https://doi.org/10.1016/j.specom.2019.10.008

Mechanisms of Tone Sandhi Rule Application by Tonal and Non-tonal Non-native Speakers

Si Chen[1, 2]*, Yunjuan He[3], Ratree Wayland[4], Yike Yang[5], Bei Li[6], Chun Wah Yuen[7]

*Corresponding author: sarahchen@polyu.edu.hk, Tel:852-59414757
[1] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China;
[2] Hong Kong Polytechnic University-Peking University Research Centre on Chinese Linguistics, Hong Kong, China; sarahchen@polyu.edu.hk
[3] Department of Modern and Classical Languages, University of North Georgia, U. S.; yunjuan.he@ung.edu
[4] Department of Linguistics, University of Florida, U. S.; ratree@ufl.edu
[5] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China; yi-ke.yang@connect.polyu.hk
[6] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China; benita.li@polyu.edu.hk
[7] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Hong Kong SAR, China; david.cw.yuen@polyu.edu.hk

Abstract

This study is the first comprehensive acoustic study to examine the acquisition of two Mandarin tone sandhi rules: the third tone sandhi and the more phonetically motivated, half-third sandhi rule by both tonal (Cantonese) and non-tonal (American English) speakers using a Wug Test. Participants were asked to form disyllables from two monosyllabic morphemes. To test for the operation of the lexical versus the computation mechanisms in sandhi rule application, both real and various types of wug (nonsense) morphemes were included. Functional data analysis revealed that Cantonese and American speakers apply the two rules similarly on both real words and wug words, suggesting that the sandhi forms are stored as part of the representation of the abstract Tone 3 (T3) category, and computation of allophonic variants is likely to be involved during production. However, in their computation of tone sandhi rules, L2 learners showed less detailed and less accurate production of tonal contours compared to native speakers, due, perhaps, to less detailed phonological representations of allophonic variants. In general, Cantonese speakers performed better than American speakers. Perceptual mapping between Mandarin sandhi T3 to existing Cantonese tone categories may be responsible for the observed pitch contours among Cantonese speakers. Finally, no phonetic bias was found in the application of the two sandhi rules among these groups of L2 learners, which is likely due to more variability in L2's speech, obscuring any differences that may exist.

Key words:
tone sandhi; wug words; tonal acquisition

1.0 INTRODUCTION

Tone sandhi is defined as a type of tonal alternation triggered by certain phonological environments (e.g., Chen 2000; Chen, 2012). Mandarin has two sandhi patterns involving Tone 3 (T3), known as third-tone sandhi and half-third sandhi (Zhang, & Lai, 2010), respectively. The half-third sandhi rule is considered to be more phonetically motivated than the third-tone sandhi rule, and is more easily and accurately acquired (Zhang, & Lai, 2010).

Numerous perceptual studies have been conducted to examine the effects of first language (L1) background on isolated Mandarin tone perception (Hallé, Chang, & Best, 2004; Hao 2012; Ning, Shih, & Loucks 2014; Wu, Munro, & Wang 2014; Zhang, Samuel, & Liu 2012; Wang 2013), and a few have examined tone perception in disyllabic contexts (Broselow, Hurtig, & Ringen, 1987; He, 2013, 2014; Hao, 2012; Hao, 2018). In contrast, relatively fewer studies have examined the effects of L1 linguistic background on lexical tone production in Mandarin monosyllables or disyllables, particularly variants of T3 in sandhi contexts. Perceptual studies revealed difficulties in Mandarin tone identification and discrimination among non-native listeners from both tonal and non-tonal linguistic backgrounds. However, it remains inconclusive whether tonal listeners have an advantage over non-tonal listeners in isolated tone perception of a foreign tonal language (see e.g., Wayland, & Guion, 2004; Hao, 2012; So, & Best, 2010). Non-tonal speakers may have difficulty in perception but not necessarily in production (Dong, Tsubota, & Dantsuji, 2013). Hao (2012) also suggests that tonal L2 speakers may not have an advantage over non-tonal L2 speakers in perception and production of Mandarin tones. Yang (2015) further points out that the established tonal categories in tonal L2 learners interfere with acquisition of Mandarin tones, and non-tonal L2 learners' production accuracy may increase once new tonal categories have been formed.

Two models have been proposed to account for perception and production of non-native phonemic categories: the Perceptual Assimilation Model (PAM; e.g. Best, 1995) and the Speech Learning Model (SLM; e.g. Flege, 1995). PAM predicts degrees of perceptual discriminability of non-native contrasts based on how they are assimilated to existing L1 categories, driven by articulatory features. SLM focuses on perceived differences between L1and L2 sounds and the ability to form new phonetics categories. For SLM, production accuracy may depend directly on perceptual ability.

Only a few studies have examined the acquisition of allophonic variations at the segmental level, and L1 acquisition theories based on the notions of transfer and markedness are typically used to explain the results. For example, Shoemaker (2014) attributed superior acquisition of English prevocalic glottalic stop over aspiration as a word boundary cue among French speakers to the universally more salient nature of the glottal stop over aspiration. Macleod & Fabiano-Smith (2015) examined acquisition of stop-spirant alternations in Spanish-English and stop-affricate alternation in French-English bilingual children and suggested that the higher error rates in the production of velar spirants in comparison to the bilabial spirants among Spanish-English bilingual children is due to the fact that the velar place of articulation is more marked cross-linguistically than the bilabial place of articulation On the other hand, the faster rate of stop-affricate allophone acquisition among French–English bilinguals over French monolinguals was attributed to the use of articulatory knowledge from English. The results are consistent with the Processing Rich Information from Interactive Representations (PRIMIR) model, which suggests that children make use of statistical regularities to process and store information during phonological acquisition (Werker & Curtin, 2005). Curtin, Byers-Heinlein and Werker (2011) further proposes that the interactions between a bilingual child's two languages should be accounted for, and a bilingual child may create a unique phonological representation.

At the supra-segmental level, allophonic representations of tones have been explored more among native speakers. Evidence has been found to support the multiple variant representation account, suggesting that the allophonic variant of the third sandhi tone is stored together with the canonical representation of T3 (Chen 2011, Li & Chen 2015). Fewer studies have examined the acquisition of variants of T3 by second language learners. Zhang (2017) found that the beginner and intermediate American learners tended to overproduce the full-T3 to substitute the half third sandhi tone variant, whereas the advanced learners could finally acquire half third sandhi tone due to greater exposure to the target language. The current study fills this research gap by examining alternations of lexical tones among non-native speakers. To this end, production of Mandarin Tone 3 in two sandhi contexts by native speakers of English and native speakers of Cantonese are compared to each other and to native Mandarin speakers' production.

1.1 Mandarin and Cantonese tones

Mandarin Chinese has four lexically contrastive tones, with Mandarin Tone 1 having high-level pitch (55), Tone 2 having high-rising pitch (35), Tone 3 having low-dipping pitch (213) and Tone 4 high-falling pitch (41). The Chao tone numbers used in describing each tone reflect the starting and ending point of the speaker's voice pitch on a scale from 1 to 5, where 1 indicates the lowest pitch of the speaker and 5 the highest (Chao, 1948). The four contrastive tones are commonly illustrated with the syllable [ma]: [ma]1(55) 'mother'; [ma]2(35) 'hemp'; [ma]3(213) 'horse'; [ma]3(41) 'to scold'. The pitch tracks of the four tones averaged over four syllables produced by seven speakers are given in Figure 1.

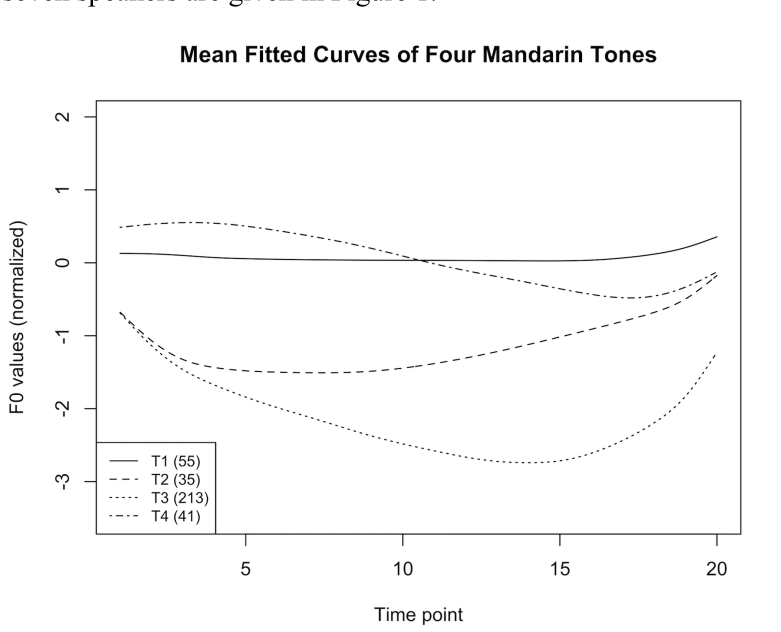


Figure 1. Four time-normalized Mandarin tones based on the mean F0 values calculated from four syllables (/fu/, /ji/, /ʃɤ/ and /ʃi/) produced by seven female speakers.

With three level tones and three contour tones in open syllables, the tone system in Cantonese is more complex than in Mandarin. These are illustrated in [fu] syllables: fu1 (55/53) 'to call'; fu2 (25) 'bitter; fu3 (33) 'rich'; fu4 (21) 'to hold', fu5 (23) 'woman' and fu6 (22) 'rotten'. The additional three tones on the checked syllables are viewed as counterparts of three level tones in open syllables (Bauer, & Benedict, 1997; Matthews, & Yip, 1994): The pitch tracks of the six tones averaged across four open syllables produced by seven female speakers are given in Figure 2.

