

Chunsheng Yang

Acquiring the pitch patterns of L2 Mandarin Chinese

Abstract: This study examines the acquisition of utterance-level pitch patterns in Mandarin Chinese by American second language (L2) learners. It is an exploratory study with the goal of identifying the utterance-level prosody in L2 Mandarin Chinese. The focus of this study is not on the pitch patterns of individual learners but those of subject groups. The analysis shows that the pitch patterns between two syntactic structures for the same tone sequence vary with the tone sequence and the subject group. The biggest difference between first language (L1) and L2 Mandarin Chinese lies in the frequency of target undershoot in L2 speech. The infrequent tone target undershoot in L2 speech, especially among the intermediate learners, was attributed to the incomplete acquisition of L2 prosody. It was argued that the infrequent tone target undershoot may render L2 speech more staccato or robot-like, which contributes to the perception of a foreign accent in L2 Mandarin Chinese.

Keywords: prosody acquisition, Mandarin Chinese, L2 learners, F0, tones, tone target undershoot

Chunsheng Yang: E-mail: chunsheng.yang@uconn.edu

1 Introduction

The acquisition of second language (L2) segments has been extensively examined in the literature (Aoyama, Fledge, Guion, Akahane-Yamada, and Tsuneo 2004, Best 1995, Flege, Bohn and Jang 1997, Fledge, MacKay and Meador 1999, Hojen and Flege 2006, Ingram and Park 1997, Sheldon and Strange 1982, Tsukada, Birdsong, Bialystok, Mack, Sung and Fledge 2005, among many others). Prosodic acquisition (i.e., the acquisition of stress, intonation, and so on) in many languages other than English (McGory 1997, Archibald 1995, 1997, 1998a, b), however, is a rather under-studied area (Gut 2003, Nguyen et al. 2008). With respect to Mandarin Chinese, many studies were conducted on the acquisition of tones by L2 learners, due to the fact of Chinese being a tone language. However, very limited studies have examined the acquisition of Mandarin prosody beyond the level of lexical tones (Chen 2000, Viger 2007, Yang and Chan 2010).

This study investigates the acquisition of utterance-level pitch patterns by American L2 learners. Complementing the studies on the acquisition of tones, this study will have important pedagogical implications for the teaching of Chinese as a Second Language (CSL).

2 Mandarin and English prosody

2.1 Mandarin prosody

Typologically speaking, Mandarin Chinese is a tone language, in which tones are lexically specified, namely the fundamental frequency (F0)¹ or pitch pattern over a syllable can be used to distinguish the meaning of words. There are four lexical tones in Mandarin Chinese. They are often referred to as Tone 1 (high level tone, referred to as T1 hereafter), Tone 2 (mid-rising tone, referred to as T2 hereafter), Tone 3 (low-dipping tone, referred to as T3 hereafter), and Tone 4 (high-falling tone, referred to as T4 hereafter). Chao (1930) designed a five-level numerical scale for representing pitch height, with “1” being the lowest pitch value and “5” the highest pitch value within a speaker’s pitch range. In this scale, T1 through T4 are represented as “55”, “35”, “214”, and “51”, respectively. The low-dipping variant of T3 (“214”) only occurs in isolation or at utterance-final position. At non-utterance-final position, T3 often surfaces as a low tone (“21” or “22”). Moreover, when there are two consecutive T3’s, the first T3 often undergoes tone change, namely tone sandhi, and becomes a rising tone. Thus, the numerical values for T3 in actual speech can be “214”, “21”² (or “22”), or “35”. In the Pinyin system, the official Romanization system used in the People’s Republic of China since 1958, and widely adopted throughout the world, lexical tones are marked with iconic diacritics above the letters representing vowels, as in ā, á, ǎ, à. Table 1 lists the four lexical tones and their names, together with the tone values in numbers and tone diacritics in the examples.

In addition to the four lexical tones, there is also a neutral tone (轻声 qīngshēng) in Mandarin Chinese. The neutral tone (T0), which occurs on an unstressed (atonic) short syllable at non-initial position in a word or phrase, must

1 Fundamental frequency (F0) is the rate of the vocal fold vibration of a speaker per second. Pitch is the perceived fundamental frequency. Even though fundamental frequency and pitch are not exactly the same, especially at higher frequency, they are used interchangeably in this paper.

2 The low tone (“21”) also frequently surfaces at utterance-final position in Beijing Mandarin.

Table 1: The four lexical tones in Mandarin Chinese

Tone	Tone name	Tone values	Example
T1	High level tone	55	mā, 妈, ‘mother’
T2	Mid rising tone	35	má, 麻, ‘hemp’
T3	Low dipping tone	214	mǎ, 马, ‘horse’
T4	High falling tone	51	mà, 骂, ‘to scold’

be preceded by at least one syllable that carries one of the four lexical tones (Lin 2007: 98).

Besides tones, intonation, the F0 pattern over a phrase or an utterance, is also an integral part of Mandarin phonology. In this sense, F0 plays a dual role in Mandarin Chinese, namely, representing both lexical tones and intonation.

2.2 English prosody

Within the framework of Autosegmental-Metrical (AM) phonology (Pierrehumbert and Beckman 1986, Beckman 1996, Ladd 1996, 2008), English prosody can be described with three tone³ events: pitch accent, phrase accent, and boundary tone. Pitch accent falls on stressed syllables, marking the prominence of the syllables (H*, L*, L+H*, etc.). With respect to prosodic phrasing in English, there are two levels of phrases, namely, intermediate phrase (ip) and intonational phrase (IP). An intermediate phrase has one pitch accent or more, together with a phrase accent (H-, L-). The phrase accent has a scope over the entire prosodic phrase (Pierrehumbert & Hirschberg 1990: 302). The phrase accents are associated with coherence and cohesion in the discourse. There are two types of phrase accents in English. The high phrase accent (H-) indicates that “the current [prosodic] phrase is to be taken as forming part of a larger composite interpretive unit with the following phrase,” whereas the low phrase accent (L-) emphasizes the separation of the current phrase from the subsequent phrase. An intonation phrase may have one or more intermediate phrases and ends with a boundary tone, which can be either high (H%) or low (L%).

³ Tone here, different from the tone in Mandarin Chinese, is the specific pitch pattern on a syllable (or syllables).

