Acta Linguistica Hungarica Vol. **62** (2015) 1, 35–62
DOI: 10.1556/ALing.62.2015.1.2

Prosodic challenges faced by English speakers reading Mandarin

Hua-Li Jian
Faculty of Technology Art and Design, Oslo and Akershus University College of Applied Sciences, Oslo
Hua-Li.Jian@hioa.no

Abstract: This study compares the prosodic characteristics of L2-Mandarin as spoken by L1-English speakers using L1-Mandarin utterances. The acoustic correlates examined include individual tonal realizations, interactions of tones in sequence, durational features and intensity envelopes. L2-Mandarin users realize the contour tones RISE and FALL with both rising and falling pitch, and produce the second tone of disyllabic words with more varied pitch. L2-users employ larger vowel durations, syllable durations and larger variation over vowel intervals in sequential pairs than L1-Mandarin users. Both user groups show similar intensity envelopes. Implications of this study include tailoring language training programs that counterbalance L1 influences.

Keywords: English L1; Mandarin L2; suprasegmentals; acoustic correlates; L2 speech production

1. Introduction

Few studies have examined L2 speech productions in tone languages such as Mandarin by native English speakers. Most research has focused on English as L2 by different languages. The present study attempts to fill this gap by examining and comparing the prosodic aspects realized by tone-language versus non-tone language speaker groups.

This study compares prosodic patterns in Mandarin speech produced by L1-English and L1-Mandarin speakers by examining acoustic features, including the fundamental frequency (F0), duration, and intensity. In Mandarin, pitch differences are used to distinguish lexical words. Mandarin speakers mainly employ pitch and duration to achieve prosodic variation, with loudness being a secondary feature (Chao 1980). Although F0 is considered the dominant acoustic correlate, contours of amplitude and duration may also contribute to tonal distinctions (e.g., Whalen & Xu 1992).

Mandarin employs four tones to denote meaning. Table 1 and figure 1 show the pitch contours of the tones denoted in tonal values using the

5-point scale (based on Chao 1930, with 1 representing the lowest and 5 the highest pitch): tone 1 (HIGH, tonal value 55) with high level pitch, tone 2 (RISE, tonal value 35) low-rising, tone 3 (LOW, tonal value 213 or 214) low-dipping, and tone 4 (FALL, tonal value 51) high-falling. A sandhi occurs when a LOW is followed by another LOW, where the first LOW is changed to RISE. The duration of the four tones differs, RISE being the longest and FALL the shortest (Lin 1985).

Table 1: Mandarin tone systems (based on Chao, 1930)

	Characteristics	**Tone contour**
Tone 1	High	55
Tone 2	Low-rising	35
Tone 3	Low-dipping	213/214
Tone 4	High-falling	51

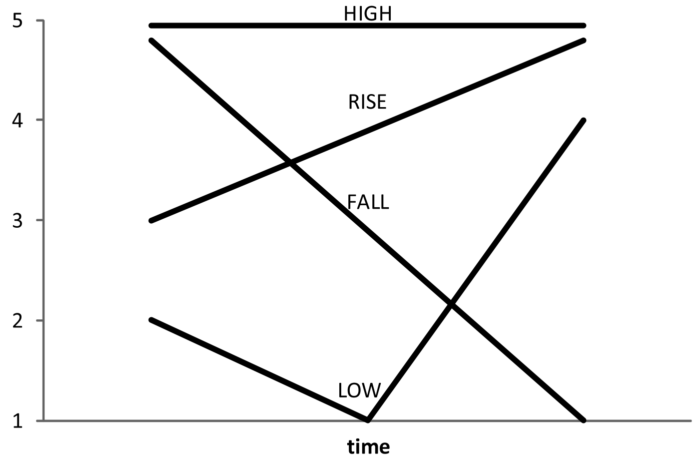

Figure 1: Tone contours

Generally, Mandarin syllables are organized into feet (Duanmu 2000; Feng 1998; Shih 1986). The size of a foot ranges from one to three syllables (Duanmu 2000; Feng 1998). As for how a foot is phonologically formed, views differ, from it being based on stress (Duanmu 2000) to actually not involving stress (Feng 1998). If it indeed involves stress, it contains a foot pattern of strong-weak (Duanmu 2000; Feng 1998) or weak-strong (Chao

1968). The Mandarin foot-related stress is different from English lexical stress. Word stress in English is lexical, functioning to distinguish words, with clear stress placement of which syllable bears the stress. The closest Mandarin equivalent to English lexical stress is reported to be the neutral tone, being lexically contrastive and phonetically similar to unstressed English syllables at least in F0 (Chen & Xu 2006; Xu & Xu 2005) and duration (Lin 1985). However, the neutral tone is a tone and not stress, and it only makes up ca. 4.6% of Mandarin morphemes (Mi 1986) or 6.7% according to Li (1981). No other equivalents of English lexical stress are known (Chen 2000; Duanmu 2000).

Mandarin grouping-related stress has been explored acoustically. Kochanski et al. (2003) quantitatively measured prosodic strength by manipulating F0-contours. Prosodic strength was defined by the degree of realization of tones within contextual influences, viz., the more realized, the greater prosodic strength there is. Prosodic alternating patterns tend to occur and with a clear trend of strong–weak (i.e., as trochees) in disyllabic words, consistent with previous studies (Duanmu 2000; Feng 1998). Shih (1986) found the same strong–weak pattern to occur at higher-level four-syllable words, the first disyllabic unit showing greater prosodic strength than the second disyllabic unit, hence also displaying a hierarchical frame.

Tonal contrasts have been observed to be one of the most difficult areas for non-tone language speakers. Speakers with a non-tone language background do not perceive tones categorically and are therefore unable to align tones with associated syllables or words (Hallé et al. 2004; Repp & Lin 1990; Fox & Qi 1990). Also, segmental variations related to articulatory constraints may cause the tonal expressions of non-tone speakers to deviate from ideal realizations, including carryover tonal variations, F0 peak delays, and long transition of F0 at syllable boundaries (Xu 1999; Xu & Wang 2001). Learning Mandarin lexical tones for English speakers is challenging because of the specific alignment between F0 and segmental features (Shen 1989). Their learning has also been found to be affected by pitch use for English stress and intonation systems (White 1981; Broselow et al. 1987), as well as for affective characterization (Ross et al. 1988). Contrastively, pitch contours in Mandarin are used to differentiate word meanings; intonation is often indicated by adding boundary tones after lexical tones, not as pitch fluctuation on lexical words (Duanmu 2004). Mandarin also generally shows a greater scope of pitch variation over the course of a sentence than English (Chen et al. 2001). The first research question therefore addresses the tonal production of the two groups: What are the differences between L2-Mandarin and L1-Mandarin tonal realizations, with reference to F0 variations, particularly among disyllabic words (i.e., tone sequences)?

Mandarin is classified as a syllable-timed language (e.g., Grabe & Low 2002; Lin & Wang 2007), where syllable durations are close to equal. Comparatively, stress-timed English is more varied in length due to a mix of stressed and unstressed syllables. However, Mandarin duration patterns are reported to be position-specific, similarly to lexical stress languages such as English (Xu & Wang 2009). Syllable lengthening and shortening occur in both languages. In English, the duration of a word increases in proportion to the increasing number of syllables, and varies with stress patterns for words of the same size (Nakatani et al. 1981). Constituent-initial and constituent-final lengthening reported in English (Cooper et al. 1977) also occur in Mandarin, where the last syllables in three- and four-syllable phrases show the largest duration and the initial syllable being the second largest (Xu & Wang 2009). A similar effect of polysyllabic shortening in English (Klatt 1976; Turk & Shattuck-Hufnagel 2000), where individual syllables shorten as the number of syllables in a syllable group increases, also occurs in Mandarin (Xu & Wang 2009) for all-rising, all-falling, and all-high sequences among one- to four-syllable words. Although English polysyllabic shortening may be explained by word- or phrase-level lengthening (Nakatani et al. 1981), Mandarin syllabic shortening was reported to be even stronger than English shortening effect (Xu & Wang 2009) in that (a) the final syllables of disyllabic words were much shortened in Mandarin but not in English; and (b) the inserted medial syllables were found to be much shorter than initial syllables in di-, tri- and quadra-syllabic words; contrastively, English medial syllables were found to be slightly more reduced than onset syllables. Xu (1999) observed a short–long duration pattern for Mandarin disyllabic words, irrespective of focus or sentence position. This study focuses on basic durational differences in Mandarin sentences read by the two speaker groups since lengthening and shortening effects occur both in Mandarin and English speech. The second research question examines the durational feature realized by the two groups: Do L1-English speakers exhibit more durational variations in Mandarin utterances than L1-Mandarin users?

