonetic similarities (Elliot, 1991; Miracle, 1989; Sun, 1998; Jin, 1996; etc.) (2) “the similar and different uses of pitch in English and Mandarin have been posited as a source of transfer in language learners” (Sun, 1998, p.12), and the transfer might be negative as well as positive. (3) The position where the lexical tone occurs is considered a crucial variable in the perception and production of MC lexical tones (Miracle, 1989; Lu, 1992; Sun, 1998, among others). (4) Universal phonological constraints, together with L1 transfer, significantly affect the production of L2 learners in terms of lexical tones, as proposed in Zhang (2013).

1.3 Focus

1.3.1 Focus

Though a strict and unified definition of focus hasn’t been well established in the literature, generally speaking, focus refers to a special emphasis or prominence given to a certain constituent of a sentence, and it is usually determined by pragmatic or discourse contexts instead of requirement of the grammar. The focused parts normally provide new, special or contrastive information (Bolinger, 1958, etc), as illustrated in the conversation below:

(a) I remember that apple is your favorite fruit.

(b) No, mango is my favorite.

In this case, the speaker of sentence (b) intends to contrast “mango” with other possible alternatives to the constituent, such as “apple” or “watermelon”. The focus is thus placed on the word “mango”, which carries information not shared by speaker (a). On the contrary, the rest of the sentence, sometimes referred as “presupposition” (Jackendoff, 1972), are shared by both the speaker and the hearer with no necessity to emphasize.

The focus described above is placed on an individual word “mango”, thus considered a “narrow focus”, contrasting with a notion of “broad focus”, which is firstly proposed by Ladd (1980). “Broad focus” is the focus on whole constituents or sentences instead of individual words. Broad focus normally occurs when the speaker believes that the whole constituent or sentence is new information to the hearer. This thesis only discusses narrow focus, thus the term “focus” only refers to narrow focus without further clarification.

Focus differs from stress. The former is on a sentence or post-lexical level, determined by pragmatic contexts, whereas the latter “belongs to the phonological component of the grammar and is responsible for the manifestation of focus” (Jin, 1996). Because of the existence of broad focus and phenomena like “contextual deaccenting” (Ladd, 1980), focus and stress do not have a straightforward one-to-one relationship, where the focused constituent is stressed and the stressed component is the focus of the sentence. This idea has been proved by studies of Halliday (1967), Chomsky (1970), Schmerling (1976), Ladd (1980), etc.

Since the relevant languages for the present study are English, the first language (L1) of the subjects, and MC, the second language (L2) or the target language of the subjects, previous results on focus realization in these two languages will be introduced in the following two sections.

1.3.2 Focus in English

Besides Xu (2005), previous studies have reached a general consensus on the realization of focus in English: “presence of pitch accents on focused constituents and absence of pitch accents on post-focal constituents”, as summarized in Chen & Gussenhoven (2008).

Pitch accent, first introduced by Bolinger (1958), is defined as “a local feature of pitch contour – usually but not invariably a pitch change and often involving a local maximum and minimum – which signals that the syllable with which it is associated is prominent in the utterance” (Ladd, 1996). In the British nuclear tone tradition, or the IPO model, the focus is mainly characterized as a High Fall nucleus, with a following nuclear tail and a preceding head (Cruttenden, 1997; O’Connor & Arnold, 1961). Based on the framework of the Autosegmental-Metrical (AM) theory, the focus of a declarative sentence in English contains H* (or L+H*) pitch accent followed by a L- phrase accent and an L% boundary tone, as summarized in Xu (2005).

As mentioned before, Xu (2005) reevaluated and disagreed with the prevailing view of seeing English focus as a pitch accent. He believes that the British nucleus tradition and the AM theory characterize focus of English primarily in *form*, rather than an intonational *function*, which he proposes is the nature of English focus and is fundamentally the same as focus of MC. A brief summary of his argument will be given in section 1.4.1.

For English, focus is directly associated with the surface variation of f0 contour, compared to MC where local lexical tones affect the variation as well as focus. From the perspective of f0 movement, it is observed that in a short declarative sentence, “a narrow focus is realized by expanding the pitch range of the on-focus stressed syllables, suppressing the pitch range of post-focus syllables, and leaving the pitch range of pre-focus syllables largely intact” (Xu, 2005).

1.3.3 Focus in Mandarin

Since the acoustic manifestation of focus is the manipulation of f_0 , which in Mandarin also carries the contrast between four lexical tones, the surface f_0 curve is not able to show implementation of focus directly in MC. Therefore, the realization of focus in MC is always explored by investigating its interaction with lexical tones.

One of the pioneer studies on the interaction between intonation (including focus) and lexical tones is by Chao (1968). He describes the relationship between MC lexical tones and intonations as “small ripples riding on big waves”: If the lexical tone contours glide in the same pitch direction as the global intonation, they would be intensified. On the contrary, when the directions differ, the two counteract with each other.

A well-known “Range Grid” model proposed by Garding (1987) is considered the first to formalize variation of pitch range during the realization of lexical tones under varying intonational contexts. In this model, focus is realized as expansion and compression of the grid, which affects the pitch range of the lexical tones. The notion of variation of pitch range introduced by focus has been widely adopted in following studies, including this thesis.

Some recent models are reviewed and tested by Chen & Gussenhoven (2008), including the PENTA (parallel encoding and target approximation) model, proposed by Xu (2001) and a physiologically motivated model, the Soft TEMplate Mark-up Language, proposed in Kochanski and Shih (2003). These models, together with the study of Chen & Gussenhoven, all confirm the substantial expansion of pitch range and longer duration of the syllable under focus. However, Chen & Gussenhoven argued that the previous models failed to predict several details of the focus effect, including the fact that “the expansion of f_0 range was tone-intrinsic” and the expansion over the F tone is larger than the H and R tones.

Besides expansion of pitch range under focus, several studies discovered the abrupt fall of pitch value in the syllable immediately following the focused word (Liao, 1994; Tseng, 1981; Garding, 1987; Xu, 1999; among others).

Some other acoustic parameters, such as duration and intensity, have been examined as well. The duration of the syllable under focus is significantly increased (Shih, 1988; Liao, 1994; Shen, 1992; Jin, 1996; Xu, 1999; among others), however, it has been found that “intensity is not a reliable cue for stress” in MC (Jin, 1996; among others).

One crucial fact about the interaction between focus intonation and lexical tones in MC is that the identities of the four lexical tones are preserved (Ladd, 1996; Flemming, 2008; Zhang, 2013): “a focus neither replaces the tone directly under it, nor eliminates the tones after it” (Xu, 2005).

1.3.4 Second language acquisition of focus in Mandarin

Compared to SLA studies regarding segmental inventories, lexical tones and tone sandhi rules of MC, the amount of SLA studies on the perception and production of L2 intonation is extremely small, not to mention those specifically designed for the acquisition of focus. This situation is aligned with lack of intonation or prosody instruction in language classes of MC. In fact, studies on perception and production of L2 intonation in general, not just about MC, are far from enough. According to a survey, only 9 studies were found to investigate intonation in all the journals of SLA study of past 25 years (Mennen, 2006).

The most recent and in-depth study on English L2 production of Mandarin focus is found in the dissertation of Zhang (2013), where she explored the influence from L1 prosodic structures

of English, Japanese and Korean on the L2 production of sentence-level prominence (narrow focus), and demonstrated a significant effect from universal phonological constraints as well.

Her study differs from the current one mainly in that Zhang focused on acquisition of lexical tones, therefore focus is only introduced to induce error patterns in L2 production of lexical tones. However, some of her observations proved to be extremely helpful in analyzing and explaining results of the current study.

1.4 Studies of Xu

1.4.1 Xu (2005) on English focus

Xu studied phonetic realization of English focus in short declarative sentences by conducting detailed measurement over the f0 curves under various focus conditions (without focus or focus placed in different positions). He argues that focus in English should be seen as a function that manipulates the pitch range of the f0 curve, instead of a form, like a High-Fall nucleus tone, or a H* pitch accent, that is “anchored” onto the f0 curve regardless of the local f0 shape in the neutral condition. He disagrees that the nucleus pitch accent will be the only intonational component once placed onto a word or syllable. The crucial evidence he found to support the idea is there is “local f0 movement largely independent of focus” and focus has a “three-zone pitch range manipulation” pattern that is independent of other factors as well. The so-called “local f0 movement” is mainly characterized by the position of the f0 peak, which Xu believes is not determined by focus.

In the study, Xu compares the phonetic realization of focus in English with that of Mandarin multiple times and suggests both languages realize focus in a “fundamentally similar” pattern. If this is the case, then for English L2 learners of Mandarin, there should be no deviation in terms

of realizing Mandarin focus intonation. This prediction largely inspired the current study.

1.4.2 Xu (1999) on Mandarin focus and lexical tone

Xu (1999) conducted a detailed analysis over the effects of tone and focus on the formation and alignment of F0 contour based on native production of Mandarin. His main conclusion is that “lexical tones and focus are implemented in parallel. More specifically, tone identities are implemented as local F0 contours while focus patterns are implemented as pitch range variations imposed on different regions of an utterance” (Xu, 1999).

The experiment design, methodology and main results of this study serve as reference for this thesis, which is required by a comparative analysis between the production of L2 and native speakers.

1.5 Summary and overview

This thesis is organized as follows: Chapter 2 explains the methodology of the experiment. Then in Chapter 3, the results on the production of the L2 subjects and the comparison between the L2 subjects and the native speakers will be presented in detail. In Chapter 4, discussion based on the results will be made, both answering the research questions and further investigating the problems found in the results. Finally, Chapter 5 presents overall conclusions of the study, as well as points out directions for future studies.

CHAPTER 2

METHODOLOGY

This study not only aims to do descriptive analysis over production of L2 speakers, but also focuses on comparative analysis between current L2 data and data of native speakers from Xu (Xu, 1999). For the purpose of that, experiment design of the present study is largely inspired by and intentionally follows the study of Xu (1999) on native speakers. The biggest change is made over recording materials: only two instead of three syllables carry varying focus conditions, considering the feasibility of the study and the fact that the change will not affect experiment results significantly given the purpose of the present study.

This chapter introduces details of subjects, materials, recording procedures, measurements of the study.

2.1 Subjects

Eight native speakers of American English, four males and four females, aged between 18 to 21, participated in this study. All of the subjects speak American English as their first language and are monolingual.

All subjects are undergraduate students enrolled in 2014 Spring semester of University of North Carolina at Chapel Hill, taking the Chinese language course (CHIN 201 or 204) in the Department of Asian Studies.

They have been studying Chinese for no less than 2 years and have learned lexical tones of Mandarin since the beginning of their study of the language. The language course they were taking (CHIN 201 or 204) is considered intermediate level, and the primary textbook they were using is Integrated Chinese (Level 2 Part 1) by T. Yao et al (2005).

All subjects participated in the experiment voluntarily.

2.2 Materials

Sentences used in the experiment are composed of three words. The first word “mao mi” is disyllabic: “mao” with H tone is used as the first syllable consistently, and “mi” with four alternating lexical tones: H, R, L and F, is the second syllable. Some combinations are nonsense words created solely for the purpose of the study. The second word is monosyllabic, alternating between “mo” (H tone), “na” (R tone) and “mai” (F tone). The third word, “mao mi” is disyllabic and both syllables are produced with H tone consistently. Therefore, only the lexical tones of the second and third syllables alternate, and are under direct examination for the study. Table I below shows all the words in Pinyin with English translations used in the study. In total, $4 \times 3 = 12$ sentences with all possible combinations of syllables are produced by each subject.

Table 2-1. Words with varying lexical tones used as recording materials

Word 1	Word 2	Word 3
HH māomī ‘kitty’	H mō ‘touches’	HH māomī ‘kitty’
HR māomí ‘cat-fan’	R ná ‘takes’	
HL māomǐ ‘cat-rice’	F mài ‘sells’	
HF māomì ‘cat-honey’		

In order to study the realization of focus, these 12 sentences with different focus conditions are elicited using precursor questions asked by the researcher, as well as hints in reading materials for the subjects. The critical focus conditions include: (1) neutral, which means no focus is placed on any word, (2) focus on word 1 (the second syllable), and (3) focus on word 2 (the third syllable). For each sentence, different precursor questions are used to elicit target sentences with particular focus conditions. Table 2-2 lists precursor questions in English translations for the sentence “mao (H) mi (H) mo (H) mao (H) mi (H)”. All together, $12 \times 3 = 36$ individual sentences with different combinations of lexical tones and focus conditions are produced by each subject.

Table 2-1. Examples of precursor questions

Precursor question (English translation)	Target sentence (Pinyin)
“What is kitty doing?”	māomī mō māomī (neutral)
“Who is touching kitty?”	<u>māomī</u> mō māomī (focus on word 1)
“What is kitty doing to kitty?”	māomī <u>mō</u> māomī (focus on word 2)

Each sentence will be repeated five times randomly by each subject, therefore, the number of sentences recorded for this study is $36 \times 8 \times 5 = 1440$.

During the experiment, the subjects will see each target sentence present in Pinyin on a digital tablet individually and randomly. The lexical tones are shown in numeric number following each syllable for the ease of reading and differentiating between alternating tones (“1” for the high tone, “2” for the rising tone, “3” for the low tone and “4” for the falling tone). The

word under focus is underlined and present in the red color. A sample sentence is given below to illustrate what subjects would see during the experiment:

103. mao1 mi4 na2 mao1mi1

In Xu (1999), the third word alternates between “mao (H) mi (H)” and “ma (L) dao (H)”, and additional focus condition is placed on the third word. This is the only difference between this study and his in terms of the recording materials. This change significantly reduces the amount of data and still makes it possible to examine the interaction between focus and lexical tones. For the ease of direct comparison, the number of subjects, the choice of words, the precursor questions and the number of repeats, all follow Xu (1999).

2.3 Recording

The recording was conducted in the soundproof phonetic lab located in Room 103 of Dey Hall at the University of North Carolina at Chapel Hill.

Before recording, the subjects did a questionnaire on their age, how long they have been learning Mandarin Chinese and the highest level of Chinese class they have taken at the university. Then a short instruction on the experiment is given. The instruction is composed of two parts: First, the researcher showed each syllable in Pinyin with their meanings and how they are combined to form three words. The subjects are asked to read the individual words and sample sentences with alternating tones in order to get familiar with the recording materials. Secondly, the researcher explained what focus is and the three critical focus conditions for each sentence. The corresponding precursor questions are also introduced and explained. Then the subjects practiced realizing focus in Mandarin. No instruction on how to produce focus in Mandarin was given.

The recording was made after the subjects were familiar with the sample sentences and feeling comfortable producing focus according to the precursor questions and hints.

During the recording, a microphone-headset and a MacBook Air laptop with Version 10.9.3 of OS X were used, along with Praat of Version 5.3.63 in all default settings. The subject clicked the screen of an iPad 3 to view each individual sentence, and read it after the researcher asked the precursor question in Mandarin.

After the recording, the subjects received ten dollars as compensation.

2.4 Measurements

Critical measurements for the study include duration, pitch range, f0-max and f0-min of the second and third syllables. For the ease of visual inspection and comparison, time-normalized f0 curves are created as well.

Praat of Version 5.3.63 along with ProsodyPro (Xu, 2013), a Praat script written by Xu were used to segment the second and third syllables, and then generate duration, f0-max, f0-min and time-normalized f0 values, etc. for individual subject and across all subjects.

After running ProsodyPro in Praat, a window pops up and displays the waveform and spectrogram of the current sound together with optional pitch track and formant tracks in the spectrogram panel. Then the syllables were labeled according to both the waveforms and spectrograms. The syllables were considered to begin where the amplitude expanded significantly and abruptly and end where it contracted. The boundaries are rather clear according to the spectrograms as well: the formants darken significantly within the nucleus. Figure 1 below demonstrates how the nucleus of the syllables were segmented and labeled in Praat. Segmentation was conducted as consistently as possible.

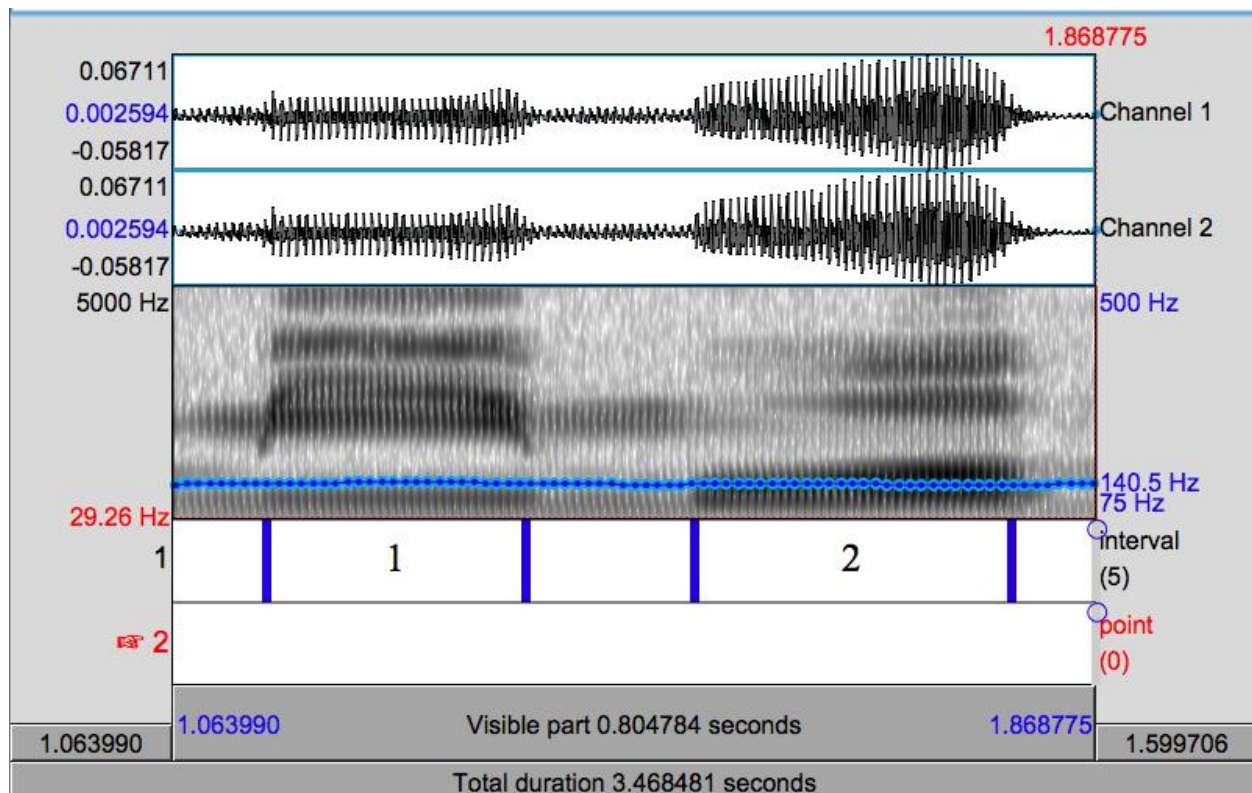


Figure 1. An example of segmentation of the second and third nucleus in Praat

After individual segmentation of each sound file was done, ProsodyPro was run again and by checking “Process all sounds without pause” and setting the repetition number as “5”, it automatically generated a series of measurements averaged across all repetitions for each identical sentence, including mean f0-max, mean f0-min, mean duration, mean normalized f0, etc. After measurements for each subject were collected, by running the script again with all relevant files place together in a fold, the measurements across all subjects were automatically generated.