3 Previous studies on the acquisition of Mandarin prosody

Most studies on the acquisition of Mandarin phonology focus on the acquisition of lexical tones (Bent 2005, Chen 1997, 2000, Miracle 1989, Shen 1989, White 1981, and so on). Very few studies to date examine the acquisition of utterance-level prosody of Mandarin Chinese by L2 learners. To the best of my knowledge, Chen (2000) and Viger (2007) are the only studies that examine American L2 learners' acquisition of utterance-level prosody in Mandarin Chinese. Chen (2000) analyzed the tone errors above the word level by asking judges to evaluate the accuracy of tones in spontaneous speech produced by advanced L2 learners. Chen proposed three mechanisms to account for the tone errors in his study: 1) transfer of English intonation and sentence stress; 2) avoidance strategy (i.e., avoiding some tones), and 3) consistent substitution of a particular tone for another. Chen identified cases of interference from English intonation and stress, and some developmental patterns that had nothing to do with the first language. Even though Chen (2000) claimed that his analysis of tone errors went beyond word-level tone errors, his analysis did not differ significantly from previous analyses of tone errors in L2 Mandarin Chinese. The fact that tone errors tend to occur more frequently when tones are concatenated in an utterance than in isolation suggests that the analysis of the utterance-level prosody should go beyond the analysis of tone errors at the word level. Viger (2007) examined the acquisition of Mandarin Chinese utterance-level prosody by American L2 learners by comparing the pitch patterns at the end of utterances in L1 and L2 Mandarin. Viger found that L2 Mandarin Chinese exhibited a striking absence of the utterance-level prosody that occurred in L1 Mandarin. For example, L1 Mandarin speakers produced a global raise of pitch throughout Mandarin echo and yes/no questions, but that was not exhibited in the speech of L2 learners. Viger also found the lack of transfer of global English prosodic contours into L2 Mandarin speech. Viger, however, did find some transfer effects on utterance-final syllables, namely, L2 Mandarin learners imported a final rise on the final syllable in Mandarin echo and yes/no questions. A follow-up question to Viger's study is whether there is any L1 English prosodic interference at other positions of an utterance in L2 Mandarin.

To bridge gaps in the previous research, this study investigates the acquisition of pitch patterns of Mandarin Chinese by American L2 learners.

4 Methodology

4.1 Stimuli

Read speech was used in order to control for the tones and syntactic structures in the data collected. When preparing the stimuli, we used all four lexical tones in Mandarin Chinese. Even though the four lexical tones could be combined in various ways, we only used three types of tone sequences consisting of these tones, namely compatible tone sequences, conflicting tone sequences and the other tone sequence. Drawing on Xu's study (1994), in a compatible tone sequence, the target⁴ at the offset of the preceding tone and the target at the onset of the following tone are identical (i.e., both are H targets or both are L targets), whereas, in a conflicting tone sequence, the target at the offset of the preceding tone and the target at the onset of the following tone are different (i.e., if one is a H target, the other is a L target; alternatively, if one is a L target, the other is a H target).

Examples for each type of tone sequence are given below.

(1) Compatible tone sequence

T2T4 alternating sequence: e.g., 罗燕 谈论 名利。
Luó Yàn tán lùn míng lì
 Surname talks fame and profit
 “Luo Yan talks about fame and profit.”

(2) Conflicting tone sequence

T2 sequence: e.g., 刘明 来 拿 羊毛。
Líu Míng lái ná yángmáo
 Surname come get wool
 “Wang Ming comes to get wool.”

⁴ Tone targets refer to the phonological components of tones. T1 has high target only; T2 has low and high targets; T3 has low target or both low and high targets; and T4 has high and low targets. Tone target undershoot refers to the phenomenon that the target in a tone is not realized or only partially realized. For example, a rising tone (T2) may surface as a level tone, due to the low target undershoot.

(3) The other tone sequence

T1 sequence⁵: e.g., 殷安 轻 摸 猫咪。
Yīn Ān qīng mō māomī.
 Surname gently touch kitty
 “Yin An gently pets a kitty.”

For each specific tone sequence, two pairs of sentences were used. The two sentences in each pair are identical segmentally and tonally, but differ syntactically. For example, the sentence in (4) (referred to as SS-33 hereafter) and that in (2) (referred to as SS-24 hereafter) are such a pair. Throughout this paper, “SS” refers to “syntactic structure”. The two numbers that follow “SS” code the number of the syllables (the first for the subject and the second for the predicate). Hence, “SS-24” refers to the sentence that has a disyllabic subject and a quadrisyllabic predicate, while “SS-33” refers to the corresponding sentence in the pair that has a trisyllabic subject and a trisyllabic predicate.

(4) Syntactic Structure-33 [SS-33] [刘明莱]_{subject} [拿 羊毛]_{predicate}。
 [Liú Mínglái]_{subject} [ná yángmáo]_{predicate}
 Surname get wool
 “Liu Minglai gets wool.”

In preparing the sentences, efforts were made to make sure that all the segments in the sentences were sonorants. Only when it was impossible to use only sonorants were obstruents used. Nonetheless, the use of obstruents was kept to the minimal so that the F0 track could be continuous. Altogether there were 10 pairs of sentences used in this study. All 20 sentences were embedded in conversational scenarios in recording. One example of a scenario is given in (5).

(5) Míngtiān de gōngzuò ān pái le ma
 A: 明 天 的 工 作 安 排 了 吗 ?
 ‘Is tomorrow’s work arranged for?’
 Ān pái le Wáng Míng lái ná yángmáo
 B: 安 排 了。 王 明 来 拿 羊 毛。
 ‘Yes. Wang Ming will come to get the wool.’

⁵ T1 only has a high (H) target. Thus, the T1 sequence is a compatible tone sequence as well. However, there is only one target involved in T1, while there are two targets (H and L) in all the tones in both the compatible tone sequences and conflicting tone sequences. That is why the T1 sequence falls within the category of “other tone sequences” in this study.

Wáng Míng lái ná yángmáo. Hǎo wǒ qù zhǔnbèi yíxià
 A: 王明 来拿羊毛? 好, 我去准备一下。
 “Wang Ming comes to get the wool? Ok. I will go to prepare for it.”