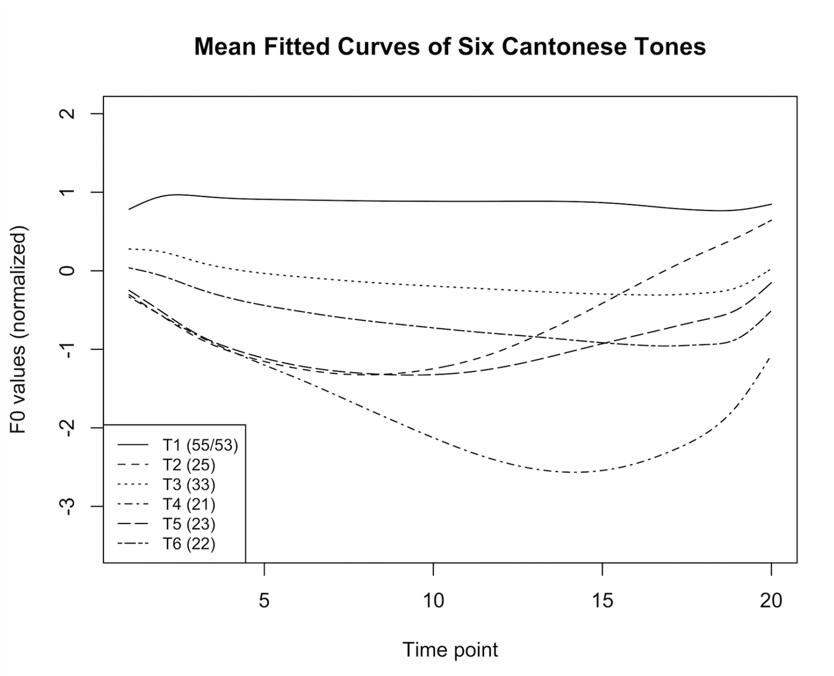


Figure 2. Six time-normalized Cantonese tones based on the mean F0 values calculated from four syllables (/fu/, /ji/, /sɛ / and /si/) produced by seven female speakers.

1.2 Tone Sandhi in Mandarin Chinese and Cantonese

In Mandarin Chinese, two sandhi patterns exist: the third-tone sandhi and the half-third sandhi rules, which are synchronically productive in both disyllables and phrases (Cheng, 1986; Chao, 1948, 1968; Zhang, & Lai, 2010). As exemplified in (a) below, Mandarin Tone 3 (213) becomes Mandarin Tone 2 (35) when followed by another 213; but 213 becomes 21 when followed by any other Mandarin tones (b).

Mandarin tone sandhi

a. T3(213) → T2 (35)/___ T3(213) (third-tone sandhi)
ni213 xɑu213 → ni35 xɑu213 "hello"

b. T3(213) → 21/___{T1(55), T2(35), T4(51)} (half-third sandhi)
mei 213 t^{hj}æn55 → mei 21 t^{hj}ɛn 55 "every day"
mei 213 k^{hw}oo35 → mei 21 k^{hw}ɔɔ 35 "the United States of America"
mei 213 l^{j}ii51 → mei 21 l^{j}ii51 "beautiful"

This more traditional proposal treated the full-T3 (214) as the underlying form of T3, and the two variants are the third sandhi tone (35) in rule (a) and half third sandhi tone (21) in rule (b). Zhang (2017) noted that the question about the underlying form of T3 is still open, where the half third sandhi tone may be taken as the underlying form, but the traditional proposal is well accepted in the field of teaching Chinese as a second language. Even though both sandhi processes are considered phonological involving tonal alternation patterns not predictable by tonal coarticulation, the half-third sandhi is considered more phonetically motivated. Specifically, the half-third sandhi involves the truncation of the second half of the pitch contour, consistent with the phonetic mechanism of a tone change in short syllables, which exhibits post-lexical rather than lexical characteristics (e.g., being independent of syntactic brackets, allophonic and applied across the board) and does not correspond to a historical sandhi pattern (Zhang, & Lai, 2010). In contrast, while contextual pitch modification (i.e., tonal coarticulation) occurs in Hong Kong Cantonese, there is a lack of tone sandhi patterns in the proper sense. Cantonese is reported to have tone alternation in restricted situations, such as in compounds and reduplicated expressions. For example, *yauh* "right" bears a low level tone, but the level tone changes to a rising tone in a compound. This tone alternation is often referred to as tone change instead of tone sandhi (Matthews, & Yip, 1994) because unlike Mandarin tone sandhi rules, where it is applied as long as the phonological environment is met, Cantonese tone change is not widely applicable, and only occurs in certain situations.

1.3 Computation versus lexical mechanisms in tone sandhi application

With regards to tone sandhi, the question remains whether the sandhi contour is encoded and stored lexically as part of the morpheme (lexical mechanism) or as part of the abstract tonal category and contextually compiled (computation mechanism) (Nixon et al, 2015; Zhang, Xia and Peng, 2015). Specifically, lexical mechanism refers to the encoding and accessing of the surface tone sandhi contours stored as part of the morpheme. Evidence in support of the lexical mechanism in sandhi rule application comes from findings that tone sandhi rules only apply to real words (Hsieh, 1970, 1975, 1976; Zhang, Lai, & Sailor, 2011), and that production of tone sandhi in real words and wug words are acoustically different (Zhang & Peng, 2013; Chen & Li, 2017; Chen & Li, submitted). The computation mechanism, on the other hand, refers to the encoding and compiling of the sandhi and non-sandhi forms (phonological alternation) stored as part of the abstract tonal category based on the phonological contexts. Evidence for this mechanism comes from both behavioral and neurophysiological studies (e.g., Politzer-Ahles & Zhang, 2015; Zhang, Xia, & Peng, 2015). For instance, Politzer-Ahles & Zhang (2015)'s finding suggested that Mandarin disyllabic words bearing a two-T3 string are stored as the two-T3 string underlyingly and are turned into surface forms before articulation. This result is consistent with findings from an ERP study (Zhang, Xia & Peng, 2015) in which phonological encoding of Mandarin T3 was more effortful in sandhi than in non-sandhi context. Moreover, Nixon, Chen and Schiller (2015) offers evidence to suggest that both phonemic (category) and sub-phonemic (context-specific) representations are activated in Sandhi word production by native Mandarin speakers.

In the current study, sandhi tone production in real words and wug words by native and second-language learners of Mandarin were acoustically compared to reveal similarities and differences in tone sandhi applications and the underlying mechanism involved.

1.4 Effects of L1 experience on tone production

A few production studies have investigated tone production in isolated tones and disyllabic tonal combinations by L2 learners. English speakers are found to use a narrower pitch range than do native speakers (Chen, 1974; Tu, Hsiung, Wu, & Sung, 2014; White, 1981). For the production of isolated tones by American speakers, it is generally reported that T1 and T4 are easier than T2 and T3 in both perception and production (Lee, Tao, & Bond, 2008; Wang, Jongman, & Sereno, 2003; Elliott, 1991; Sun, 1997; Chen, 1997; Leather, 1990), though Shen (1989) found a different hierarchy and Miracle (1989) found evenly distributed errors across tones. There were 12 possible error types in the production of Cantonese speakers (e.g., T1-to-T2, T1-to-T3, T1-to-T4, T4-to-T3), where T2-to-T3 and T3-to-T2 confusions were the most highly significant (Hao, 2012). The production error types summarized from Wang et al. (2003), Miracle (1989) and Hao (2012) are compared in Table 1. Wang et al. (2003) pointed out that American speakers tended to mispronounce T1 as T2 or T4 and T4 as T1 or T3, and they also confused T2 with T3. Miracle (1989) described the errors in detail for each mispronounced tone by American English speakers and Hao (2012) listed all the errors committed by Cantonese speakers. Note that there is little agreement on the mistakes made by American learners. The observed variability in the reported error patterns may be due to various factors such as differences in proficiency, the mode of instruction, and the nature of the input across studies.

Table 1 ***Mispronunciation by American and Cantonese speakers***

Mandarin Tones	American Speakers		Cantonese speakers
	Mispronounced as (Wang et al., 2003)	Tonal register and contour errors (Miracle, 1989)	Significantly mispronounced as (e.g. Hao, 2012)
T1	T2, T4	Too low; falling tone	T4
T2	T3	High beginning of the tone; falling tone or level tone	T3
T3	T2	Too high; Rising tone	T2
T4	T1, T3	Mid beginning of the tone; N.A	T1

Both English and Cantonese speakers have most problems with perception and production with Mandarin T2-T3 pair, and Cantonese speakers have additional difficulty with the T1-T4 pair (Hao, 2012; So, & Best, 2010). An experiment of perceptual assimilation of Mandarin tones to Cantonese tones showed that both T1 and T4 were mapped onto overlapping Cantonese tones, and hence, were difficult to distinguish. However, T2 and T3 were mapped onto different Cantonese tones, and thus cannot be explained by PAM. Instead, the difficulty may lie in acoustic similarity (Hao, 2012).

Error patterns have also been identified in studies of disyllabic tones. Hao (2012) points out that the production of monosyllables is more accurate than that of disyllables. Error rates were the highest on the initial syllable, lower on the final position of disyllables, and lowest in monosyllables (Hao, 2012). Similar results were found in English speakers (He, 2013, 2014) and Japanese speakers (Dong et al, 2013).

2.0 The current study

Although production of isolated tones and tonal combinations have been investigated, third tone sandhi pairs were usually excluded from the above studies involving disyllabic tones (Hao, 2012; He, 2014; Dong, Tsubota, & Dantsuji, 2013). Some preliminary studies report that the computation mechanism might be involved in tone sandhi application by tonal speakers (Chen, He, Yuen, Li, & Yang, 2017). It is yet unknown whether tonal and non-tonal speakers perform differently in producing the two Mandarin tone sandhi rules on both real words and different types of wug words.

In this study, we examined the ability to produce Mandarin tones in sandhi contexts among a group of tonal (Cantonese) speakers and a group of non-tonal (American English) speakers who are learners of Mandarin using a wug test paradigm (Berko, 1958). This paradigm, widely used to test productivity of morphophonological rules and alternations (e.g., Hsieh 1970, 1975, 1976; Bybee, & Pardo, 1981; Wang, 1993; Albright, Andrade, & Hayes, 2001; Albright, 2002; Albright, & Hayes, 2003, Pierrehumbert, 2006; Zuraw, 2000, 2007; Hayes, & Londe, 2006) allows us to test the speakers' knowledge of a phonological pattern..

The three research questions we explored are:

1. Are Mandarin sandhi tonal contours produced on real words and wug words similarly among native speakers of another tone language (Cantonese) and native speakers of non-tone language (American English)?
2. Are Mandarin allotonic tonal variations produced in both real words and wug words by Cantonese and American English speakers equally as accurate as native Mandarin speakers?
3. Is the production of the third-tone and the half-third tone sandhi equally as accurate in real words and wug words among Cantonese and American English speakers? Is there a phonetic bias in rule applications?