For statistics analysis, both SPSS and SAS were used to conduct necessary evaluations and generate critical results. SAS mostly verified results generated from SPSS analysis.

CHAPTER 3

RESULTS

This chapter presents the result of the experiment: how the subjects realized focus and lexical tones in their production. In section 3.1, with lexical tones being the controlled factor, realization of focus is reflected in differences in f_0 , pitch range, and duration when focus and non-focus contexts are compared. In section 3.2, in order to find out how the lexical tones were realized under each focus condition, focus conditions are held consistent and the f_0 and slopes of the four lexical tones are examined. Both sections utilize graphic display of f_0 curves for direct visual inspection, and then statistic analysis is adopted for further examination and verification. Then, comparative results are shown in section 3.3 and 3.4. Section 3.3 gives comparison on mean results between English L2 subjects of the current study and the native speakers of Xu's study (Xu, 1999), showing how production of the L2 subjects deviated from that of the natives. After that, section 3.4 compares crucial measurements such as pitch range, f_0 -max and f_0 -min between individual subject and the native speakers, aiming to check if there's any idiosyncratic behavior significantly deviant from the mean values.

3.1 Realization of Focus

This section aims to answer the first research question: Are the English L2 subjects able to realize focus under the influence of lexical tones?

Since English is not a tonal language and it has been proved that lexical tones are extremely challenging to English L2 learners of Mandarin (Miracle, 1989, Wang 2006, among others), it is likely that under the pressure of producing accurate lexical tones of MC, the subjects might fail to make any significant adjustment in their production to suggest focus intonation or different positions of the focus. It should be noted that their production doesn't have to be correct, or native-like, in terms of focus realization. Any consistent pattern that significantly differs variant focus conditions from the neutral condition should be considered a sign of successful focus realization. Measurements that are most likely to reflect focus realization include maximum f0 value (f0-max), minimum f0 value (f0-min), pitch range and duration, because English focus involves adjustment over these measurements, as well as Mandarin focus.

3.1.1 F0

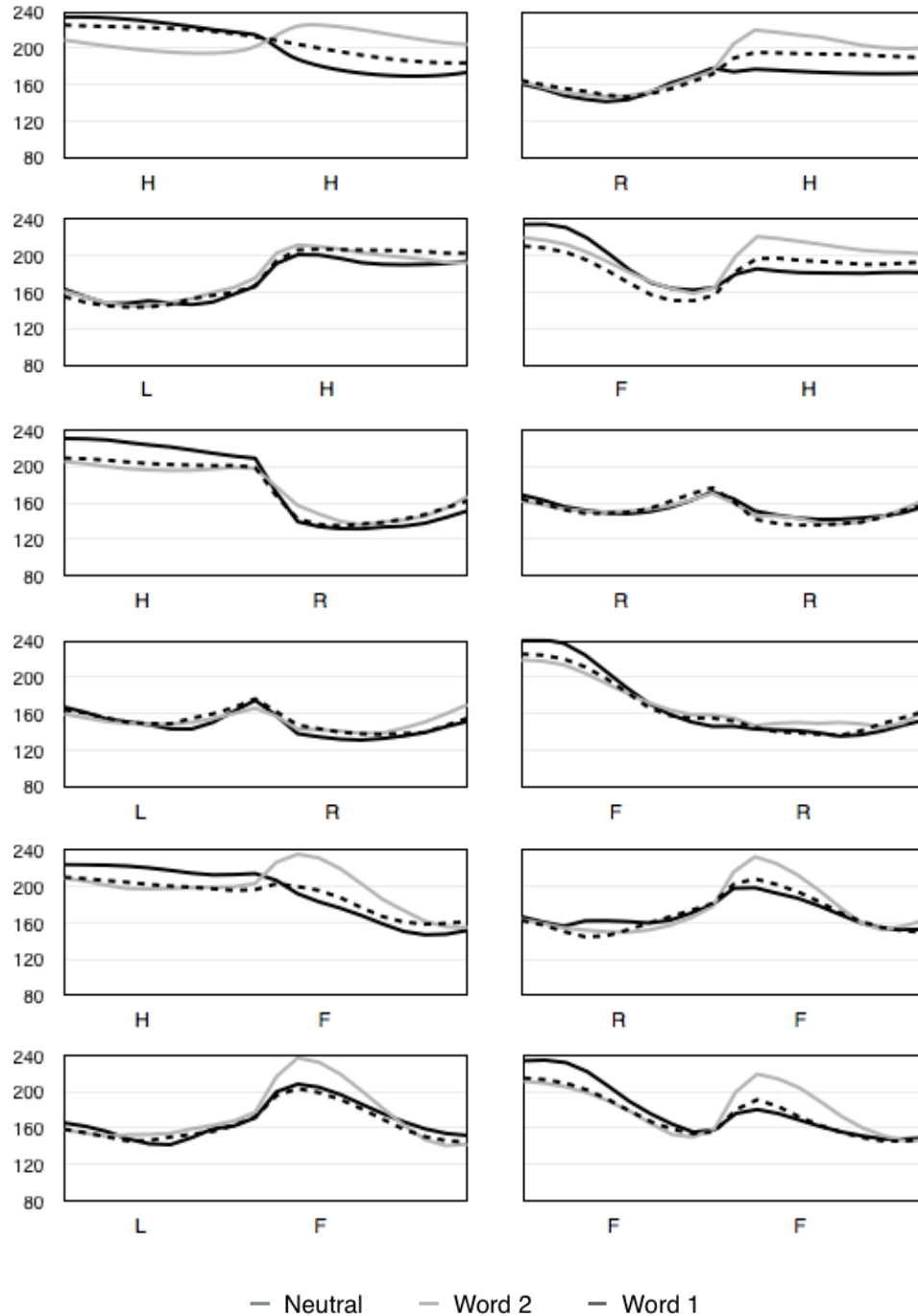
3.1.1.1 F0: visual observation over f0 curves

Focus, as an intonation type, is manifested in f0 contour. Visual observation over f0 curves provides a preliminary answer to the question of focus realization. For each identical sentence, there should be three f0 curves, each corresponding to a distinct focus condition: focus on word 1, focus on word 2, and neutral (no focus). If the f0 curves with focus, either on the first or second word, do not overlap with the one in the neutral condition, then it suggests that focus is somehow realized.

Figure 2 below has 12 panels that show the f0 curves of each test sentence averaged across all subjects. For each panel, the lexical tones are the same, thus only focus condition affects the contour of f0 curves. For the ease of discussion, each panel will be mentioned in terms of its

column and row. For example, panel C2R4 means the one on the second column and the fourth row.

Figure 2. Realization of focus produced by L2 shown in f0 curves. (Under each panel, the lexical tones for word 1 and word 2 are marked as H, R, L or F respectively. Lines in different shades and styles represent f0 curves in three distinct focus conditions: neutral (no focus), focus on word 1 and focus on word 2. Unit of the y-axis is Hz.)



The realization of focus can be most obviously seen in C1R1, where both syllables (as well as others that are not shown) bear the same lexical tone, the H tone, therefore the least interference from lexical tones is introduced. As can be seen in this particular panel, the three curves did not overlap: When focus is placed on the first word, f0 of the focused syllable is raised a bit and that of the post-focus syllable is lowered a lot. When focus is placed on the second word, f0 of the pre-focus syllable is somewhat lowered, while that of the focused syllable is substantially raised. It is clearly shown in this panel that the subjects managed to produce focus of Mandarin by adjusting the height and shape of f0 contour when lexical tones are least introduced in their speech and the pattern seems to be raising the f0 value of the focused syllable.

After checking on all 12 panels, it can be seen that focus is realized in most sentences (C2R3 and C1R4 are exceptions where all three curves almost overlapped). Based on visual inspection, some general patterns can be concluded: Compared to the neutral condition, (1) the f0 peaks are more or less raised when the syllable is under focus; (2) the height of the f0 curves right *after* focus is lowered in most cases, although the lowering is not very substantial (the solid black line over word 2 are mostly below the dotted lines, except in C2R3 and C1R6); (3) The height and shape of f0 curves *before* focus are barely changed in most panels (the grey lines over word 1 overlap with the dotted lines, except in C1R1, C2R2).

Besides general patterns, interesting details are observed as well. When looking at different lexical tones, it seems that the f0 curves representing three distinct focus conditions almost coincide in syllables that carry the R and L tones, such as word 1 in panel C2R1, word 1 in C1R2 and both syllables in C2R3. Contrastively, more substantial adjustment of f0 curves can be

observed over syllables bearing the H and F tones. Thus, it seems that focus is easier to realize when a syllable carries the H or F tone.

Another observation is that the f0 curves are more likely to overlap with each other in the first syllable shown in the panel regardless of the lexical tone and the focus condition. It is possible that focus on word 2 is easier to produce than on word 1. Another possibility is that the focus was being realized on the first syllable of word 1 and so is simply not measured here.

Overall, based on visual observation of f0 curves in Figure 2, there is no doubt that the English L2 subjects realized focus of MC in their production.

3.1.1.2 F0: statistic analysis

In order to verify and further analyze the patterns observed earlier, maximum and minimum f0 values of word 1 and word 2 in three different focus conditions were measured. In Table 3-1, the left four columns show mean maximum and minimum f0 values of tones in word 1 and word 2 as indicated in the headings of the rows. Differences in mean f0 values between different focus conditions are displayed in the right six columns, which are calculated the way suggested by the headings of the columns.

The purpose of Table 3-1 is to show whether focus significantly affects the f0 curves, including raising and lowering the curves in different focus conditions as observed in the panels of Figure 2. In order to prove statistic significance, two-tailed paired t-test with a probability level of 0.05 is adopted. In Table 3-1, if the probability value is less than 0.05, which suggests that the difference is significant, the number of that difference is marked in bold.

As can be seen in the table, all the maximum f0 values are higher when the word is produced under focus compared to neutral conditions, even though in Word 1 the maximum f0 of

the R tone is not statistically significant. The mean difference ranges from 12 to 35 Hz. This verifies the pattern of raising f0 curve under focus observed before.

Table 3-1. Realization of Focus. (In the left four columns, each number is the f0 value (maximum or minimum) of a particular tone, in a specific position (word 1 or 2), under a particular focus condition. “Pre” is short for “the pre-focus condition”, where the syllable is placed before the focus and “post” represents “the post-focus condition”, where the syllable occurs right after the focus. “Neutral” means no focus is placed on any word and “Focus” indicates that the syllable is produced under focus. Numbers in the remaining six columns suggest variation in f0 value between conditions. The unit of all the numbers is Hz. Significant numbers are marked as bold.)

Focus Effect										
	Neutral	Pre	Focus	Post	Focus - Neutral	Focus - Pre	Focus - Post	Pre - Neutral	Pre - Post	Neutral - Post
Word 1										
H Tone - Max	227	221	246		20 (p<0.05)	25 (p<0.05)		-6 (p=0.45)		
R Tone - Max	197	192	209		12 (p=0.09)	17 (p<0.05)		-5 (p=0.53)		
F Tone - Max	234	238	264		30 (p<0.001)	26 (p<0.05)		3 (p=0.67)		
R Tone - Min	144	149	141		-3 (p=0.71)	-8 (p=0.31)		5 (p=0.53)		
L Tone - Min	140	142	135		-4 (p=0.57)	-7 (p=0.34)		2 (p=0.7)		
F Tone - Min	151	149	145		-6 (p=0.44)	-4 (p=0.62)		-2 (p=0.78)		
Word 2										
H Tone - Max	215		239	208	24 (p<0.05)		31 (p<0.001)			7 (p=0.25)
R Tone - Max	182		198	174	17 (p<0.05)		25 (p<0.05)			8 (p=0.22)
F Tone - Max	225		260	219	35 (p<0.001)		41 (p<0.001)			6 (p=0.41)
R Tone - Min	130		129	133	-1 (p=0.85)		-4 (p=0.54)			-3 (p=0.68)
F Tone - Min	141		141	142	0 (p=0.98)		-1 (p=0.91)			-1 (p=0.92)

In contrast to overall substantial adjustment over mean maximum f0 value, the mean minimum f0 values under focus are not significantly lowered compared to those in neutral conditions, though the absolute values are often lower. This suggests that in the production of L2 subjects, focus is realized mainly by raising f0.

Mean maximum and minimum f0 values of words produced before focus are not significantly different from those in neutral conditions, as shown in the “Pre-Neutral” column, which shows that focus doesn’t affect the portion before the focused syllable.

No difference exists between f0 values under neutral and post-focus conditions. The mean maximum values are only marginally smaller in post-focus conditions, as shown in the “Neutral-Post” column.

Therefore, the visually observed general pattern is verified: under focus condition, the peak of f0 curve is substantially raised; f0 curves of the post- and pre-focus syllables are not significantly different from those in the neutral condition.

In terms of different lexical tones, it can be seen that the R and L tones show less significant difference between different focus conditions. In word 1, not all the maximum and minimum f0 values of the R tone are significantly adjusted under focus. In word 2, the maximum f0 value is significantly raised, however the minimum f0 value is not significantly lowered. For the L tone, no significant difference is made over the minimum f0 value when produced under focus. In contrast, the H and F tones in both word 1 and word 2 show significant difference over maximum f0 value between focus and neutral conditions.

The pattern of easier focus production in word 2 seems verified as well: The differences of mean maximum f0 values between focus and neutral conditions in word 2 are relatively bigger than those in word 1.

To sum up, statistic analysis verifies general patterns and detailed observations made from Figure 2. There is no doubt that the English L2 subjects successfully realized focus while producing lexical tones. They achieved focus intonation of MC by adjusting f0 values according to different focus conditions. Different lexical tones and focus positions seem to affect the realization as well.

3.1.2 Pitch Range

Significant adjustment over pitch range is another effective indicator of successful realization of focus. Having measured the f0-max and f0-min respectively, pitch range shows an overall adjustment over the f0 contour.

Table 3-2 below displays pitch ranges of both word 1 and word 2 in different focus conditions. Pitch range is the difference between the highest and the lowest f0 values. In the six columns on the right, the differences of pitch range between different focus conditions are computed the way suggested in the headings of the columns. Same as in Table 3-1, a two-tailed paired t-test is used to show statistic differences and a significance level of 0.05 is adopted.

Table 3-2. Pitch ranges in different focus conditions. (In the left four columns, each number is the pitch range in a specific position (word 1 or 2), under a particular focus condition. “Pre” is short for “the pre-focus condition”; “Post” represents “the post-focus condition”; “Neutral” means no focus is placed on any word and “Focus” indicates that the syllable is produced under focus. Numbers in the remaining six columns suggest variation in pitch range between conditions. The unit of all the numbers is Hz. Significant numbers are marked as bold.)

Pitch Range										
	Neutral	Pre	Focus	Post	Focus - Neutral	Focus - Pre	Focus - Post	Neutral - Pre	Neutral - Post	Pre - Post
Word 1	54	52	74		20 (p<0.001)	22 (p<0.001)		2 (p=0.61)		
Word 2	53		75	49	21 (p<0.001)		25 (p<0.001)		4 (p=0.29)	

It is clear that pitch range is significantly greater when the word is produced under focus than in the neutral condition. Both in word 1 and word 2, the difference is around 20Hz. (The difference in word 2 is slightly bigger.) It seems that pitch range in post-focus condition is somewhat lowered compared to that in the neutral condition, however, it is proven not statistically significant. Difference between pre-focus and neutral conditions is minimal, which again proves focus barely affects the syllable before it.

3.1.3 Duration

In both English and Mandarin, the focused constituent is produced with a longer duration. Even if lexical tones might bring difficulty in realizing focus for the English L2 subjects, the adjustment over duration is highly expected. Table 3-3 displays how different focus conditions affect the duration of the two syllables. Again, a two-tailed paired t-test was used to show statistic significance of the differences, with a probability level of 0.05. Significant differences are marked in bold.

Table 3-3. Durations in different focus conditions. (In the left four columns, each number is the duration of the syllable in a specific position (word 1 or 2), under a particular focus condition. “Pre” is short for “the pre-focus condition”; “Post” represents “the post-focus condition”; “Neutral” means no focus is placed on any word and “Focus” indicates that the syllable is produced under focus. Numbers in the remaining six columns suggest variation in duration between conditions. The unit of all the numbers is ms. Significant numbers are marked as bold.

Duration										
	Neutral	Pre	Focus	Post	Focus - Neutral	Focus - Pre	Focus - Post	Neutral - Pre	Neutral - Post	Pre - Post
Syllable 2 of Word 1	287	283	336		50 (p<0.001)	53 (p<0.001)		3 (p=0.67)		
Syllable 3 (Word 2)	292		367	274	75 (p<0.001)		92 (p<0.001)		17 (p<0.001)	

It can be seen that when the syllable is produced under focus condition, its duration is significantly greater compared to any other focus conditions. No significant difference is found in the duration of syllables produced in neutral and pre-focus conditions. Duration of the syllable right after focus decreases significantly compared to that in the neutral condition.

By looking at changes in durations under different focus conditions, it again proves that the English L2 subjects did successfully differentiate the focused constituents from the neutral ones.

3.1.4 Summary

Display of f0 curves, detailed measurements and statistic analysis of pitch range and duration suggest that in the production of L2 subjects, focus is realized mainly in two ways: First, the overall pitch range increases under focus, while it does not change significantly before and after focus. Secondly, duration of the syllable under focus is significantly longer than any other conditions. It can be primarily concluded that on average subjects successfully realized focus in their production of Mandarin during the experiment. Whether or not their way of producing focus is the same as that of native speakers will be discussed later in section 3.3 & 3.4.