The scenarios containing the 20 target sentences⁶ were randomized, with one practice scenario and one attention-check scenario⁷ placed at the beginning and at the end respectively. Each line of script in a scenario was coded as either A or B, as shown in (5). In order to facilitate the recording, each subject was only responsible for either line A or line B in the whole process of recording. There were two blocks of recording scenarios and they were all the same except for the line coding. Each block consisted of 44 scenarios, with the same 22 scenarios repeated twice. The two blocks differed only in the line coding, namely the same lines in the first block had the opposite line coding as compared to those in the second block. Thus, each and every subject produced two renditions for every single target sentence. In data analysis, only the second rendition was used, except for four cases in which the second rendition had severe disfluency and the first rendition was used.

When the reading scenarios were presented in recording, the person’s names were underlined. The reason for doing so was to make sure that the subjects were aware that the underlined part was a proper noun of personal name. Due to the particular design in this experiment, it was not very evident whether the first two or the first three syllables formed the name of a person.

4.2 Subjects

Ten intermediate-level and ten advanced-level American L2 learners were recruited to participate in the study. These two different levels of learners, intermediate and advanced, were recruited to examine whether there was any learning effect between them. The L2 learners were recruited from the American L2 learners in a mid-western state university. The control group consisted of ten Beijing Mandarin speakers who were recruited from the Chinese community at the same university. The ten native Mandarin speakers were all born and grew up in Beijing

⁶ See Appendix for the list of the experimental recording materials used in this study.

⁷ A practice scenario was placed at the beginning of all the scenarios containing the target sentences, in order to help subjects get ready for the recording. An attention-check scenario was placed at the end of all the scenarios containing the target sentences to make sure that subjects were still attentive during the recording of the penultimate scenario.

before coming to study in the United States. All subjects reported no speech or hearing problems. They were paid \$10 for their participation in the recording.

4.3 Recording procedure

The recording was conducted in a sound-attenuated studio. Prior to recording, the researcher gave instructions to the subjects. Then the subjects signed the consent form. The subjects could ask any question related to the study. They could withdraw from the recording any time without any penalty.

In the process of recording, two subjects (all the recording pairs were matched for language status, namely a native speaker pairing with another native speaker, and a learner pairing with another learner at the same level) worked together in the recording studio. The two subjects sat comfortably behind their microphones. Two microphones were placed side by side behind a 13.3" MacBook Pro laptop. The recording materials were presented through PowerPoint slides on the laptop. The investigator sat in front of the recording control panel outside the recording studio. The progress of the slides was controlled by the investigator through a remote control. If a subject misread one syllable or syllables segmentally, or had severe disfluency in his/her recording, he/she would be asked to read the whole scenario again with his/her partner. The recording input was digitized at 44.1 kHz with a 24-bit resolution.

There was a five-minute break between recordings of the two blocks. The whole recording lasted less than 30 minutes. On each PowerPoint slide, the *pinyin* Romanization was placed above each character to facilitate learners' character recognition. However, the font size (size 20) of the *pinyin* Romanization was intentionally smaller than that of the characters (size 32), thus enabling the subjects, especially the learners, to focus on the characters and not the *pinyin* Romanization.

4.4 Data transcription

The speech data were transcribed in accordance with the Pan-Mandarin ToBI transcription system (Peng et al. 2005). Only four tiers of the eight tiers proposed in the Pan-Mandarin ToBI system were transcribed in this study. The four tiers are: morpheme (in Chinese character), pinyin Romanization (with citation tones), tone targets (H and L targets on each syllable), and break indices. All the transcription criteria were kept consistent throughout the transcription.

To guarantee the reliability of break transcription, breaks in the T2 and T4 sequences were tested for intra- and inter-transcriber reliability. For the intra-transcriber reliability test, the breaks in T2 and T4 sequences were transcribed by the author twice, with a one-and-a-half month's interval between the two transcriptions. The transcription agreement rate between them was 96.8%. For the cases of discrepancy between the two transcriptions, the original utterance was listened to again and the final transcription was determined. For the inter-transcriber reliability test, another transcriber who was a Chinese doctoral student of phonetics, was hired to transcribe the same subset of data independently. The transcription agreement rate between her transcription and the author's transcription was 93.5%. For the cases of discrepancy between the two transcribers, the two transcribers discussed those cases and came up with a final transcription.

4.5 Acoustic measurement

The F0 values were measured on all syllables in different tone sequences. The F0 values were measured by running a Praat script (Boersma and Weenink 2009) by Xu (2005–2010). This script measured the ten F0 values on the voiced portion of each labeled syllable by dividing the voiced portion into ten equal intervals. The F0 values obtained this way are time-normalized (i.e., the duration of the tone-carrying syllables was treated as the same across all subjects) and can represent the F0 contours of the whole utterance.

5 Data analysis of pitch patterns in L1 and L2 Mandarin Chinese

In the analysis of pitch patterns, we made sure that SS-24 and SS-33 for the same tone sequence had the same breaks at all positions except for the syl-2 and the syl-3 positions. In addition to controlling for the breaks in SS-24 and SS-33, we also controlled for tones and used only the utterances that had the correct surface tones (sandhi tones were considered correct). That is to say, only utterances with the correct surface tones and correct breaks were used in the analysis. The mean time-normalized F0 values were obtained by averaging all the F0 values at each syllable position in the utterances with correct tones and breaks produced by all subjects in each group. The mean time-normalized F0 values on each syllable for SS-24 and SS-33 in the three subject groups were plotted, with male and female

subjects separated. In these figures, the x-axis displays the syllable position; the y-axis displays the F0 values (Hz). The legend on the right of each figure shows the different syntactic structures.

5.1 F0 patterns in the other tone sequence

The F0 patterns in the T1 sequence were analyzed first. Figures 1 to 3⁸ display the F0 contours of the T1 sequence for the three male subject groups.

