The Effect of Tonal Context on Second Language Learners' Mandarin Tone Production

Yen-Chen Hao
University of Tennessee
E-mail: yhao6@utk.edu

ABSTRACT

This study investigates the effect of tonal context on L2 learners' production of Mandarin tones. Ten English-speaking learners of Mandarin were asked to imitate native productions of disyllables encompassing all possible tone combinations. Their productions were evaluated by two native judges. The results show that the learners' error patterns vary depending on the tonal context. Specifically, their T2→T3 error is most common when *followed* by T1, while the T3→T2 error is most frequently observed when *preceded* by T4. F_0 measurements were conducted on the T2T1 and T4T3 sequences produced by the native talker and selected learners. The comparison reveals that the learners overly lower the onset and valley of their T2 when the following tone has a high onset, and raise T3 to a higher offset when the preceding tone ends in a low pitch. These findings suggest a dissimilatory effect of tonal context on L2 learners' production.

Keywords: L2 Mandarin tone production; tonal context; tone dissimilation.

1. INTRODUCTION

Second language (L2) learners of Mandarin have often been reported to have difficulty with mastering the four tones: T1 (high-level or 55), T2 (high-rising or 35), T3 (low-dipping or 214), and T4 (high-falling or 51). However, existing studies on L2 learners' tone production have not reached a consensus about the most challenging tones and the sources of difficulty. For example, Shen [6] found that T4 and T1were least accurately produced by L2 learners. Wang and colleagues [8], on the other hand, found that learners were most accurate with T1 and least accurate with T3 both before and after perceptual training. The most common error was consistently T3 being misproduced as T2. Chen similarly reported that T2 and T3 were generally harder to produce than T1 and T4, and the learners most often substituted the target tones with level tones [3]. Regarding the sources of difficulty, interference from the learners' native language intonation, and physiological constraints such as narrower pitch range or inapt command of pitch movement have commonly been proposed [1, 2, 6, 9].

One possible explanation for the divergent findings in the previous studies is that the effect of tonal context is not taken into account. While most of these studies examined L2 learners' tone production in different prosodic contexts, they rarely explored how the preceding and following tones influence the learners' accuracy and whether certain tone sequences are more difficult. Nevertheless, prosodic context should be given more importance considering that its effect has been clearly demonstrated in native production. For instance, Shen [7] analysed Mandarin speakers' production of tri-syllabic words and found that the high offset of T1 and T2 frequently raised the onset pitch of the following tone. Additionally, the high onset of T4 would raise the ending pitch of the preceding tone, while the low onset of T2 and T3 had no effect, suggesting that certain tonal contexts generated an assimilatory effect. In another study, Xu [10] examined native productions of Mandarin disyllables and revealed differences in the carry-over and anticipatory effect. For the former, the offset of the first syllable consistently assimilated the onset of the second syllable. As for the anticipatory effect, it appeared to be dissimilatory albeit subtle.

To the best of the author's knowledge, thus far only one study, which compared English speakers' imitation of Mandarin tones in monosyllables and tri-syllables, has explored the effect of tonal context in non-native production [1]. Specifically, while T3→T4 error was most common in monosyllables, T3→T2 error became most frequent in tri-syllables, illustrating a positional variation in error types. However, since only one tri-syllabic context (following T1 and preceding T4) was examined in [1], it is unclear whether the results can be generalized to other tonal environments. In view of this gap in the literature, the present study aims to investigate English-speaking learners' tone production in different contexts, and address the following questions: 1) Do L2 learners make different errors in different prosodic positions? 2) In what ways does the learners' production differ from the native norm?

2. METHODS

2.1 Participants

Ten English-speaking learners of Mandarin (8 male, 2 female) participated in this study. Their mean age was 22.9 years (SD = 6.4), and their length of studying Mandarin was 2.68 years (SD = 1.91).

2.2 Stimuli

Four syllables /jo/, /ma/, /waŋ/, /ji/ were selected because they are sonorants and can combine with all four tones. These syllables were paired to form two disyllabic non-words /waŋ.yi/ and /jo.ma/, and the tones for each initial and final syllable were systematically varied to include all of the 16 possible tone combinations, resulting in 32 tokens in total. Under the Mandarin T3 sandhi rule, the underlying T3T3 sequence surfaces as T2T3. As a result, there are 15 instead of 16 distinctive tonal contours.

One female native speaker of Mandarin produced all 32 non-words, which were randomized into one block. To ensure that the stimuli are good exemplars of Mandarin tones, two Mandarin speakers identified the tones of the stimuli. One listener achieved 100% and the other achieved 98% accuracy, indicating that the tones of the stimuli were clearly intelligible to native Mandarin speakers.

