Mechanisms of Tone Sandhi Rule Application by Non-native Speakers

Si Chen[1], Yunjuan He[2], Chun Wah Yuen[1], Bei Li[1], Yike Yang[1]

[1]Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China
[2] Department of Modern and Classical Languages, University of North Georgia, U. S.

sarahchen@polyu.edu.hk, yunjuan.he@ung.edu, david.cw.yuen@polyu.edu.hk,
benita.li@polyu.edu.hk, yike.yang@polyu.edu.hk

Abstract

This study is the first to examine acquisition of two Mandarin tone sandhi rules by Cantonese speakers. It designs both real and different types of wug words to test whether learners may exploit a lexical or computation mechanism in tone sandhi rule application. We also statistically compared their speech production with Beijing Mandarin speakers. The results of functional data analysis showed that non-native speakers applied tone sandhi rules both to real and wug words in a similar manner, indicating that they might utilize a computation mechanism and compute the rules under phonological conditions. No significant differences in applying these two phonological rules on reading wug words also suggest no bias in the application of these two rules. However, their speech production differed from native speakers. The application of third tone sandhi rule was more categorical than native speakers in that Cantonese speakers tended to neutralize the sandhi Tone 3 more with Tone 2 produced in isolation compared to native speakers. Also, Cantonese speakers might not have applied half-third tone sandhi rule fully since they tended to raise f0 values more at the end of vowels.

Index Terms: Mandarin tone sandhi, speech production, Cantonese L2 learners, wug words, functional data analysis

1. Introduction

Previous studies examined Mandarin tone production on monosyllables and disyllables by tonal and non-tonal language learners. There are four tones in Mandarin with tone numbers (1 standing for the lowest tone value, and 5 for the highest) [1]:

Tone 1 (T1) ma high-level (55) "mother"
Tone 2 (T2) ma high-rising (35) "hemp"
Tone 3 (T3) ma low-dipping (213) "horse"
Tone 4 (T4) ma high-falling (51) "to scold"

Isolated tone acquisition by American speakers is generally believed to be easier for T1 and T4 than T2 and T3 in perception and production [2], [3], though Shen [4] showed a different difficulty hierarchy where T4 and T1 are more problematic for American learners than T2 and T3, and Miracle [5] found evenly distributed errors across tones. Moreover, both groups of English and Cantonese speakers have most problems with perception and production in the tone pair T2-T3, and Cantonese speakers show additional difficulty with the pair T1-T4. Error patterns were also identified in production of disyllabic tones, and that of monosyllables was better than disyllables for both groups [6].

Models have been proposed to account for speech perception and production of non-native speakers such as the Perceptual Assimilation Model (PAM) [7]. PAM illustrates how a listener discriminates non-native phonemes by assimilating sounds with respect to the similarity of their own native system. A non-native contrast can be acquired successfully if it matches contrast in their native language, whereas difficulty arises if the non-native contrast is assimilated to the same category of their language. PAM has been used to account for acquisition of Mandarin tones [6]. However, Hao [6] pointed out PAM failed to account for one confusing tone pair of Mandarin T2 and T3 for Cantonese speakers. Cantonese has six distinct tones, where three tones on the checked syllables are viewed as counterparts of three level tones [8]:

Tone 1 (T1) si high-level (55/53) "teacher"
Tone 2 (T2) si high-rising (25) "history"
Tone 3 (T3) si mid-level (33) "test"
Tone 4 (T4) si mid-low falling (21) "time"
Tone 5 (T5) si mid-low rising (23) "market"
Tone 6 (T6) si mid-low level (22) "matter"

Based on the perceptual assimilation of Mandarin tones to Cantonese tones, Mandarin T2 is mapped onto Cantonese T2, T3 and T5, and Mandarin T3 is mapped onto Cantonese T4. Therefore, the pair of Mandarin T2 and T3 should belong to the "categorized-uncategorized" cases in PAM, where two sounds are assimilated to two categories, and discrimination should be good. However, the pair remains confusing for Cantonese learners.