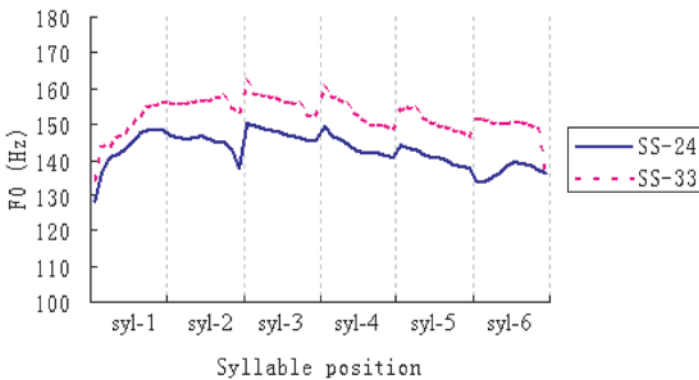


Fig. 1: F0 contours in the T1 sequence: Male native speakers

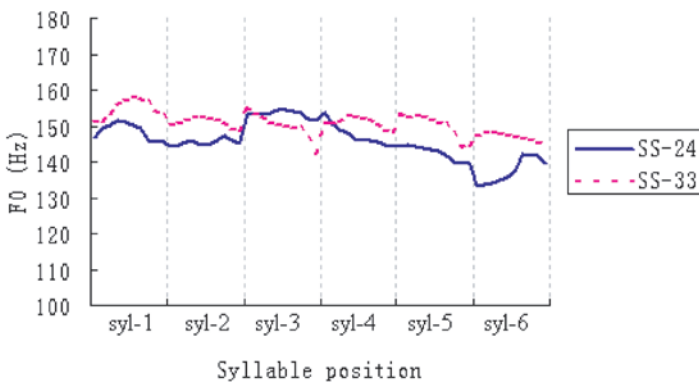


Fig. 2: F0 contours in the T1 sequence: Male advanced learners

⁸ Due to space limit, the F0 contours in the three female groups were not presented. However, they have similar patterns to the male groups.



Fig. 3: F0 contours in the T1 sequence: Male intermediate learners

Before we discuss the F0 contours in the T1 sequence, it is worth pointing out that the discussion in this paper is mainly based on the comparison and contrast of the surface F0 contours of the same tone sequences across subject groups. From Figures 1–3 it can be seen that there is overall F0 declination in the F0 contours of the T1 sequence by each subject group and that there is no significant difference in the F0 contours between SS-24 and SS-33 and across all subject groups, except that the F0 contours are often higher in SS-33 than in SS-24. However, there are some minor differences across groups. For the native group, the F0 contour in SS-33 is consistently higher than that in SS-24, whereas for the two learner groups, only portion of the F0 contour in SS-33 is higher than that in SS-24. It is worth mentioning there are individual differences in the surface F0 patterns across groups, especially in the two learner groups. However, due to the focus on the overall pitch pattern in L1 and L2 Mandarin Chinese, the F0 difference among individuals in a group will not be further explored in this paper.

5.2 F0 patterns in the compatible tone sequences

Two pairs of compatible tone sequences (i.e., the T2T4 alternating sequence and the T4T2 alternating sequence) were used in this study. Figures 4–6 display the F0 contours in the T2T4 sequence for the three male subject groups.

Figures 4–6 show that F0 patterns in SS-24 and SS-33 are very similar in all three subject groups. However, the F0 contours produced by the native group are fairly smooth in the sense that the F0 contours are composed of regular falls and rises with little local irregularities, while the F0 contours by the two learner groups are not as smooth as those by the native group. This difference could be

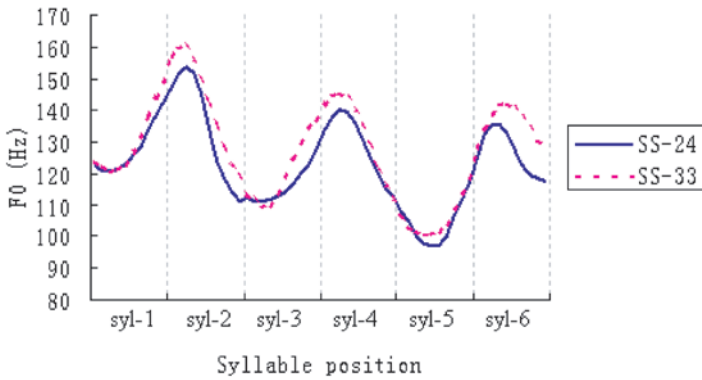


Fig. 4: F0 contours in the T2T4 sequence: Male native speakers

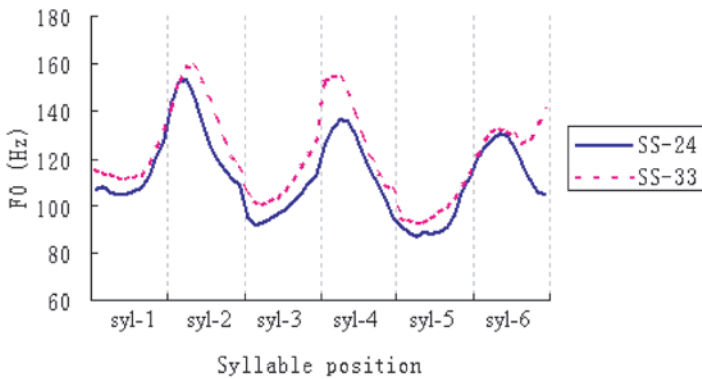


Fig. 5: F0 contours in the T2T4 sequence: Male advanced learners

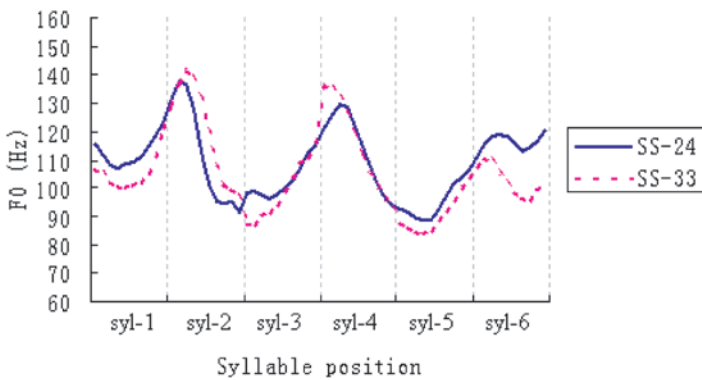


Fig. 6: F0 contours in the T2T4 sequence: Male intermediate learners

attributed to the removal of some utterances containing tone errors in the learner groups when the F0 contours were plotted. The F0 patterns in the T4T2 sequence are similar to the patterns in the T2T4 sequence, namely the F0 patterns in SS-24 and SS-33 are similar in all three subject groups.