With regards to the first research question, there exists a debate of whether the sandhi contour is stored lexically as part of the morpheme or as part of the tonal category among native Mandarin speakers. As pointed out by Nixon et al. (2015), "If the sandhi contour is processed as part of a purely abstract Tone 3 category, effects should be equal for all morphemes. However, if the contour is processed by exemplar, morphemes which rarely occur in sandhi contexts should see significant attenuation of the contour effect relative to morphemes that frequently occur in sandhi contexts" (p. 501). However, it is currently unknown if and how the two mechanisms are involved in sandhi tone production by learners of Mandarin from different L1 backgrounds. Similarity in sandhi tone production in real words and wug words would suggest that sandhi contours may be represented as part of the abstract T3 category and computed on-line in contexts. However, according to Clahsen & Felser (2006), "The syntactic representations adult L2 learners compute during comprehension are shallower and less detailed than those of native speakers" (p.3). If this 'shallow' hypothesis also applies to the phonology domain of language, it is possible that the representations of sandhi tone among Cantonese and American English speakers may also be less detailed and, therefore, their conversion to motor programs for production is also less accurate than that of native Mandarin speakers. However, with relatively greater native experience with tone production, Cantonese speakers may be more accurate in both establishing and translating sandhi tone representations to articulation than American English speakers, hence the hypothesis for our second research question. For the third research question, it has been claimed that the half-third sandhi is more phonetically motivated as it involves the truncation of the second half of the pitch contour, a mechanism deemed phonetically natural of a tone change in short syllables (Zhang, & Lai, 2010). Thus, half-third sandhi tone production may be more accurate than third-tone sandhi production among both Cantonese and American English speakers. However, various different results were reported for L2 learners. Zhang (2013) found that American learners tend to perform better in the third tone sandhi rule, whereas Yang (2016) found the half-third sandhi rule to be easier for American learners. The current study used Cantonese and American learners to test whether there is a phonetic bias in rule applications in second language acquisition.

2.1 Methodology

2.1.1 Subjects

A total of 23 learners were recruited in this study, 12 of which were native Cantonese (NC) speakers (age: 22.5 ± 2.15, Mean ± SD; starting age: 18.17 ± 2.21; years of learning: 2.33 ± 0.89) and 11 were native English (NE) speakers (age: 22 ± 1.79; starting age: 18.27 ± 2.49; years of learning: 3.0 ± 0.92). The age of participants starting to learn Mandarin and the length of their Mandarin study were controlled so they were similar between NC and NE groups. All the Cantonese and American speakers were intermediate learners of Mandarin. Based on their self-reported proficiency scores [scale 1(low proficiency)-5(high proficiency)] on four aspects of reading, listening, speaking and writing, we calculated proficiency scores for both groups [NC (mean(sd) = 3.38(0.75); NE (mean(sd) = 3.21(0.45)]. Regarding instructions, the third sandhi rule is introduced as a linguistic rule to American speakers and is also introduced to Cantonese speakers in the classroom. Twelve native Mandarin (NM) speakers (age: 26.25 ± 4.61), who had lived in Beijing for most of their lives (22.83 ± 3.43), were included as the control group. No participants reported speaking or hearing problems. We followed the procedure of power analysis for linear mixed effects models to achieve an effect size of -0.05 with two variables (group and word type), requiring 20 subjects in total to achieve the power of 80%. So in our experiment, we recruited 35 subjects in total.

All participants were paid for their participation and signed informed consent forms approved by the Human Subjects Ethics Sub-committee of the Hong Kong Polytechnic University and the University of North Georgia IRB (Internal Review Board). The Cantonese participants were recorded wearing a GMH C 8.100 D headset at the speech lab of the Hong Kong Polytechnic University. The American participants were recorded by a Sanako SLH07 headset in a quiet room of University of North Georgia.

2.1.2 Stimuli

Based on previous studies (Zhang, & Lai, 2010; Hsieh, 1970, 1975, 1976; Zhang, & Peng, 2013), we used six types of stimuli as listed below. The first type is real words, and all the other types are wug words.

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 (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 (e.g. Zhang, & Lai, 2010), we improved their research design by including more types of wug words, controlling for the vowels in the first syllable of all types of diagrams by making them similar to avoid the intrinsic F0 effect from vowels on pitch contours (Hombert, 1977; House, & Fairbanks, 1953). Pairs with aspirated and unaspirated onsets were not included to avoid consonant perturbation effect on F0 values (Xu, & Xu, 2003). Similar to Zhang and Lai (2010), all the chosen diagrams and individual characters were highly frequent, selected from character and diagram frequency corpus (Da, 2004). We also ensured that disyllabic wug words were not real words in any tonal combinations to avoid neighborhood effects.

Filler words consisting of real words and wug words were also included to avoid revealing the purpose of the experiment. In total, there were 192 target stimuli and 192 filler words with all possible tonal combinations. Twenty-four blocks of the stimuli (four blocks for each type) were presented to the participants, where the order of presentation was counterbalanced across participants. We recorded all the monosyllabic stimuli read by a native Mandarin speaker born in Beijing, and T3 was pronounced as full third tones with a falling-rising F0 contour. We used E-prime to present monosyllables and their characters (if available) and phonetic symbols (pinyin) along with sounds. Participants heard two monosyllables presented to them with 800ms between the two monosyllables. Participants were given instructions and demonstrated how to put the two monosyllables together to form a disyllabic word in Mandarin using some filler words. They were instructed to speak at a normal speaking rate and to self-correct when necessary. Practice sessions and instructions with examples were also offered before the experimental session. Before the experiments, all participants familiarized themselves with all the real words and monosyllables in wug words. We also tested them by letting them write down the characters and translation to make sure that they knew the meaning of all the real words.

2.1.3 Acoustic and statistical analyses

The vowel portions of the recordings were first segmented, and F0 values were extracted at 20 normalized time points using the ProsodyPro Praat script (Xu, 2013). After normalizing F0 values, we applied functional data analysis (Chen, Zhang, McCollum, & Wayland, 2017; Ramsay, & Silverman, 2005, 2009) to model F0 contours and compare pairs of contours, such as contours produced by native versus non-native speakers. Levitin, Nuzzo, Vines and Ramsay (2007) argued that traditional statistics such as ANOVA, repeated measures ANOVA and GLM-based models are inferior to functional data analysis as they can only answer simple questions with some assumptions violated. In modelling F0 curves, growth curve analysis and functional data analysis are two current statistical methods used to test statistical differences in pairs of F0 contours (e.g. Li & Chen, 2016; Chen, Zhang, McCollum, & Wayland, 2017; Mirman, 2014). Growth curve analysis can be used to fit a pair of F0 curves, and statistically test if the two curves are significantly different in slope and intercept. Functional t-tests fit F0 curves using basis functions instead of polynomials and can further specify the regions where two curves are significantly different.

Specifically, we fit pairs of normalized F0 contours with the following model:

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

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 $N(0, \sigma^2)$. For each pair of surface F0 contours, we chose 20 break points and used four B-spline basis functions to fit surface F0 curves. We made use of a generalized cross-validation measure (GCV) to determine the best values for the smoothing parameter λ. After fitting F0 contours, we proceeded to conduct functional t-tests to test whether there were differences between any pair of F0 contours. Two hundred random samples were simulated to calculate observed t-statistic, point-wise 0.05 critical value and maximum 0.05 critical value. Statistical differences between two F0 contours were indicated when observed t-statistics exceeded critical values.

3.0 Results

3.1 Performance of Cantonese learners

3.1.1 Third tone sandhi rule

F0 contours of the first sandhi T3 in the disyllabic tonal context T3 + T3 both on real and wug Mandarin syllables showed significant differences between Beijing Mandarin speakers and Cantonese speakers. Figure 3 plots the mean fitted F0 curves produced by two groups of speakers.

For the sandhi T3 on real syllables, Cantonese speakers tended to produce lower F0 values than Mandarin speakers. The mean fitted values on several points within the significant areas are reported in Table 2. The results showed significant differences by functional t-test from around the middle part to the end (observed t-statistic exceeding maximum 0.05 critical value in about 45% ~ 98% of the vowel).

Table 2 ***The mean fitted values of the sandhi T3 on real syllables and wug syllables within their significant regions***

Percentage (real)	Mandarin speakers	Cantonese speakers	Percentage (wug)	Mandarin speakers	Cantonese speakers
40%	-0.16	-0.37	20%	0.11	0.13
50%	-0.17	-0.47	40%	-0.30	-0.32
60%	-0.07	-0.44	60%	-0.17	-0.27
70%	0.13	-0.29	80%	0.49	0.27
80%	0.57	0.11	100%	0.79	0.51
90%	0.84	0.39			

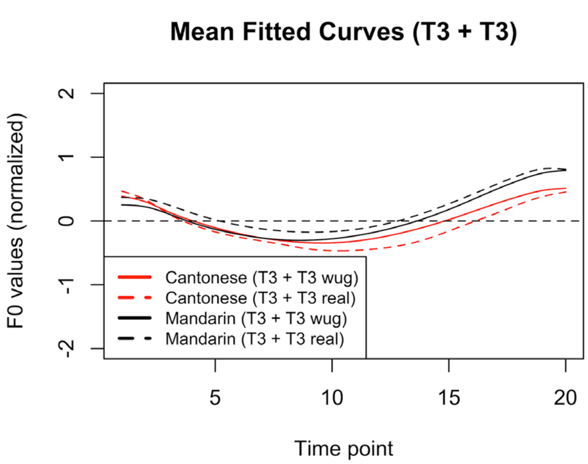


Figure 3. Mean fitted curves of the first sandhi T3 in T3 + T3 between Mandarin and Cantonese speakers. The dotted lines stand for real words, and solid lines for wug words. The red lines represent Cantonese speakers and the black lines Mandarin speakers.

In addition, Cantonese speakers were also significantly different from Mandarin speakers in producing the sandhi T3 on wug syllables in general (*AO+AO, AO+AG, AG+AO, AG+AG and Pseudo words), as plotted in Figure 3. Based on functional data analysis, 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 1% ~ 4%, 56% ~ 100% of the vowel). Similar to real syllables, Cantonese speakers showed lower F0 values toward the end of the vowel, reported in Table 2.