3.2 Realization of lexical tones

Having proven that the English L2 subjects successfully realized focus, it is crucial to check whether they were producing the lexical tones of MC at the same time, as the native speakers do, because there is possibility that they put so much effort in emphasizing the focused constituents, most likely as pitch accent, that they failed to realize lexical tones of MC. This is where deviation is more likely to occur, considering difficulty of Mandarin lexical tones and negative

transfer effect from L1. The result might answer the second research question on the deviation in production.

Since the predicted error entails the L2 subjects having difficulty realizing lexical tones under focus, more specifically, by replacing at least some of the lexical tones with a pitch accent, straightforward visual observation of f_0 curves of the tones under varying focus conditions should be helpful. If each lexical tone were consistently associated with a distinct f_0 curve in *all* focus conditions, then it indicates Mandarin lexical tones were realized under focus. Consistency is crucial: any identifiable pattern of f_0 contour observed in the neutral condition should be found in focus conditions as well. Notice, again, the shape of the f_0 contours does not have to be exactly the same as that of the native speakers, what matters is whether consistency exists throughout all focus conditions. If the pattern observed in the neutral condition can no longer be preserved and identified under focus, for example, the distinct f_0 curves in the neutral condition tend to overlap and become hard to differentiate from each other under focus, then it might imply that the L2 subjects failed to realize lexical tones.

Slope of the tones is another good indicator of lexical tone realization. On average, in the neutral condition, slope of each lexical tone produced by native speakers should be different. Particularly, the slope of the L and F tones should be significantly lower than that of the H and R tones. Given the L2 proficiency level of the subjects, they should be able to at least differentiate the four tones in the neutral condition, therefore, if they successfully realized lexical tones, differences between slope of the lexical tones should be found in all focus conditions.

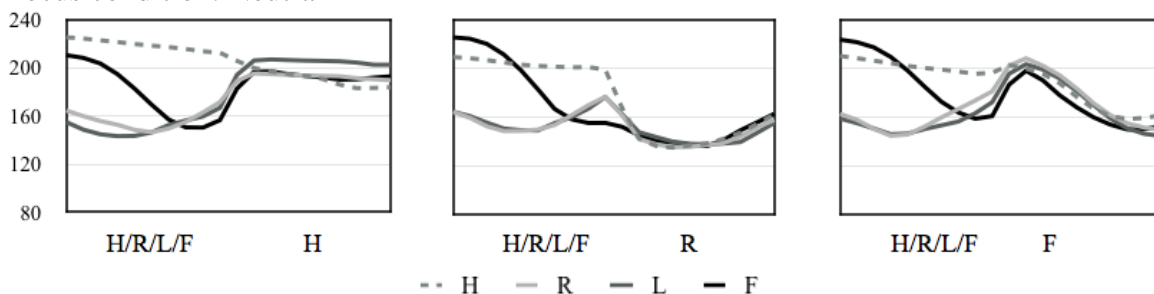
3.2.1 F0

3.2.1.1 F0: visual observation over f0 curves

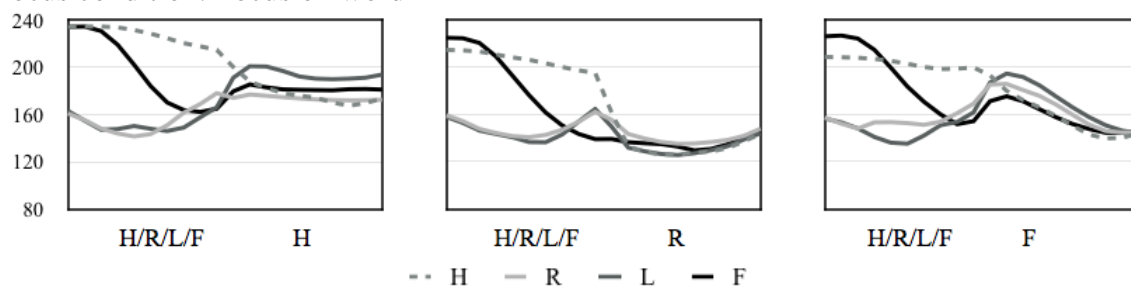
Each panel in Figure 3 below shows f0 curves of word 1 (the second syllable) when the following syllable carries the same lexical tone and the focus condition is held consistent. Panels on the same row share the same focus condition: neutral, focus on word 1 or focus on word 2. Within each row, each panel only differs in the lexical tone of the following word. Each f0 curve in word 1 corresponds to a lexical tone produced under the specific condition. In Figure 4, f0 curves of word 2 are under examination, therefore lexical tones of word 1 are controlled. Four f0 curves corresponding to the lexical tones were represented with lines in different shades and styles, as indicated in the legend.

Figure 3. Realization of lexical tones of word 1 (the second syllable). (Lexical tones are labeled as H, R, L or F right below the syllable. “H/R/L/F” represents varying lexical tones in the position. Focus conditions are listed above each row. Lines in different shades and styles represent f0 curves when word 1 carries varying tones. Unit of the y-axis is Hz.)

A. Focus condition: Neutral



B. Focus condition: Focus on word 1



C. Focus condition: Focus on word 2

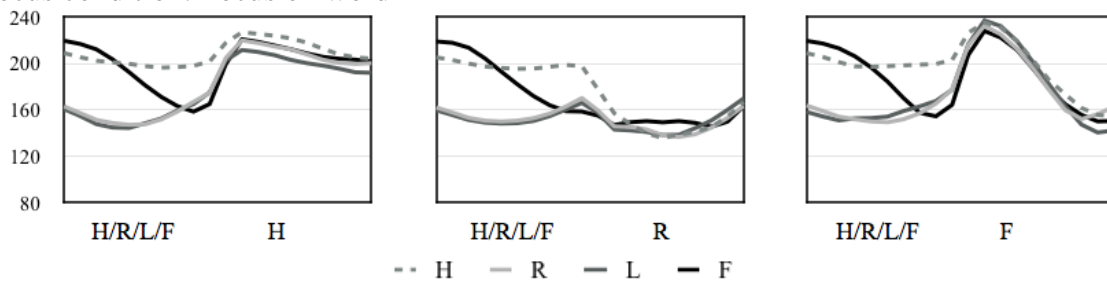
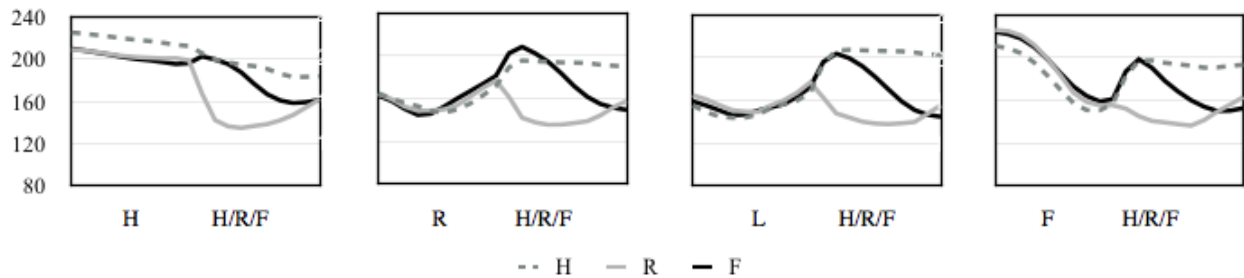
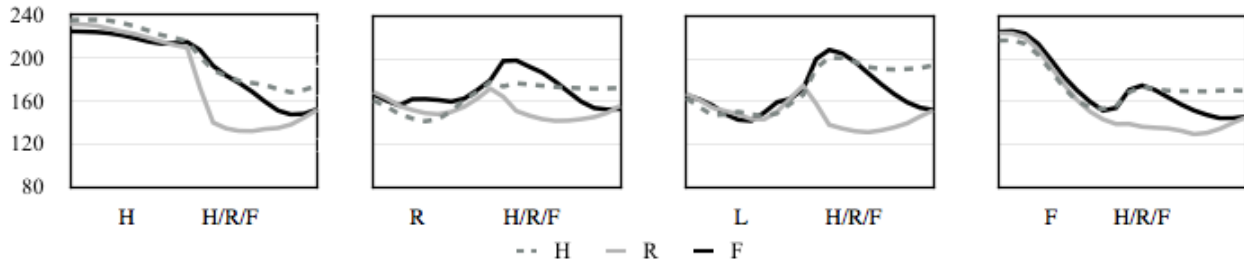


Figure 4. Realization of lexical tones of word 2 (the third syllable). (Lexical tones are labeled as H, R, L or F right below the syllable. “H/R/L/F” represents varying lexical tones in the position. Focus conditions are listed above each row. Lines in different shades and styles represent f0 curves when word 2 carries varying tones. Unit of the y-axis is Hz.)

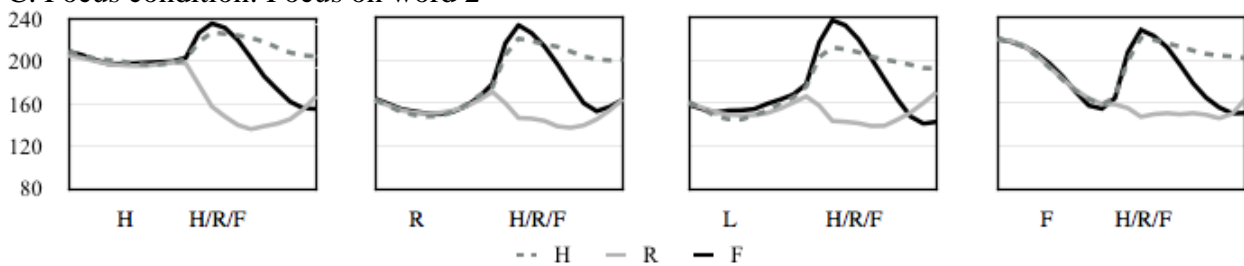
A. Focus condition: Neutral



B. Focus condition: Focus on word 1



C. Focus condition: Focus on word 2



Based on visual observation of Figure 3 and Figure 4, it is clearly shown that the subjects successfully realized lexical tones in all the focus conditions. In each panel, regardless of the focus condition, within the syllable with varying lexical tones, distinct f0 curves corresponding to each lexical tone can be observed consistently. The dotted grey lines, representing the H tone, remain in the high pitch range compared to the other lines, which shows that the H tone was successfully produced with the interference of focus. Another easily observed lexical tone under all focus conditions is the F tone. The black solid line representing the f0 curves of the F tone, always start from a high pitch level and fall rather drastically into a low pitch level, resulting in a typical ‘falling’ f0 contour of the F tone. F0 curves of the R and L tones are similar to the ones in Figure 2, where no obvious differentiation could be seen between the two. As shown in Figure 3, the second darkest solid line, representing the L tone, and the light grey line, representing the R tone, almost always overlapped in the panels. Both lines start from a mid-low pitch level, then more or less lowers in pitch height, and gradually raises to a mid pitch range. This particular observation indicates two facts: First, as mentioned in the previous section, the L2 subjects had difficulty differentiating the R and L tones in the production. Secondly, they managed to produce a rather consistent f0 contour shape for the tones (starting from a low pitch level, moving downward and then upward, forming a dipping “U” shape) which is obviously different from the others when produced under focus.

These facts all suggest that the L2 subjects managed to produce lexical tones, though possibly not in correct forms, because tone identities are consistently and clearly identified and preserved in all focus conditions.

3.2.1.2 F0: statistic analysis

By utilizing statistic tools, Table 3-4 below proves that in the production of the L2 subjects, lexical tones indeed significantly affect the formation of f0 curves. In other words, the distinct f0 curves observed above are caused by different lexical tones. The tone identities and the f0 contours are significantly associated with each other.

Table 3-4 has two main sections, A and B, separated by a horizontal thick black line. The upper section A shows mean maximum pitch values of syllable 2 (the second syllable of word 1) and syllable 3 (word 2) with varying lexical tones, under different focus conditions, as indicated by the headings, across all subjects. For example, number 227, in the leftmost cell on the first row in section A means the averaged maximum pitch value of the second syllable with H tone in the neutral condition is 227 Hz. Similarly, the lower B section shows mean minimum pitch values of these syllables with certain lexical tone in conditions suggested by the column and row headings.

Table 3-4. Realization of lexical tones reflected in maximum and minimum f0 of each syllable. (The vertical thick solid line separates results of different syllables. The horizontal one separates results of maximum and minimum values. Headings of the columns indicate focus conditions of each syllable, varying from “Neutral” (no focus), “W1” (focus is placed on word 1 or the second syllable) to “W2” (focus is on word 2, which is the third syllable of the sentence). Lexical tones of each syllable are indicated in the headings of the rows. Unit of the numbers is Hz.)

			Syllable 2			Syllable 3		
	Focus Conditions		Neutral	W1	W2	Neutral	W1	W2
A. F0-Max mean	Tone of Syllable 2	H	227	247	221	207	199	235
		R	197	209	192	208	201	234
		L	198	210	196	210	205	228
		F	234	264	237	204	196	233
	P=		0.046	0.004	0.016	0.476	0.528	0.046
	Tone of Syllable 3	H	213	230	213	216	208	239
		R	216	230	210	182	174	199
		F	213	238	212	225	219	260
	P=		0.425	0.402	0.691	0.011	0.004	0.012
B. F0-Min mean	Tone of Syllable 2	H	206	214	199	155	143	159
		R	144	139	149	153	155	157
		L	139	135	142	155	156	156
		F	151	145	149	152	149	159
	P=		0.034	0.029	0.029	0.915	0.584	0.475
	Tone of Syllable 3	H	161	160	160	191	179	204
		R	160	158	159	130	133	128
		F	160	159	159	141	142	141
	P=		0.928	0.868	0.650	0.007	0.014	0.061

In order to show the effect of lexical tones, or in other words, to show whether lexical tones significantly determined the maximum and minimum pitch values (which largely determines the shape of the f0 contours), a two-factors repeated ANOVA analysis was conducted in SPSS version 22. The lexical tones of syllable 2 and syllable 3 were considered independent variables. Significant probability values are shown at the bottoms of the corresponding columns. If the probability value is under 0.05, then it suggests that the lexical tone of the particular syllable significantly affects the maximum or minimum value of the f0 curve.

As can be seen from the table, regardless of the focus condition, the maximum and minimum f0 means of the second syllable are significantly affected by the lexical tones of the

same syllable. Likewise, for the third syllable, lexical tones significantly affect the maximum and minimum f0 means of the syllable in almost all situations, except when focus is placed on word 2 or the third syllable, the probability value is marginal.

Table 3-4 also suggests that lexical tones have greater impact on the change in maximum and minimum f0 values than the focus conditions.

When syllable 2 is under focus (in the column “W1”), the difference in maximum f0 between the F and R tones is the biggest, as much as 55 Hz, which means changing the tone identity of the syllable causes the largest variation in maximum f0 value of 55 Hz. For the third syllable (in the column “W2”), the difference is as much as 61 Hz between the F and R tones when the syllable was produced under focus. Similarly, the largest difference in minimum f0 value, 79 Hz, can be observed between the H and L tones in the second syllable, and for the third syllable, the number is 76 Hz, between the H and R tones.

The change in maximum and minimum f0 values brought by the effect of focus can be calculated in the same fashion: For the second syllable, the largest difference in maximum f0 can be seen in the F tone between focus on the second syllable and the neutral condition, as much as 30 Hz. (On the “F” row of the “Tone of Syllable 2” section, between “W1” and “Neutral”.) As for the third syllable, the biggest variation is 41 Hz between focus on “W2” and “W1” in the F tone (the “F” row of the “Tone of Syllable 3” section). In terms of the minimum f0 value, for the second syllable, the biggest difference, 15 Hz, is seen in the H tone between focus on the second and third syllable. The number is as much as 25 Hz for the third syllable.

Therefore, it seems that while lexical tones and focus both significantly affect the formation of f0 curves in the subjects’ production, when comparing the change in the maximum and minimum f0 values brought by these two factors, lexical tones have greater impact.

3.2.2 Slope

One effective method to determine whether the L2 subjects realized and preserved lexical tones under focus is to look at the values of slope. For the purpose of the current study, the slope is calculated by dividing the difference between the final and initial f0 values for each relevant syllable with the duration of the syllable. Since the measurements used in the calculation come from time-normalized f0 values averaged across all sentences and subjects, the duration can be seen as the same, or “1”. Therefore the value for the slope is essentially the difference between the initial and final f0 values.

Table 3-5 below shows slope of four lexical tones in four different focus conditions: neutral, under focus, pre-focus and post-focus. With a two-tailed paired t-test in SPSS and a significant probability value of 0.05, it is shown that the slope of the F tone is significantly negative, as expected. Marginally positive slopes of the R and L tones under focus indicate that the L2 subjects managed to produce these lexical tones instead of replacing them with a potential falling accent with significant negative slope. Even though the slope of the H tone is somewhat slightly negative under focus, it clearly differs from the F tone in absolute number. The lowering of the H tone under focus might suggest difficulty in preserving the tone while producing focus. The overall result concluded from Table 3-5 is consistent with those from Figure 3, Figure 4 and Table 3-4.

Table 3-5. Slope of lexical tones in four focus conditions. (The focus conditions and lexical tones are indicated in the headings. “H” “R” “L” and “F” represent “High tone” “Rising tone” “Low tone” and “Falling tone” respectively. The unit of the slope is Hz.)

	Neutral				Under Focus				Pre-focus				Post-focus		
	H	R	L	F	H	R	L	F	H	R	L	F	H	R	F
Slope	-6	4	13	-55	-11	5	3	-76	-7	11	13	-58	-5	-7	-44
P=	0.29	0.41	0.07	<0.001	0.06	0.35	0.64	<0.001	0.35	0.12	0.08	<0.001	0.41	0.23	<0.001

3.2.3 Summary

The subjects consistently produced lexical tones regardless of the focus conditions, though the H and F tones were comparatively more accurate than the R and L tones and little differentiation was made between the R and L tones. The crucial implication of these results is that tone identities were largely preserved in the ‘under focus’ condition without predicted substitution. In fact, lexical tones affected the subjects’ final production more substantially than the focus intonation.

3.3 Comparison between L2 subjects and native speakers

One of the research questions is to see how the subjects’ production deviated from that of the native speakers. It will be answered in this section with three sets of comparison. The first compares focus realization in terms of duration, pitch range, maximum and minimum f0 values. The second compares focus realization over each lexical tone. The last one focuses on realization of lexical tones only.