5.3 F0 patterns in the conflicting tone sequences

Unlike the F0 patterns in the T1 sequence and in the compatible tone sequences, the F0 patterns in the conflicting tone sequences are different between SS-24 and SS-33 and across subject groups. Figures 7, 9, and 10 display the F0 contours for the T2 sequence in the three male subject groups.

Figure 7 displays the F0 contours in the T2 sequence by the male native speakers. It can be seen from the figure that, due to tone co-articulation (i.e., the effect of adjacent tones on the realization of a specific tone), the low targets on the syl-4 in SS-24 and on the syl-2 in SS-33 were not realized, namely, low target undershoot occurred. Figure 8 presents the F0 contour of an utterance containing the T2 sequence produced by a male native speaker. The low target undershoot can be clearly observed on the syl-2 position in Figure 8.

Figure 9 displays the F0 contours in the T2 sequence by the male advanced learners. It can be seen from the figure that there is low tone target undershoot at the syl-2 position in SS-33 and at the syl-4 position in SS-24, patterning similarly to the native group. Note that the rising contour of the T2 at the syl-2 position is somewhat maintained, even though perceptually the tone at the syl-2 position would approximate to a high flat tone. However, an examination of the productions of the T2 and the T4 sequences by individual advanced learners showed that

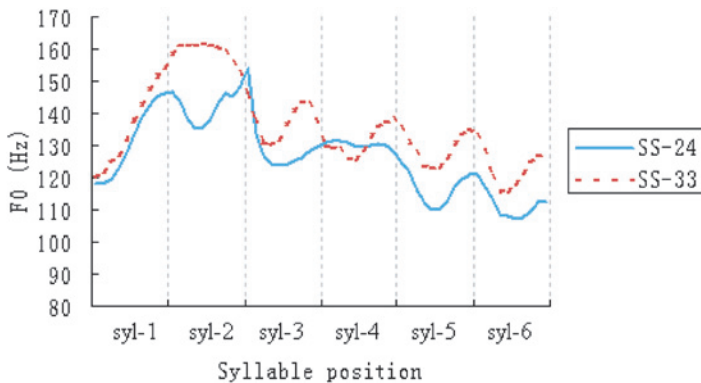


Fig. 7: F0 contours in the T2 sequence: Male native speakers

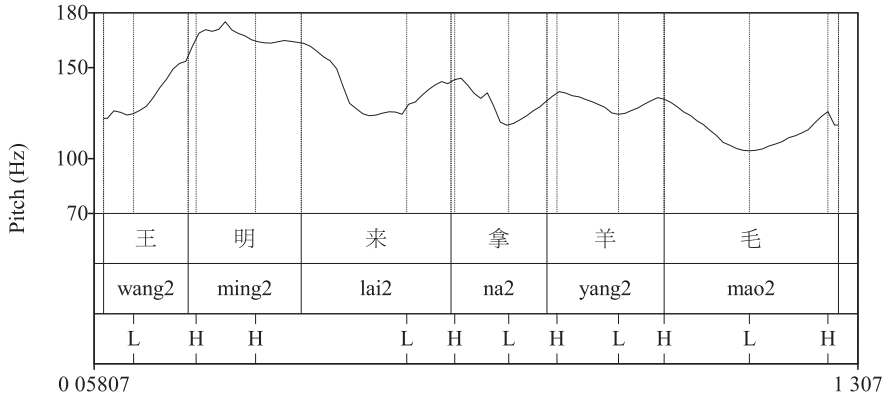


Fig. 8: F0 contour in the T2 sequence (SS-33): An utterance produced by a male native speaker

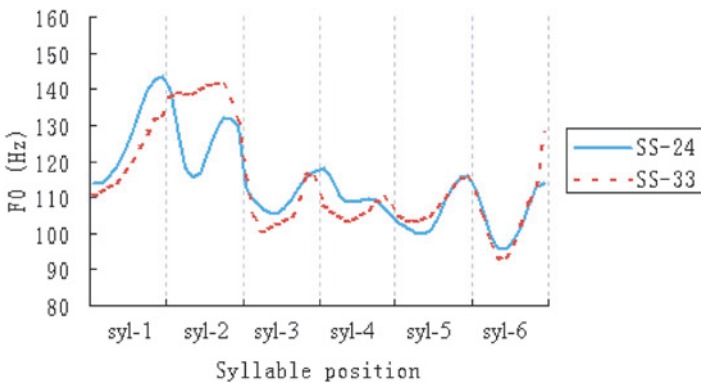


Fig. 9: F0 contours in the T2 sequence: Male advanced learners

the advanced learners did not produce the tone target undershoot as frequently as the native speakers.

Figure 10 displays the F0 contours in the T2 sequence produced by the male intermediate learner group. As shown in the figure, there is low target undershoot at the syl-2 position in SS-33. There is slight target undershoot at the syl-4 position in SS-24 and at the syl-5 position in SS-33. However, the rising F0 contours are clearly maintained at those two positions.

The similar F0 patterns were also found in the T4 sequence. Namely, low target undershoot occurred more frequently in the native group and the advanced learner group than in the intermediate learner group. Specifically, the target

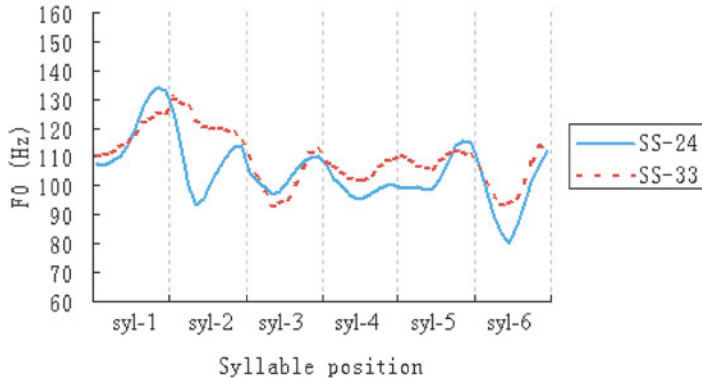


Fig. 10: F0 contours in the T2 sequence: Male intermediate learners

undershoot tended to occur more often at the syl-2 position in SS-33 and the syl-4 position in SS-24, namely, the phrase-medial positions of each prosodic phrase.