The role of intensity in phonological and phonetic classification often overlaps with F0 and duration. F0 is considered the primary acoustic cue for Mandarin tones, but duration of a syllable and amplitude contour may also contribute to lexical tonal information with their consistent variation across tone categories (Gandour 1983; Howie 1976; Liu & Samuel 2004; Whalen & Xu 1992). The intensity mean obtained over the syllable has also been identified as a possible correlate of English stress differences (Fry 1958; Beckman 1986). Amplitude and duration were further noted as a primary parameter in cuing stress in American English, with F0 having a secondary role (Silipo & Greenberg 2000; Greenberg 1999). Similarly, loudness and duration were used to mark prominent syllables in utterances in British and Irish English, with F0 as a secondary cue (Kochanski et al. 2005). Kochanski et al. (2005) found that measurements of amplitude change as an approximation to steady-state perceptual loudness, rather than as overall intensity or loudness, was a reliable correlate of stress in English, contradicting the view that intensity is relatively ineffective correlate (Sluijter & van Heuven 1996; Sluijter et al. 1997). Only syllable intensity peaks are examined in this study since intensity is believed to function as a secondary acoustic cue in Mandarin. The third research question is: What are the intensity realization differences in L2-Mandarin versus L1-Mandarin utterances?

2. Method

2.1. Participants

The participants included a control group of eight L1-Mandarin speakers and a group of eight Mandarin learners of L1-English. L1-Mandarin participants whose speech samples were treated as the reference were female graduate students (mean age 26) from Cheng Kung University. The L2-Mandarin group consisted of male students (mean age 28) from Cheng Kung University and Tainan University. The L2-Mandarin participants were at an intermediate level of proficiency. Ideally, differences in age and gender should be controlled to create a more balanced sample for comparison. Unfortunately, the recruitment of English L1 participants was challenging as there were not many such individuals available for the experiment. Nevertheless, gender and age have been found to have limited effect in previous studies of verbal ability (see Hyde & Linn 1988). In the present study, the measurements were normalized to counterbalance potential gender and age biases. In tone production, learners' English L1 or L2-Mandarin proficiency is likely to have a stronger impact on the results than possible age or gender differences.

2.2. Materials

A Mandarin version of the fable "The North Wind and the Sun" (from the *Handbook of the International Phonetic Association*) was chosen for the experiments as it is widely used in studies on prosody, making comparisons to other works easier. The reading passage contained 41 phrases comprising 18.1% HIGH, 20.3% RISE, 21.0% LOW and 40.6% FALL words. All the sentences were declarative, thereby avoiding the intonational characteristics of interrogative and exclamation sentences. The Mandarin passage included *pin-yin* phonetic transcriptions to assist the L2-Mandarin users. The original English text was also included to help the learners familiarize themselves with the content.

2.3. Procedure

Before reading the Mandarin passage, the learner group was allowed to familiarize themselves with the English version. The learners were assisted with the first reading trial. The second reading was used for recording. Mispronounced utterances were not analyzed. The speech was digitally recorded using a laptop computer with a build-in microphone. The participants were asked to sit still at a fixed position, with a constant distance from the microphone. T-tests were used to verify the significance level of difference between the speaker groups.

2.4. F0

F0 slopes were examined to determine if the pitch patterns were level, rising, or falling. The F0 range was measured for each vowel as the difference between the initial pitch value and the final pitch value and then computing the F0 slope by dividing the F0 range by the vowel duration, see (1) (t_{start} signals the starting time of the vowel and t_{end} its end; $\text{F0}(t_{\text{start}})$ and $\text{F0}(t_{\text{end}})$ are the start-F0 and end-F0 measurements, respectively). If the F0 slope is greater than zero, the pitch pattern is rising, and if it is less than zero, falling.

(1) $\text{F0}_{\text{slope}} = \frac{\text{F0}(t_{\text{end}})-\text{F0}(t_{\text{start}})}{t_{\text{end}}-t_{\text{start}}}$

F0 realizations of tones in sequence were observed and interpreted by examining contextual effects both within each tone sequence and across immediate sequential boundaries. The contextual analysis scheme employed by (Xu 1997) was chosen in this study as the majority of morphemes in the speech material occurred in non-boundary positions and mostly consisted of disyllabic morphemes. Tonal effects included pre-tonal and post-tonal influences, the former addressing tonal effect of the first syllable on the second syllable of the tone sequence, the latter focusing on the following syllable's effect on its preceding syllable. Pre-targets and post-targets denote the immediate neighboring syllables of the tone sequence. Both HIGH and LOW were considered static pitch targets, both static in the sense of there being no or little pitch movement. LOW has the characteristics of a low level tone when uttered rapidly. RISE and FALL were considered dynamic pitch targets because of the pitch movement.

2.5. Duration

The durational realization of the two speaker groups was examined by means of direct comparison in vowels and syllables. Syllable and vowel durations were measured as the interval between the beginning and end points of the F0 curves.

The Variability Index (VI) proposed by Chen & Chung (2008) was adopted, see (2) (X_i is the ith syllable or vowel, E_i is the mean of the ith syllable or vowel over the L1-Mandarin utterances (the control group), K is the number of syllables or vowels in the sentence). Large VI values indicate that syllable or vowel duration deviates from the norm.

(2) $$\mathrm{VI} = \sum_{i=1}^{K} \frac{(X_i - E_i)^2}{K}$$

Rhythmic characteristics were estimated using the normalized Pairwise Variability Index (nPVI) (Low et al. 2000; Grabe & Low 2002). The nPVI was computed by dividing the absolute difference in duration between each pair of successive measurements with the mean duration of the pair and then transform the ratio into a percentage (3):

(3) $$\mathrm{nPVI} = 100 \times \left[\sum_{k-1}^{m-1} \left| \frac{d_k - d_{k+1}}{(d_k + d_{k+1})/2} \right| /(m-1) \right]$$

In (3), m is the number of intervals – vocalic or intervocalic – in the text, and d is the duration of the kth item. Vocalic intervals were defined as the stretch of signal between vowel onset and vowel offset, characterized by vowel formants. A vocalic interval may contain a monophthong, a diphthong, or, in some cases, two or more vowels spanning the offset of one word and the onset of the next. The intervocalic interval was defined as the segment between the vowel offset and vowel onset, irrespective of the number of consonants included.

The nPVI is related to the vowel dimension. Common English utterances consist of both full and reduced vowels, resulting in considerable variability in vowel realizations. Mandarin exhibits less vocalic variability as vowel duration reduction is uncommon. Stress-timed languages tend to show higher vocalic nPVI and intervocalic nPVI than syllable-timed languages.

2.6. Intensity

The intensity peak values were measured for all the words in each sentence and for disyllabic phrases. To compare the intensity distributions of the two groups, the means and variances were elicited using log-values of the intensity observations. The reported values are in dB, but intensity is treated in Pascal internally in Praat (the software used for the acoustic analyses) according to the documentation. The peak dB values were first converted to linear scale by taking the exponential, the average of the linear intensity values were computed and the final linear mean was converted to dB by taking the logarithm.

3. Results

3.1. F0

F0 was normalized using semitones to adjust for individual differences in pitch range; the results are presented in figure 2.

Syllables with HIGH tone occurred mostly in non-boundary positions with its high level pitch pattern. The F0 slopes for L2-Mandarin were all greater than, but close to, zero, showing non-falling pitch, acoustically close to level. The mean F0 slope for L2-Mandarin was slightly larger than that for L1-Mandarin, not significantly different ($t(102) = 0.34$, $p> .05$). However, the spread in slope was much larger for L2-Mandarin than L1-Mandarin, suggesting that L2-Mandarin HIGH pronunciation was less consistent than that of L1-Mandarin.