3.3.1 L2 vs Native: Realization of focus

It has been shown in section 3.1 that focus significantly affected the duration, pitch range, maximum and minimum f0 values in L2 production. By comparing these measurements in the current study to those reported in Xu's study (Xu, 1999), one could see how focus was realized differently in the L2 subjects' production. Since the absolute values of these measurements are not comparable given different subjects and experiment designs, it is better to look at the percentage of the extent of adjustment each focus condition brought to the measurements.

In order to compare the effect of focus, values of each measurement, including duration, pitch range, maximum and minimum f0 values, in neutral conditions are divided by the difference between varying focus conditions (including under focus, pre-focus, post-focus) and the neutral condition, which shows to what extent different focus conditions adjusted the neutral values of the measurements. Therefore, in Figure 5-8 and the following graphs, "Focus - Neutral (%)" "Pre-focus - Neutral (%)" and "Post-focus - Neutral (%)" mean the percentage of variation each focus condition brings to the particular measurements, with their values in neutral condition as reference. The explicit calculation process, take "Focus - Neutral (%)" as an example, is as follows:

"Focus - Neutral (%)" = (Value in focus condition - Value in neutral condition)/Value in neutral condition

The results shown in Figure 5, 6, 7 and 8 were averaged across all subjects, sentences and lexical tones. Data of native speakers were from Table III, IV and V of Xu(1999).

Figure 5 below shows variations in duration under different focus conditions, made by L2 subjects and the native speakers. It can be seen that both L2 subjects and the native speakers extensively elongated the duration of the syllable under the focus condition. Almost no

adjustment was made over duration in the pre-focus condition. In the post-focus condition, the duration of the syllable slightly shortened in both groups. Interestingly, it seems that slightly more variation over duration was made in the production of the L2 subjects, indicating that the L2 subjects had no difficulty realizing focus in terms of duration.

Figure 5. L2 versus native speakers: Extent of variation in duration under various focus conditions. The number of percentage each bar corresponds to shows the extent of increase (when the number is positive) or decrease (when the number is negative) of the duration under a specific focus condition compared to the duration in the neutral condition. The contrast between each focus condition with the neutral condition is indicated below each pair of comparison between L2 and native speakers. For example, “Focus – Neutral (%)” means to the extent of variation in duration under focus condition compared to neutral. Numbers of the y-axis represent percentage.)

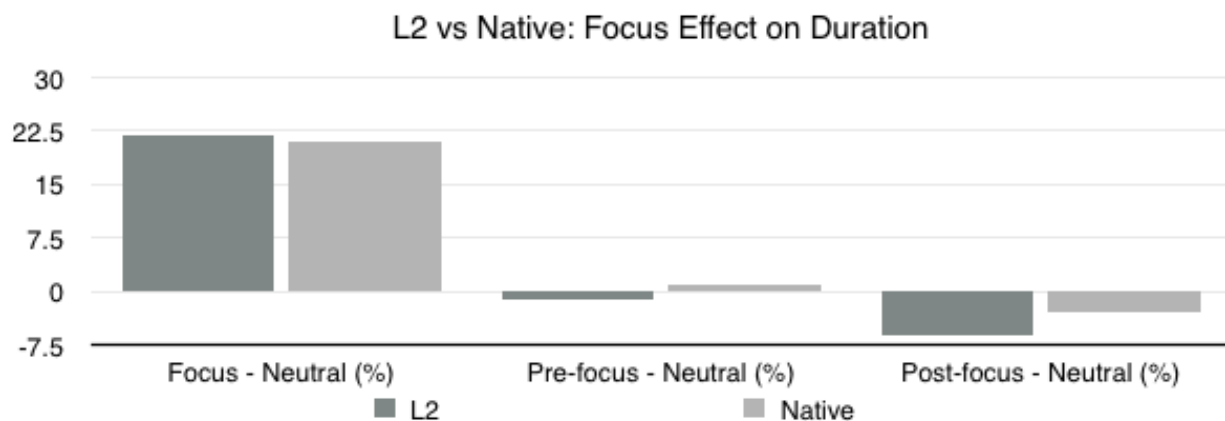


Figure 6 below compares levels of variation in focus effect on the pitch range. It clearly shows that both groups enlarged the pitch range of the syllable under focus rather drastically (around 40% for L2 subjects, 80% for the native speakers). Nearly no variation was made to pitch range of the syllable before focus. Pitch range of the syllable right after the focus was shortened a little in production of both groups. This figure indicates that even though the L2

subjects managed to adjust pitch range according to varying focus conditions in the same fashion as the native speakers, but the extent of variation was not as large as that of the native speakers.

Figure 6. L2 versus native speakers: Extent of variation in pitch range under various focus conditions. The number of percentage each bar corresponds to shows the extent of increase (when the number is positive) or decrease (when the number is negative) of pitch range under a specific focus condition compared with pitch range in the neutral condition. The contrast between each focus condition with the neutral condition is indicated below each pair of comparison between L2 and native speakers. For example, “Focus – Neutral (%)” means the extent of variation in pitch range under focus condition compared to neutral. Numbers of the y-axis represent percentage.)

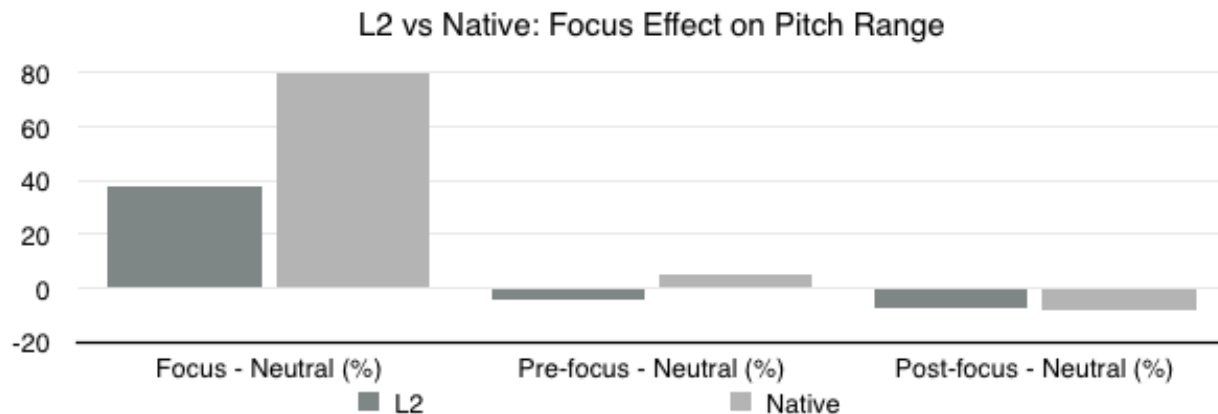


Figure 7 below compares variations in the maximum f0 values brought by different focus conditions. In the pre-focus condition, none of the groups performed much change over the maximum f0 value. Differences between the L2 subjects and the native speakers can be observed over the other two focus conditions. In the focus condition, even though both groups extensively raised the maximum f0 values, the L2 subjects performed less extent of variation. Similarly, in the post-focus condition, the L2 subjects marginally decreased the maximum f0 value, not as drastically as the native speakers did.

Figure 7. L2 versus native speakers: Extent of variation in maximum f0 value under various focus conditions. The number of percentage each bar corresponds to shows the extent of increase (when the number is positive) or decrease (when the number is negative) of maximum f0 value under a specific focus condition based on the value in the neutral condition. The contrast between each focus condition with the neutral condition is indicated below each pair of comparison between L2 and native speakers. For example, “Focus – Neutral (%)” means the extent of variation in maximum f0 value under focus condition compared to neutral. Numbers of the y-axis represent percentage.)

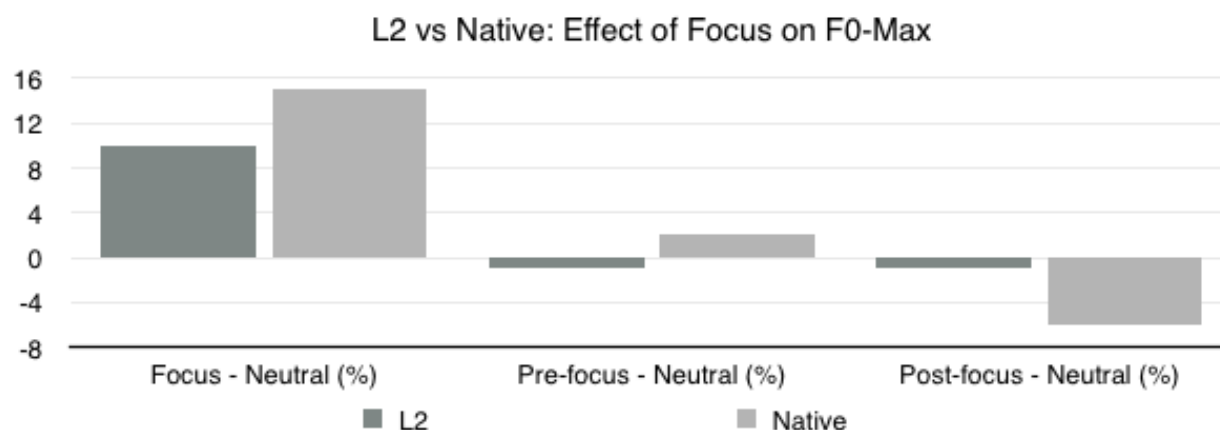
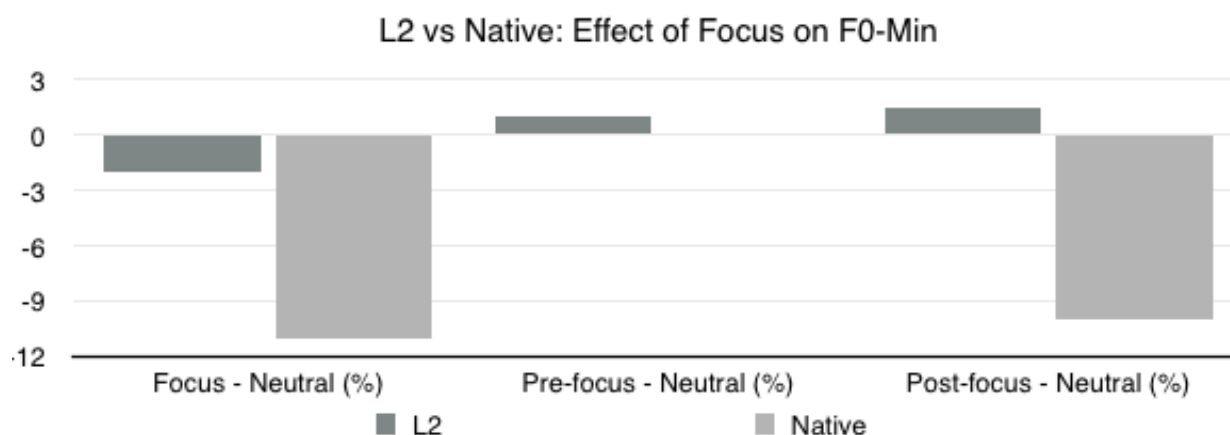


Figure 8 below compares the effect of focus over the minimum f0 values between two groups. The most striking result is that the L2 subjects barely lowered any minimum f0 values under any focus condition. In contrast, the native speakers drastically lowered the minimum f0 values of the syllables under focus and post-focus conditions. This comparative result explains one reason why less variation was made over the overall pitch range in the production of the L2 subjects: they had less difficulty raising the maximum f0 values according to focus conditions, but had much more difficulty lowering the minimum f0 values.

Figure 8. L2 versus native speakers: Extent of variation in minimum f0 value between various focus conditions. The number of percentage each bar corresponds to shows the extent of increase (when the number is positive) or decrease (when the number is negative) of minimum f0 value under a specific focus condition based on the value in the neutral condition. The contrast between each focus condition with the neutral condition is indicated below each pair of comparison between L2 and native speakers. For example, “Focus – Neutral (%)” means the extent of variation in minimum f0 value under focus condition compared to neutral. Numbers of the y-axis represent percentage.)



In sum, Figure 5, 6, 7 and 8 confirm the fact that the L2 subjects managed to realize focus by adjusting the duration, overall pitch range, maximum and minimum f0 in the same direction or tendency as the native speakers did, but the extent of variation was less, especially over the minimum f0 values.

3.3.2 L2 vs Native: Realization of focus over four lexical tones

In order to compare how different focus conditions caused different extent of variation to the maximum and minimum f0 values of each lexical tone, Figure 9, 10 and 11 were made and each of them presents a comparison under a particular focus condition as indicated in the titles. The process of calculation is the same as the previous section: only the extent of variation in percentage is concerned instead of the absolute values. The data of the native speakers were from Table VI of Xu (1999).

Figure 9 below compares when the syllable is under focus, how much variation each lexical tone had in terms of the maximum and minimum f0 values. The H, R and L tones produced by the native speakers experienced a larger extent of rise in maximum f0 values compared to the L2

subjects. Noticeably, the F tone produced by the L2 subjects showed more variation than the natives. In terms of the minimum f0 values, the overall extent of variation was not as big as that of the maximum values, however, it is clear that the H, R and L tones produced by the L2 subjects had much less variation over the minimum f0 values. The largest difference in variation between two groups could be found over the L tone. The native speakers modified the maximum and minimum values of the L tones most substantially across four tones when the syllable is under focus, however, the L2 subjects made the least adjustment. It might suggest that the L2 subjects experienced more difficulty producing the L tone with focus – they couldn't modify the f0 curve of the L tone as freely as the natives do when focus interferes, and as easy as that of the other lexical tones in the same focus condition produced by the L2 subjects.

Figure 9. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under focus condition compared to neutral condition. The left and right figures show result of maximum and minimum f0 values respectively. In each figure, each pair of bars shows comparison between L2 and native speakers in terms of the extent of variation of each lexical tone, as indicated under the pair. “H”, “R”, “L” and “F” means “High tone”, “Rising tone”, “Low tone”, and “Falling tone” correspondingly. Numbers of the y-axis represent percentage.)

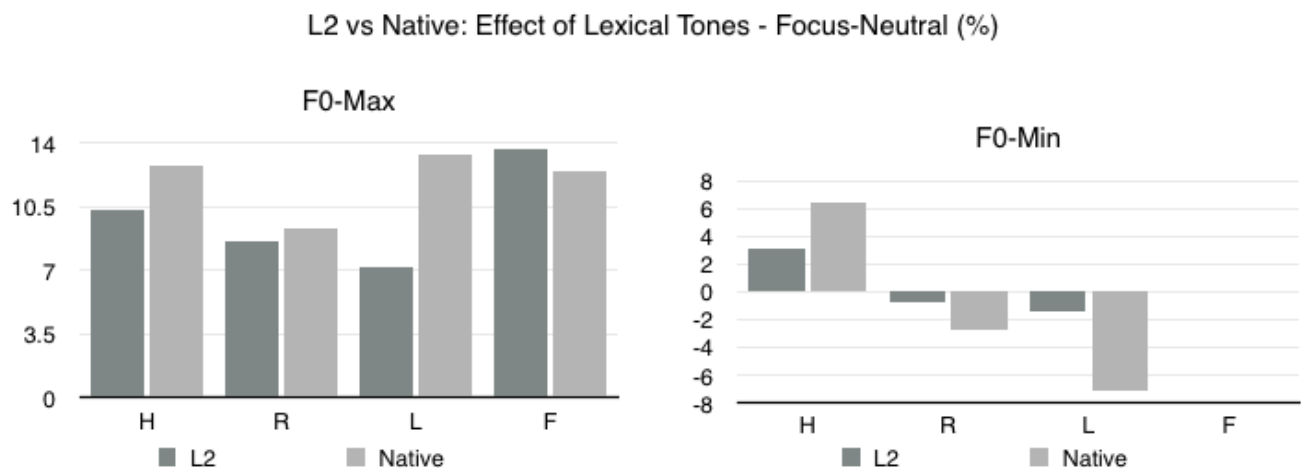


Figure 10 below compares how the maximum and minimum f0 values of four lexical tones varied under the pre-focus condition. The variations are not substantial, given the largest

absolute percentage is less than 3%. Therefore, a comparison between two groups over four lexical tones might not be meaningful. This figure again confirms the fact that for the syllable right before the focus, the difference in f0 values is marginal. However, a general tendency to lower both the maximum and minimum f0 values seems to exist for most of the tones.

Figure 10. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under pre-focus condition compared to neutral condition. The left and right figures show result of maximum and minimum f0 values respectively. In each figure, each pair of bars shows comparison between L2 and native speakers in terms of the extent of variation of each lexical tone, as indicated under the pair. “H”, “R”, “L” and “F” means “High tone”, “Rising tone”, “Low tone”, and “Falling tone” correspondingly. Numbers of the y-axis represent percentage.)

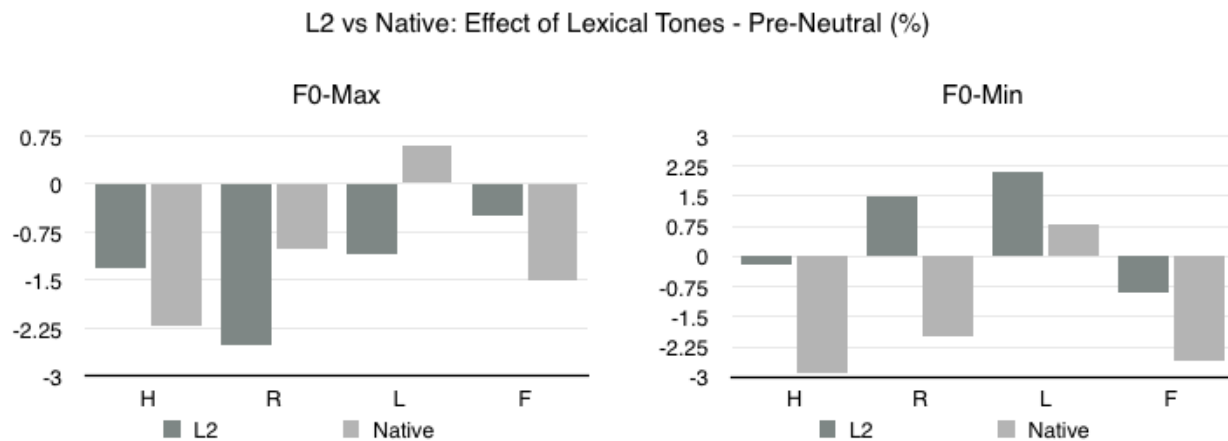


Figure 11. L2 versus native speakers: Extent of variation in maximum and minimum f0 values under post-focus condition compared to neutral condition. The left and right figures show result of maximum and minimum f0 values respectively. In each figure, each pair of bars shows comparison between L2 and native speakers in terms of the extent of variation of each lexical tone, as indicated under the pair. “H”, “R”, “L” and “F” means “High tone”, “Rising tone”, “Low tone”, and “Falling tone” correspondingly. Numbers of the y-axis represent percentage.)