2.3 Procedure

The participants were seated in a quiet room and listened to the stimuli through headphones (SONY MDR-Q22LP). Their task was to repeat after each stimulus without any orthographic prompt. Their responses were recorded using a digital recorder (SAMSUNG YV150).

2.4 Analysis

The learners' productions were evaluated by two native speakers of Mandarin. The inter-rater agreement was 90%, and the evaluations from both judges were used to compute the accuracy and error rates. The tone sequences that incurred the highest error rates were further examined in an acoustic analysis.

3. RESULTS

3.1 Native listeners' evaluation

The native listeners' identification of the learners' tone productions is presented as confusion matrices. Table 1 summarizes the word-initial productions separated by the four following tones, while Table 2 illustrates the word-final tone productions. The rows mark the target tones, and the columns label the learners' responses as identified by the native listeners. The accurate responses are shaded. Note that both T2 and T3 responses for the T3 target in the sandhi position were considered accurate because the judges may treat the T2T3 and T3T3 sequences as identical.

As can be seen in Table 1, the learners' most common mistake is the misproduction of T2 as T3. However, the error rates vary considerably depending on the following tone. The learners were more likely to make T2→T3 error when the following tones are T1 and T4. In contrast, their productions of T3 in the same contexts are highly accurate, suggesting a unidirectional T2→T3 bias when the following tones start with a high pitch.

Table 1. The learners' production of the word-initial tones followed by the four Mandarin tones

| Following tone | ___T1 | | | | ___T2 | | | | ___T3 | | | | ___T4 | | | |
Response / Target	*T1*	*T2*	*T3*	*T4*	*T1*	*T2*	*T3*	*T4*	*T1*	*T2*	*T3*	*T4*	*T1*	*T2*	*T3*	*T4*
T1	1	0	0	0	0.93	0	0	0.08	0.95	0	0	0.05	0.95	0.05	0	0
T2	0.05	0.23	0.73	0	0.1	0.85	0.05	0	0.13	0.78	0.05	0.05	0	0.68	0.33	0
T3	0	0	1	0	0	0	0.95	0.05	0	0.85	0.1	0.05	0	0	1	0
T4	0	0	0	1	0.05	0	0	0.95	0.08	0	0.05	0.88	0.03	0.03	0.03	0.93

Table 2. The learners' production of the word-final tones preceded by the four Mandarin tones

| Preceding tone | T1___ | | | | T2___ | | | | T3___ | | | | T4___ | | | |
Response / Target	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4	T1	T2	T3	T4
T1	1	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0
T2	0	0.8	0.2	0	0	0.98	0.03	0	0.08	0.88	0.05	0	0	0.9	0.1	0
T3	0	0.33	0.68	0	0	0.23	0.78	0	0	0.23	0.78	0	0.03	0.6	0.38	0
T4	0	0	0.03	0.98	0	0	0.03	0.98	0	0	0	1	0	0	0	1

Table 2 shows a different pattern. In the final position, the learners most frequently misproduced T3 as T2, especially when following T4. On the other hand, the T2→T3 error always occurs at a lower frequency than the T3→T2 error in the same context. This demonstrates an asymmetrical bias from T3 to T2 in the word-final position.

Repeated Measures ANOVA were conducted on each position to examine the effect of the adjacent tone on the learners' accuracy. With regard to the initial position, the Adjacent tone is found to be marginally significant ($F(3, 27)=2.66$, $p=0.068$), while the Target tone and the interaction are both significant ($F(3, 27)=15.42$; $F(9, 81)=9.48$). Paired-sample t-tests reveal that the accuracy for T2 is lower when followed by T1 than by the other three tones ($t(9)=6.23$; 4.71; 4.63; ps<0.01), suggesting that the T2T1 sequence is particularly hard for L2 learners.

As for the final position, the Adjacent tone ($F(3, 27)=7.01$), Target tone ($F(3, 27)=22.11$), as well as the interaction ($F(9, 81)=4.60$) are all significant factors. Paired t-tests show that when the target tone is T2, its accuracy is lower when following T1 than when following T2 ($t(9)=3.28$, $p=0.01$). When the target is T3, it receives the lowest accuracy when following T4 than the other three tones ($t(9)=2.70$; 4.00; 4.71, ps<0.05).