About 88.6% RISE words occurred at medial (non-boundary) positions (31 words out of 35). The mean F0 slope for both groups was greater than zero signaling rising patterns. A t-test showed that the F0 slope values

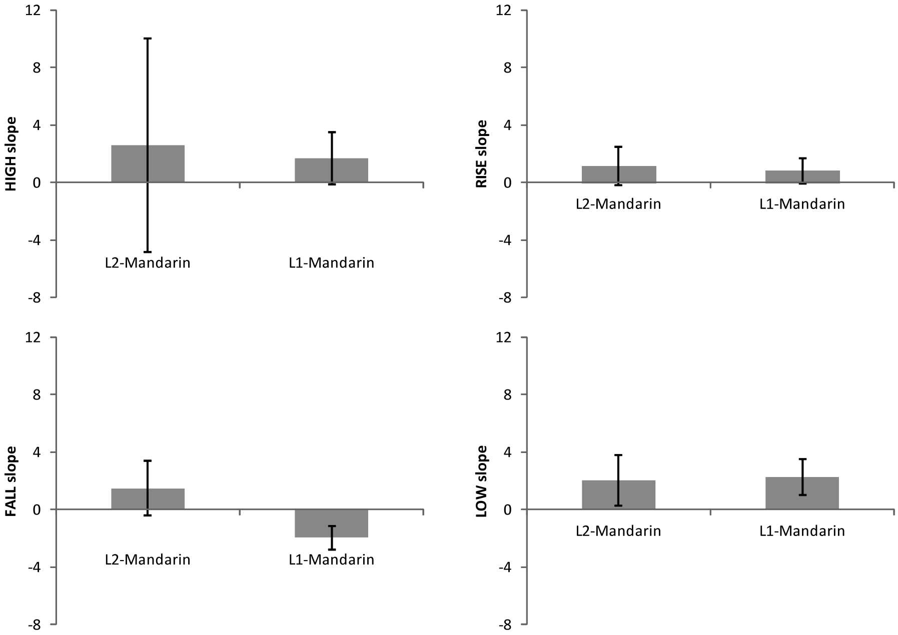

Figure 2: F0 slopes of the four tones by the two speaker groups

in RISE for the two groups were not significantly different ($t(166) = 0.33$, $p > .05$). L2-Mandarin utterances could be expected to show similar pitch patterns in RISE to those of L1-Mandarin, as rising contours are common in standard English yes–no question intonation (Pierrehumbert & Hirschberg 1990; Ladd 1996). However, nearly half of the RISE words were produced with falling patterns (232 lexical RISE words, 29 words × 8 L2 participants).

The canonical LOW tone contour was 213, the pitch value of the initial F0 point being lower than the end point. Hence, the F0 slope value of LOW should be greater than zero, as realized in both groups, with the L2-Mandarin also being aware of the initial fall followed by a rise. A t-test showed that the F0 slope values for LOW in both groups were not significantly different ($t(142) = 1.2$, $p > .05$).

FALL was the most frequent tone in the passage with most of the FALL words located at non-boundary positions. L2-Mandarin utterances could be expected to have similar pitch patterns in FALL to those of L1-Mandarin, as falling contours were common in English neutral declarative intonation (Pierrehumbert & Hirschberg 1990; Ladd 1996). The mean F0 slope of the L1-Mandarin utterances was less than zero, signaling a falling pitch. Five

L2-Mandarin users exhibited rising F0 slopes. There were a total of 464 lexical words (58 words × 8 L2 participants) with FALL, and roughly half of the words were produced with a rising tone resulting in a mean rising slope. A t-test confirmed that the slope values for the two groups were significantly different ($t(294) = 5.77$, $p < .05$).

3.2. Tone sequences

There were 16 unique tone pairs. LOW was excluded from this analysis as the L2-Mandarin speakers generally did not master tone sandhi. There were six tone-pair sequences in the passage excluding LOW, namely HIGH–RISE, HIGH–FALL, RISE–HIGH, RISE–FALL, FALL–RISE, and FALL–FALL. The tone sequences discussed in the following sections were reconstructed from the mean F0 values with normalized time. Normalization of time allowed the data from the various speakers to be combined despite the speakers' individual pitch range and speed of articulation. As observed, the learner group had difficulty producing RISE and FALL, each being realized in two patterns as falling and rising. The learners realized them as expected in some tonal sequences, but not as expected in others. RISE was realized as expected for the HIGH–RISE, RISE–FALL, and FALL–RISE sequences; FALL was realized as expected for the RISE–FALL and FALL–RISE sequences.

The HIGH–RISE sequence (see figure 3a) realized tones from static pitch target (HIGH) to dynamic target (RISE). The pitch difference between the offset of HIGH and its following dynamic RISE target was small, that is, 8 Hz and 9 Hz for the L1-Mandarin and L2-Mandarin, respectively. The dynamic target (RISE) was realized in 6 Hz by the L1-Mandarin (208–214 Hz), versus 39 Hz (162–201 Hz) by the L2-Mandarin. Both groups realized this sequence expectedly as a high tone followed by a rising tone (HIGH–RISE), with the L2-Mandarin realizing a much larger rise for the second RISE syllable, showing more variation on the second syllable than that of L1-Mandarin.

The carryover effect of the first syllable on the second syllable revealed a speaker difference. The L2-Mandarin showed a steeper rising F0 contour of the second syllable compared to that of L1-Mandarin, with a more immediate rising portion at the onset than that of L1. This may be interpreted as a less natural realization of L2 since the affected transitional portion of F0 onset which is normally reflected in the flatter F0 shape is missing (cf. Xu 1997). As for the anticipatory effect of the following tone on its preceding tone, the raising effect of RISE on its preceding HIGH could be seen to exist on both L1 and L2 realization as HIGH was higher than in

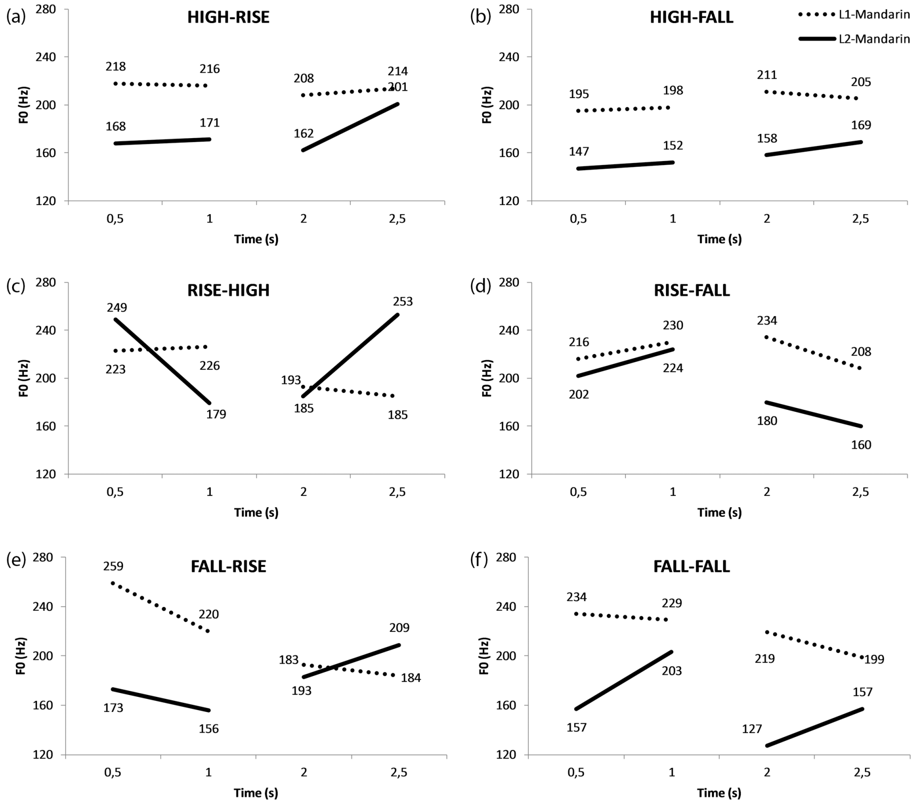

Figure 3: Tone sequence pair of L2-Mandarin (solid line) vs. L1-Mandarin (dotted line). (a) HIGH–RISE, (b) HIGH–FALL, (c) RISE–HIGH, (d) RISE–FALL, (e) FALL–RISE, and (f) FALL–FALL.

the other sequence with HIGH (HIGH–FALL), particularly for L1 realisation. In other words, the L2 raising effect was comparatively weaker than the L1 raising effect. The observed anticipatory raising was consistent with previous studies by Xu (1997), Shih (1986), and Shen (1990).