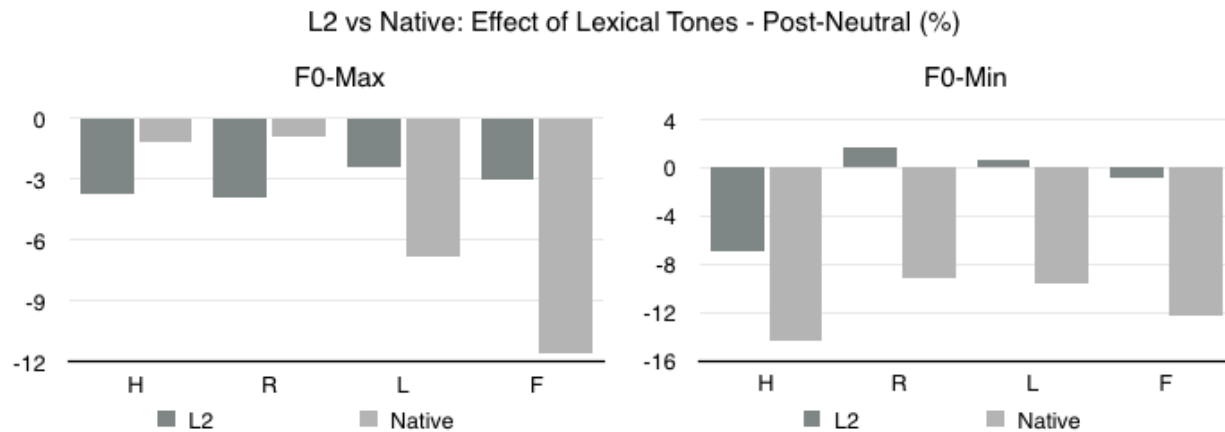


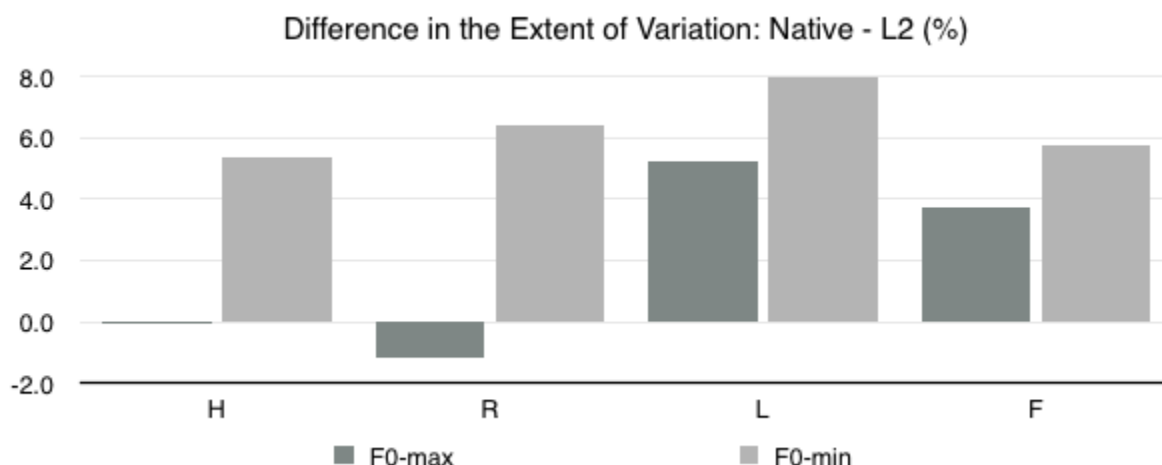
Figure 11 above compares in the post-focus condition, how each lexical tone varied in maximum and minimum f0 values. The L2 subjects lowered the maximum f0 value of the H and R tones a little bit more, but failed to lower that of the L and F tones as drastic as the native speakers did. The biggest difference is over the F tone, where the L2 subjects only lowered about 3% of the maximum f0, the native speakers decreased the value about 12%. Looking at the minimum f0 values, the difference between the L2 subjects and the native speakers is even more obvious: For each lexical tone, the native speakers dropped the minimum f0 values over 8%, even as much as 15% for the H tone. However, the L2 subjects barely adjusted the minimum f0 values of the lexical tones. They managed to lower the minimum f0 of the H tone, but much less than 15% of the native speakers. For the R and L tones, they even raised the minimum value slightly.

This figure shows that compared to the native speakers, the L2 subjects had difficulty lowering both the maximum and the minimum f0 values in post-focus conditions. When producing lexical tones of the syllable right after the focus, the L2 subjects made little change to the overall f0 curve. Whereas for the native speakers, even though they didn't change much over

the maximum f0 values of the H and R tones, they substantially lowered those of the other tones, resulting in an overall lowering in four lexical tones in terms of pitch.

To summarize, Figure 9, 10 and 11 indicate that, compared to the native speakers, the L2 subjects realized focus differently in terms of the extent of variation over the relevant measurements of the tones. Figure 12 below directly shows the difference in the extent of variation between the two groups. The differences are calculated by deducting the percentage of the L2 subjects from that of the native speakers, and are averaged across all focus conditions except the pre-focus condition, where neither group exhibited substantial variation. The bigger the difference is, the larger the extent of variation was made by the native speakers.

Figure 12. Difference in the extent of variation between native speakers and L2. (“Native – L2 (%)” indicates that the difference is calculated by subtracting the percentage of L2 from that of native speakers across focus conditions. A positive percentage shows a larger extent of variation in native speakers’ production. The results are grouped according to lexical tones. “H”, “R”, “L” and “F” means “High tone”, “Rising tone”, “Low tone”, and “Falling tone” correspondingly. Both maximum and minimum f0 values are considered, indicated by different shades of the bars. Numbers of the y-axis represent percentage.)



Now it can be clearly seen that among the four lexical tones, for the L2 subjects, it was most difficult to realize the L tone by adjusting the maximum and minimum f0 values. The native

speakers raise or lower the relevant f0 values to a much larger extent over the L tone. Similarly, the L2 subjects did not vary the relevant f0 values of the F tone as much as the native speakers did, but better than the L tone. As for the H tone, the L2 subjects adjusted the maximum value to the same extent as the native speakers. Interestingly, they even made slightly more variation in the maximum f0 value of the R tone. However, in terms of the minimum f0 values, like the F and L tones, much less variation was found over the H and R tones.

3.3.3 L2 vs Native: Realization of lexical tones

Section 3.2 implies that the L2 subjects did realize lexical tones with interference of focus intonation. Exactly how their realization of lexical tones deviates from the native speakers will be shown in the following comparisons, each presenting result under a consistent focus condition, which ranges from neutral, under focus, pre-focus to post-focus. These comparative results are essentially different from those in the previous section: section 3.3.2 shows the effect of focus on each lexical tone, therefore it only looks at the degree of variation; this section treats each focus condition as a controlled factor and it only cares about how four lexical tones were realized. Admittedly, the main purpose of this study is to examine phonetic realization of focus, it is still beneficial to check realization of lexical tones. On the one hand, it might provide insight to the interaction between focus and lexical tones produced by the L2 subjects; on the other hand, it will serve as reference for discussion on L2 instruction of MC.

Two methods are used to compare the realization of lexical tones. The first method is by visually comparing the f0 curves of the lexical tones produced by both groups. Though the method may not be accurate, it is straightforward in showing rough shapes of the curves. The second method, which is more accurate, is by looking at the mean maximum and minimum f0 values of each lexical tone under each focus condition, which represent the peaks and valleys of

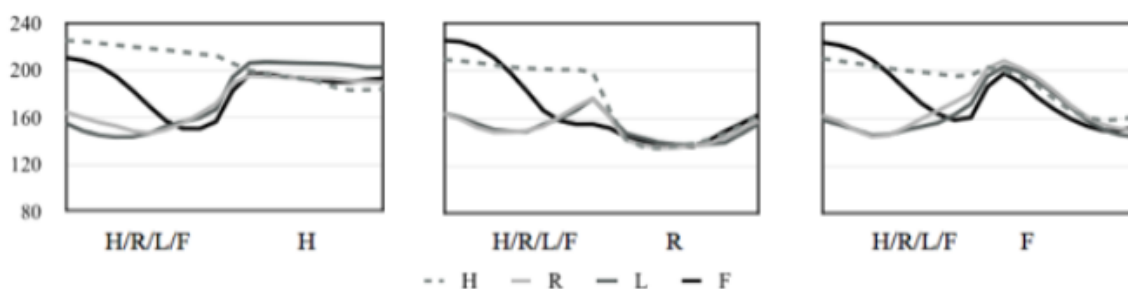
the f0 curves respectively of each lexical tone. The first method is only applied to the neutral condition since tonal identities have been proved to preserve in other focus conditions.

3.3.3.1 L2 vs Native: F0 curves of the lexical tones in the neutral condition

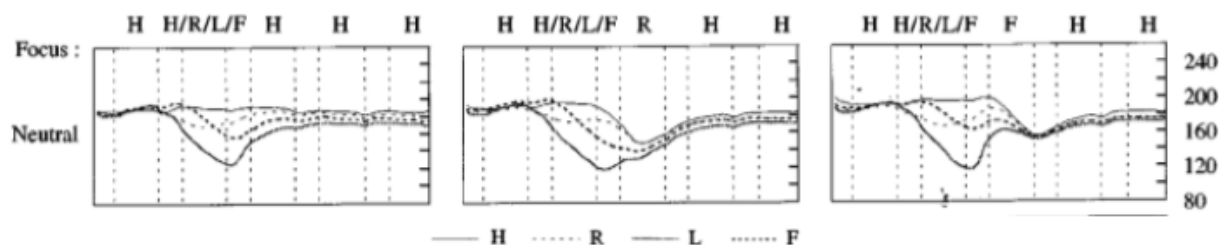
Figure 13 below compares f0 curves of the second syllable with varying lexical tones produced by the L2 subjects and the native speakers in all neutral conditions. The f0 curves of the native speakers are from Figure 4 of Xu (1999). In Xu's study, he presented overall f0 contour of the whole sentence with five syllables, but for the purpose of comparison, only f0 curves over the second and third syllables are concerned.

Figure 13. L2 vs Native: F0 curves of the second syllable in the neutral condition. (The figures on the second row are from the study of Xu (Xu, 1999) on the production of native speakers. Lexical tones are labeled as H, R, L or F right below the corresponding syllables. "H/R/L/F" represents varying lexical tones in the position. Lines in different shades and styles represent f0 curves when word 1 carries varying tones. Unit of the y-axis is Hz.)

L2:



Native:

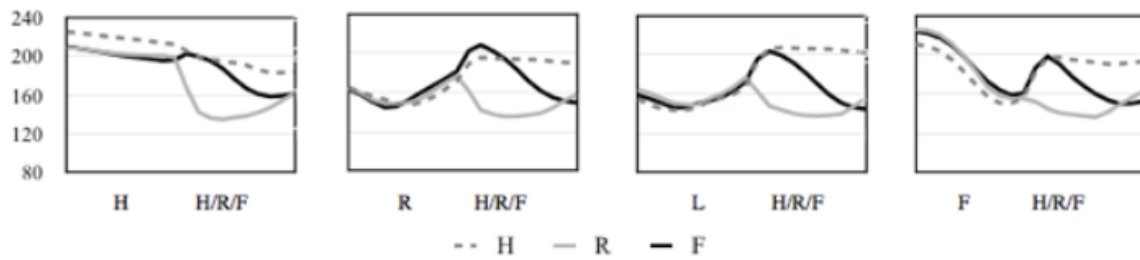


Comparing the lines within the duration of the second syllable, one can easily find out that the most noticeable deviation lies in the shape of the L tone produced by the L2 subjects. In the lower panels of Figure 13, the densest dotted lines, representing f0 curves of L tone, start from a relatively low pitch range compared to the other three lexical tones, and then drop rapidly to a pitch range much lower than the minimum f0 of the F tone. Whereas in the production of the L2 subjects, the L tones, represented by the second darkest solid lines, always start from almost the same pitch level as the R tones, dip a little in height, and then end in a slightly higher pitch than the F tones. F0 curves of the L tones produced by L2 subjects keep overlapping the curves of the R tones, which leads to nearly no differentiation between the L and R tones in the L2 production. The shapes of the H, R and F tones are relatively similar between the two groups.

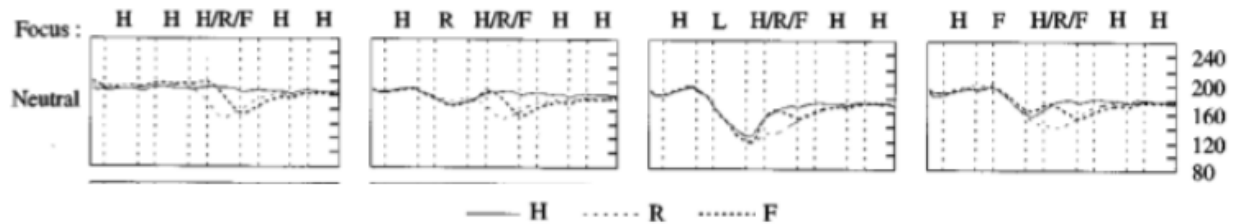
Figure 14 below shows a comparison between f0 curves of the third syllable with varying lexical tones in the neutral condition. The lower four panels are from Figure 5 of Xu (1999).

Figure 14. L2 vs Native: F0 curves of the third syllable in the neutral condition. (The figures on the second row are from the study of Xu (Xu, 1999) on the production of native speakers. Lexical tones are labeled as H, R, L or F right below the corresponding syllables. “H/R/L/F” represents varying lexical tones in the position. Lines in different shades and styles represent f0 curves when word 2 carries varying tones. Unit of the y-axis is Hz.)

L2:



Native:

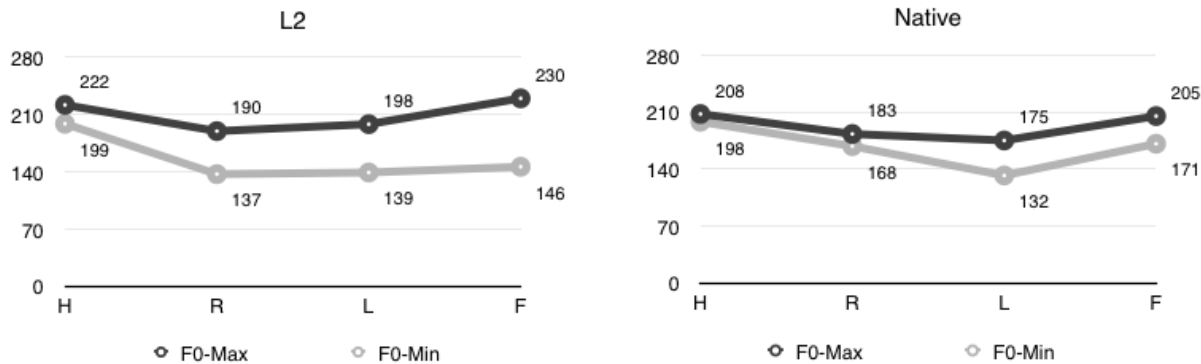


Comparing the lines within the range of the third syllable, one could observe similarity in the shape of the f0 contours between the two groups: F0 curves of the H tones are relatively flat, placed in a higher pitch range consistently. The F tones all start from a high pitch and later fall into a lower pitch range. F0 curves of the R tones all show a slight “dipping” form close to the onset of the syllable. When the second syllable carries the L tones, the starting points of the f0 curves of the third syllable are much higher in the upper panel (carryover effect), which again proves that the L2 subjects failed to lower the L tones as much as the native speakers did.

3.3.3.2 L2 vs Native: Mean maximum and minimum f0 of the lexical tones

Figure 15 below compares the mean maximum and minimum f0 values of each lexical tone in neutral conditions. The absolute numbers are not for direct comparison, but the differences between the maximum and minimum values could serve to depict the shape of the curves observed in the previous section more accurately.

Figure 15. L2 vs Native: Mean f0-max and f0-min of four lexical tones in neutral condition. (The two figures show results of L2 and native speakers separately. In each figure, the four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively. The dots in black and grey colors indicate averaged results of maximum and minimum f0 values correspondingly. Unit of the numbers is Hz.)



Based on Figure 15, it seems that in the neutral condition, there was least deviation in the pitch range of the H tone between two groups, therefore H tones produced by L2 and native speakers are all high and level in pitch contour. Larger differences between the maximum and minimum f0 values were found in the L2 subjects' production of the R, L and F tones. This result suggests that, though variation in these values with the effect focus is less compared to the native speakers, as shown in the previous section, when produced in the neutral condition, lexical tones possess larger pitch range in the production of L2 subjects.

Another deviation worth mentioning is that the difference between maximum and minimum f0 values is the largest in the L tone for the native speakers, whereas for the L2 subjects the pitch range of the L tone is similar to that of the R tone, and the largest difference was found in the F tone. This particular deviation confirms the observation that f0 curves of the L tones of the native speakers are much noticeably different from those of the L2 subjects.

In Figure 16-18, the observed pattern in the neutral condition is easily identified. This again confirms that lexical tones were realized regardless focus condition. Deviation from the native speakers observed in the neutral condition can be found in Figure 16-18 as well.

Figure 16. L2 vs Native: Mean f0-max and f0-min of four lexical tones under focus. (The two figures show results of L2 and native speakers separately. In each figure, the four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively. The dots in black and grey colors indicate averaged results of maximum and minimum f0 values correspondingly. Unit of the numbers is Hz.)