Repeated measures ANOVA were also conducted on the 12 error types with different adjacent tones. In the initial position, the factor Adjacent tone only approached significance ($F(3, 27)=2.66$, $p=0.68$), while the Error type was significant ($F(11, 99)=13.55$), and so was the interaction ($F(33, 297)=10.26$). Post-hoc analyses confirmed that the T2→T3 error rate was significantly higher than the majority of the error types. When it comes to the final position, the Adjacent tone ($F(3, 27)=7.22$), the Error type ($F(11, 99)=21.61$), and the interaction ($F(33, 297)=4.73$) were all found to be significant. Post-hoc analyses indicated that the T3→T2 error rate was significantly higher than all the other error patterns except for the T2→T3 error. In summary, the statistical tests support the observation that L2 learners' tone productions are prone to different error types in different tonal contexts.

3.2 Acoustic analysis

The Mandarin listeners' evaluation shows that some of the learners' tone productions are easily confusable with other tones as judged by native listeners, yet it does not reveal the specific ways in which their productions differ from the native norm. To address this question, f_0 measurements were conducted on the tone sequences that were judged to be least accurate: T2T1 and T4T3. For each tone combination, two learners who received 100% error rate in the native listeners' evaluation were selected to be compared to the native speaker (NS) who produced the stimuli. Their productions were time-normalized to the average duration of the NS, and then converted to the pitch values ranging from 0 to 5 with the following formula commonly adopted in the literature [1, 4, 8]:

(1) $T=[(\log X-\log L)/(\log H-\log L)]\times 5$

in which X is the pitch in Hz, and L and H stand for the lowest and highest pitch of the speaker.

The normalized pitch values of the T2T1 sequence is plotted in Figure 1 while the T4T3 values are plotted in Figure 2. The NS's productions are represented by a black solid line, and the two learners' are plotted with a grey solid line and a dashed line respectively.

Figure 1. F_0 measurement of the T2T1 sequence produced by the NS and two learners
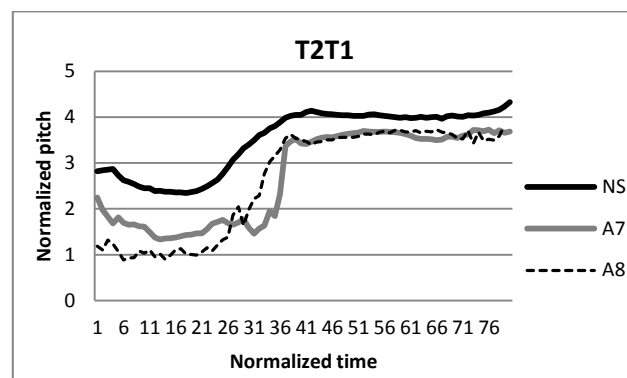
Figure 2. F_0 measurement of the T4T3 sequence produced by the native talker and two learners
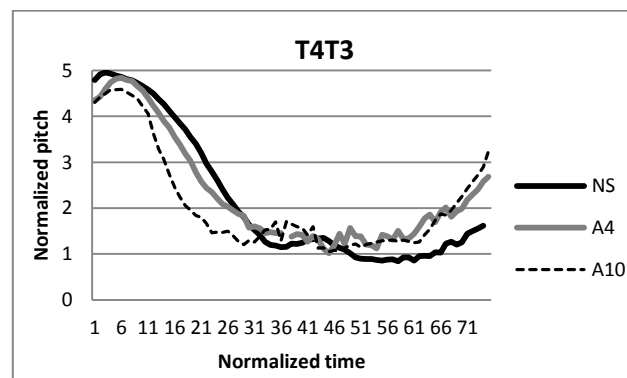
As can be seen in Figure 1, the learners' T2 productions have a lower onset (normalized pitch value=2.25 & 1.19) than the NS (2.82) and also dip to a lower pitch (1.42 & 0.90 vs. 2.35) before rising

for the onset of the following T1. Hence even though the overall contour shape of the learners' T2 is similar to that of the NS, their pitch height is considerably lower. This explains native listeners' confusion of these productions as T3, which occupies a low register. As for the production of the T4T3 sequence, the two learners dip almost as low as the NS (1.02 & 1.05 vs. 0.85) for T3, yet their pitch rises earlier than the NS by 13 normalized time points. In addition, their T3 offset is higher than the NS by 1.1 and 1.6 respectively. The relative early rise and high ending pitch of the learners' T3 productions are very likely the causes for the native listeners' T3→T2 identification.