Although many studies investigated Mandarin tone acquisition, they generally excluded tone sandhi pairs in their analysis of disyllabic tones, and rarely explored acquisition of tone sandhi rules. There are two tone sandhi rules in Mandarin:

a. T3(213) → T2 (35)/___ T3(213) (third-tone sandhi)
  lau 213 - pan 213 → lau 35 - pan 213 "boss"
b. T3(213) → 21/___{T1(55), T2(35), T4(51)} (half-third sandhi)
  lau 213 - şi55 → lau 21 - şi55 "teacher"
  lau 213 - liəŋ 35→ lau 21 - liəŋ 35 "old age"
  lau 213 - xua 51→ lau 21 - xua 51 "old saying"

The third-tone sandhi rule describes the phenomenon that T3 becomes T2 when followed by another T3. The half-third tone sandhi rule describes the phenomenon that T3 changes from the tone value 213 to 21 when followed by T1, T2 or T4. For native speakers, there is a debate between a lexical and computation mechanism in applying third tone sandhi [9]. The view of lexical mechanism believes that the encoding of tone sandhi involves lexical representations, thus only applying to real words as represented in inconsistent application in Taiwanese [10], whereas the view of computation mechanism

believes that the application of tone sandhi involves computation based on the phonological context as in Mandarin.

In addition to the debate on the two mechanisms, the application of two tone sandhi rules (third-tone sandhi and half-third sandhi) also exhibits differences among native speakers. Zhang and Lai [11] examined the application of two tone sandhi rules in Mandarin using wug words. They showed that the contour shapes of the third sandhi tones (the first T3 in T3+T3) were significantly different between real and wug words, and the wug words had a lower and later turning point and longer duration. They proposed that these results indicate a bias toward phonetically less motivated patterns. Yang [12] examined the application of the two tone sandhi rules by intermediate American L2 learners, and found the learners were able to apply phonological rules in general, though they may commit some errors. Both the effects of tone sandhi rules and lexical context were detected since the learners performed better in applying for the half-third sandhi rule than the third tone sandhi rule, and the American learners applied both rules better on familiar words.

The current study examines acquisition of Mandarin tone sandhi rules by Cantonese speakers, who have a different linguistic background from non-tonal language speakers. It aims to answer three research questions. First, do non-native speakers apply the Mandarin third tone sandhi and half-third tone sandhi? Second, do they apply them differently from the native speakers? Third, how does the lexical context (real words vs. wug words) affect the application of these two phonological rules?

2. Method

2.1. Participants

A total of 24 participants were recruited in this study, where 12 of them were native Mandarin speakers (age: 26.25 ± 4.61(Mean ± SD)) who had lived in Beijing for most of their lives (22.83 ± 3.43), 12 of them were native Cantonese (NC) speakers (age: 22.5 ± 2.15; age of acquiring Mandarin: 18.17 ± 2.21; years of learning: 2.33 ± 0.89). The participants were recruited from the Hong Kong Polytechnic University. All participants were paid for their participation, and signed informed consent forms approved by the Human Subjects Ethics Sub-committee at the Hong Kong Polytechnic University. No participants reported speaking or hearing problems. The participants were recorded by a GMH C 8.100 D headset at the speech lab of the Hong Kong Polytechnic University.

2.2. Stimuli

Based on experimental designs described in the literature [11], [13], we used six types of stimuli listed below. The first type consisted of real words, and all the other types were wug words as in "wug" tests [14]:
1. Real disyllabic words (AO-AO, where AO stands for actual occurring morphemes);
2. Non-occurring sequences consisting of real morphemes (*AO-AO);
3. Sequences of a real morpheme and a syllable of an accidental gap, which has a legal syllable, but bears a tone not allowed on that syllable in Mandarin (AO-AG, where AG stands for accidental gap);

4. Sequences of a syllable of an accidental gap and a real morpheme (AG-AO)
5. Sequences of two syllables of accidental gaps (AG-AG)
6. Sequences of two pseudo words, where the combination of vowels and consonants do not exist in Mandarin, but the vowel and consonant components of each syllable do exist.