Figures 7, 9 and 10 also show that for both the native and advanced learner groups, the tone target undershoot tended to occur in both the first and the second prosodic phrases, while for the intermediate learner group, the tone target undershoot tended to occur only in the first prosodic phrase. The positional pattern of the tone target undershoot in the three subject groups shows the learning effect on the advanced learner group.

6 Summary and discussion

In this section, we will summarize the findings in this study and provide a general discussion.

In this study, we examined the pitch patterns in three types of tone sequences, namely, the compatible tone sequences, the conflicting tone sequences, and the other tone sequence. The analyses of pitch patterns showed that there was difference in the realization of tones in different tone sequences across subject groups. In the T1 sequence and the compatible tone sequences, there is nearly no difference in pitch patterns between SS-24 and SS-33 and across subject groups. In the T1 sequence, the pitch patterns of SS-24 and SS-33 are almost flat, with some declination. In the compatible tone sequences, the pitch patterns of both SS-24 and SS-33 are alternating falling and rising patterns, without tone co-articulation. However, the pitch patterns between SS-24 and SS-33 in the conflicting tone sequences are different across groups. More specifically, the native group and the advanced learner group tended to produce more tone co-articulation or tone

target undershoot than the intermediate learner group. The difference in tone target undershoot in different tone sequences shows that tone targets were realized differently in different phonetic contexts across subject groups. In a compatible tone sequence, the target at the offset of a preceding tone and the target at the onset of the following tone are identical. Thus, there is little possibility of tone target undershoot. That is why the surface F0 contours between SS-24 and SS-33 are almost the same in the compatible tone sequences and in the T1 sequence in all subject groups. However, in a conflicting tone sequence, the target at the offset of a preceding tone and the target at the onset of a following tone are different. The sudden change of the tone targets on the adjacent syllables posed more difficulty in F0 production for the learners, especially the intermediate learners, than for the native speakers. It is worth pointing out that learners in both the intermediate and advanced groups produced some tone errors in such sequences.⁹ The sudden change of tone targets on the adjacent syllables in the conflicting tone sequences posed difficulty for the native speakers as well. However, the difficulty in F0 production led to the frequent target undershoot in the native speech. However, even if the conflicting tone sequences posed great difficulty in F0 production for the learners, especially for the intermediate learners, they still attempted to maintain the underlying tone targets and produce each tone as fully as possible, with infrequent tone target undershoot. The difference in the conflicting tone sequences across subject groups suggests that the learners, especially the intermediate learners, have not fully acquired the F0 production of L2 Mandarin Chinese. The finding of the difference in tone co-articulation in different tone sequences also warranted the use of different tone sequences in that without the further categorization of tone sequences, it would be nearly impossible to identify the above patterns.

The similar patterning in F0 patterns between the native group and the advanced learner group shows the learning effect in L2 prosody, namely as learners' proficiency improves, their prosodic production becomes more native-like.

Our analysis of pitch patterns in L1 and L2 Mandarin shows that the main difference between L1 and L2 Mandarin lies in realization of the tone targets. A follow-up question to ask is whether the infrequent target undershoot in the L2 speech matters. The production of tone target undershoot is the result of interpolation of the tone targets on the adjacent syllables. In natural speech, especially in the speech by Beijing Mandarin speakers, such tone target undershoot is very prevalent. Figure 11, cited from Wu (1996), shows the F0 contour of an interroga-

⁹ The patterns of these tone errors will be reported in another paper.

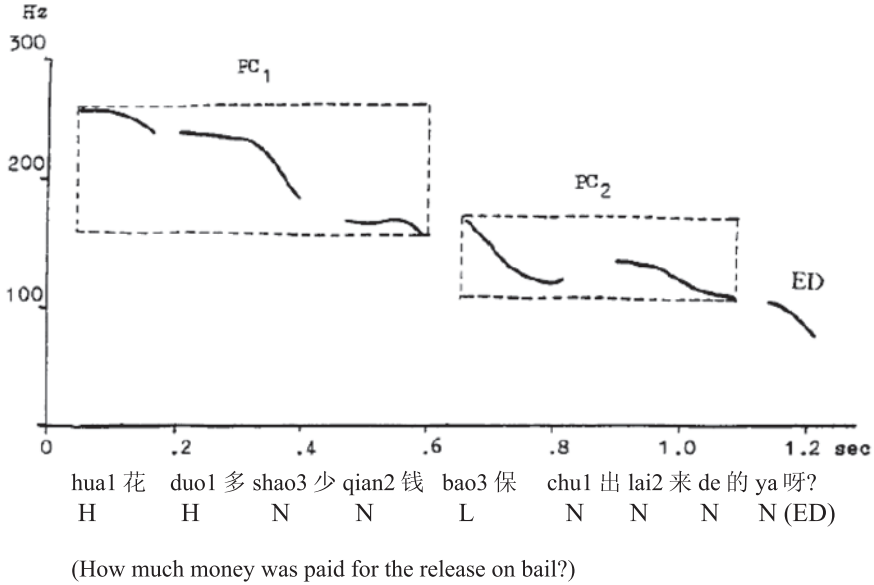


Fig. 11: An interrogative with a sequence of neutral tones spoken by a Beijing actress in a movie

tive utterance with an ending in down-drift, spoken by a Beijing actress in a movie.

The tone targets in five out of the nine syllables are undershot, as marked by N (neutral). In standard Mandarin, such tone target undershoot occurs frequently as well, although not so frequently as in the Beijing Mandarin. Thus, the lack of or the infrequent production of tone target undershoot does matter. L2 speech with infrequent or little tone target undershoot would sound staccato and unnatural, which, in turn, would lead to the perception of a foreign accent. However, a perception study is required to confirm such a hypothesis.