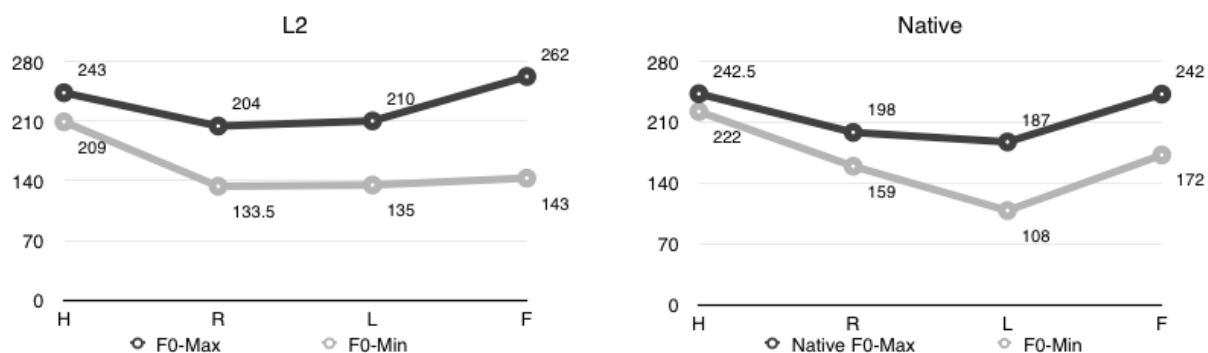


Figure 17. L2 vs Native: Mean f0-max and f0-min of four lexical tones in pre-focus condition. (The two figures show results of L2 and native speakers separately. In each figure, the four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively. The dots in black and grey colors indicate averaged results of maximum and minimum f0 values correspondingly. Unit of the numbers is Hz.)

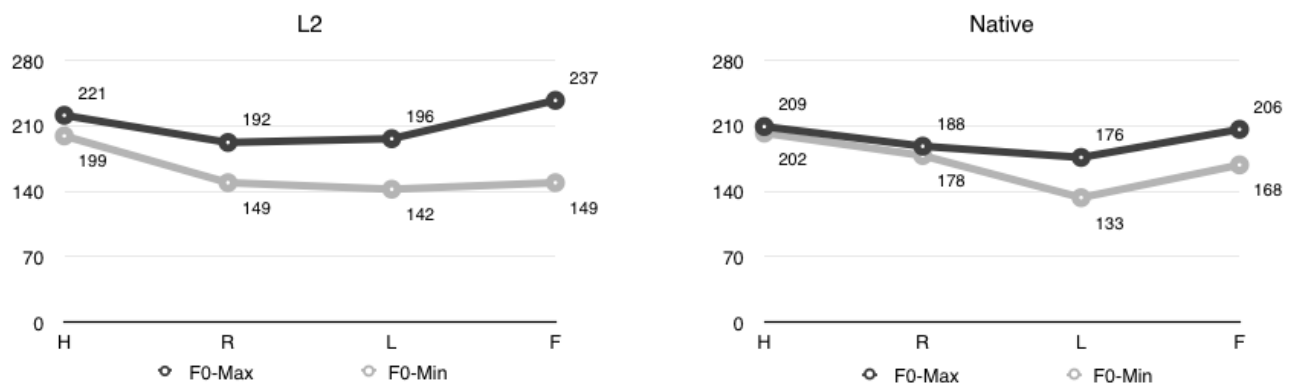
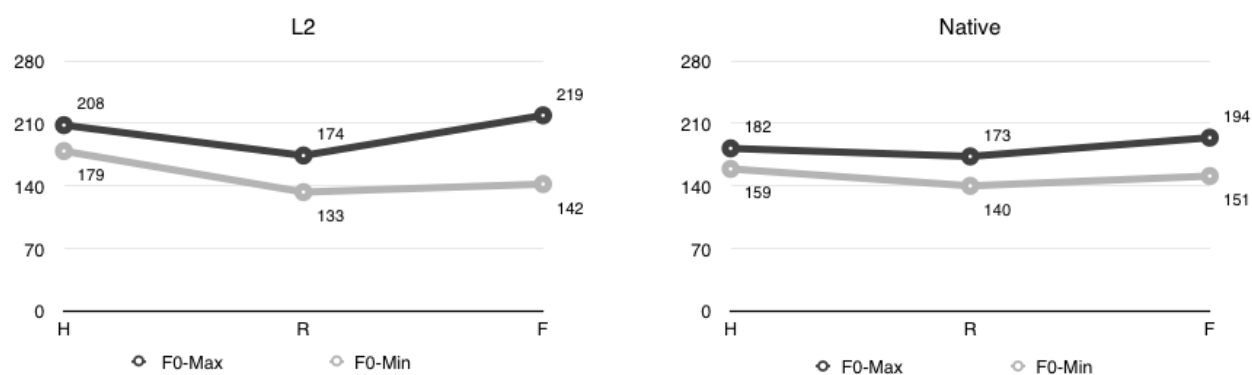


Figure 18. L2 vs Native: Mean f0-max and f0-min of four lexical tones in post-focus condition. (The two figures show results of L2 and native speakers separately. In each figure, the four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively. The dots in black and grey colors indicate averaged results of maximum and minimum f0 values correspondingly. Unit of the numbers is Hz.)



3.3.4 Summary

Table 3-6 below roughly summarizes the results of all comparisons made between the L2 subjects and the native speakers, in terms of the realization of focus and lexical tones.

Table 3-6. L2 vs Native: Realization of focus and lexical tones

L2 compared to the native speakers		
Focus	Duration	same ways of adjustment and slightly larger in extent
	Pitch Range	same ways of adjustment but less in extent
	F0-Max	same ways of adjustment but less in extent
	F0-Min	marginal adjustment and much less in extent
	H	similar in the variation of f0-max; much less variation in f0-min
	R	slightly more variation of f0-max; much less variation in f0-min
	L	least extent of variation of f0-max; least extent of variation in f0-min
	F	much less variation both in f0-max and f0-min
Tones	H	similar in height and shape
	R	similar in shape but larger difference between f0-max and f0-min
	L	different in shape and height
	F	similar in shape but larger difference between f0-max and f0-min

3.4 Comparison between subjects

3.4.1 Individual subject vs. Native speaker: Realization of focus

Previous results are calculated from averaged measurement across all eight subjects, therefore these results only suggest average performance of L2 learners over focus and lexical tones of Mandarin Chinese. Although the subjects were all enrolled in Chinese class of the same level, according to the questionnaire and judgement of the researcher (a native speaker of Mandarin), the eight subjects were not equally proficient in speaking Mandarin, especially producing lexical tones. It is interesting to examine the realization of focus by looking at results of each subject, which would test the previous results based on mean values and help investigate possible correlation between the ability to produce focus and lexical tones.

The ability to realize Mandarin focus is determined by the extent of adjustment over duration, pitch range, maximum f0 and minimum f0 values under different focus conditions. The native speakers elongate the duration, expand the pitch range, raise the maximum f0 and lower the minimum f0 values under the focus condition. And in the post focus range, they would compress the pitch range, lower the maximum and minimum f0 values. Not much deviation takes place in the pre-focus condition compared to the neutral condition. The closer to the way the native speakers produce the focus, the better the ability of the L2 subject to realize focus is.

The subjects' ability to produce lexical tones of Mandarin is roughly judged by the researcher. Both accuracy and consistency are concerned. Among the eight subjects, six of them could pronounce lexical tones rather accurately without interplay of the focus intonation. The other two, however, were not as good when producing lexical tones in the neutral condition and some inconsistency was observed in varying focus conditions.

Figure 19-22 below show the extent of adjustment or variation over duration, pitch range, maximum and minimum f0 values made by each subject under varying focus conditions. S1-8 represents eight subjects. Among them, S1 and S3 are the two not so good at Mandarin lexical tones. The results of the native speakers are shown as reference on the right end.

Two major conclusions could be made based on these figures. First of all, though on average, the L2 subjects realized focus in a similar fashion as the native speakers did, big differences exist between subjects. For example, S2 and S7 made relatively less variation in duration, pitch range and the maximum f0 value. Some subjects, such as S8, made extremely large expansion over the pitch range. As can be seen in Figure 19, most of the L2 subjects didn't lower the minimum f0 value as much as the native speakers did, as mentioned before, however, S1 made quite drastic variation in the under-focus and pre-focus ranges. Secondly, no apparent correlation could be established between ability of realizing focus and producing accurate lexical tones. S1 and S3 are not noticeably worse in adjusting duration, pitch range, maximum and minimum f0 values than the others. In fact, the variations made in duration, maximum and minimum f0 values by S1 were quite similar to those of the native speakers. On the other hand, subjects that were able to produce relatively accurate and consistent lexical tones, such as S2, made much less variation over the relevant measurements.

Figure 19. Comparison between subjects: Variation in duration under varying focus condition. (Unit of the y-axis is percentage. S1-S8 represents each subject. “Native” represents results of the native speakers. “Focus – Neutral” equals the amount of increase (if the number is positive) or decrease (if the number is negative) in duration of the syllable between focus and neutral conditions. Different shades of the bars show different contrast of focus condition as suggested.)

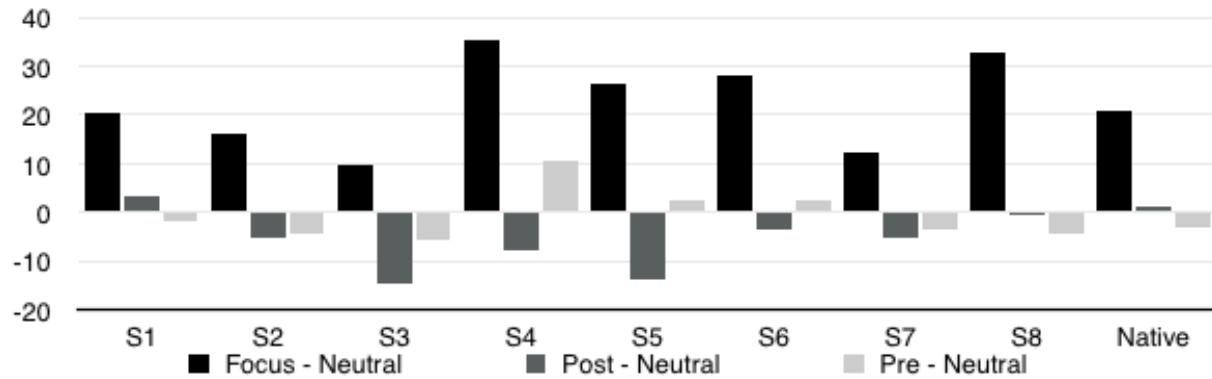


Figure 20. Comparison between subjects: Variation in pitch range under varying focus conditions. (Unit of the y-axis is percentage. S1-S8 represents each subject. “Native” represents results of the native speakers. “Focus – Neutral” equals the amount of increase (if the number is positive) or decrease (if the number is negative) in pitch range of the syllable between focus and neutral conditions. Different shades of the bars show different contrast of focus condition as suggested.)

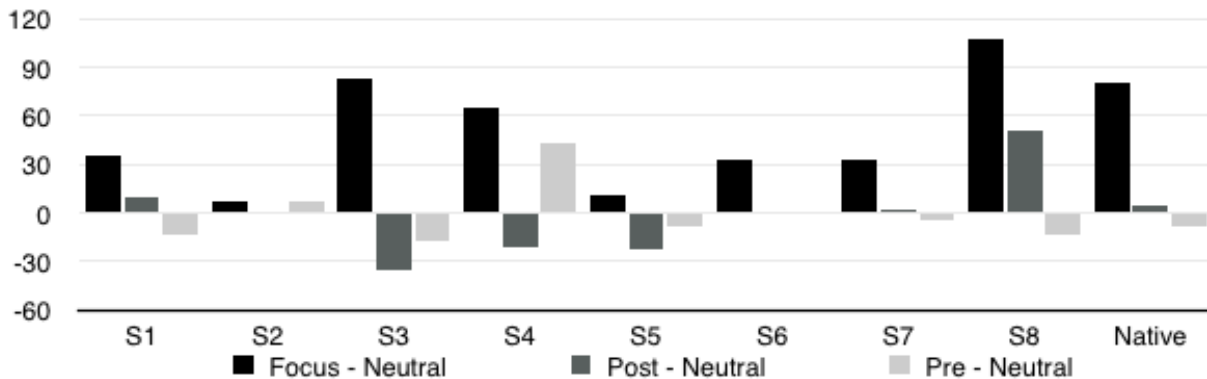


Figure 21. Comparison between subjects: Variation in f0-max under varying focus conditions. (Unit of the y-axis is percentage. S1-S8 represents each subject. “Native” represents results of the native speakers. “Focus – Neutral” equals the amount of increase (if the number is positive) or decrease (if the number is negative) in f0-max of the syllable between focus and neutral conditions. Different shades of the bars show different contrast of focus condition as suggested.)

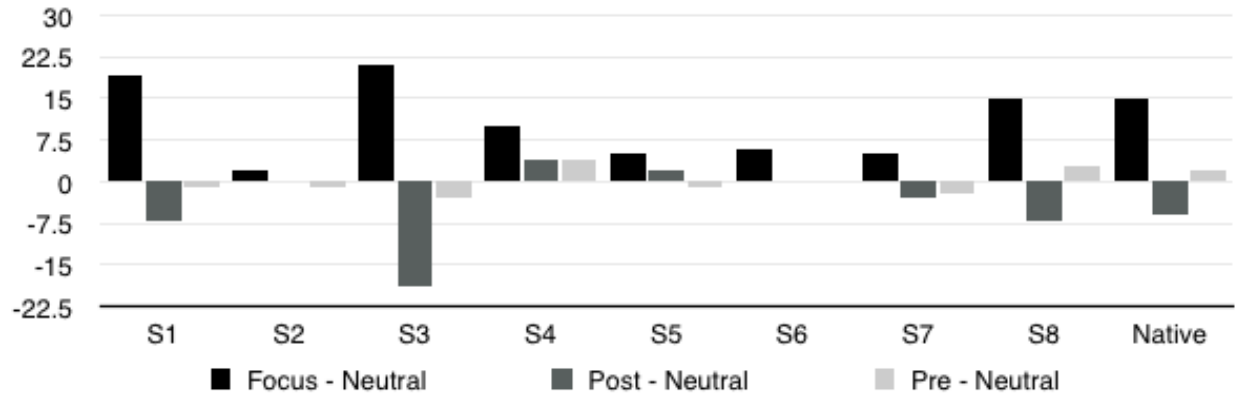
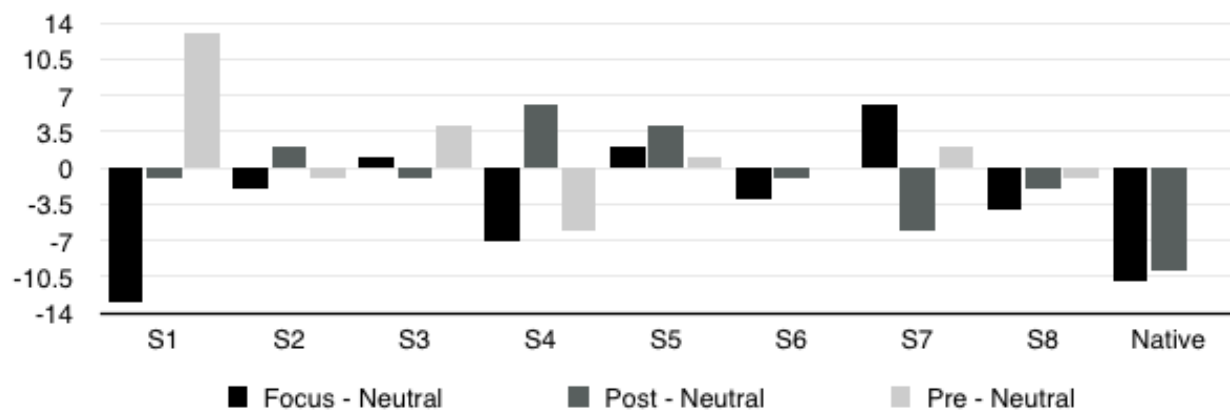


Figure 22. Comparison between subjects: Variation in f0-min under varying focus conditions. (Unit of the y-axis is percentage. S1-S8 represents each subject. “Native” represents results of the native speakers. “Focus – Neutral” equals the amount of increase (if the number is positive) or decrease (if the number is negative) in f0-min of the syllable between focus and neutral conditions. Different shades of the bars show different contrast of focus condition as suggested.)



3.4.2 Individual subject vs. Native speaker: Slope

As shown in section 3.2.2, based on average values of the relevant measurements, under varying focus conditions, the slopes of the R and L tones are mostly positive, and the slope of the H tone is not significantly negative, which implies that the L2 subjects preserved identities of

different tones therefore successfully realized lexical tones under varying focus conditions. However, after checking on individual production, the result of a particular subject, S3, is strikingly opposite to the average result regarding the slope of lexical tones, which might have crucial implication on the nature of English focus intonation.

As discussed in the introduction, one purpose of this thesis is to test Xu's theory on the nature of English focus. Xu opposed the traditional belief of English focus as a pitch accent and proposed that it is essentially the same as the Mandarin focus, which manipulates f_0 instead of replacing neutral f_0 with a pitch accent (Xu, 2005). Based on Xu's understanding, it should be predicted that no pitch accent would be found in English L2 production of Mandarin focus and tonal identities should be preserved the same as in the neutral condition across all L2 subjects. However, consistent substitution pattern found in the interlanguage of S3 shown below refutes Xu's prediction, thus confirms English focus is indeed a pitch accent. The performance of S3 is extremely crucial in this study. On the one hand, it shows problems in producing Mandarin L2 due to negative transfer effect of English. On the other hand, it reveals true property of L1 (English) of the L2 learners.

As shown in Table 3-7 below, when word 2 is under focus, S3 produced a syllable with significantly negative slope value regardless of the lexical tones. It is found that she did this with extremely few exceptions, as shown in Figure 24. When word 1 is under focus, the situation is not significantly different from the averaged result in Table 3-5: only the slope of the F tone is significantly negative. Considering the fact that she was actually able to produce the lexical tones, at least differentiate the H, R tones from the F tone in the neutral condition in terms of slope, it is likely that this particular subject failed to preserve the identities of the H and R tones under focus, specifically by replacing the lexical tones with a F tone that is similar to English pitch accent in

f0 contour. Interestingly, the possible substitution mostly took place in the position of word 2.

When the focus was placed on the first word, she changed the H, R and L tones for a few times.

Table 3-7. Average slope of S3. (The focus conditions and lexical tones of each word are indicated in the headings. “H” “R” “L” and “F” represent “High tone” “Rising tone” “Low tone” and “Falling tone” respectively. The unit of the slope is Hz.)