Compared to previous studies, we improved the design by controlling the vowels in the first syllable of all types of diagrams to be the same in order to avoid the intrinsic f0 effect from vowels [15]. Pairs with aspirated and unaspirated onsets were not included to avoid consonant perturbation effect on f0 values [16]. Similar to Zhang and Lai [11], all the chosen diagram and individual characters were highly frequent words selected from character and diagram frequency corpus [17]. We also ensured that disyllabic wug words were not real words in any tonal combinations to avoid neighborhood effects.

Filler words consisting of real and wug words were also used in order not to reveal the purpose of the experiment. In total, there were 192 target stimuli and 192 filler words with all possible tonal combinations. All the monosyllabic stimuli used to elicit speech production were read by a native Mandarin speaker born in Beijing, and T3 was pronounced as a full tone with a falling-rising f0 contour.

2.3. Procedure

The software E-prime was used to present monosyllables with their characters (if available) and phonetic symbols (pinyin) along with sounds. Following the procedure in Zhang and Lai [11], participants heard two monosyllables presented to them with 800ms in between, and pronounced the two monosyllables as one disyllabic word. In a practice session, participants were given instructions and demonstrated how to put the two monosyllables they heard together to form a disyllabic word in Mandarin. In total, there were 24 blocks (six types of words * four tonal combinations) of the stimuli, and the order of presentation was counterbalanced across participants. They were instructed to speak at a normal speaking rate, and could self correct when necessary.

2.4. Acoustic and statistical analysis

The vowel portions of the recordings were first segmented, and f0 values were extracted at 20 normalized time points using the ProsodyPro Praat script [18]. After normalizing f0 values, we applied functional data analysis[16] to model f0 contours and compared pairs of contours such as contours produced by Mandarin and Cantonese speakers. We fit pairs of normalized f0 contours with the following model:

$$y_i(t_j) = f_i(t_j) + \varepsilon_{ij} \qquad (1)$$

where $y_i(t_j)$ is the normalized f0 value at time point t_j for the utterance i by each individual and $i = 1, \ldots, n$ and $j = 1, \ldots, m$. The error term ε_{ij} follows a normal distribution N(0, σ^2). For each pair of surface f0 contours, 20 break points were first chosen, and four B-spline basis functions were used to fit surface f0 curves. After fitting f0 contours, we proceeded to conduct functional t-tests to test whether there were differences between any pair. Two hundred random samples were used to calculate observed t-statistic, point-wise 0.05 critical value and maximum 0.05 critical value. The observed t-statistics exceeding critical values indicate statistical significance.

3. Results

3.1. Third tone sandhi rule

Significant differences were found between Beijing Mandarin speakers and Cantonese speakers in f0 contours of the first T3 (also called sandhi T3) in the disyllabic tonal context T3 + T3 on real syllables. The mean fitted curves of two groups of speakers are shown in Figure 1. The dotted lines stand for real words, and the red line represents Cantonese speakers and the black line represents Mandarin speakers. In general, Cantonese speakers have lower f0 values. From around the middle part to the end of the f0 contours, there were significant differences by functional t-test (observed t-statistic exceeding maximum 0.05 critical value in about 46% ~ 95% of the vowel, and exceeding point-wise 0.05 critical value in about 37% ~ 100% of the vowel).
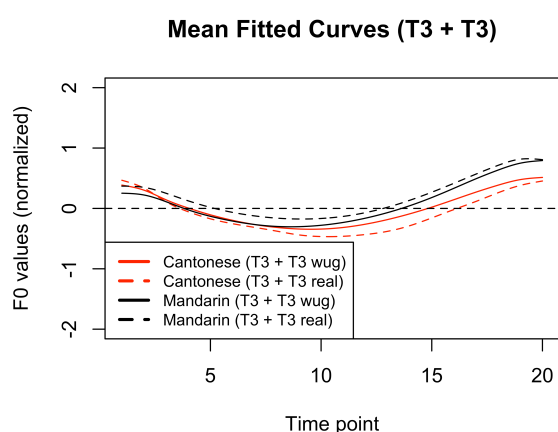