Yang (2012) argues that tones are perceived at the phonological level and produced at the phonetic level. Thus, it takes L2 learners longer time to acquire phonological features of tones. Likewise, tone target undershoot takes place as a result of the production of the phonological tone targets, which is at the phonetic level as well. The different patterns in tone target undershoot in L1 and L2 Mandarin show that the native speakers and the advanced learners tended to produce the target undershoot, resulting in the loss of tone targets, while the intermediate learners tended to produce less target undershoot and to maintain the underlying tone targets. Thus, the finding in this study has interesting implications for the teaching of Chinese as a second language. Tone production has been attached

great importance in the teaching of Mandarin Chinese. However, the findings in this study seem to suggest that the learners may be facing a dilemma: on the one hand, to produce tones accurately and maintain tone targets; on the other hand, to produce tone target undershoot and possibly lose some tone targets. Since tones in Mandarin Chinese can distinguish lexical meanings, tone accuracy definitely should be prioritized in the teaching of Chinese as a second language. But, does it mean that tone target undershoot or the utterance-level prosody should be ignored or that tone target undershoot will be naturally acquired as learners' proficiency improves? As learners' proficiency progresses and approximates the native-level, tone target undershoot is likely to occur. However, this will be a very long process and may even never take place. Thus, it is important to help learners realize the importance of tone target undershoot in natural speech. The T3 sandhi rule, namely a T3 becomes a rising tone when it precedes another T3 in the same prosodic domain, is taught in all Chinese textbooks. The tone target undershoot is also a sandhi phenomenon, but a phonetic sandhi, unlike the phonological sandhi as in the T3 sandhi. In the literature, a widely discussed phonetic sandhi pattern is that a T2 becomes T1 when it occurs between a T1 or T2 and a non-neutral tone (Chao 1968, Lin 2007). However, research has found that such a T2 sandhi most frequently occurs in prosodically weak positions, such as word-medial syllable in casual fast speech (Xu 2007: 17). One example of such T2 sandhi is *yóu* in *cōng yóu bǐng* (葱油饼 “scallion pancake”) becomes *yōu* in casual fast speech. The target undershoot is very similar to the T2 phonetic sandhi. Thus, since the tone target undershoot is such a prevailing pattern in the L1 speech, it should be included in the CSL teaching and student's awareness of the tone target undershoot as well as the utterance-level prosody should be developed. However, learners should be taught about the phonetic conditions for such tone target undershoot. Similar to the T2 sandhi, such tone target undershoot mainly occurs in casual or relatively fast speech. Without the inclusion of the phonetic environments of such tone target undershoot, learners may get confused when they do not hear such tone target undershoot in slow speech, including teacher talk.

Regretfully, utterance-level prosody in Chinese has received little attention in both CSL teaching and research, while the utterance-level prosody in English has been widely researched (McGory 1997 and references therein). In this sense, the CSL field may start to pay attention to the utterance-level prosody in both research and teaching.

The next question to ask is how to teach utterance-level prosody, such as tone target undershoot. Exercises consisting of phrases (such as the phrases in this study) or short sentences should be prepared to train both learners' prosodic production and perception. On the other hand, it would be of great necessity for the

CSL instructors to avoid using the slow teacher's talk and try to use speech at the normal speed so that learners can get used to the speech that they will encounter in actual communication with other native speakers. It is expected the training of the utterance-level prosody will complement the training of lexical tones in isolation.

7 Conclusion

In this study, we examined the acquisition of pitch patterns in Mandarin Chinese by American L2 learners. It was found that L2 learners, especially the intermediate learners, did not frequently produced tone target undershoot in utterances. Even though it is an empirical issue to examine whether the infrequent tone target undershoot contributes to the perception of a foreign accent in L2 speech, the findings in this study suggest that L2 training should go beyond the mono- or disyllabic phrases. It is expected that by training L2 learners on the F0 production at the phrase/sentence level can L2 speech further approach the native or the near-native patterns.

Acknowledgments: This paper is part of my doctoral dissertation at The Ohio State University. I would like to thank my advisor Dr. Marjorie K.M. Chan and dissertation committee members, Dr. Mary Beckman, Dr. Mineharu Nakayama, and Dr. Cynthia Clopper, for their advising and guidance.

References

- Aoyama, Katsura, James E. Flege, Susan G. Guion, Reiko Akahane-Yamada, and Tsuneo Yamada. 2004. Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /r/ and /l/. *Journal of Phonetics*, 23, 233–250.
- Archibald, John. 1995. The acquisition of stress. In J. Archibald (Ed.), *Phonological Acquisition and Phonological Theory* (pp. 81–109). Hillsdale, NJ: Lawrence Erlbaum.
- Archibald, John. 1997. The acquisition of English stress by speakers of tone languages: Lexical storage versus computation. *Linguistics*, 35, 167–181.
- Archibald, John. 1998a. Metrical parameters and lexical dependency: Acquiring L2 stress. In S. Flynn & G. Martohardjono (eds.), *The Generative Study of Second Language Acquisition* (Vol. 14, pp. 279–301). Mahwah, NJ: Lawrence Erlbaum.
- Archibald, John. 1998b. *Second Language Phonology*. Amsterdam: John Benjamins.
- Beckman, Mary E. 1996. The parsing of prosody. *Language and Cognitive Processes*, 11 (1/2): 17–67.
- Bent, Tessa. 2005. *Perception and Production of Non-native Prosodic Categories*. Ph.D dissertation. Northwestern University.

- Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp. 171–204). Timonium, MD: York Press.
- Boersma, Paul and David Weenink. 2009. *Praat: Doing phonetics by computer* (Version 5.1.05) [Computer program]. Retrieved May 2009, from <http://www.praat.org/>
- Chao, Yuen-Ren. 1930. A system of tone letters. *Le Maître Phonétique*, 45: 24–27.
- Chao, Yuen-Ren. 1968. *A Grammar of Spoken Chinese*. University of California Press.
- Chen, Qinghai. 1997. Toward a sequential approach for tonal error analysis. *Journal of Chinese Language Teachers Association*, 32, 1: 21–39.
- Chen, Qinghai. 2000. *An Analysis of Mandarin Tonal Errors in Connected Speech by English-speaking American Adult Learners: A Study at and above the Word Level*. Ph.D dissertation. Brigham Young University.
- Ellis, Nick. 2003. Constructions, chunking and connectionism: The emergence of second language structure. In C. Doughty & M. H. Long (Eds.), *Handbook of Second Language Acquisition* (pp. 63–103). Malden, MA: Blackwell.
- Flege, James, Ocke-Schwen Bohn, & Sunyoung Jang. 1997. The effect of experience on nonnative subjects' production and perception of English vowels. *Journal of Phonetics*, 25, 437–470.
- Flege, James, Ian MacKay, and Diane Meador. 1999. Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America*, 106, 2973–2987.
- Gut, Ulrike. 2003. Non-native speech rhythm in German. Paper presented at the 15th ICPhS, Barcelona.
- Hojen, Anders and James Flege. 2006. Early learners' discrimination of second-language (L2) vowels. *Journal of the Acoustical Society of America*, 119, 3072–3084.
- Ingram, John C. L. and See-Gyoon Park. 1997. Cross-language vowel perception and production by Japanese and Korean listeners of English. *Journal of Phonetics*, 25, 343–370.
- Ladd, D. Robert. 1996. *Intonational Phonology*. Cambridge: Cambridge University Press.
- Ladd, D. Robert. 2008. *Intonational Phonology*. New York: Cambridge University Press.
- Lin, Yen-Hwei. 2007. *The Sounds of Chinese*. Cambridge, UK: Cambridge University Press.
- MacWhinney, Brian. 2001. The competition model: the input, the context, and the brain. In P. Robinson (Eds.), *Cognition and Second Language Instruction* (pp. 69–90). Cambridge, England: Cambridge University Press.
- McGory, Julie T. 1997. *Acquisition of Intonational Prominence in English by Seoul Korean and Mandarin Chinese Speakers*. Unpublished Ph.D., Ohio State University.
- Miracle, W. Charles. 1989. Tone production of American students of Chinese: A preliminary acoustic study. *Journal of Chinese Language Teachers Association*, 24, 3: 163–176.
- Nguyen, T. Anh-Thu, John C. L. Ingram and J. Rob Pensalfini. 2008. Prosodic transfer in Vietnamese acquisition of English contrastive stress patterns. *Journal of Phonetics*, 36: 158–190.
- Peng, Shu-hui, Marjorie K.M. Chan, Chiu-yu Tseng, Tsan Huang, Ok Joo Lee, and Mary E. Beckman. 2005. Toward a Pan-Mandarin System for Prosodic Transcription. In Sun, S. (ed.) *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press. Pages 230–270.
- Pierrehumbert, Janet and Julia Hirschberg. 1990. The Meaning of Intonational contours in the Interpretation of Discourse, in P. Cohen, J. Morgan, and M. Pollack (eds). *Intentions in Communication* (pp. 271–311). Cambridge, MA: MIT.
- Pierrehumbert, Janet and Mary Beckman. 1988. *Japanese Tone Structure*, Linguistic Inquiry Monograph 15, MIT Press, Cambridge.

- Sheldon, Amy and Winifred Strange. 1982. The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence that Speech Production Can Precede Speech Perception. *Applied Psycholinguistics*, 3.3, 243–61.
- Shen, Xiaonan Susan. 1989. Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association*, 24, 3: 27–47.
- Tsukada, Kimiko, Davird Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and James Flege. 2005. A developmental study of English vowel production and perception by native English adults and children. *Journal of Phonetics*, 33, 263–290.
- Viger, Tanya Lynn. 2007. *Fundamental Frequency in Mandarin Chinese and English: Implications for Second Language Speakers*. Ph.D dissertation. The City University of New York.
- White, Caryn. 1981. Tonal pronunciation errors and interference from English intonation. *Journal of Chinese Language Teachers Association*, 16, 2: 27–56.
- Xu, Yi. 1994. Production and perception of co-articulated tones. *Journal of the Acoustical Society of America*, 95: 2240–2253.
- Xu, Yi. 2005–2010. ProsodyPro.praat. Available from: <http://www.phon.ucl.ac.uk/home/yi/ProsodyPro/>
- Xu, Lei. 2007. *Phonological Variation and Word Recognition in Continuous Speech*. Ph.D dissertation. The Ohio State University.
- Yang, Bei. 2012. The gap between perception and production of tones by American learners of Mandarin – An intralingual perspective. *Chinese as A Second Language Research*, 1, 1: 33–53.
- Yang, George Chunsheng and Marjorie K.M. Chan. 2010. The perception of Mandarin Chinese tones and intonation by American learners. *Journal of the Chinese Language Teachers Association*, 45, 1: 7–36.

Appendix

A complete list of the tone sequences used in the study

1. 邬安英修飞机。Wū Ānyīng xiū fēiji.
“Wu Anying repairs planes.”
2. 邬安应修飞机。Wū Ān yīng xiū fēiji.
“Wu An should repair planes.”
3. 殷安青摸猫咪。Yīn Ānqīng mō māomī.
“Yin Anqing pets a kitty.”
4. 殷安轻摸猫咪。Yīn Ān qīngmō māomī.
“Yin An gently pets a kitty.”
5. 王明来拿羊毛。Liú Míng lái ná yángmáo.
“Wang Ming comes to get wool.”
6. 王明来拿羊毛。Liú Míng lái ná yángmáo.
“Wang Minglai gets wool.”
7. 刘明来游云南。Liú Míng lái yóu Yúnnán.
“Liu Minglai travels in Yunnan.”

8. 刘明来游云南。Liú Míng lái yóu Yúnnán.
“Liu Ming comes to travel in Yunnan.”
“Li Weixiang buys a wild bird.”
9. 陆蔚用慢用药。Lù Wèi yòng màn yòng yào.
“Lu Wei uses the slow medicine.”
10. 陆卫用卖孕药。Lù Wèiyòng mài yùnyào.
“Lu Weiyong uses the contraceptive medicine.”
11. 魏丽要卖腊肉。Wèi Lì yào mài làròu.
“Wei Li wants to sell bacon.”
12. 魏立耀卖腊肉。Wèi Lìyào mài làròu.
“Wei Liyao sells bacon.”
13. 南梦来卖燃料。Nán Mèng lái mài ránliào.
“Nan Meng comes to sell fuels.”
14. 南梦莱卖燃料。Nán Mènglái mài ránliào.
“Nan Menglai sells fuels.”
15. 罗燕谈论名利。Luó Yàn tánlùn mínglì.
“Luo Yan talks about fame and profit.”
16. 罗彦坛论名利。Luó Yàntán lùn mínglì.
“Luo Yantan talks about fame and profit.”
17. 陆岩练习育苗。Lù Yán liànxí yùmiáo.
“Lu Yan practises growing seeds.”
18. 陆言练学育苗。Lù Yánliàn xué yùmiáo.
“Lu Yanlian practices growing seeds.”
19. 孟岩爱读外文。Mèng Yán ài dú wàiwén.
“Meng Yan likes studying foreign languages.”
20. 孟言艾读外文。Mèng Yán'ài dú wàiwén
“Meng Yanai studies foreign languages.”