The pre-targets had one high-offset (RISE) and one low-offset (LOW), possibly neutralizing the mean F0 of the second RISE syllable, with L2 showing a higher rise than L1. The two post-targets were both FALL. If the post-target effect did exist, its high-onset would cause the maximum F0 of the second syllable (its preceding syllable) to lower more than a low-onset post-target, hence a dissimilatory effect, according to Xu (1997). The smaller magnitude of L1 second syllable RISE than that of L2 may be an indication of lowering triggered by the high-onset of the post-target FALL. Such a post-target lowering effect on its preceding syllable may not be strong in L2 speech. One possible factor may be the lacking maturity or naturalness of L2 on both cross-syllabic tonal connection and tonal-segmental alignment (both consonantal and vocalic segments).

The HIGH–FALL sequence (see figure 3b) stretched from a static pitch target (HIGH) to dynamic target (FALL). The pitch difference before the dynamic fall was 13 Hz for L1-Mandarin and 6 Hz for L2-Mandarin. The FALL was realized in a 6 Hz fall by L1-Mandarin (211–205 Hz) versus a rise of 11 Hz (158–169 Hz) by L2-Mandarin. Thus, the L1-Mandarin realized the HIGH–FALL with a moderate FALL, where the following FALL was higher than the preceding HIGH. This L1 realization could be explained as anticipatory lowering (Xu & Wang 2001; Xu 1997). The following FALL with its high F0 onset lowered the preceding HIGH. The results suggested that the L2 users had difficulty producing a dynamic rise or fall following a static HIGH. The slope of the unexpected realization of FALL as RISE by L2 was not as steep as the actual RISE in the HIGH–RISE sequence.

The result agreed with Xu's (1997) findings that the effect of pre-tonal offset was stronger on the following HIGH and RISE than on the following FALL or LOW. Both speaker groups realized the first syllable HIGH of HIGH–FALL sequence with similar F0 contour (with L2 exhibiting slightly more movement than L1), but with a variation on the second FALL syllable (L1's 6-Hz small fall versus L2's 11-Hz rise). Comparatively, the L2 showed more variation in the second RISE syllable (HIGH–RISE sequence) than in the second FALL syllable (HIGH–FALL sequence). The effect of the post-tonal onset of FALL on the preceding HIGH was not noticeable in this study for either speaker groups, as was also reported by Xu (1997). The HIGH–FALL sequence was bordered by pre-targets having a low offset (one FALL, two LOW, and one neutral tone). Since the offsets of the pre-targets were low, the mean F0 of the second FALL syllable was not likely to be raised. The low offsets suggested that the pre-target effect on the first HIGH syllable was nonexistent. Although the mean F0 of the second FALL appeared higher than that of the first HIGH syllable in the constructed illustration, the raising effect may be caused by the following post-target onset. Post-targets indicated a tie of onset values (two high onsets of FALL; two low onsets of one RISE and one neutral tone). The second syllable revealed an F0 directional difference between speakers' realizations, a small L1 fall versus a small L2 rise. Xu (1997) found that the minimum F0 of FALL was not affected by carryover assimilation or anticipatory dissimilation; the raising effect of low-onset post-targets on the maximum F0 of the second FALL could be reasoned to be likely greater than the lowering effect of high post-target onset. The L1's realization of the FALL by employing higher onset portion than the preceding HIGH could reflect such a post-target raising effect.

The RISE–HIGH sequence (see figure 3c) involved a motion from a dynamic pitch target (RISE) to static target (HIGH), where RISE began immediately after the rime of the first syllable. The data showed that L1-Mandarin RISE began high, followed by a much lower HIGH. The pitch difference before the static HIGH was 33 Hz by L1-Mandarin versus 6 Hz by L2-Mandarin. The RISE was realized in 3 Hz by L1-Mandarin (223–226 Hz), versus a fall of 70 Hz (249–179 Hz) by L2-Mandarin. The L1 users produced the second syllable (HIGH) with a slight fall (193–185 Hz) while L2 users unexpectedly realized it with a 68 Hz rise (185–253 Hz). This RISE–HIGH sequence appeared difficult for the L2 users, with both syllables realized unexpectedly.

L1-Mandarin realized the following HIGH lower than the first syllable's slight RISE. The pre-tonal effect on the second syllable was not noticeable since the slight rise (or near-level) of the first syllable of L1 would more likely cause the following syllable to rise moderately or be level rather than a fall; thus, an assimilatory carryover effect was not observed. A cross-syllabic pre-target effect on the first syllable may appear since the pre-target neutral tone had a slight fall from mid to low, or simply a mid level tone, hence assimilating a near-level rise for the first syllable. The post-target FALL, having a high onset, might show an anticipatory effect on its preceding HIGH tone by lowering its F0 and causing a small fall, as seemingly was the case. The L2-Mandarin unexpectedly realized the disyllable as FALL–RISE. There was a noticeable immediate adjacent pre-tonal effect. Since nearly half of the L2 rising tones were realized as falling tone, the complementary contour being realized (RISE being realized as FALL), the realized falling contour then caused the following HIGH to rise. This carryover effect with the assimilatory pattern was consistent with Xu's (1997) finding, although it was illustrated by L2's unexpected FALL–RISE in this study.

The RISE–FALL sequence (see figure 3d) involved a turn from a dynamic target (RISE) to another dynamic target (FALL). The pitch difference between the two targets was 4 Hz for the L1-Mandarin and 44 Hz for L2-Mandarin. RISE was realized at 14 Hz by the L1-Mandarin (216–230 Hz) versus 22 Hz (202–224 Hz) by L2-Mandarin and FALL with L1-Mandarin 26 Hz (234–208 Hz) versus L2-Mandarin 20 Hz (180–160 Hz). Thus, L2-Mandarin exhibited larger rise, but smaller fall than the L1-Mandarin. The L2-users achieved both dynamic targets, but the rise and fall pitch difference was much larger than that of the L1-Mandarin (44 Hz vs. 4 Hz). Further, the L2-users realized fall with a much lower F0 than the L1-users. This observation suggested that L2-users had difficulty producing this rise–fall sequence, especially the fall, but the difficulty may also start at the transitional turning point from the high offset of rise to the high onset of fall. The large pitch boundary difference of the L2-Mandarin between rise and fall (44 Hz) could be explained as a brief gestural rest after the deliberate endeavor of sustaining an acceptable rising pattern before resuming to perform the second syllable fall, hence possibly resulting in this apparent downstep event where the second H (high onset of fall) was much lower than the first H (high offset of rise) (cf. Xu & Wang 2001). Note that downstep was also reported in non-tone languages, including English (Pierrehumbert 1980). This difference may also be caused by alignment difficulties involving segmental features (cf. Shen 1989).

The effect of pre-tonal offset was claimed to be larger on the following high and rise than on the following fall or low. Further, tones with low F0 offset such as fall and low were less likely than those with high offset to show carryover effects throughout the vowel length (Xu 1997). The rise–fall sequence showed a similar trend of realization by the two speaker groups. The differences lay mainly in the magnitude of rise or fall: L1 revealed a steeper fall than L2 at the second syllable while L2 showed a steeper rise than L1 at the first syllable. The more striking difference was the F0 difference transitioning the two tones. The L1 showed a more gradual 4 Hz rise from the high offset of rise, which then linked up to the high onset of the following fall. The L2 conversely revealed a steep fall from the high offset of rise to the following fall. Such a drastic transitional difference within a common tone sequence reflected a less stable F0 maneuver realized by the L2. The overall L1 realization of the rise–fall sequence was closer to Xu's (1997) observation than what was the case for L2.