	Word 1								Word 2					
	Neutral				Under Focus				Neutral			Under Focus		
	H	R	L	F	H	R	L	F	H	R	F	H	R	F
Slope	-10	47	45	-94	-81	8	-8	-182	-1	-5	-107	-85	-71	-149
P=	0.41	<0.001	<0.001	<0.001	<0.001	0.52	0.47	<0.001	0.89	0.62	<0.001	<0.001	<0.001	<0.001

Figure 23. Slope of word 1 (the second syllable) produced by S3. (Unit of the y-axis is Hz. Each bar shows the slope of the tone in an individual utterance. The four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively.)

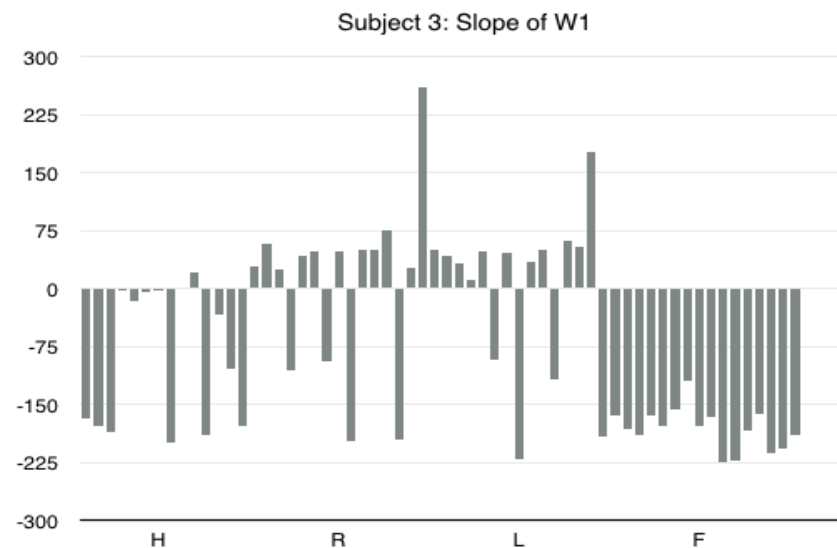


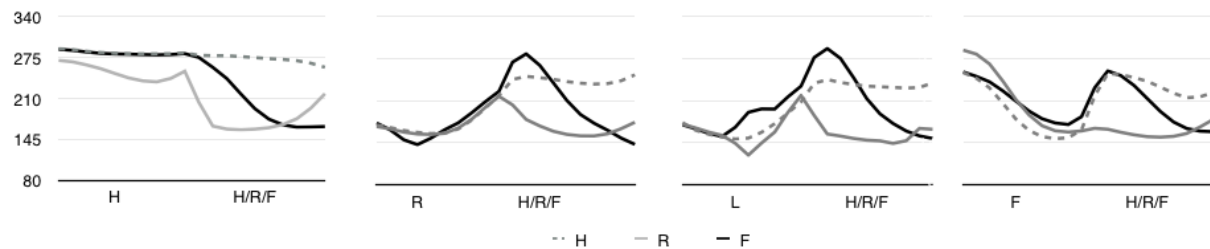
Figure 24. Slope of word 2 (the third syllable) produced by S3. (Unit of the y-axis is Hz. Each bar shows the slope of the tone in an individual utterance. The four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively.)



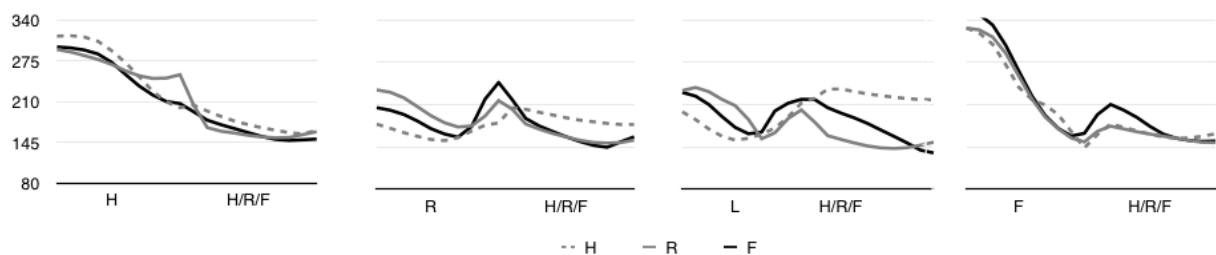
To confirm S3 did replace the lexical tones with a pitch accent, only looking at the slope of lexical tones is not sufficient, because in the current study, “slope” is only determined by the initial and final f0 values, there is possibility that the H and R tones still preserved their characteristics in contour that are fundamentally different from the F tone while the relevant f0 values dropped. However, result shown in Figure 25 excludes this possibility. In Figure 25, it is clearly shown that in the neutral condition, the H tone is consistently high in pitch value; the R tone starts in the mid pitch range, immediately declines a bit and forms a dipping shape; the F tone begins from or raises to a high pitch range and rapidly drops to a very low pitch level. However, as observed in the other two rows of panels where focus interplays, characteristics that represent the identity of the H and R tones are hard to find. Instead, all of the curves form a drastic falling shape as the F tone. In some panels, the f0 curves almost overlap entirely. Based on Figure 25, now it is safe to say that this subject indeed failed to realize lexical tones with focus and produced pitch accent instead.

Figure 25. Realization of lexical tones of the word 2 by S3. (Lexical tones are labeled as H, R or F right below the syllable. “H/R/L/F” represents varying lexical tones in the position. Focus conditions are listed above each row. Lines in different shades and styles represent f0 curves when word 2 carries varying tones. Unit of the y-axis is Hz.)

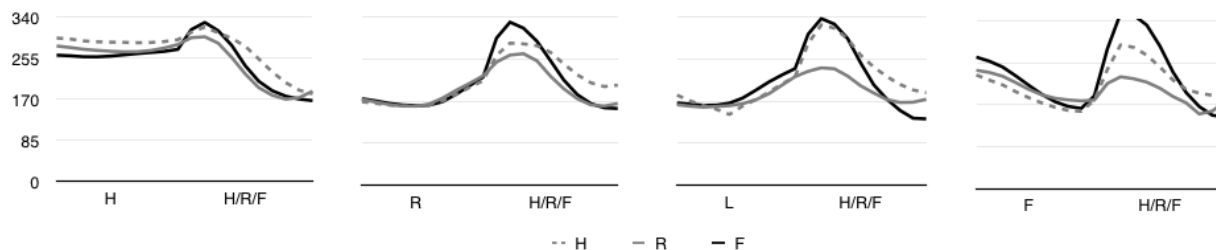
A. Neutral condition



B. Focus on word 1



C. Focus on word 2



3.5 Summary

This chapter presents in detail how focus and lexical tones were realized by the L2 subjects. Section 3.1 and 3.2 showed results on focus and lexical tones respectively with both direct visual observation over f0 curves and statistic analysis. Section 3.3 and 3.4 further compared the results

with the native speakers and between the individual subjects, in order to find out deviations from the native speakers, and specific differences between the subjects.

The overall conclusions of the experiment are: 1) Overall, the L2 subjects were able to realize focus in the same fashion as the native speakers do but less in the extent of variation. 2) Most of the L2 subjects produced four lexical tones under varying focus conditions without substituting any with a pitch accent. 3) Four lexical tones brought different levels of difficulty to the L2 subjects under varying focus conditions. 4) The eight subjects varied in their ability to realize focus and this variation is not correlated with their performance in producing lexical tones. Critically, subject 3 was not able to realize lexical tones of word 2 under the focus condition.

CHAPTER 4

DISCUSSIONS

This chapter aims to give detailed answers to each of the research questions based on results reported in the previous chapter and attempts to explain some of the phenomena. The controversy over “what an English focus really is” brought by Xu’s study will be resolved next. At last, pedagogical implication will be briefly discussed.

4.1 Answers to research questions

4.1.1 Research question 1

The first research question on focus realization regardless of the underlying lexical tones has been generally answered in section 3.1 with averaged results across all subjects. The L2 learners successfully realized focus intonation by significantly increasing the duration of the focused constituent by 22%, expanding its pitch range by 38%, raising the pitch peak by 10% and marginally lowering the minimum pitch value by 2%. At the same time, these measurements of the syllable right before the focused constituent barely changed. The duration, pitch range, maximum and minimum f0 of the syllable immediately following the focused constituent decreased by 6%, 7%, 1%, -1.5% respectively. Generally speaking, the most significant contrast lies in the focused constituent.

This pattern is further confirmed in detailed comparison between individual subjects and native speakers in section 3.4.1. According to results in section 3.4.1, not all subjects adjusted these measurements equally and significantly, but the direction or tendency of adjustment is largely the same especially for the focused constituent.

In conclusion, by manipulating duration, pitch range, maximum and minimum pitch values, the subjects indeed differentiated a sentence with narrow focus from the same sentence in declarative intonation. The hypothesis for this research question is proven correct.

4.1.2 Research question 2

The second research question is the core of this thesis: deviations in the L2 production from the native speakers. It also attempts to explore causes of the deviations.

Compared to *whether* focus intonation was realized alone, it is more important and meaningful to investigate *how* focus intonation was realized.

In MC, implementation of the focus intonation is closely related to realization of the lexical tones: they must be realized simultaneously. Though the surface height and contour of the lexical tones are affected by the intonation compared to those in the neutral condition, crucially, their identities have to be preserved regardless of the focus condition. This nature of Mandarin focus requires the investigation of deviations in L2 production from two perspectives. First, deviation in focus realization, regardless of the lexical tones. Second, deviation in tone realization under varying focus conditions. Without the second perspective, one cannot identify whether the L2 learners realized focus intonation in the same way as the native speaker do, in another word, whether they were really producing Mandarin focus intonation, other than the mostly likely

English focus. According to the hypothesis, most deviation should be found in the realization of lexical tones while they were producing focus intonation.

4.1.2.1 Deviation in focus realization regardless of lexical tones

Comparative result in section 3.3.1 answers the first part of the question. It shows that the L2 subjects, like the native speakers of MC, increased the duration, enlarged the pitch range, raised the pitch peak, and lowered the minimum pitch value of the focused constituent. However, the extent of variation in pitch range, maximum and minimum pitch value in L2 production was found less than that of the native speakers. The least extent of variation is found in the minimum f0 value. Therefore, it seems that even though they “knew” what to do to show narrow focus, they failed to realize it as extensively as the native speakers and they had most difficulty lowering the minimum f0 value when the native speakers did it extensively.

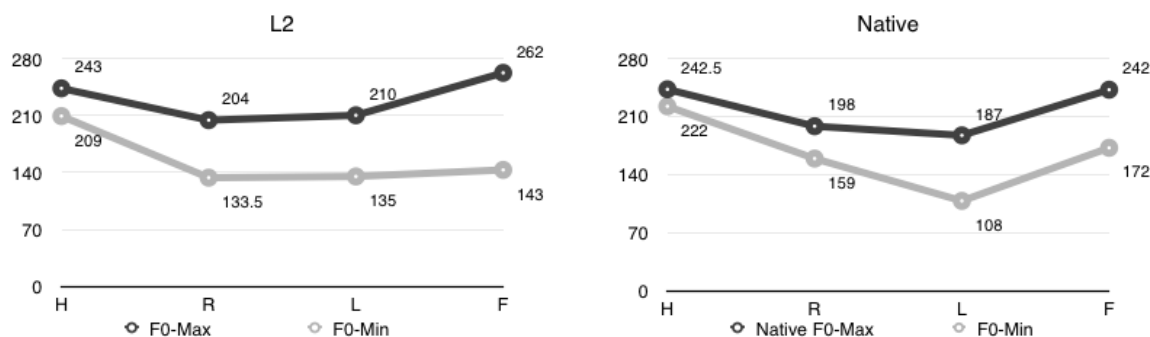
Can this deviation be accounted by difference in the nature of English and Mandarin focus? The deviation is definitely not due to the fact that English focus is a pitch accent, because the pitch accent of English in narrow focus condition is manifested as longer duration, higher f0 peak, and larger pitch range, which is the same as Mandarin in terms focus realization when lexical tones are not concerned.

Is it possible that English focus intrinsically has a more limited extent of expansion over pitch range than Mandarin focus? In Xu 2005, he measured duration, maximum and minimum f0 in focused and neutral condition based on several English declarative sentences produced by eight native speakers of American English. The maximum and minimum f0 were raised by 34% and 0.03% respectively. The pitch range should be enlarged due to significantly higher peak and the barely changed minimum pitch. The duration is increased by 14%. These numbers are not

completely comparable to those in this thesis, but it suggests that the English L2 learners should be able to make substantial adjustment to the pitch range in the same direction. Unfortunately, it's hard to find reliable quantitative evidence for average extent of variation in these measurements for English focus, as well as for Mandarin. Whether less extent of adjustment over Mandarin pitch range produced by English L2 subjects is due to less extent of variation in pitch range of English focus is unknown.

Is the deviation a result of language-specific difference in overall pitch range? Is it possible that native speakers of Mandarin are able to expand the pitch range of the focused syllable more than the English L2 learners of Mandarin do because Mandarin might have an overall larger pitch range than English? The answer is probably yes. Previous studies (S. Chen, 2005; G. T. Chen, 1972; Xue et al., 2002; Mang, 2001), examining the pitch range of Mandarin or comparing it with that of English, have suggested that Mandarin has a relatively larger pitch range. In a most recent study (Keating & Guo, 2010), the result is somewhat confirmed again, but only in single word utterances. As shown in Figure 16, the height of lexical tones in the neutral condition produced by the English L2 learners are comparable, or even a little higher, than the native speakers. Therefore, relatively limited overall pitch range might be a reason why the English L2 learners failed to expand the pitch range as much as the native speakers of Mandarin do.

Figure 16. L2 vs Native: Mean f0-max and f0-min of four lexical tones under focus. (The two figures show results of L2 and native speakers separately. In each figure, the four lexical tones “High” “Rising” “Low” and “Falling” are represented as “H” “R” “L” and “F” respectively. The dots in black and grey colors indicate averaged results of maximum and minimum f0 values correspondingly. Unit of the numbers is Hz.)



4.1.2.2 Deviations in realization of lexical tones under varying focus conditions

The hypothesis for this question is that the L2 subjects would fail to realize lexical tones under focus and specifically they would replace the lexical tones with tones that have intrinsic high pitch level, including the H and R tones, with the F tone, because the pitch contour of the F tone is almost the same as the pitch accent of English in the focus intonation.

Result in section 3.2 based on averaged data contradicts this prediction. It is shown that, on average, the L2 subjects preserved identities of the lexical tones. The four lexical tones produced by English L2 subjects deviated from native production, for example, the R and L tones were hard to differentiate. However, the contour of the H, R and L tones are significantly different from that of the F tone. More importantly, this difference that shows the identity of each lexical tone was found across all focus conditions.

Does this result prove that the nature of English focus and Mandarin focus are the same, as predicted by Xu? The answer is no. First of all, result of S3 with negative transfer effect from English focus confirms the traditionally believed fundamental difference, which will be discussed in detail in section 4.2. Secondly, the fact that most of the subjects managed to produce lexical tones under focus might be due to other possible reasons.

One possibility is that as L2 learners with intermediate proficiency of Mandarin, they somehow have already acquired the intonation without explicit L2 instruction. Since intonation is embedded in natural speech, nearly all of the native production of MC they have been exposed to carry sentence-level intonations. It can be imagined that in classroom environment, narrow focus is produced when the instructor is introducing new vocabulary, contrasting correct forms with errors, or having conversations similar to the experiment materials in the present study. Therefore, they have unconsciously acquired the idea that the focus intonation in Mandarin never conflicts with lexical tones, or, lexical tones should be realized regardless of sentence intonation.

Another possibility is that the experiment design unintentionally makes the subjects pay much attention to the lexical tones. The lexical tones of the second and third syllables varied constantly during the experiment. Under the pressure to produce lexical tones accordingly, it is possible that the subjects didn't put as much effort to realize focus intonation. Result of a particular subject, S2 provides some evidence. In Figure 19-22, compared to other subjects and the native speakers, she made the least adjustment to the relevant measurements. It seems that she might have failed to produce significant focus intonation. On the other hand, the lexical tones she produced were comparatively accurate across all focus conditions. Thus, success in preserving lexical tones under focus is not result of identical nature of English and Mandarin focus.

Now it has been proven that in terms of preserving tonal identity under focus intonation, the English L2 subjects did not differ from the native speakers. However, deviations in production in each focus condition still exist. In another word, the English L2 subjects differed from the native speakers not in *whether* the lexical tones were realized in varying focus conditions, but *how* they were realized.

Figure 13-18 show that, when observed in the same focus condition, the overall f0 contour of the H, R and F tones are similar between two groups. The biggest difference lies in the L tone, as mentioned in section 3.3.3.1. The L2 subjects produced the L tone the same way as the R tone, whereas the native speakers lower the minimum f0 of the L to a pitch level even lower than the minimum f0 of the F tone, forming a rather drastic dip.

This deviation in the L tone is not surprising. From a perspective of intrinsic characteristic of the L tone, it is the most vulnerable among the four in terms of preserving the complete underlying f0 contour. The canonical lexical form of the tone is only realized in isolated and stressed syllables or the end of the utterances. Tone sandhi of the L tone adds more challenge for the L2 learners to master the L tone. From the perspective of language acquisition, it has been proved in both L1 and L2 acquisition studies that the L and R tones are the last to be acquired. Therefore, the L tone is relatively hard to realize in the first place than the H and F tones.

Reason for confusion between the R and L tones is easily explained by the intrinsic similarities between the two. It has long been observed the f0 contour of the R and L tones are similar. Both tones have a portion of rise in the f0 curve, and the rise of the R tone only starts from the middle of the duration. When a L tone is produced without a fully realized dip in the contour, it would presumably look similar to the R tone. Same error pattern of confusing the L tone with the R tone can be found in previous L2 studies on Mandarin lexical tones (Miracle, 1989; Sun, 1998). The confusion error even exists in L1 acquisition (Chao, 1951; Li & Thompson, 1977; Clumeck, 1980).