Production of the sandhi T3 in T3 + T3 on wug and real syllables was also compared for both Mandarin and Cantonese groups respectively, and yet neither group showed statistical differences in the third-tone sandhi application on wug vs. real syllables.

3.1.2 Half-third tone sandhi rule

Figure 4 plots the mean fitted curves in producing the first T3 in T3 + T1, T3 + T2 and T3 + T4 from the Mandarin and Cantonese groups on both real and wug syllables. On both real and wug syllables, Mandarin speakers produced lower F0 values for a bigger portion of the contour toward the end. On real syllables, a smaller portion of significant differences was found toward the end of surface F0 contours between two groups of speakers (observed t-statistic exceeding maximum 0.05 critical values in about 82% ~ 96% of the vowel) as shown in Figure 5. In contrast, Cantonese speakers stopped lowering F0 earlier and started to initiate a rising portion earlier than Mandarin speakers, though Mandarin speakers showed a steeper slope of rising toward the end of the contour.

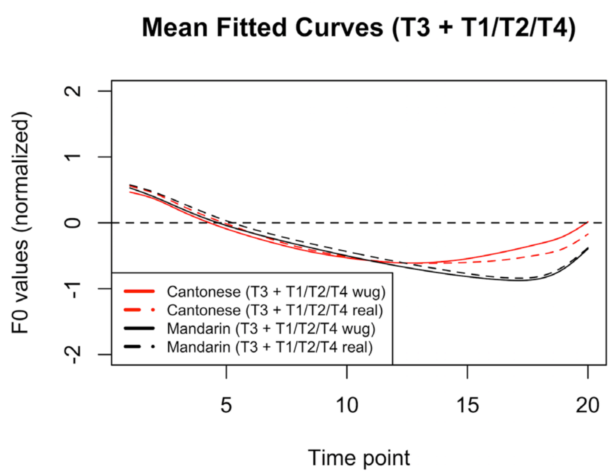


Figure 4. Mean fitted curves of the first T3 in T3 + T1/T2/T4 between Mandarin and Cantonese speakers. The dotted lines stand for real words, and solid lines stand for wug words. The red lines represent Cantonese speakers and the black lines represent Mandarin speakers.

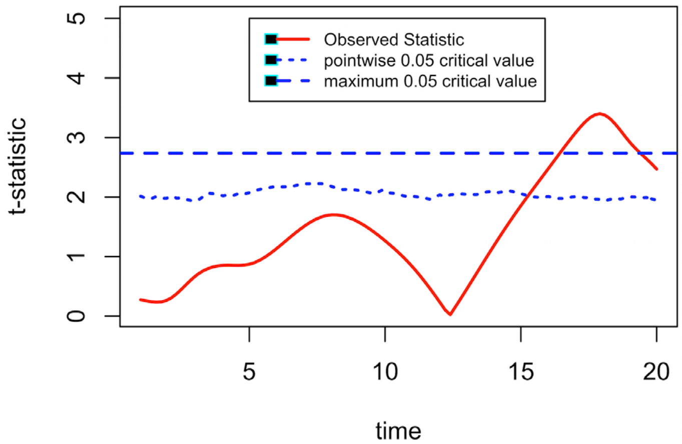


Figure 5. A functional t-test of real syllables (T3 + T1/T2/T4) between Mandarin and Cantonese speakers. The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

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 4 plots the mean fitted values of the half-third sandhi tone on wug words produced by the two groups of speakers. The F0 contours showed significant differences in two parts (observed t-statistic exceeding maximum 0.05 critical value in about 18% ~ 42% and 61%~100% of the vowel), plotted in Figure 6. Compared to Mandarin speakers, Cantonese speakers produced lower F0 values for the first half, and then they raised the pitch contour earlier than Mandarin speakers.

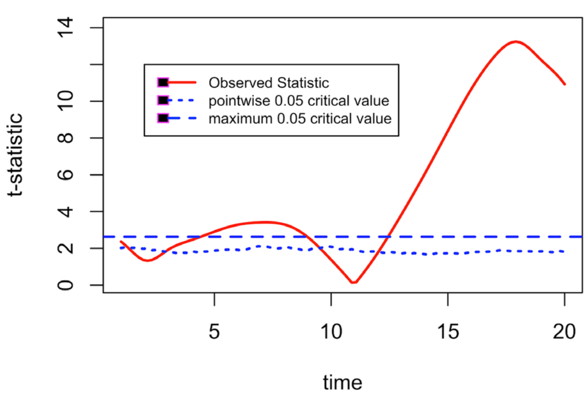


Figure 6. A functional t-test of wug syllables (T3 + T1/T2/T4) between Mandarin and Cantonese speakers. The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

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 28% ~ 36% of the vowel). The mean fitted F0 curves have slightly lower values for wug syllables than for real syllables. Similarly, we compared tone production on wug syllables and real syllables by Cantonese speakers, which showed no statistical significance (observed t-statistic did not exceed maximum 0.05 critical value). For wug syllables, Cantonese speakers tended to raise F0 contours more toward the end.

3.2 Performance of American speakers

3.2.1 Third tone sandhi rule

Significant differences were found on F0 contours of the first sandhi T3 in T3 + T3 on real syllables and wug syllables between Beijing Mandarin speakers and American speakers. Figure 7 plots the mean fitted curves on real and wug words produced by American and Beijing speakers.

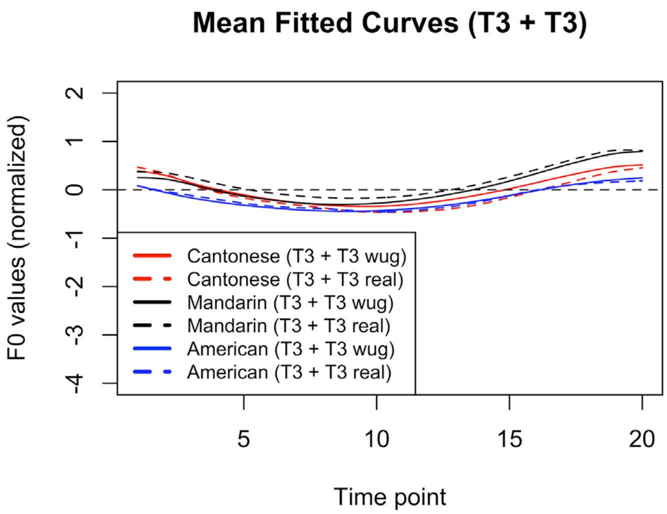


Figure 7. Mean fitted curves of the first sandhi T3 in T3 + T3 produced by Mandarin, Cantonese and American speakers. The dotted lines stand for real words, and solid lines for wug words. The red lines represent Cantonese speakers, the black lines Mandarin speakers and the blue lines American speakers.

Table 3 reports the mean fitted values on several time points throughout vowels produced by the two groups of speakers for a comparison. It can be seen that for the sandhi T3 on real syllables, American speakers tended to produce much lower F0 values. There were significant differences based on a functional t-test (observed t-statistic exceeding maximum 0.05 critical value in about 2%~26%, 39% ~ 100% of the vowel), as shown in Figure 8.

Table 3 ***The mean fitted values based on functional data analysis of real and wug syllables T3 + T3 within their significant regions***

Percentage (real)	Mandarin speakers	American speakers	Percentage (wug)	Mandarin speakers	American speakers
20%	0.12	-0.2	20%	0.11	-0.17
40%	-0.16	-0.39	40%	-0.30	-0.44
60%	-0.07	-0.40	60%	-0.17	-0.36
80%	0.57	0.08	80%	0.49	0.08
100%	0.81	0.18	100%	0.79	0.25

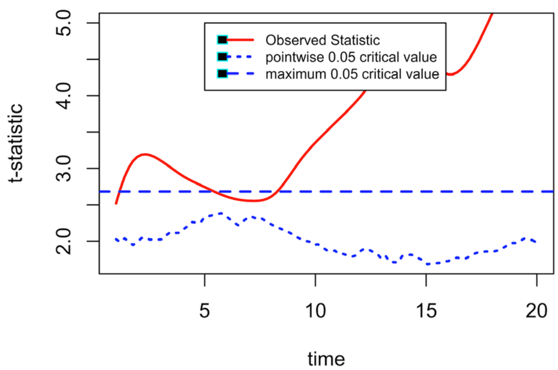


Figure 8. A functional t-test of real syllables (T3 + T3) between Mandarin and American speakers. The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

In addition, American speakers were significantly different from Mandarin speakers in producing the sandhi T3 in the context T3 + T3 on wug syllables (*AO+AO, AO+AG, AG+AO, AG+AG and Pseudo words), as shown in Figure 7. Significant differences were found over the entire F0 contours (observed t-statistic exceeding maximum 0.05 critical value in 100% of the vowel), as shown in Figure 9. Similar to real syllables, American speakers showed lower F0 values than Beijing Mandarin speakers. However, their production of the sandhi T3 in T3 + T3 on real vs. wug syllables did not show any statistically significant differences.

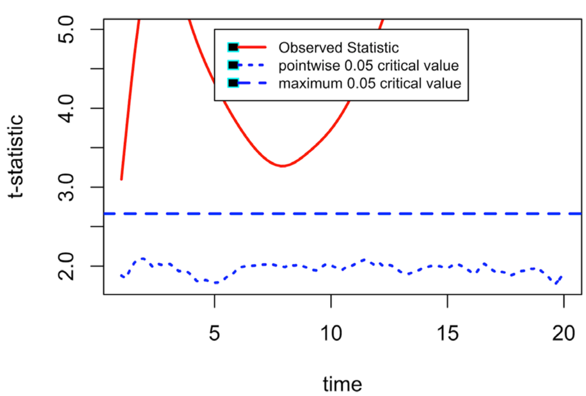


Figure 9. A functional t-test of wug syllables (T3 in T3 + T3) between Mandarin and American speakers; The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

3.2.2 Half-third tone sandhi rule

Figure 10 plots the mean fitted curves in producing the first T3 in T3 + T1, T3 + T2 and T3 + T4 from the Mandarin and American groups. Mandarin speakers produced lower F0 values through most of the vowels. A large portion of significant differences was found between Beijing and American speakers (observed t-statistic exceeding maximum 0.05 critical value in about 31% ~ 100% of the vowel) as shown in Figure 11.