The results show that the immediate carryover effect of high offset pre-targets (high and rise) on the first syllable rise may be comparatively larger than the low offset pre-targets (fall, low, and the neutral tone), although non-significant, as was in Xu's study (Xu found no significant carryover effects of pre-targets on both the first and second syllables of tone sequences). Given that L2 realization showed a larger rise on the first syllable than L1 realization, it might be likely that L2-speakers produced steeper (high offset) pre-target rise than L1-speakers. A potential raising effect appeared to be stronger for L1 than for L2, since L1 realization showed a much higher mean F0 for the second syllable FALL than L2 realization. Only one post-target had high onset (FALL); the others had low onset (one RISE, one LOW, and one neutral). Given that L1's second syllable FALL revealed comparatively a larger maximum F0 than that of L2, one possible factor may be the less-than-expected L2 post-target tonal productions. The L1 realisation for the RISE–FALL sequence where the high onset of FALL appeared to reduce the range of the preceding RISE also corresponded to the previous results using authentic stage speech (Kratochvil 1984).

The FALL–RISE sequence (see figure 3e) involved a move from one dynamic target (FALL) to another dynamic target (RISE). The L1-Mandarin realized this sequence with a lowered pitch in RISE due to the falling trend of the previous FALL, which was consistent with Xu's previous study (1997). The pitch difference between the two dynamic targets was similar for both groups (−27 Hz for L1 and +27 Hz for L2). The FALL was realized in 39 Hz by the L1-Mandarin (259–220 Hz) versus 17 Hz (173–156 Hz) by the L2-Mandarin; thus, both groups realized certain degrees of fall with the L1 fall being larger. The L2-users employed RISE with a larger range (183–209 Hz) compared to the L1-users, who realized FALL–RISE as FALL–LOW, since RISE was similar to LOW. The L1 realization thus seemingly resembled a downstep event (cf. Xu & Wang 2001). In this instance, the onset of RISE within the FALL–RISE sequence (ideally HL–MH) became even lower. The following RISE might have been affected by the preceding FALL: the onset M of RISE being lowered by the preceding L of FALL and the H offset of RISE further lowered or declined. The unexpected fall or level off of RISE of the L1 group might also have been induced by the tone following RISE. A rising tone can become falling or "downward gliding" when preceded by a high tone and followed by a low tone (Wu 1984). Such occurrence was also observed in the FALL–RISE sequence of L1, considering that RISE was preceded by FALL (having a high onset) and in four out of six cases followed by post-targets having low tones, that is, three LOW and one neutral tone. Conversely, L2-Mandarin utterances in this FALL–RISE sequence appeared to realize the canonical RISE.

The carryover effect from the preceding syllable has been reported to be stronger than the anticipatory effect from the following syllable. Also, the pre-tonal offset has been observed to affect the following tones having high offset (such as HIGH and RISE) up to the whole vowel length (see Xu 1997). Comparatively, for this tone sequence, the post-target effect (having low onset) on the second syllable RISE might be stronger than the pre-tonal effect (low offset of fall), and as mentioned, both the low-onset post-targets and the pre-tonal fall had effects on rise. The pre-target carryover effects on the first and second syllable of the tone sequence may not be strong since all pre-targets had low F0 offset (five fall and one low).

The fall–fall sequence (see figure 3f) realizes a dynamic target (fall) followed by another dynamic target (fall). The series showed that the second target began lower than the first target for both L1-users and L2-users. The pitch difference between the two dynamic targets was 10 and 76 Hz for the L1-users and L2-users, respectively. The initial fall was realized in 5 Hz by L1 (234–229 Hz) versus a 46 Hz rise (157–203 Hz) by L2. The second fall was realized as a L1 20-Hz fall (219–199 Hz) vs. a L2 30-Hz rise (127–157 Hz). The L2-users showed a large range of rise with 46 and 30 Hz of a rise–rise pattern. Still, both groups showed the downstep event (cf. Xu & Wang 2001) in their respective fall–fall and rise–rise patterns, with the second H being comparatively lower than the first H in the disyllabic sequences.

L1 and L2 F0 contour realizations contrasted for the fall–fall sequence. L1 fall–fall revealed a downstep event, where the second high was lower than the first high. Both falls showed similar falling contour, with the second fall having a larger magnitude (20 Hz, 219–199 Hz) than the first fall (5 Hz, 234–229 Hz). The L1 realization in this sequence was consistent with Xu's study (1997) in terms of contextual disyllabic F0 contour movement. The second fall was less likely to show carryover effect from the offset of the preceding tone since the second fall had a low offset. As known, the anticipatory effect is generally small (Xu 1997) and dissimilatory in nature. Fall as the first syllable had the beginning portion raised (see Xu 1997). This could also partially explain why the first fall in the results herein appeared to have a higher F0 beginning portion than the second fall. The F0 variation within the first fall was smaller than that of the second fall. This reduction of the F0-falling range of the preceding fall has been explained as an anticipatory effect exerted by the high onset of the second fall. This result was also consistent with the previous finding employing L1 authentic stage speech (Kratochvil 1984) and L1 meaningless material (Xu 1997).

Of the post-targets, four had high onset (four high) and two had low onset (one rise and one neutral). The post-target highs can be interpreted as being subject to a small anticipatory effect as their high onset did not cause the maximum F0 of the preceding tone to raise, but rather lowered it. High offset (two high and one rise) and low offset (two fall and one neutral) pre-targets may have an equaling effect on mean F0 measurements. Still, the high offset pre-targets would likely have a stronger effect than low offset pre-targets on both the first and second FALL syllables, and might also raise the onset of its immediately following first FALL, as shown by Xu (1997). This was also observed in L1 realization of the present study.

L2 realized FALL–FALL as RISE–RISE, but a downstep event (cf. Xu & Wang 2001) could also be seen where the H of the second RISE was lower than the H of the first RISE. The offset of the second RISE in this study was comparatively much lower than that of the first RISE. In Xu's (1997) results, using L1 meaningless syllables, the difference between the two offsets of RISE–RISE sequence was much smaller. As observed in Xu's study, tones with high offsets were more likely than tones with low offset to show carryover effects of pre-tonal offset throughout the vowel length. With the L2 RISE–RISE realization, the carryover effect from the offset of the first RISE (pre-tonal) on the second RISE was likely present as the second RISE had a high offset. Nevertheless, noticeably the offset value of the second RISE was as low as the onset of the first RISE, although still being much higher than its own onset.

3.3. Duration

Figure 4 shows the VI observations. The mean VI of the L2-Mandarin syllables was larger than that of the L1-Mandarin. A t-test confirms that this difference was statistically significant ($t(7) = 15.22$, $p < .05$).

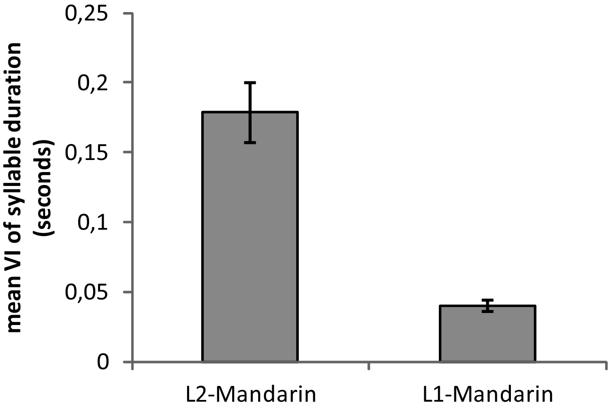

Figure 4: The mean VI value of syllable duration. Error bars show SD.

Figure 5 reveals that the VI of L2-Mandarin vowel durations was larger than that of L1, suggesting that L2-Mandarin users produced longer vowels or that they spoke more slowly than the L1 users. A t-test confirmed that the VI values were significantly different ($t(7) = 6.51$, $p < .05$).