Besides the most obvious deviation in the L tone, according to Figure 15-18, it can be seen that the H, R and F tones produced by the L2 subjects had a greater pitch range than those produced by the native speakers. This deviation can be found across all focus conditions. This

might imply a reason for the limited extent of variation in realizing Mandarin focus by the English L2 learners: Given the fact that “the English and Mandarin speakers have the same physical capabilities with respect to the rate of vocal fold vibration” (Keating & Guo, 2010), it’s reasonable to imagine that for the L2 subjects, when producing lexical tones in the neutral condition, they’ve already put so much effort to maximize distinguishable characteristics of the tones that further expansion is hard to realize. Whereas the native speakers spent so little effort in the neutral condition (in Figure 15, pitch range of the R and L tones are unexpectedly small) that larger extent of expansion is possible to produce whenever focus interplays. This conjecture can be indirectly confirmed by an experiment on Mandarin focus, where increase in pitch range from “Emphasis” to “MoreEmphasis” condition is not substantial because the lexical tones have been fully realized, therefore “maximally distinguishable from each other” (Chen & Gussenhoven, 2008). In the L2 production, the lexical tones have been realized to an extent like the “Emphasis” condition for the native speakers, physiological limitation on muscles controlling the pitch peak and valley made further raising or lowering relatively hard for the L2 subjects.

4.1.3 Research question 3

The third question focuses on tone-specific focus effect. Did focus intonation affect differently on four lexical tones? Did different lexical tones bring different levels of difficulty to the L2 learners in the production of focus?

In the hypotheses that tonal identities are hard to preserve, answer to this question was the pattern of substitution: The H and R tones should be easily replaced, whereas the L and F tones are easier to preserve. However, the average data have shown that the expected substitution didn’t occur. Therefore answer to this question now focuses on different extent of variation

brought by focus intonation between four lexical tones. The larger the extent of variation is, the easier to realize focus while producing the lexical tone.

The answer can be found in section 3.1: variation in pitch range, maximum and minimum f_0 of the R and L tones is less than the H and F tones. Based on Figure 15 and 16, compared to the neutral condition, the extent of expansion over pitch range with focus for the H, R, L and F tone are 187%, 32%, 27% and 42% respectively for the L2 subjects. Therefore, the English L2 subjects experienced more difficulty realizing focus while producing the R and L tones with respect to adjustment over pitch range. Since the L tone was produced almost the same as the R tone in the production of the L2 subjects in this experiment, it is the R tone that brings most difficulty.

There are some possible explanations for this difficulty. First, it's possible that for the English L2 learners, the R (or L) tone is already relatively hard to produce in the neutral condition, thus when focus interplays, the subjects spent more effort in realizing the f_0 contour of the R tone than expanding the pitch range. This is confirmed by the Tonal Markedness Scale (TMS) analyzed in Zhang (2007, 2013). It has been well proved that the R tone is the most disfavored tone ($*T2 > *T4 > *T1$) universally. Because most effort has been contributed to realizing the f_0 contour of the most disfavored R tone, limited adjustment according to focus conditions could be realized at the same time.

Second, the R tone has an intrinsically more restricted range of expansion under focus, meaning the native speakers would perform less expansion over the pitch range of R tone than the F tone. The idea of “tone-intrinsic expansion of f_0 range” is put forward by Chen & Gussenhoven (2008). Data of native speakers in their experiment shows “lexical tones exhibited different limits on the extent to which their F_0 range could be expanded, with HL [the F tone]

expanding more than LH [the R tone] and H [the H tone].” However, based on values from Xu’s experiment, shown in Figure 15 and 16, the pitch range of the H, R, L and F tones expanded 100%, 160%, 84% and 106% respectively. In Xu’s study (Xu, 1999), focus turns out to have more effect over pitch range of the R tone, which is contradictory to Chen & Gussenhoven’s prediction. The possibility of tone-intrinsic pitch range expansion needs future study to confirm.

In sum, answer to the third research question is that though all four lexical tones were realized under focus, focus intonation affected them differently. The R and L tones brought most difficulty to focus realization for the L2 subjects.

4.2 The nature of English focus

Though the predicted substitution error wasn’t found in average data, it did occur to a particular subject, S3. When producing with focus intonation, she replaced the H and R tones with a f0 contour the same as the F tone, which starts from a high pitch range and rapidly declines to a low pitch level. This is exactly the pattern of English pitch accent under the same intonation context, which is summarized as High Fall or H* or L+H* in traditional theories of English intonation.

Zhang’s dissertation (2013) suggests a similar result. She predicted substitution patterns based on the AM theory and found out that for monosyllabic focused constituents, “the accuracy rates of each tone type are 70% for T1, 45% for T2, 25% for T3 and 90% for T4”, and “the error pattern shows that English speakers are best at doing T4 for monosyllabic focus expression.” Her correct prediction confirms the nature of English focus intonation as a pitch accent.

Why did Xu reach the wrong conclusion? I think it’s because he confused the association between English pitch accent and the local stress with the relationship between the Mandarin

focus and the lexical tones. He proved that a local pitch peak existed in the neutral condition, which is independent from any intonation. He then showed that a significantly raised pitch peak under focus aligned with the local pitch peak without focus. So he concludes that focus is essentially a pitch range manipulation *function* that adjusts local pitch value accordingly, instead of a rigid *form* of pitch accent like L+*H followed by a L- phrase accent. However, Xu failed to exclude the possibility that the pitch accent of English focus only co-occurs with the local stress, or the local metrically strong syllable. Whether the higher pitch peak is a result of rising or just a form of a pitch accent cannot be determined from Xu's study. Now with evidence from English L2 production of Mandarin focus, the nature of English focus is proven indeed to be a pitch accent, or else no substitution error and less deviation should be found.

4.3 Pedagogical implication on L2 instruction

Result of the current study provides pedagogical implication for L2 instruction on both Mandarin lexical tones and Mandarin focus intonation.

As for teaching Mandarin lexical tones, the instructors should emphasize the low pitch level of the L tone. As suggested in Zhang (2013), since the complete 214 contour of the L tone only takes place in rather limited context, the default L tone is better introduced as a low tone (21) instead of a contour ending in high pitch level (214), which may avoid the confusion between the R and L tones found in the present study.

As for teaching Mandarin focus intonation, what can be inferred from this study are as follows. First, the L2 learners are able to acquire and realize Mandarin focus without instruction. In personal conversations, some of the subjects admitted that they never consciously thought about intonation in producing Mandarin before and they never received any instruction on that.

However, they successfully realized Mandarin focus and produced lexical tones at the same time. Secondly, to avoid potential negative transfer of English pitch accent, it is necessary to emphasize the identities of lexical tones must be preserved regardless of sentence intonation. Thirdly, based on individual comparison to the native speakers in section 3.4, no positive correlation is found between the ability to realize focus and the ability to produce accurate lexical tones. High proficiency in Mandarin lexical tones can't predict better performance in realizing Mandarin focus and vice versa. S1 is another subject that is bad in lexical tones. However, the extent of variation on duration, pitch range, maximum and minimum f0 values are quite similar to the native speakers. On the other hand, S2 almost failed to produce focus intonation while better pronouncing the lexical tones. A teaching plan for Mandarin focus instruction should consider this phenomenon. Lastly, in order to reach a more significant focus effect, the students shouldn't exaggerate their pronunciations in the neutral condition in the first place. It is possible that as the proficiency level of the L2 increases, the L2 learners would be less likely to over exaggerate their pronunciations, thus more natural intonation could be produced. Future studies on this possibility based on subjects from varying proficiency level might be able to provide reliable evidence.

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

This chapter begins with final conclusions in section 5.1 and then in section 5.2 points out limitations of the current study and some directions for future study.

5.1 Overall conclusions

First of all, the three research questions have been addressed in the experiment as well as the comparative analysis. The English L2 learners realized Mandarin focus by increasing the duration, raising the maximum f0 and expanding the pitch range of f0 contour. At the same time, most of them preserved identities of the lexical tones observed in the neutral condition. Among the four lexical tones, the F and H tones brought less difficulty than the R tone in realizing focus. Regardless of the focus condition, the L2 learners didn't significantly differentiate the R and L tones. Deviation in L2 production compared to the native speakers mainly lies in less extent of variation of relevant measurements under focus. Another interesting observation in the neutral condition is that lexical tones produced by the English L2 learners seemed to possess a wider pitch range, compared to that of the native speakers, as if the L2 learners were exaggerating their production before any focus intonation interplays.

Second, negative transfer effect from L1 (English) to L2 production found in the interlanguage of a particular subject, S3, resolves the controversy over the nature of English

focus. Contrary to traditional theory that English focus is a pitch accent, Xu believes English focus is fundamentally the same as Mandarin focus: it manipulates the pitch range, maximum and minimum f_0 values and duration. However, result of S3, where a consistent pattern of typical English pitch accent substitutes Mandarin lexical tones, disproves Xu's idea and shows that the nature of English focus is indeed a pitch accent.

5.2 Limitations of the current study and directions for future study

The present study is limited in the scope of experiment design and selection of the subjects mostly for the convenience of comparative analysis. It should be considered a preliminary study on the L2 production of mandarin focus. A more comprehensive and in-depth study should consider the following aspects:

First, the context for intonation realization should be more spontaneous. In the present study, the focus intonation is motivated by a precursor question and the focused constituent is underlined in the recording material. The subjects mostly read what is underlined in the sentence. In a more natural and spontaneous environment without reference to recording materials, the Mandarin focus produced by the English L2 subjects might show other patterns in realization.

Second, subjects with different levels of proficiency in Mandarin should be recruited for a comprehensive study. It is shown in this study that there might be no positive correlation between proficiency in MC and the ability to realize focus. Whether this is true can be confirmed with a larger number of subjects with wider range of proficiency level of MC. Besides, deviation from the native production can be further investigated and compared between each proficiency group.

Third, more measurements could be adopted for accurate comparison over the f_0 contours. The measurements used in this study, including duration, pitch range, maximum and minimum

f0, and slope, might not be able to depict the details of the f0 curves, though they are sufficient to address particular research questions of the study. Other measurements, such as the location of the pitch peak and valley, the size, speed and slope of the rise or fall of the contour, are able to give a fuller account of the realization of focus and lexical tones.

Fourth, the recording material could be more complex in terms of the combination of lexical tones. In this study, tone sandhi of the L tone is avoided. Lexical tones of the first, fourth and fifth syllables are consistently the H tone. By complicating adjacent local lexical tones, more interesting results might be found with respect to focus realization, as well as realization of lexical tones.

Last but not least, the position of the focus in a sentence might have potential influence on the focus realization. This is briefly mentioned in the study: it seems that extent of variation in word 2 is a bit greater than in word 1. For the S3, substitution took place with much less exception in word 2. According to phonological constraints that influence tonal production discussed in Zhang (2013), position should be a significant factor.

The direct implication of this study is for L2 instruction or SLA study of Mandarin intonation. It will be interesting to see how L2 learners from other tonal or non-tonal language background study focus intonation of MC. Thus general or language-specific instruction strategy could be developed in the future. Beyond the scope of learning MC, some questions in a broad realm of language acquisition inspired by results of the current study can be solved in the future. For example, the acquisition sequence of segmental and suprasegmental features. It has been generally accepted that vowels and consonants are learned before and better than lexical tones. The acquisition sequence of the intonation, such as focus, compared to them needs further

exploration, based on the correlation that lacks positive correlation between performance in lexical tone and focus intonation observed in this study.

The result of this thesis also sheds light on studies on language prosody in general.

First, from a methodological perspective, this study provides a method for exploring prosodic features of a language. Just as the nature of English focus as pitch accent was confirmed as a negative transfer in the interlanguage, if controversy over properties of a certain prosodic feature of a language exists, a study on the interlanguage with this language as L1 might be able to reveal some insights by looking at its potential transfer effect. This method of studying L1 using related interlanguage might be useful in general, not only in prosodic features.

Second, from the perspective of focus, as a type of intonation, this study suggests that even though languages share this same type of intonation, it is realized differently. In Mandarin, it is a communicative *function* that manipulates phonetic features, such as duration, pitch range, pitch peak, etc. In English, it is a *form* of pitch accent with higher pitch peak, wider pitch range and longer duration. Based on this implication, some questions need to be addressed in future studies: 1) How does focus intonation work in other languages? Can its realizations be categorized as either a function, or a form, or even something else? 2) Given that focus can be realized as a function or a form, how is this property correlated with other prosody features of the language? Is it possible that focus of tonal languages tends to be a function as in Mandarin, because tonal features need to be preserved to differentiate lexical meanings, whereas focus of non-tonal languages such as English, is realized as a form, because surface f0 directly reflects the sentence intonation? Is the nature of focus predictable given other prosodic properties of the language?

In general, this thesis opens up new research directions for studies on L2 instruction, SLA of MC, SLA of language intonation, and language prosodic features.

REFERENCES

- Bolinger, D. (1958). A theory of pitch accent in English. *Word* 14: 109-49.
- Chao, Y-R. (1930). "A system of tone-letters". *Le maitre phonetique* 45.24-27. (Reprinted in English orthography in *Fangyan* 1980, 2: 81-83.)
- Chao, Y. R. (1951). The Cantian idiolect: An analysis of the Chinese spoken by twenty-eight-months-old child. In W.J.Fischel (ed.), *Semitic and Oriental Studies* (pp 27-44). Berkeley: University of California Press.
- Chao, Y-R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, M. (2000) *Tone sandhi: patterns across Chinese dialects*. (Cambridge Studies in Linguistics 92.) Cambridge: Cambridge University Press.
- Chen, Y., & Gussenhoven, C. (2008). Emphasis and tonal implementation in Standard Chinese. *Journal of Phonetics*, 36(4), 724-746.
- Cruttenden, A. (1997). *Intonation* (2nd ed.). Cambridge: Cambridge University Press.
- Elliot, C.E (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawaii Working Papers in ESL*, 10(2), 177-204.
- Flemming, E. (2008) The role of pitch range in focus marking. Slides from a talk given at the Workshop on Information Structure and Prosody, Studiecentrum Soeterbeeck (2008). <http://web.mit.edu/flemming/www/paper/FlemmingSoeterbeeck.pdf>
- Gårding, E. (1987). Speech act and tonal pattern in Standard Chinese: constancy and variation. *Phonetica*, 44(1), 13-29.
- Halliday, M. A. K. 1967. *Intonation and grammar in British English*. Mouton.
- Jackendoff, Ray. 1972. *Semantic interpretation in generative grammar*. Cambridge, MA: MIT Press.
- Jin, S. (1996). Acoustic Study of Sentence Stress in Mandarin Chinese.
- Ladefoged, P., & Johnson, K. (2012). *A course in phonetics*. Cengage learning.
- Ke C. (2012) Research in Second Language Acquisition of Chinese: Where We Are, Where We Are Going. *Journal of the Chinese Language Teachers Association*, 47, 43-113
- Keating, P., & Kuo, G. (2012). Comparison of speaking fundamental frequency in English and Mandarin. *The Journal of the Acoustical Society of America*, 132(2), 1050-1060.

- Kochanski, G., & Shih, C. (2003). Prosody modeling with soft templates. *Speech Communication*, 39(3/4), 311–352.
- Ladd, D. R. (1980). *The structure of intonational meaning: evidence from English*. Bloomington: Indiana University Press.
- Ladd, D. R. (1996). *Intonational phonology*. (Cambridge Studies in. Linguistics 79.) Cambridge: Cambridge University Press.
- Liao, R. (1994). *Pitch contour formation in Mandarin Chinese: a study of tone and intonation* (Doctoral dissertation, The Ohio State University).
- Li, C. N., & Thompson, S. A. (1978). The acquisition of tone. In V. A. Fromkin (Ed.), *Tone: a linguistic survey*. New York: Academic.
- Lu, J-M (1992). On the perception and production of Mandarin tones by adult English- speaking learners. *Dissertation Abstracts International* 53(7), 2350A (University Microfilms No.9233909)
- Miracle, C (1989). Tone Production of American Students of Chinese: A Preliminary Acoustic Study. *Journal of the Chinese Language Teachers' Association*, Volume XXIV: No.3, October 1989, pp.49-65
- O'Connor, J. D., & Arnold, G. F. (1961). *Intonation of colloquial English*. London: Longmans.
- Pierrehumbert, J.B. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.
- Pierrehumbert, J. and Hirschberg, J. (1990) The Meaning of Intonation in the Interpretation of Discourse, in P. Cohen, J. Morgan, and M. Pollack, (eds). *Intentions in Communication*, MIT Press: Cambridge, MA.
- Rasier, L., & Hiligsmann, P. (2007). Prosodic transfer from L1 to L2. Theoretical and methodological issues. *Nouveaux cahiers de linguistique française*, 28, 41-66.
- Shen, Xiaonan S (1992). On tone sandhi and tonal coarticulation. *Acta Linguistica Hafniensia* 24:131-152
- Shih, C. (1988). Tone and intonation in Mandarin. *Working Papers of the Cornell Phonetics Laboratory*, 3, 83-109.
- Sun, Sylvia Henel (1998). *The Development of a Lexical Tone, Phonology in American Adult Learners of Standard Mandarin Chinese*. University of Hawaii Press.
- Tseng, C. (1981). *AN ACOUSTIC PHONETIC STUDY ON TONES IN MANDARIN CHINESE*. (Order No. 8215629, Brown University). *ProQuest Dissertations and Theses*
- Wan, I-Ping and Jaeger, Jeri (1998). Speech errors and the representation of tone in Mandarin Chinese". *Phonology* 15 (1998) 417-461

- Wang, Y. (2006). L2 Acquisition and Processing of Mandarin Tone. Chapter in E. Bates, L.H. Tan, and O. Tseng (eds.) *Handbook of Chinese Psycholinguistics*. Cambridge University Press
- Xu, Y. (1999). Effects of tone and focus on the formation and alignment of f₀ contours. *Journal of Phonetics*, 27, 55–105.
- Xu, Y., & Emily Wang, Q. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication*, 33(4), 319-337.
- Xu, Y., & Xu, C. X. (2005). Phonetic realization of focus in English declarative intonation. *Journal of Phonetics*, 33(2), 159-197.
- Xu, Y. (2013). ProsodyPro — A Tool for Large-scale Systematic Prosody Analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), Aix-en-Provence, France. 7-10.
- Yip, M. (1980) *The Tonal Phonology of Chinese*, PhD dissertation, MIT. Published (1990) Garland Publishing, New York.
- Yip, M. (2001). Tonal features, tonal inventories and phonetic targets, in *UCL Working Papers in Linguistics* 13, C. Iten and A. Neeleman, editors, Department of Phonetics and Linguistics, University College London.
- Zhang, H. (2007) A phonological study of second language acquisition of Mandarin Chinese tones. Master's thesis, University of North Carolina at Chapel Hill.
- Zhang H. (2013) The Second Language Acquisition of Mandarin Chinese Tones by English, Japanese and Korean Speakers[D]. University of North Carolina at Chapel Hill,