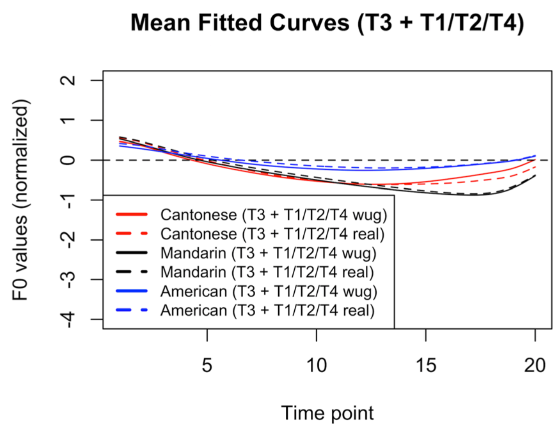


Figure 10. Mean fitted curves of the first sandhi T3 in T3 + T1/T2/T4 between Mandarin and American speakers. The dotted lines stand for real words and solid lines for wug words. The red lines represent American speakers and the black lines Mandarin speakers.

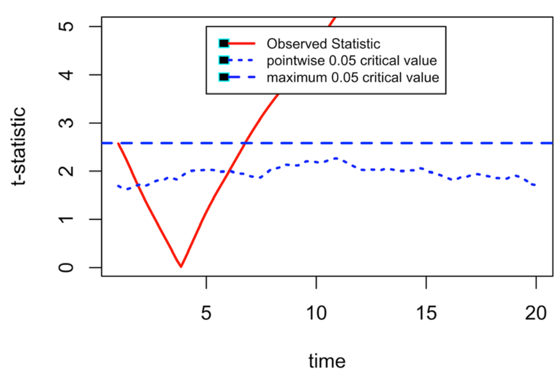


Figure 11. A functional t-test of real syllables (T3 in T3 + T1/T2/T4) between Mandarin and American speakers; The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

In producing the half-third sandhi tone on wug syllables (*AO+AO, AO+AG, AG+AO, AG+AG and Pseudo words), American speakers also significantly differed from Mandarin speakers, as shown in Figure 10. The F0 contours showed significant differences in two parts (observed t-statistic exceeding maximum 0.05 critical value in about 1% ~ 8% and 21%~100% of the vowel), plotted in Figure 12. American speakers produced significantly higher F0 values compared to Beijing Mandarin speakers. In addition, our results showed no statistical differences between wug and real syllables of the first T3 in T3 + T1/T2/T4.

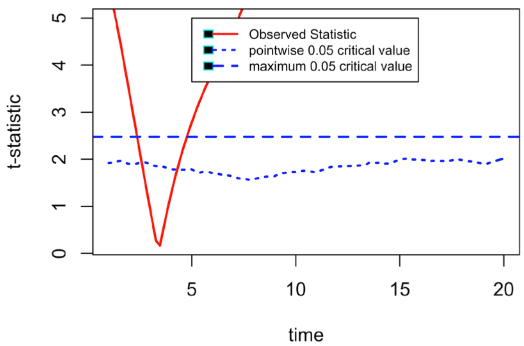


Figure 12. A functional t-test of wug syllables (T3 + T1/T2/T4) between Mandarin and American speakers. The observed statistics are represented by a solid line, point-wise 0.05 critical values are indicated by a dotted line, and the maximum 0.05 critical values are represented by dashed lines.

3.3 Acquisition of allophonic variants by Cantonese and American learners

Based on the modelling from functional data analysis, we presented the plots of two allophonic variants by Beijing, Cantonese and American speakers in Figure 13-15. Figure 13 suggests that the tonal contours of the two allophonic variants (half third sandhi tone and third sandhi tone) produced by native Mandarin speakers can be clearly differentiated, where the half third sandhi tone exhibited a falling slope for the most part of the contour and the third sandhi tone was raised to a great extent. However, compared to the native speakers, the tonal contours of the two variants produced by Cantonese speakers did not differ to the same extent, where the half sandhi tone still exhibited a rising portion and the third sandhi tone did not rise high enough. The tonal contours of the two variants produced by American speakers were even less distinctive, where the two contours remain relatively flat. A comparison among the three figures suggests a ranking of distinction among the three groups: American speakers < Cantonese speakers < Beijing speakers.

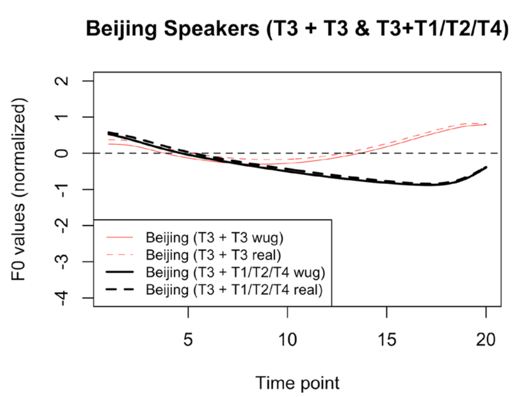


Figure 13. A comparison of two allophonic variants produced by Beijing speakers

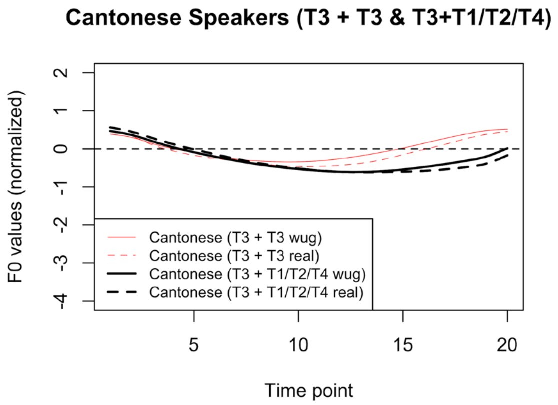


Figure 14. A comparison of two allophonic variants produced by Cantonese speakers

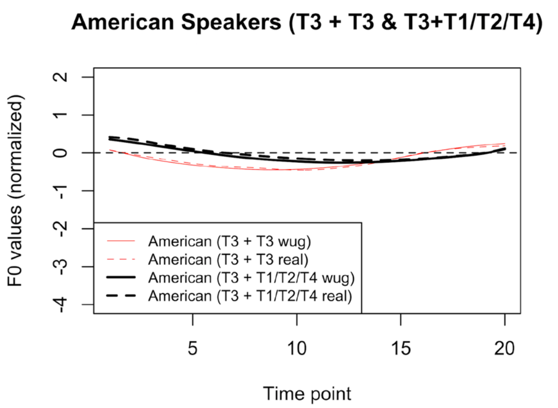


Figure 15. A comparison of two allophonic variants produced by American speakers

4.0 Discussion

4.1 Tone sandhi rule applications by Cantonese and American learners

The current study aims to answer three research questions, and our results made three empirical contributions to the field. The first question is whether pitch contours realized on the Mandarin third- and half-third sandhi tones are similar in real and wug words among native Cantonese and American English speakers. To answer this question, we compared the third- and the half-third sandhi tonal contours produced by both groups of speakers on real and different types of wug words. Overall, the analyses yielded a non-significant difference in pitch contour realizations on real and wug words for both types of sandhi tones and for both groups of non-native speakers, suggesting that the same underlying mechanism is likely involved in their tone sandhi production. According to Nixon et al. (2015), equal effects should be found for both real and Wug morphemes if sandhi and non-sandhi forms are parts of the abstract T3 category. Therefore, similarities in pitch contours realized on real and wug morphemes by L2 learners suggest that surface sandhi contours are encoded as part of the abstract representation of Tone 3 category rather than lexically represented as part of a morpheme.

The second question we asked was whether Cantonese and American English speakers' production of sandhi tones are equally as accurate as that of native speakers on real words and wug words. In regards to this question, despite similarities in their sandhi tone production in real and wug words, both Cantonese and American English production of the sandhi tones remain non-native like.

Specifically, Cantonese speakers' production of the third sandhi tones significantly differed from those of Mandarin speakers. In general, Cantonese speakers tended to produce lower F0 values than Mandarin speakers did. Significantly different regions of F0 contours were found in the middle part to the end of vowels for both real syllables and wug syllables. The produced tonal contours of the third sandhi tone variant failed to rise as high as those produced by native speakers. To explain the lower F0 values produced by the Cantonese speakers, we compared Cantonese T2 and T5 pitch contours and Mandarin T2 and T3 pitch contours. As shown in Figure S1 and Table S7, Cantonese T2 and T5 showed higher F0 values over the whole contour than Mandarin T2 and T3, thus a direct transfer from the Cantonese T2 and T5 as the source of the exhibited lower F0 values is not readily evident. However, there might be some influence from the rising contours of existing tonal categories. Moreover, according to Hao (2012), Mandarin T3 is mapped mostly to Cantonese T4 (21), suggesting that when comparing Mandarin tones to Cantonese tones, Cantonese speakers pay more attention to the low-falling portion of the Mandarin T3. This is likely the reason why their third tone sandhi contour exhibited lower F0 values.

Cantonese speakers' production of the half-third sandhi tone on wug syllables and real syllables were also compared to those of Mandarin speakers. On real and wug syllables, F0 contours produced by Cantonese speakers had an earlier rising contour toward the end with higher F0 values than Mandarin speakers. Two possible explanations can be offered to account for the F0 rising pattern. First, it is possible that the first Mandarin T3 in the half-third sandhi context is mapped to either Cantonese T2 or T5, the two Cantonese rising tonal categories. Second, it is also possible the first Mandarin T3 remains to be perceptually linked to Cantonese T4, and the observed raising F0 at the end of the contour is due to coarticulatory effects exerted by the following tones. Further experimentation is needed to choose between these two explanations.

Similarly, significant differences were found on F0 contours of the third sandhi tone on real and wug syllables produced by Beijing Mandarin speakers and American English speakers. The portion of significant differences on both real and wug syllables was much bigger than for Cantonese speakers, where almost the entire F0 contour differed from Mandarin speakers. American English speakers produced lower F0 values than both Mandarin and Cantonese speakers. Although they were instructed to raise F0 contours to produce the third sandhi tone, there are no existing tonal categories in their L1—especially rising tones—to be mapped to, which may lead to their retention of low F0 values in the underlying T3 form and thus more incomplete computation compared to Cantonese speakers.