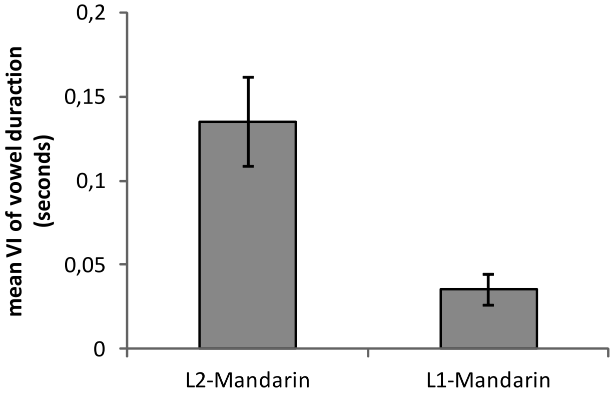

Figure 5: The mean VI value of vowel duration

The results showed that L2-Mandarin utterances were closer to L1-Mandarin for the shortest sentence *(Then the Sun shone out warmly)* compared to other sentences in terms of VI measures of both syllable and vowel duration. L1-Mandarin users uttered syllables with a similar duration across all the sentences, except this shorter sentence with larger vowel durations, possibly slowing down to prepare for closing of the passage.

Table 2 shows the nPVI measurements. The L2-Mandarin was expected to show higher vocalic nPVI compared to L1-Mandarin if their vowel durations were more varied for consecutive pairs. Languages with simple syllable structures, such as Mandarin, tend to exhibit lower intervocalic nPVI values. The results confirmed a small difference in nPVI for the two groups, with a vocalic nPVI of 43.6 for L2-Mandarin and 36.2 for L1-Mandarin. Both values were closer to French (43.5) than English (57.2) using measurements by (Grabe & Low 2002) as reference. However, L2-Mandarin showed similar intervocalic nPVI (58.7) to L1-Mandarin (59.2). These values were more similar to Spanish (57.7) than English (64.1).

Table 2: Vocalic and intervocalic nPVI for the two groups

nPVI	L2-Mandarin	L1-Mandarin
Vocalic	43.6	36.2
Intervocalic	58.7	59.2

3.4. Intensity

The mean and variance of the intensity peaks for all the words by both groups were computed, but no noticeable differences were observed.

4. Discussion

4.1. F0

Individual tonal realizations show that both groups used similar F0 slope for HIGH and RISE. Both groups also realized positive F0 slope for LOW, with the L2 exhibiting more variations. Moreover, L2 users did not adhere to tone sandhi with LOW–LOW sequences. FALL was associated with the largest variations in F0 slope, with L2-Mandarin realizing roughly half of the words with rising pitch. There may be several explanations as to why L2 users unexpectedly realized FALL. First, tone "errors" do occur in natural settings even among L1 users, including Thai (Gandour 1977) and Mandarin (Wan 2007), and can occur as freely as segmental errors (Wan 2007). Observably, learners exhibited more production errors than L1 speakers. Further, a falling tone residing in non-boundary positions may be difficult to realize for non-tone language speakers. Since many Mandarin words are disyllabic, a non-tone speaker may not fully exercise a falling pitch onto the non-boundary phrase until towards the end of the sentence, as in unmarked English intonation. L2-Mandarin users may have purposely tried not to employ the English way of marking a prominent word, which includes a combination of greater intensity and HIGH or LOW or a combination of these (Büring 2003). The falling pattern thus did not fully surface. Moreover, it has also been found that falling tones are the most frequently misproduced in substitution errors in L1 tone production in Thai (Gandour 1977) and Mandarin (Wan 2007). Thus, the tendency of non-tone L1 speakers in misproducing Mandarin FALL as RISE confirms the same substitution tendency.

L2-Mandarin users exhibited a positive mean F0 slope for all the four tones, indicating that rising pitch was used in different tonal environments. The dynamic tones RISE and FALL were more varied than the static tones HIGH and LOW. These L2-Mandarin results correspond to the findings in L1 tone production that contour tones are more prone to production errors than static tones in Mandarin (Wan 2007) and also that contour tones are four times more likely to be wrongly produced than static tones in Thai (Gandour 1977). L2-Mandarin users had difficulties producing Mandarin falling and rising tones despite the common occurrence of rise and fall in their L1-English pitch accents. During exposure to L2 sounds, learners' L1 categories are expanded and reinforced by integrating with unlikeness or uniqueness of a L2 sound in relation to the closest L1 sound (MacKay et al. 2001; Flege 1987; Flege & Liu 2001). In the context of Mandarin tone, this involves its association with lexical items in underlying representation and its alignment with the syllable that carries it. Also, tone in Mandarin is an essential part of the phonological organization of the lexicon. Further, unlike lexical stress in English, a Mandarin tone is not an organizing factor in phonological syntagmatic planning, and hence tone and stress should not be stored at the same place during speech production planning (Wan 2007). One possibility is that the L2 group in this study was unable to separate the representations of tones and stress during the reading session. Moreover, stress in English is linked to phrasal prosody, where the nuclear (most stressed) syllable follows from the syntactically-defined phrase intonation (Beckman 1986). Mandarin differs from the intonation system of Germanic languages such as English in that it has no prosodically prominent tonic syllable at phrase end. All these factors may possibly contribute to the unexpected tonal production of the non-tone speakers.

4.2. Tone sequences

Differences between L1-and L2-Mandarin productions may shed light on areas of production difficulty for L2 users. Overall, contextual variation in L2 realization was larger and less expected compared to that of L1 realization. Boundary F0 difference between the two tones of the tonal sequence investigated show that L2 realization tended to be wider and in opposite direction compared to L1. In particular, this observation applied to boundary F0 differences between L2 RISE–FALL, FALL–RISE, and FALL–FALL, because of instances of unexpected L2 realization of dynamic tones including RISE and FALL. Although rising and falling F0 contours also exist in L1 English, L2-Mandarin speech realization requires training of a different kind.

The most different realizations between L1 and L2 are rise–high and fall–fall. These two sequences require special training for L2 users. The second syllable high in L1 rise–high is much lower than expected, possibly because of the need to prepare for the following post-target high-falling tone. As observed by Xu (1997), high onset post-target causes its preceding tone to lower its F0 contour. Fall–fall sequences are challenging for L2 users, their distinctive F0 pattern as rise–rise (instead of fall–fall) may reflect a production difficulty already observed in the F0 slope measurement. L2 speakers produced roughly half of the falling tones with a rising tone resulting in a mean rising slope.

Central questions are thus why fall–fall is challenging for L2 users and why is fall–fall realized as rise–rise? Is this due to unfamiliarity of this tone sequence? Dynamic tones were challenging for L2-Mandarin users, with the second syllable being impacted by the first syllable. Did L2-Mandarin users over-produce rise because it is easier to realise? The substitution of rise for fall is surprising in view of the literature that considers fall easier than rise to produce (Zhang 2002) and perceive (Hombert 1975). The falling and high level tones are considered easier to acquire than rising and low-dipping tones (Li & Thompson 1976). The Mandarin tonal system lacks the syntagmatic relationship that English stress and accent play in discourse association and organization (Wan 2007). Mandarin tones are aligned with lexical items and are tightly associated with the segments and syllables. One explanation is therefore that L2-Mandarin users did not obtain the same firmness of affiliation between tone and its hosting syllable, complicated by the distinct English usage of stress and accent in discoursal context, leading to contextual pitch switch. Alternatively, Mandarin learners may find it difficult to produce the ideal version of a sequence such as fall–fall where it requires a great jump in pitch which is rare in English. Tone combinations closer to the possible intonational patterns of non-tone languages are likely to be easier for foreign Mandarin learners to master. Certain pitch jumps or drops required in the ideal tonal realization would result in greater differences between foreign Mandarin users and L1-Mandarin speakers.

fall–rise also posed challenges for L2 users. L1 realization showed a subtle modification of rise as a gliding fall contour because of its pre-tonal and post-target context. L2 users understandably may have such difficulty because of this specific tonal context. The high–fall sequences revealed that L2 learners struggle realizing the second fall. The carryover effect from the previous syllable would more likely affect the second syllable that has high offset; it is thus more likely that FALL as the second syllable would not be affected (since FALL has a low offset). The L2 users' realiziation of the second FALL as rise is unexpected, given that the pre-target effect is weak (mostly low offsets) and that post-targets would be neutralized because of balanced number of high- versus low-onsets. One may speculate that the pre-tonal HIGH of L2 utterances may affect the following FALL, since the HIGH F0 contour was realized as a low tone and not high as expected. It may then become hard to execute a high falling FALL after such low F0 realization and thus instead realized as a small rise. This FALL, realized as a rise by L2 users, may be a production difficulty. Note that FALL was believed to be difficult for foreign learners to acquire (Shen 1989), contrasting with the view of Li & Thompson (1976).