5. DISCUSSION

This study examines English-speaking learners' production of Mandarin tones in different tonal contexts. Mandarin listeners' evaluation shows that the participants made significantly more T2→T3 errors than other error types in the initial position particularly when followed by T1. In the final position, on the other hand, their T3→T2 error is most prominent when the preceding tone is T4. The acoustic analysis comparing the learners' productions to those of the NS reveals that the learners' T2 tends to have a lower onset and valley than the NS, while their T3 rises earlier and ends at a higher offset. These suggest that the learners have more difficulty producing intelligible T2T1 and T4T3 sequences, and their productions resemble the pitch contour of T3T1 and T4T2 combinations, respectively.

The current findings are compatible with the observation in previous studies that T2 and T3 are more confusable for the learners [3, 8], which is probably because these two tones share similar onset values and falling-rising contour [5]. Interestingly, this study shows that the learners are biased toward producing T3 in one position and T2 in another, demonstrating that they are capable of producing intelligible T2 or T3, yet have difficulty maintaining the contrast in certain tonal contexts. Regarding their T2→T3 bias when followed by T1, it shows the learners' tendency to substitute a mid-high contour with a low-high contour, the latter involving a wider pitch range and more drastic pitch movement. Hence the learners' tonal errors do not seem to result from their narrower pitch range, contrary to [2], or the inability to raise or lower their pitch rapidly within one syllable, as suggested in [1], but rather from their imprecise implementation of more subtle pitch changes. The observation that this error occurs most frequently when followed by T1 and T4 suggests that the high

onset of the following tones probably contributes to lowering the onset and valley of the preceding T2. In other words, the tonal errors may be a result of tone dissimilation.

As for the prevalent word-final T3→T2 errors when the preceding syllable carries T4, the learners' most conspicuous deviation from the NS is that their T3 rises earlier and to a higher ending pitch. This similarly illustrates that the learners replace a smaller pitch movement with a larger one. Different from the T2T1 sequence in which the learners overshoot the falling portion, they amplify the final rise of T3 to a point that it sounds more like T2 to native listeners. Since the T3→T2 error occurs most commonly following T4, the low offset of T4 could very likely be the trigger for the extensive rising in the learners' T3 production, resembling a dissimilatory effect.

This study yields inconsistent findings with those in [7], which predominantly found assimilation in tone coarticulation. Tone dissimilation was reported in [10], which showed that Mandarin speakers' pitch peak of T1 and T4 became higher when followed by the low tone, T3, than by other tones. Yet the dissimilatory effect was very mild and only anticipatory in [10], in contrast to the magnitude and directionality observed in this study. The divergent findings between the present study and these previous ones are expected due to the many obvious differences in the elicited data, including non-native vs. native production and an imitation vs. a reading task. Nonetheless, the overall trend seems to suggest that the effect of tonal context on production is more substantial for L2 learners than for native speakers.

To conclude, the current findings suggest that the learners' tonal errors in production are generally a result of substituting a moderate pitch movement with a more dynamic one. The direction of their pitch change is conditioned by the adjacent tones, and the effect is dissimilatory in nature. This shows the learners' tendency to increase the contrast between the first and second tone in disyllables. Whether this effect is manifested in specific contexts or across the board will require more extensive analyses on the learners' productions in the future. Furthermore, different tasks may also be employed to tease apart the effect of perception and production, which are intertwined in the imitation task adopted in the current study. Nonetheless, this study demonstrates a clear effect of tonal contexts on L2 learners' production, which provides some insight for future investigation on L2 suprasegmental acquisition.

6. REFERENCES

[1] Bent, T. (2005). *Perception and production of non-native prosodic categories.* Doctoral dissertation, Northwestern University.

[2] Chen, G. T. (1974). The pitch range of English and Chinese speakers. *Journal of Chinese Linguistics* 2, 159-171.

[3] Chen, Q. (1997). Toward a sequential approach for tonal error analysis. *Journal of the Chinese Language Teachers Association* 32, 21-39.

[4] Ladd, D. R., Silverman, K., Tolkmitt, F., Bergmann, G., & Scherer, K. (1985). Evidence for the independent function of intonation contour type, voice quality, and F0 range in signaling speaker affect. *Journal of the Acoustical Society of America* 78, 435-444.

[5] Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese Tones. *Journal of the Acoustical Society of America* 102, 1864-1877.

[6] Shen, S. X. N. (1989). Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association* 24, 27-47.

[7] Shen, S. X. N. (1990). Tonal coarticulation in Mandarin. *Journal of Phonetics* 18, 281-295.

[8] Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America* 113, 1033-1043.

[9] White, C. M. (1981). Tonal perception errors and interference from English intonation. *Journal of Chinese Language Teachers Association* 16, 27-56.

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