More differences were also found between Mandarin and American English speakers on both real and wug syllables for the half-third sandhi. American English speakers produced higher and flatter F0 values than both Mandarin and Cantonese speakers. Unlike Cantonese, American English does not have any existing tonal categories that may affect American speakers' speech production of half-third sandhi tones. Therefore, they may have paid more attention to the truncation of the underlying T3 (from 213 to 21), but failed to articulate an allophonic tone with a falling slope as native speakers did, which may be due to their poor perception of tonal contours (e.g. Gandour & Harshman, 1978), leading to less accurate allophonic tonal representations.

In general, the distinction of tonal contours in the third sandhi and half third sandhi tone produced by L2 learners is less discernible than native speakers. Our results thus support the "shallow" hypothesis, initially proposed for less detailed syntactic representations in L2 learners (Clahsen & Felser, 2006). The current study found similar results in the application of tone sandhi rules, where L2 learners tended to show less detailed and less accurate production of tonal contours, due perhaps to less detailed phonological representations of allophonic variants.

In addition, the acquisition of allophonic variants can be viewed as acquisition of phonetic systems based on statistical learning. In first language acquisition, Maye, Gerken & Werker (2003) found that infants are sensitive to statistical distributions of phonetic variation, which can affect whether they can discriminate speech sounds. Werker and Curtin (2005) proposes a developmental framework for Processing Rich Information from Multidimensional Interactive Representations

(PRIMIR), claiming that phonological acquisition is based on organization of rich information in speech along interactive planes, including initial biases, regularities of acoustic input and statistical learning. Peperkamp, Le Calvez, Nadal, and Dupoux (2006) further proposed that statistical analysis of the distribution of segments can be a good indicator of complementary allophonic distribution in artificial copra. However, distribution plus linguistic constraints (i.e. constraint1: allophones and the default segment are similar phonetically; constraint 2: allophones should be more similar to its context than the default segment) can lead to more successful learning of allophonic rules than distribution alone. In second language acquisition, learners also need to learn the similarities and dissimilarities of allophonic variants based on statistical distributions (Shea & Curtin, 2014), which can account for acquisition of allophonic tones. In the similarity stage, learners need to learn similarities between allophonic variants, namely third sandhi tone, half third sandhi tone and underlying T3. In the dissimilarity stage, since there is no co-occurrence of these allophonic variants in the same contexts, the learners postulate phonological alternation, and they learn to compute the differences among the variants.

However, American and Cantonese learners did not learn the dissimilarities of allophonic variants well enough to make a native-like distinction in the two allophonic sandhi tones. Nonetheless, Cantonese speakers differentiated the two tonal contours better than American speakers. Better performance among Cantonese speakers may be attributed to L1 transfer since Cantonese speakers do have existing L1 tonal categories. In producing the third sandhi tone, Cantonese speakers might have shown positive transfer from both existing Cantonese rising tonal contours in their native language. American speakers, however, did not have existing rising tonal categories in their L1 to be mapped to and were less sensitive to change in pitch contours, leading to retained low F0 values in the underlying T3 form and less accurately computed sandhi forms. Similarly, in producing the half-third sandhi tone, Cantonese speakers may have mapped the Mandarin T3 to the rising tonal categories or their T4 (21) whose contour is more closely matched to the low-falling portion of the Mandarin T3. In contrast, American English speakers did not map it to existing tonal categories and failed to perceive the falling tonal contours, so they produced the tonal variant with a flatter contour. Our results are consistent with the study supporting L1 transfer in acquiring allophones at the segmental level (Shoemaker, 2014; MacLeod & Fabiano-Smith, 2015), and we have further showed evidence supporting L1 transfer in learning allophones at the suprasegmental level.

4.2 Phonetic bias in rule applications

The third question we asked was whether the production of the third-tone and the half-third tone sandhi is equally accurate in real words and all types of wug words among Cantonese and American English speakers. As previously discussed, the half-third sandhi rule is considered to be phonetically more natural than the third-tone sandhi rule. According to Zhang and Lai (2010), a more phonetically motivated sandhi rule may be realized more accurately on wug words than on real words. Their own findings and those of Chen & Li (2017) obtained from native Mandarin speakers in their application of the two sandhi rules are consistent with this suggestion. However, we find little evidence to suggest the presence of the phonetic bias in sandhi rule application among non-native speakers.

Specifically, we found that Cantonese speakers produced similar tonal contours between real vs. wug words for both rules. Compared to native Mandarin speakers, they produced significantly different tonal contours of the third sandhi tone on all types of wug words except for AGAO (percentage of significant differences: mean = 0.14; sd = 0.11). In applying the half-third sandhi tone rule, only T31a on *AOAO and AOAG was not significantly different from native Mandarin speakers (percentage of significant differences: mean = 0.29; sd = 0.11). No significant differences were found in the proportions as indicated by a Wilcoxon signed-rank test ($V = 3$, $p = 0.31$). For American English speakers, their production also did not differ on either real or wug words for the two rules. They produced different F0 contours from native speakers on all types of wug words for the third T3 (percentage of significant differences: mean = 0.5; sd = 0.25) and the half third sandhi tone (percentage of significant differences: mean = 0.44; sd = 0.09). There was no significant difference in the proportion according to a Wilcoxon signed-rank test ($V = 9$, $p = 0.81$). These results suggest a lack of bias in the application of a phonetically more motivated tone sandhi rule by tonal and non-tonal non-native speakers. However, variations in L2 productions of sandhi tones may have also obscured any differences in the applications of these two types of sandhi rules. In addition, the results regarding ease of learning between tone sandhi rules by L2 learners are not consistent. Zhang (2013) reported that learners tend to perform better in the third tone sandhi rule, whereas Yang (2016) found the half sandhi rule was easier for learners. Similar to studies on Mandarin phonemic tone production, many factors (e.g., differences in proficiency, the mode of instruction, and the nature and the amount of input) may affect the results in the studies of allophonic tones. Future studies are called for to examine how phonetic bias interacts with other factors in affecting application of tone sandhi rules.

5.0 Conclusions

The current study examined applications of two Mandarin tone sandhi rules by tonal and non-tonal speakers. The results showed that the computation mechanism was likely to be involved in tone sandhi rule applications by both the Cantonese and the American English speakers. However, their surface F0 contours were still significantly different from those of native speakers. The sandhi tone contours produced by Cantonese speakers likely resulted from the perceptual mapping of the Mandarin T3 in both sandhi contexts to existing Cantonese tones whereas American speakers' production seems to be more psycho-acoustically based with greater attention paid to low F0 values in the tonal contour of citation T3 in computing its sandhi forms. Finally, no phonetic bias in sandhi rule application was observed for either group due perhaps to a large individual variation.

6.0 Acknowledgement

This study is supported by grants from Faculty of Humanities [1-ZVHH], [1-ZVHJ] and Department of Chinese and Bilingual Studies [88C3], the Hong Kong Polytechnic University.

Appendix

S1. AO+AO

Tones	Chinese diagram	Transcription		Tones	Chinese diagram	Transcription	
3+1	每天	məi	t^{hj}æn	3+3	美好	məi	xɑu
	转身	ʈʂ^{w}æn	ʂən		转脸	ʈʂ^{w}æn	l^{j}æn
	产生	ʈʂ^{h}æn	ʂəŋ		反感	fæn	kæn
	首先	ʂəu	ɕæn		手指	ʂəu	ʈʂʐʐ
	整天	ʈʂəŋ	t^{hj}æn		整理	ʈʂəŋ	l^{j}ii
	可惜	k^{h}ɤɤ	ɕii		可以	k^{h}ɤɤ	ii
	假装	tɕaa	ʈʂ^{w}aŋ		假使	tɕaa	ʂʐʐ
	水晶	ʂ^{w}əi	tɕəŋ		水果	ʂ^{w}əi	k^{w}ɔɔ
3+2	美国	məi	k^{w}ɔɔ	3+4	美丽	məi	l^{j}ii
	转移	ʈʂ^{w}æn	ii		转动	ʈʂ^{w}æn	t^{w}uŋ
	反而	fæn	ər		反正	fæn	ʈʂəŋ
	手足	ʂəu	ts^{w}uu		手术	ʂəu	ʂ^{w}uu
	整洁	ʈʂəŋ	tɕee		整个	ʈʂəŋ	kɤɤ
	可能	k^{h}ɤɤ	nəŋ		可怕	k^{h}ɤɤ	p^{h}aa
	假如	tɕaa	ʐ^{w}uu		假币	tɕaa	p^{j}ii
	水平	ʂ^{w}əi	p^{hj}əŋ		水面	ʂ^{w}əi	m^{j}æn

S2. *AO+AO

Tones	Chinese diagram	Transcription		Tones	Chinese diagram	Transcription	
3+1	每聪	məi	ts^{hw}uŋ	3+3	美朵	məi	t^{w}oo
	转跟	ʈʂ^{w}æn	kən		转很	ʈʂ^{w}æn	xən
	产参	ʈʂ^{h}æn	ts^{h}æn		反撒	fæn	saa
	首插	ʂəu	ʈʂ^{h}aa		手怎	ʂəu	tsən
	整催	ʈʂəŋ	ts^{hw}əi		整早	ʈʂəŋ	tsau
	可慌	k^{h}ɤɤ	x^{w}aŋ		可散	k^{h}ɤɤ	sæn
	假森	tɕaa	sən		假朵	tɕaa	t^{w}oo
	水黑	ʂ^{w}əi	xəi		水滚	ʂ^{w}əi	k^{w}ən

3+2	美别	məi	p^{j}ee	3+4	美特	məi	t^{h}ɤɤ
	转仇	ʈʂ^{w}æn	ʈʂ^{h}əu		转不	ʈʂ^{w}æn	p^{w}uu
	反存	fæn	ts^{hw}ən		反脆	fæn	ts^{hw}əi
	手能	ʂəu	nəŋ		手否	ʂəu	fəu
	整狼	ʈʂəŋ	laŋ		整哈	ʈʂəŋ	xaa
	可勾	k^{h}ɤɤ	kəu		可害	k^{h}ɤɤ	xai
	假林	tɕaa	l^{j}əŋ		假略	tɕaa	l^{ɥ}ee
	水朝	ʂ^{w}əi	ʈʂ^{h}au		水奏	ʂ^{w}əi	tsəu