Most errors occurred in L2 sequences with dynamic tones and particularly the second syllable. In cases where both tones of the sequence deviated from those of L1-Mandarin, the pitch difference was also large. L2-users employed fall and rise with larger ranges, especially for the second tone in tone pairs. The variations in fall or rise may be connected to syllable-tone alignment. It is possible that L2-Mandarin speakers use the syllable boundary as the reference point for alignment to gain sufficient time to reach the pitch targets (Xu & Wang 2001). However, the L2-Mandarin users may segment syllables differently to the L1-Mandarin speakers, potentially in terms of duration, thereby causing proportional differences in fall and rise. L2-Mandarin users may have delayed their tonal realizations, and thus certain misalignment of F0 peaks or troughs may have occurred. The methodology employed would not be able to detect such characteristics. Important for foreign learners, tones must be timely aligned with their associated syllables in spite of having carryover variations, as observed in L1-Mandarin utterances (Xu 1998; 1999).

In responding to the first research question, the results and discussions reported herein suggest that certain production differences exist between L2-Mandarin and L1-Mandarin users in terms of the F0 feature, specifically measured in F0 slope and reflected in possible contextual effects of the tone sequences. Contours tones FALL and RISE are the most challenging for L2-users, and their F0 realizations are more varied in the second syllables of the tone sequences investigated, as well as L2-users showing less stable carryover and anticipatory effects compared to L1-users.

4.3. Duration

VI observations showed that L2-users employed larger syllable and vowel durations than L1-users. Since VI does not distinguish the speed of learners being faster or slower than the native speech rate, in hindsight, a more effective indicator could perhaps have been chosen. The durational difference between the two groups was small for the shortest sentence in the speech material. The L2-users sped up while the L1-users slowed down to adapt to the content or contextual change. It may be possible that for L2 users most Mandarin characters (if not all) have been carefully articulated (i.e., more than being neutral) and hence unintentionally lengthened. Note that prosodic prominence for specific discourse context in Mandarin and English does not function in the same way, and becomes a potential obstacle for L2-users. While speech rate would be expected to be slower for L2 users, an additional factor could be the correct processing of Chinese characters, although a *pin-yin* transcription was provided as an aid.

L2-users showed a higher nPVI in vowel intervals than L1-users. The higher L2 vocalic nPVI demonstrated larger variation over vowel intervals in sequential pairs than L1-Mandarin. The higher vocalic nPVI of learners could be due to limited proficiency in reading and speaking Mandarin, where, for instance, simple words are pronounced faster than more complicated ones.

However, both had similar intervocalic nPVI, placing the two Mandarin varieties closer to French and Spanish than to English, with syllable-timed propensity. The similarity in intervocalic nPVI could be explained by the lack of consonantal clusters in Mandarin, hence likely making it easier for L2-users to articulate consonants in a similar manner to L1-users. The results from the acoustic measures indicate that durational differences exist between the two groups. Duration observations suggest that L2-utterances vary more. It must also be emphasized that the distinction between stress-timed and syllable-timed is not necessarily that clear-cut (Dauer 1983).

In response to the second research question, L2-utterances show larger vowel duration, larger syllable duration, as well as more sequential vowel variation compared to L1-utterances.

4.4. Intensity

In terms of intensity, L2-Mandarin and L1-Mandarin users exhibit a similar realization across the sentences. Concerning the third research question, the results reveal no significant differences in intensity between the L2-utterances and L1-utterances.

4.5. Other factors

The differences between L2-Mandarin and L1-Mandarin may also be caused by other factors. The performance of L2-Mandarin could be affected by deciphering skills in reading the Mandarin passage. Despite having been assisted by the *pin-yin* transcription and the practice trial, the decoding speeds of learners were likely to be slower than those of L1-users and may vary individually. These deviations could affect the measurement of duration, where individual learners may speed up when the syllabic structure, segmental, suprasegmentals, or a combination of these, are simple and slow down when encountering more complicated ones. Such alterations could be problematic concerning pitch target and pitch alignment where tonal models had to be adhered to within reasonable time frame to sound natural. This could also partially explain why the second syllables of disyllabic words were often F0-misaligned and tonal target affected compared to L1-users. Likewise, the more varied consecutive vowel pairs of learners may be a partial mirroring of a challenging coordination of visual input and articulating output.

5. Limitation of the study and future research

The two speaker groups were close in age, but not balanced in terms of gender. Ideally age and gender should be balanced, although these have been found to have limited effect in previous studies of verbal ability (see Hyde & Linn 1988). Ideally, Mandarin tonal realization in isolation and in carrier sentences by L2-Mandarin speakers should also be conducted to obtain the complete understanding of foreign speakers' Mandarin production. Also, there is very little documented research on when and how foreign learners of Mandarin will be able to master the tone sandhi realization. More research effort involving new methods would be needed as other factors (such as age, experience, native language, target language, motivation, cognition, personality, sociocultural issues, and other idiosyncratic differences) may be involved.

6. Conclusion

This study has highlighted areas where English L1 users may find it difficult to learn L2 Mandarin prosody. The results indicate that the F0 feature is the main challenge, followed by duration, with no observed issues for intensity. The results have implications for the teaching of Mandarin as L2, specifically in terms of dynamic tones, pitch alignment and patterns in sequential tones, and rhythm practice. As an example, dynamic tones may be systematically arranged in disyllabic words to form specific tone sequences, and thereby to deal with sandhi tonal transform and expected pitch and syllable alignment. Vocalic variation can be reduced by repeated and embedded exposures of disyllabic words, in series of steady built-ups to sentence level, wherein vocalic segments are linked to legitimate consonants. All these may naturally be implemented in a concurrent manner or in separate lessons to allow maximal training for learners.

Acknowledgements

The author thanks Joyce Wu for assistance with collecting and processing the data and the anonymous reviewers for their valuable comments.

References

Beckman, Mary. E. 1986. Stress and non-stress accent. Dordrecht: Foris.

Broselow, Ellen, Richard R. Hurtig and Catherine O. Ringen. 1987. The perception of second language prosody. In G. Ioup and S. H. Weinberger (eds.) Inter-language phonology: The acquisition of second language sound system. Cambridge: Newbury House. 350–361.

Büring, Daniel. 2003. On D-trees, beans, and B-accents. Linguistics and Philosophy 26. 511–45.

Chao, Yuan Ren. 1930. A system of tone-letters. Le maitre phonetique 45. 24–27.

Chao, Yuen Ren. 1968. A grammar of Spoken Chinese. Berkeley/Los Angeles: University of California Press.

Chao, Yuan Ren. 1980. Chinese tone and English stress. In L. R. Waugh and C. H. V. Schooneveld (eds.) The melody of language. Baltimore, MD: University Park Press. 41–44.

Chen, Hsueh-Chu and Raung-Fu Chung. 2008. Interlanguage analysis of phonetic timing patterns by Taiwanese learners. Concentric: Studies in Linguistics 34. 81–100.

Chen, Mathew Y. 2000. Tone sandhi: Patterns across Chinese dialects. Cambridge: Cambridge University Press.

Chen, Yang, Michael Robb, Harvey Gilbert and Jay Lerman. 2001. A study of sentence stress production in Mandarin speakers of American English. Journal of the Acoustical Society of America 109. 1681–1690.

Chen, Yiya and Yi Xu. 2006. Production of weak elements in speech: Evidence from f0 patterns of neutral tone in standard Chinese. Phonetica 63. 47–75.

Cooper, William, Steven Lapointe and Jeanne Paccia. 1977. Syntactic blocking of phonological rules in speech production. Journal of the Acoustical Society of America 61. 1314–1320.