Mean Fitted Curves (T3 + T3)

Figure 1: *Mean fitted curves of the first T3 in T3 + T3 between Mandarin and Cantonese speakers.*

Also, for real syllables, we compared f0 contours of the sandhi T3 produced by the two groups of speakers with their speech production of T2 and T3 in isolation respectively. Consistent with previous findings [19], Beijing Mandarin speakers produced the sandhi T3 differently from both monosyllabic T2 (observed t-statistic exceeding maximum 0.05 critical value in about 4% ~ 100% of the vowel) and T3 (observed t-statistic exceeding maximum 0.05 critical value in about 10.9% ~ 100% of the vowel) in isolation. However, Cantonese speakers produced the sandhi T3 very similarly to T2 in isolation with a much smaller portion of significant difference (observed t-statistic exceeding maximum 0.05 critical value in about 6.9% ~ 31.7% of the vowel). The sandhi T3 is more different from T3 production in isolation (observed t-statistic exceeding maximum 0.05 critical value in about 21.8% ~ 25.7% and 48.5% ~ 90.1% of the vowel).

In producing the sandhi T3 in the context T3 + T3 on wug syllables (*AO+AO, AO+AG, AG+AO, AG+AG and Pseudo words), Cantonese speakers also showed significant differences from Mandarin speakers as shown in Figure 1. The solid lines stand for wug words, and the red line represents Cantonese speakers and the black line represents Mandarin speakers. Significant differences were found from around the middle part to the end of the f0 contours (observed t-statistic exceeding maximum 0.05 critical value in about 57% ~ 100% of the vowel). Similar to real syllables, Cantonese speakers showed lower f0 values toward the end of the vowel. Speech

production of the sandhi T3 in T3 + T3 on wug and real syllables was also compared for Mandarin and Cantonese groups respectively, and neither group showed statistical differences.

3.2. Half-third tone sandhi rule

For the first T3 pronounced in T3 + T1, T3 + T2 and T3 + T4 on real syllables, a smaller portion of significant differences were found toward the end of surface f0 contours between two groups of speakers (observed t-statistic exceeding maximum 0.05 critical value in about 80% ~ 100% of the vowel). The mean fitted curves from the two groups are shown in Figure 2. The dotted lines stand for real words, and the red line represents Cantonese speakers and the black line represents Mandarin speakers. Mandarin speakers produced lower f0 values toward the end of f0 contours due to the half-third tone sandhi rule. In contrast, Cantonese speakers retained a small rising portion of the half-third sandhi T3 toward the end of the vowel, indicating that they might not have fully applied half-third tone sandhi rule.
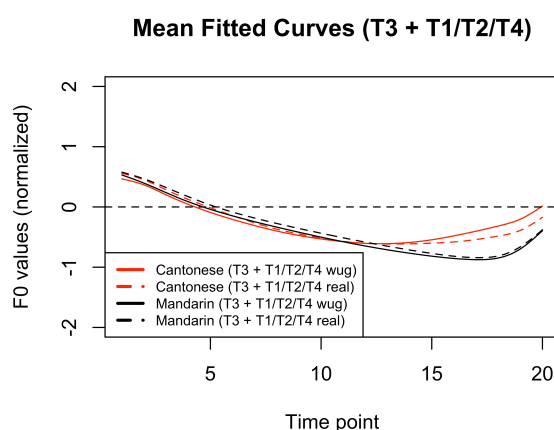

Mean Fitted Curves (T3 + T1/T2/T4)

Figure 2: *Mean fitted curves of the first T3 in T3 + T1/T2/T4 between Mandarin and Cantonese speakers.*

In producing the half-third sandhi tone rule on wug syllables (*AO+AO, AO+AG, AG+AO, AG+AG and Pseudo words), Cantonese speakers also showed significant differences from Mandarin speakers. Figure 2 plots the mean fitted curves of f0 contours produced by Cantonese and Mandarin speakers. The solid lines stand for wug words, and the red line represents Cantonese speakers and the black line represents Mandarin speakers. The f0 contours showed significant differences in two parts (observed t-statistic exceeding maximum 0.05 critical value in about 21% ~ 42% and 61%~100% of the vowel). Cantonese speakers produced lower f0 values for the first half, and then they raised the pitch contour toward the end, suggesting that they might not have applied the half-third tone sandhi fully on wug syllables either.