S3. AO+AG

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	每	den	məi	tən	3+3	美	diu	məi	t^{j}əu
	转	liang	ʈʂ^{w}æn	l^{j}aŋ		转	mie	ʈʂ^{w}æn	m^{j}ee
	产	mei	ʈʂ^{h}æn	məi		反	shuan	fæn	ʂ^{w}æn
	首	mu	ʂəu	m^{w}uu		手	sen	ʂəu	sən
	整	mai	ʈʂəŋ	mai		整	dui	ʈʂəŋ	t^{w}əi
	可	kuo	k^{h}ɤɤ	k^{hw}oo		可	cou	k^{h}ɤɤ	ts^{h}əu
	假	re	tɕaa	ʐɤɤ		假	te	tɕaa	t^{h}ɤɤ
	水	shun	ʂ^{w}əi	ʂ^{w}ən		水	run	ʂ^{w}əi	ʐ^{w}ən
3+2	美	mie	məi	m^{j}ee	3+4	美	dei	məi	təi
	转	ta	ʈʂ^{w}æn	t^{h}aa		转	zhua	ʈʂ^{w}æn	ʈʂ^{w}aa
	反	ka	fæn	k^{h}aa		反	keng	fæn	k^{h}əŋ
	手	suan	ʂəu	s^{w}æn		手	nin	ʂəu	n^{j}in
	整	dui	ʈʂəŋ	t^{w}əi		整	qiu	ʈʂəŋ	tɕ^{h}əu
	可	diu	k^{h}ɤɤ	t^{j}əu		可	ca	k^{h}ɤɤ	ts^{h}aa
	假	te	tɕaa	t^{h}ɤɤ		假	sen	tɕaa	sən
	水	cui	ʂ^{w}əi	ts^{hw}əi		水	mang	ʂ^{w}əi	maŋ

S4. AG+AO

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	hei	天	xəi	t^{hj}æn	3+3	hei	好	xəi	xau
	shuan	身	ʂ^{w}æn	ʂən		shuan	脸	ʂ^{w}æn	l^{j}æn
	pan	生	p^{h}æn	ʂəŋ		pan	感	p^{h}æn	kæn
	cou	先	ts^{h}əu	ɕæn		cou	指	ts^{h}əu	ʈʂʐ̩
	sen	天	sən	t^{hj}æn		sen	理	sən	l^{j}ii
	te	惜	t^{h}ɤɤ	ɕii		te	以	t^{h}ɤɤ	ii
	xia	装	ɕaa	ʈʂ^{w}aŋ		xia	使	ɕaa	ʂʐ̩
			t^{w}əi	tɕəŋ				t^{w}əi	k^{w}oo

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+2	dui	晶			3+4	dui	果		
	hei	国	xəi	kʰʷoo		hei	丽	xəi	lʲii
	shuan	移	ʂʷæn	ii		shuan	动	ʂʷæn	tʷuŋ
	pan	而	pʰæn	ər		pan	正	pʰæn	ʈʂəŋ
	cou	足	tsʰəu	tsʷuu		cou	术	tsʰəu	ʂʷuu
	sen	洁	sən	tɕee		sen	个	sən	kɤɤ
	te	能	tʰɤɤ	nəŋ		te	怕	tʰɤɤ	pʰaa
	xia	如	ɕaa	ʐʷuu		xia	币	ɕaa	pʲii
	dui	平	tʷəi	pʰʲəŋ		dui	面	tʷəi	mʲæn

S5. AG+AG

Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	hei	den	xəi	tən	3+3	hei	diu	xəi	tʲəu
	shuan	liang	ʂʷæn	lʲaŋ		shuan	mie	ʂʷæn	mʲee
	pan	mei	pʰæn	məi		pan	shuan	pʰæn	ʂʷæn
	cou	mu	tsʰəu	mʷuu		cou	sen	tsʰəu	sən
	sen	mai	sən	mai		sen	dui	sən	tʷəi
	te	kuo	tʰɤɤ	kʰʷoo		te	cou	tʰɤɤ	tsʰəu
	xia	re	ɕaa	ʐɤɤ		xia	te	ɕaa	tʰɤɤ
	dui	shun	tʷəi	ʂʷən		dui	run	tʷəi	ʐʷən
3+2	hei	mie	xəi	mʲee	3+4	hei	dei	xəi	təi
	shuan	ta	ʂʷæn	tʰaa		shuan	zhua	ʂʷæn	ʈʂʷaa
	pan	ka	pʰæn	kʰaa		pan	keng	pʰæn	kʰəŋ
	cou	suan	tsʰəu	sʷæn		cou	nin	tsʰəu	nʲin
	sen	dui	sən	tʷəi		sen	qiu	sən	tɕʰəu
	te	diu	tʰɤɤ	tʲəu		te	ca	tʰɤɤ	tsʰaa
	xia	te	ɕaa	tʰɤɤ		xia	sen	ɕaa	sən
	dui	cui	tʷəi	tsʰʷəi		dui	mang	tʷəi	maŋ

S6. Pseudo words

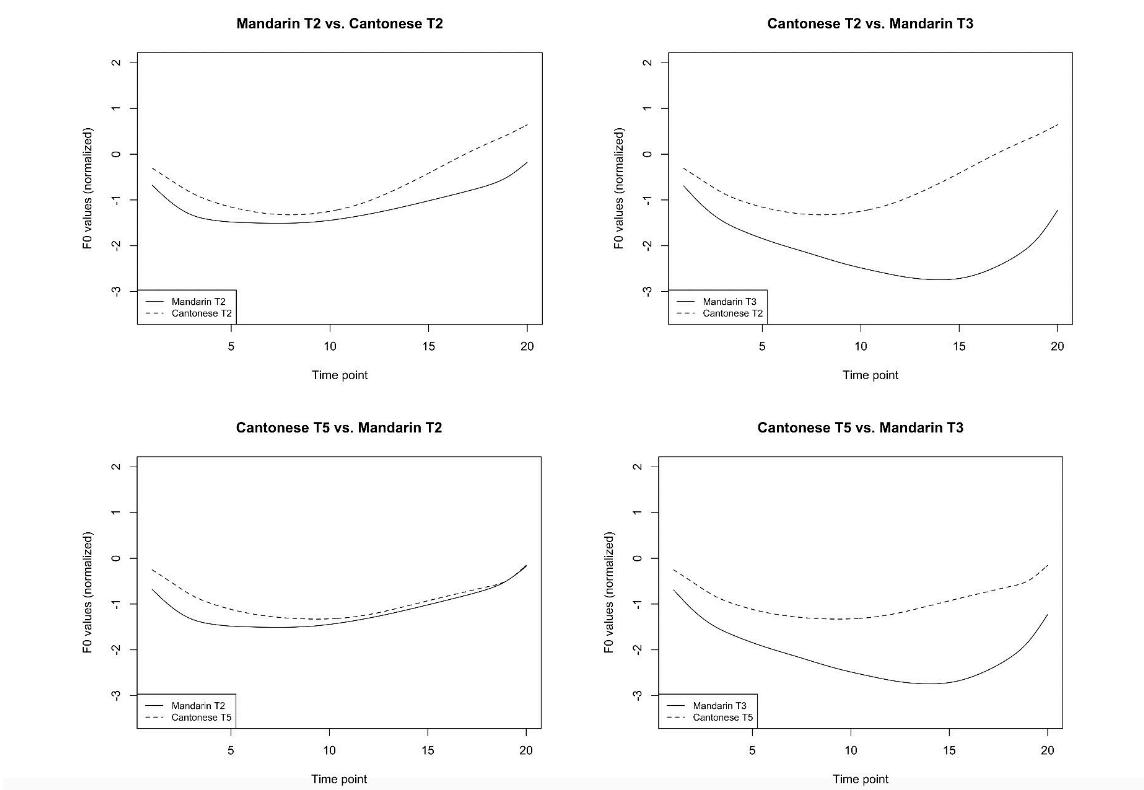
Tones	Chinese diagram		Transcription		Tones	Chinese diagram		Transcription	
3+1	lui	fai	lʷəi	fai	3+3	lui	bou	lʷəi	pəu
	fie	diang	ɦee	tʲaŋ		fie	ten	ɦee	tʰən
	ten	fie	tʰən	ɦee		ten	fai	tʰən	fai
	bou	fe	pəu	ɦɤɤ		bou	diang	pəu	tʲaŋ
	len	mui	lən	mʷəi		len	fe	lən	ɦɤɤ
	be	fiu	pə	ɦəu		be	fie	pə	ɦee
	fia	ten	ɦʲaa	tʰən		fia	mui	ɦʲaa	mʷəi
	nui	len	nʷəi	lən		nui	fong	nʷəi	ɦʷuŋ

3+2	lui	fong	lʷəi	fʷuŋ	3+4	lui	fai	lʷəi	fai
	fie	ten	ɦee	tʰən		fie	mui	ɦee	mʷəi
	ten	fe	tʰən	fɤɤ		ten	fong	tʰən	fʷuŋ
	bou	fai	pəu	fai		bou	ten	pəu	tʰən
	len	mui	lən	mʷəi		len	diang	lən	tʲaŋ
	be	fie	pə	ɦee		be	fie	pə	ɦee
	fia	bou	fʲaa	pəu		fia	bou	fʲaa	pəu
	nui	diang	nʷəi	tʲaŋ		nui	fe	nʷə i	fɤɤ

S7 A comparison of four Cantonese and Mandarin tones using functional t-tests

Tonal pairs	Significantly different regions
Cantonese T2 vs. Mandarin T2	0% to 25%, 58% to 100%.
Cantonese T5 vs. Mandarin T2	30%~100%
Cantonese T2 vs. Mandarin T3	0%~100%
Cantonese T5 vs. Mandarin T3	0%~100%

Figure S1



Reference List

Albright, A., Andrade, A., & Hayes, B. (2001). Segmental Environments of Spanish Diphthongization. In A. Albright, & T. Cho (Eds.), *UCLA Working Papers in Linguistics 7 (Papers in Phonology 5)* (pp. 117-151).

Albright, A. (2002). The Identification of Bases in Morphological Paradigms (Ph.D. dissertation). UCLA.