Dauer, Rebecca M. 1983. Stress-timing and syllable-timing reanalyzed. Journal of Phonetics 11. 51–62.

Duanmu, San. 2000. The phonology of Standard Chinese. Oxford: Oxford University Press.

Duanmu, San. 2004. Tone and non-tone languages: An alternative to language typology and parameters. Language and Linguistics 5. 891–923.

Feng, Shengli. 1998. Prosodic structure and compound words in classical Chinese. In J. L. Packard (ed.) New approaches to Chinese word formation: Morphology, phonology and the lexicon in modern and ancient Chinese. Berlin & New York: Mouton de Gruyter. 197–260.

Flege, James Emil. 1987. The instrumental study of L2 speech production: Some methodological considerations. Language Learning 37. 285–296.

Flege, James Emil and Serena Liu. 2001. The effect of experience on adults' acquisition of a second language. Studies in Second Language Acquisition 23. 527–552.

Fox, Robert Aallen and Ying-Yong Qi. 1990. Context effects in the perception of lexical tones. Journal of Chinese Linguistics 18. 261–284.

Fry, Dennis B. 1958. Experiments in the perception of stress. Language and Speech 1. 126–152.

Gandour, Jack. 1977. Counterfeit tones in the speech of southern Thai bidilectals. Lingua 41. 125–143.

Gandour, Jack. 1983. Tone perception in Far Eastern languages. Journal of Phonetics 11. 149–176.

Grabe, Esther and Ee Ling Low. 2002. Durational variability in speech and rhythm class hypothesis. In C. Gussenhoven and N. Warner (eds.) Laboratory phonology 7. Berlin & New York: Mouton de Gruyter. 515–546.

Greenberg, Steven. 1999. Speaking in shorthand – A syllable-centric perspective for understanding pronunciation variation. Speech Communication 29. 159–176.

Hallé, Pierre A., Yueh-Chin Chang and Catherine T. Best. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. Journal of Phonetics 32. 395–421.

Hombert, Jean-Marie. 1975. The perception of contour tones. In Proceedings of the First Annual Meeting of the Berkeley Linguistics Society. Berkeley, CA: Berkeley Linguistics Society. 221–232.

Howie, John M. 1976. Acoustical studies of Mandarin vowels and tones. Cambridge: Cambridge University Press.

Hyde, Janet S. and Marcia. C. Linn. 1988. Gender differences in verbal ability: A meta-analysis. Psychological Bulletin 104. 53–69.

Klatt, Dennis H. 1976. Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America 59. 1208–1221.

Kochanski, Greg, Esther Grabe, John Coleman and Bernard Rosner. 2005. Loudness predicts prominence: Fundamental frequency lends little. Journal of the Acoustical Society of America 118. 1038–1054.

Kochanski, Greg, Chilin Shih and Hongyan Jing. 2003. Hierarchical structure and word strength prediction of Mandarin prosody. International Journal of Speech Technology 6. 33–43.

Kratochvil, Paul. 1984. Phonetic tone sandhi in Beijing dialect stage speech. Cahiers de Linguistique. Asie Orientale 13. 135–174.

Ladd, D. Robert. 1996. Intonational phonology (Cambridge Studies in Linguistics 79). Cambridge: Cambridge University Press.

Li, Charles N. and Sandra A. Thompson. 1976. The acquisition of tone in Mandarin-speaking children. Child Language 4. 185–199.

Li, Wei Min. 1981. Shilun qingsheng he zhongyin [On the neutral tone and stress]. Zhongguo Yuwen [Chinese Linguistics] 1. 35–40.

Lin, Hua and Qian Wang. 2007. Mandarin rhythm: An acoustic study. Journal of Chinese Language and Computing 17. 127–140.

Lin, Mao-Ts'an. 1985. The pitch indicator and the pitch characteristics of tones in Standard Chinese. Acta Acoustica 3. 8–15.

Liu, Siyun and Arthur G. Samuel. 2004. Perception of Mandarin lexical tones when F0 information is neutralized. Language and Speech 47. 109–138.

Low, Ee Ling, Esther Grabe and Francis Nolan. 2000. Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English. Language and Speech 43. 377–401.

MacKay, Ian R. A., James Emil Flege, Torsten Piske and Carlo Schirru. 2001. Category restructuring during second-language speech acquisition. Journal of the Acoustical Society of America 110. 516–528.

Mi, Qing. 1986. A preliminary study on the teaching of neutral tone. Yuyan Jiaoxue yu Yanjiu [Language Teaching and Research] 2. 58–65.

Nakatani, Lloyd H., Kathleen D. O'Connor and Carletta H. Aston. 1981. Prosodic aspects of American English speech rhythm. Phonetica 38. 84–106.

Pierrehumbert, Janet B. 1980. The phonology and phonetics of English intonation. Doctoral dissertation. MIT.

Pierrehumbert, Janet B. and Julia Hirschberg. 1990. The meaning of intonation contours in the interpretation of discourse. In P. R. Cohen, J. Morgan and M. E. Pollack (eds.) Plans and intentions in communication and discourse (SDF Benchmark Series in Computational Linguistics). Cambridge, MA: MIT Press. 271–311.

Repp, Bruno H. and Hwei-Bing Lin. 1990. Integration of segmental and tonal information in speech perception: A cross-linguistic study. Journal of Phonetics 18. 481–495.

Ross, Elliott D., Jerold A. Edmondson, G. Burton Seibert and Richard W. Homan. 1988. Acoustic analysis of affective prosody during right-sided Wada test: A within-subjects verification of the right hemisphere's role in language. Brain and Language 33. 128–145.

Shen, Xiaonan S. 1989. Interplay of the four citation tones and intonation in Mandarin Chinese. Journal of Chinese Linguistics 17. 61–74.

Shen, Xiaonan S. 1990. Tonal coarticulation in Mandarin. Journal of Phonetics 18. 281–295.

Shih, Chilin. 1986. The prosodic domain of tone sandhi in Chinese. Doctoral dissertation. University of California, San Diego.

Silipo, Rosaria and Steven Greenberg. 2000. Prosodic stress revisited: Reassessing the role of fundamental frequency. In Proceedings of the NIST Speech Transcription Workshop. College Park, MD. 16–19.

Sluijter, Agaath M. C. and Vincent J. van Heuven. 1996. Spectral balance as an acoustic correlate of linguistic stress. Journal of the Acoustical Society of America 100. 2471–2485.

Sluijter, Agaath M. C., Vincent J. van Heuven and Jos J. A. Pacilly. 1997. Spectral balance as a cue in the perception of linguistic stress. Journal of the Acoustical Society of America 101. 312–322.

Turk, Alice and Stefanie Shattuck-Hufnagel. 2000. Word-boundary-related durational patterns in English. Journal of Phonetics 28. 397–440.

Wan, I-Ping. 2007. On the phonological organization of Mandarin tones. Lingua 117. 1715–1738.

Whalen, Douglas H. and Yi Xu. 1992. Information for Mandarin tones in the amplitude contour and in brief segments. Phonetica 49. 25–47.

White, Caryn. 1981. Tonal perception errors and interference from English intonation. Journal of Chinese Language Teachers Association 16. 27–56.

Wu, Zong Ji. 1984. Putonghua sanzizu biandiao guilu [Rules of tone sandhi in trisyllabic words in Standard Chinese]. Zhongguo Yuyan Xuebao [Bulletin of Chinese Linguistics] 2. 70–92.

Xu, Yi. 1997. Contextual tonal variations in Mandarin. Journal of Phonetics 25. 61–83.

Xu, Yi. 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. Phonetica 55. 179–203.

Xu, Yi. 1999. Effects of tone and focus on the formation and alignment of F0 contours. Journal of Phonetics 27. 55–105.

Xu, Yi and Maolin Wang. 2009. Organizing syllables into groups: Evidence from F0 and duration patterns in Mandarin. Journal of Phonetics 37. 502–520.

Xu, Yi and Q. Emily Wang. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. Speech Communication 33. 319–337.

Xu, Yi and Ching Xu. 2005. Phonetic realization of focus in English declarative intonation. Journal of Phonetics 33. 159–197.

Zhang, Jie. 2002. The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis. New York: Routledge.