A comparison between wug and real syllables of the first T3 in T3 + T1/T2/T4 produced by Beijing speakers showed statistical differences for a small portion (observed t-statistic exceeding maximum 0.05 critical value in about 27.7% ~ 35.6% of the vowel). The mean fitted f0 curves have slightly lower values for wug syllables than real syllables. Similarly, we compared tone production on wug and real syllables by Cantonese speakers, and only marginal significance was found

(observed t-statistic did not exceed maximum 0.05 critical value, but exceeded point-wise 0.05 critical value in about 90% ~ 100% of the vowel). For wug syllables, Cantonese speakers tended to raise f0 contours more toward the end, indicating that they were not likely to apply the half-third tone sandhi rule fully on wug syllables. However, the sandhi T3 produced in T3 + T1/T2/T4 was significantly different from (observed t-statistic exceeding maximum 0.05 critical value for 100% of the vowel) T3 produced in isolation, indicating that the sandhi tone did vary from the original tone, though the phonetic realization was different from native speakers.

4. Discussions and Conclusions

Our results showed that Cantonese speakers differed from Beijing Mandarin speakers in their speech production of the two tone sandhi rules both on real and wug syllables.

For the third tone sandhi rule (T3 → T2/___ T3), Beijing Mandarin speakers produced a sandhi T3 with different f0 contours from T2 produced in isolation. This result of non-neutralization to T2 is consistent with many previous studies. Myers and Tsay [19] note two existing views of Mandarin third-tone sandhi rule application: the categorical and gradient view. The categorical view believes that there is a substitution of T3 by T2, and the gradient view believes that there is only a modification of T3. Our study demonstrated that Beijing Mandarin speakers did not completely substitute T3 by T2, and non-neutralization of f0 contours was found. However, Cantonese speakers showed much fewer differences in f0 contours of the sandhi T3 from T2 produced in isolation, suggesting that they tend to neutralize more with T2 in applying this third tone sandhi rule. They might have mapped the third sandhi tone to citation T2 in their own Cantonese tonal inventory just as they map Mandarin T2 to Cantonese T2, so they produce the sandhi tone similarly to T2 produced by them in isolation. In addition, they did apply this rule to wug syllables, and the speech production of sandhi T3 did not significantly vary from that on real syllables. The speech production results show that non-native speakers might be able to acquire the third tone sandhi rule, and are capable of applying the rules with a computational mechanism on wug words. They do not merely store the lexical items and phonological representations in their mind without being able to compute phonological rules. On the contrary, they applied the third tone sandhi rule in the same fashion both on real and wug syllables.

With a better control of vowel quality and consonant perturbation effect as well as more advanced statistical techniques, our study failed to replicate Zhang and Lai [11]'s results that the third sandhi T3 in T3 + T3 is pronounced differently on wug vs. real syllables by Beijing Mandarin speakers. There might also be individual differences in applying the third tone sandhi rule on wug and real syllables, which are worth investigation in future studies.

For the half-third sandhi tone (T3(213) → 21/___{T1(55), T2(35), T4(51)}), Cantonese speakers exhibited significant differences from Mandarin speakers on producing the T3 when it is toward the end of f0 contours. Cantonese speakers tended to raise f0 values toward the end (though different from the original T3's final rising part), which were higher than Mandarin speakers. Again, they applied a similar strategy to both wug and real words with marginally significant difference. For wug words, the final raising of f0 values seems to be more dramatic than it is in real words.

In sum, non-native speakers did apply tone sandhi rules both to real and wug words. However, the speech production differed from native speakers. There was no effect of lexical context on the application of these two phonological rules, since they applied the rules in a similar manner on real and wug words. The results were different from American learners [12], which may be due to a number of factors. Yang did not conduct statistical analysis on acoustic data, but relied on transcription of the recordings. Also, learners' proficiency level, acquisition age, years of learning and linguistic background may lead to the observed differences. Future studies are called for to examine the variation, especially to compare tonal and non-tonal speakers in the acquisition of tone sandhi rules.

5. Acknowledgement

6. References

[1] Y. R. Chao, *Mandarin Primer*. Cambridge: Harvard University Press, 1948.

[2] C. Y. Lee, L. Tao, and Z. S. Bond, "Identification of acoustically modified Mandarin tones by native listeners," *Journal of Phonetics*, vol. 36, no. 4, pp. 537-563, Oct. 2008.

[3] Y. Wang, A. Jongman, and J. Sereno, "Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training," *The Journal of the Acoustical Society of America*, vol. 113, no. 2, pp. 1033-1043, Jan. 2003.

[4] X. S. Shen, "Toward a register approach in teaching Mandarin tones," *Journal of the Chinese Language Teachers Assoc.*, vol. 24, pp. 27-47, 1989.

[5] W. C. Miracle, "Tone production of American students of Chinese: A preliminary acoustic study," *Journal of Chinese Language Teachers Assoc.*, vol. 24, pp. 49-65, 1989.

[6] Y. C. Hao, "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," *Journal of Phonetics,* vol. 40, pp. 269-279, Mar. 2012.

[7] C. T. Best, "A direct realist view of cross-language speech perception," in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research,* W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 171-204.

[8] R. S. Bauer and P. K. Benedict, *Modern Cantonese phonology*. Berlin: Mouton de Gruyter, 1997.

[9] C. Zhang, Q. Xia, and G. Peng, "Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study," *Journal of Neurolinguistics*, vol. 33, pp. 149-162, Feb. 2015.

[10] J. Zhang, Y. Lai, and C. Sailor, "Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication," *Lingua,* vol. 121, no. 2, pp. 181-206, Jan. 2011.

[11] J. Zhang and Y. Lai, "Testing the role of phonetic knowledge in Mandarin tone sandhi," *Phonology,* vol. 27, no. 01, pp. 153-201, May 2010.

[12] C. S. Yang, *The Acquisition of L2 Mandarin Prosody: From Experimental Studies to Pedagogical Practice.* Amsterdam, Netherlands: John Benjamins, 2016. doi: 10.1075/bpa.1

[13] C. Zhang and G. Peng, "Productivity of Mandarin third tone sandhi: a wug test," in *Eastward flows the great river: Fetschrift in Honor of Prof. William S-Y. Wang on His 80th Birthday,* G. Peng and F. Shi, Eds., Hong Kong: City University of Hong Kong Press, 2013, pp. 256-282.

[14] J. Berko, "The child's learning of English morphology," *Word*, vol. 14, pp. 150-177, 1958.

[15] J. M. Hombert, "Development of tones from vowel height?" *Journal of Phonetics,* vol. 5, pp. 9-16, 1977.

[16] S. Chen et al., "Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal Languages," *Speech Communication*, vol. 88, pp. 17-38, Apr. 2017. doi: 10.1016/j.specom.2017.01.006

[17] J. Da. (2004). Chinese text computing: syllable frequencies with tones. [Online]. Available: http://lingua.mtsu.edu/chinese-computing/phonology/syllabletone.php

[18] Y. Xu, "ProsodyPro — A tool for large-scale systematic prosody analysis," in *Tools and Resources for the Analysis of Speech Prosody,* Aix-en-Provence, France: Laboratoire Parole et Langage, 2013. pp. 7-10. Available: http://www.homepages.ucl.ac.uk/~uclyyix/ProsodyPro/

[19] J. Myers and J. Tsay, "A formal functional model of tone," *Language and Linguistics,* vol. 4, pp. 105-138, Jan. 2003.