Albright, A., & Hayes, B. (2003). Rules vs. analogy in English past tenses: a computational/experimental study. *Cognition, 90,* 119-161.

Bauer, R. S., & Benedict, P. K. (1997). Modern Cantonese phonology. Berlin: Mouton de Gruyter.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171– 204). Timonium, MD: York Press.

Berko, J. (1958). The child's learning of English morphology. *Word, 14,* 150-177.

Broselow, E., Hurtig, R. R., & Ringen, C. (1987). The perception of second language prosody. In G. Ioup, & S. H. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 350-361). Cambridge: Newbury House Publishers.

Bybee, J. L., & Pardo, E. (1981). On lexical and morphological conditioning of alternations: a nonce-probe experiment with Spanish verbs. *Linguistics, 9,* 937-968.

Chao, Y. R. (1948). *Mandarin Primer*. Cambridge: Harvard University Press.

Chao, Y. R. (1968). *A grammar of spoken Chinese.* Berkeley: University of California Press.

Chen, S., & Li, B. (2017). Degree of incomplete computation in application of two Mandarin tone sandhi rules. *Proceedings of the International Conference on Speech Sciences* (pp. 83-84). The Korean Society of Speech Sciences, Seoul.

Chen, S., & Li, B. (submitted) Statistical modeling of application completeness of two Mandarin tone sandhi rules

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

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

Chen, S., He Y. J., Yuen, C. W., Li, B. & Yang, Y. K. (2017). Mechanisms of Tone Sandhi Rule Application by Non-native Speakers. *Proceedings of Interspeech 2017*, Stockhom, Sweden.

Chen, S., Zhang, C., McCollum, A. G., & Wayland, R. (2017). Statistical modelling of phonetic and phonologised perturbation effects in tonal and non-tonal languages. *Speech Communication*, *88*, 17-38.

Chen, M. (2000). Tone Sandhi: Patterns across Chinese Dialects. vol. 92. Cambridge University Press.

Chen, Y., Shen, R., & Schiller, N. O. (2011). Representation of allophonic tone sandhi variants. In *Proceedings of Psycholinguistics Representation of Tone*. Satellite Workshop to ICPhS, Hongkong (pp. 38-41).

Chen, Y. Y. (2012). Message-related variation: tonal variation. In A. Cohn, C. Fougeron & M. Huffman (Eds.), *Oxford Handbook of Laboratory Phonology*. pp. 103-114. Oxford: Oxford University Press.

Cheng, C. C. (1968). English stresses and Chinese tones in Chinese sentences. *Phonetica*, *18*(2), 77-88.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied psycholinguistics,* 27(1), 3-42.

Curtin, S., Byers-Heinlein, K., & Werker, J. (2011). Bilingual beginnings as a lens for theory development: PRIMIR in focus. *Journal of Phonetics*, 39, 492–504.

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

Davidson, L. (2006). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. Journal of Phonetics, 34(1), 104-137.

Dong, Y., Tsubota, Y., & Dantsuji, M. (2013). Difficulties in Perception and Pronunciation of Mandarin Chinese Disyllabic Word Tone Acquisition: A Study of Some Japanese University Students. In *Proceedings of the 27th Pacific Asia Conference on Language, Information and Computation* (pp. 143-152). Taipei: National Chengchi University.

Elliott, C. E. (1991). The relationship between the perception and production of Mandarin tones: An exploratory study. *University of Hawai'i Working Papers in English as a Second Language, 10*(2), 177-204.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Gandour & Harshman (1978) Cross-language difference in tone perception: A multidimensional scaling investigation. *Language and Speech*, 21 (1978), pp. 1-33

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, *32*(3), 395-421.

Hao, Y. C. (2018). Contextual effect in second language perception and production of Mandarin tones. *Speech Communication*, 97, 32-42.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics, 40,* 269-279.

Hayes, B., & Londe, Z. (2006). Stochastic phonological knowledge: The case of Hungarian vowel harmony. *Phonology, 23*(1), 59-104. doi:10.1017/S0952675706000765

He, Y. J. (2013). Identification of Mandarin coarticulated tones by inexperienced and experienced English learners of Mandarin. *Chinese as a Second Language Research, 2* (1), 1–21.

He, Y. J. (2014). A phonetic investigation into factors influencing Mandarin disyllabic tone production by American speakers (yingxiang meiguo xuesheng shuangyinjie hanyu shengdiao fa yin yinsude yuyin diaocha). *International Chinese Language Education (guoji hanyu jiao yu), 1,* 136-145, & 204-205.

Hombert, J. M. (1977). Development of tones from vowel height? *Journal of Phonetics, 5,* 9-16.

House, A. S., & Fairbanks, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *The Journal of the Acoustical Society of America, 25*(1), 105-113.

Hsieh, H. (1970). The psychological reality of tone sandhi rules in Taiwanese. *Chicago Linguistic Society, 6*, 489-503.

Hsieh, H. (1975). How generative is phonology? In E.F.K. Koerner (Ed.), *The transformational-generative paradigm and modern linguistic theory* (pp. 109-144). Amsterdam: Benjamins.

Hsieh, H. (1976). On the unreality of some phonological rules. *Lingua, 38,* 1-19.

Hume, E., & Johnson, K. (2003). The impact of partial phonological contrast on speech perception. In *Proceedings of the 15th international congress of phonetic sciences* (pp. 2385–2388). Barcelona, Spain.

Lee, C. Y., Tao, L., & Bond, Z. S. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*(4), 537-563.

Leather, J. (1990). Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. In J. Leather, & A. James (Eds.), *New sounds 90: Proceedings of the 1990 Amsterdam symposium on the acquisition of second language speech* (pp. 72-97). Amsterdam: University of Amsterdam.

Levitin D. J., Nuzzo R. L., Vines B. W., Ramsay J. O. (2007). Introduction to functional data analysis. Canadian Psychology, 48, 135–155.

Li, X., & Chen, Y. (2015). Representation and processing of lexical tone and tonal variants: Evidence from the mismatch negativity. *PloS one*, *10*(12), e0143097.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. Journal of Phonetics, 54, 123 – 150. doi: 10.1016/j.wocn.2015.10.002

MacLeod, A. A., & Fabiano-Smith, L. (2015). The acquisition of allophones among bilingual

Spanish–English and French–English 3-year-old children. *Clinical linguistics & phonetics*, 29(3), 167-184.
Matthews, S., & Yip, V. (1994). *Cantonese: A comprehensive grammar.* Routledge.
Miracle, W. C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of Chinese Language Teachers Association, 24*, 49-65.
Ning, L. H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication*, *63-64*, 55-69.
Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: evidence from Mandarin lexical tones. Language, *Cognition and Neuroscience*, 30(5), 491-505.
Peng, S. H. (2000). Lexical versus 'phonological' representations of Mandarin Sandhi tones. In Broe M. & J. Pierrehumbert (eds.), *Papers in laboratory phonology 5: acquisition and the lexicon* (pp. 152-167). Cambridge: Cambridge University Press.
Pierrehumbert, J. (2006). The Statistical Basis of an Unnatural Alternation. In L. Goldstein, D. H. Whalen, & C. Best (Eds.), *Laboratory Phonology VIII, Varieties of Phonological Competence* (pp. 81-107). Berlin: Mouton de Gruyter.
Ramsay, J. O, & Silverman, B. W. (2005). *Functional Data Analysis.* (2nd ed.). New York: Springer-Verlag.
Ramsay, J. O., & Silverman, B. W. (2009). Functional data analysis with R and MATLAB. Springer-Verlag, New York.
Shea, C., & Curtin, S. (2006). Learning allophonic alternations in a second language: Phonetics, phonology and grammatical change. In *Proceedings of the 8th Conference on Generative Approaches to Second Language Acquisition*, Banff, Alberta.
Shen, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association, 24*, 27-47.
Shoemaker, E. (2014). The exploitation of subphonemic acoustic detail in L2 speech segmentation. *Studies in Second Language Acquisition*, 36(4), 709-731.
So, C. K., & Best, C. T. (2010). Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. *Language and Speech*, *53*(2), 273-293.
Sun, S. H. (1997). The development of a lexical tone phonology in American adult learners of standard Mandarin Chinese. Honolulu: University of Hawai'i Press.
Tu, J. Y., Hsiung, Y., Wu, M. D., & Sung, Y. T. (2014). Error patterns of Mandarin disyllabic tones by Japanese learners. In *Fifteenth Annual Conference of the International Speech Communication Association*.
Wang, H. S. (1993). Taiyu biandiao de xinli texing (On the psychological status of Taiwanese tone sandhi). *Tsinhua Xuebao (Tsinghua Journal of Chinese Studies), 23,* 175-192.
Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal*, *97*(1), 144-160.
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,* 1033-1044.
Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones. *Language Learning: A Preliminary Report*, *54*(4), 681-712.
Werker, J. F., & Curtin, S. (2005). PRIMIR: A developmental framework of infant speech processing. *Language learning and development*, 1(2), 197-234.
White, C. M. (1981). Tonal pronunciation errors and interference from English intonation. *Journal of the Chinese Language Teachers Association, 16*(2), 27-56.
Wu, X., Munro, M. J., & Wang, Y. (2014). Tone assimilation by Mandarin and Thai listeners with and without L2 experience. *Journal of Phonetics*, *46*, 86-100.
Xu, Y. (1991). Depth of phonological recoding in short-term memory. *Memory and Cognition, 19,* 263-273.

Xu, Y. (2013). ProsodyPro — A tool for large-scale systematic prosody analysis. In Proceedings of Tools and Resources for the Analysis of Speech Prosody 2013 (pp. 7-10). Laboratoire Parole et Langage, France.

Xu, C. X., & Xu, Y. (2003). Effects of consonant aspiration on Mandarin tones. *Journal of International Phonetic Association,* 33, 165-181.

Yang, B. (2015). *Perception and Production of Mandarin Tones by Native Speakers and L2 Learners.* Springer.

Yang, C. S. (2016). The Acquisition of Second Language Mandarin Prosody: From Experimental Studies to Pedagogical Practice. Amsterdam: John Benjamins Publishing Company.

Yuan, J. H., & Chen, Y. Y. (2014). 3rd tone sandhi in standard Chinese: A corpus approach. *Journal of Chinese Linguistics*, 42(1), 218-237.

Zhang, C. C., & Peng, G. (2013). Productivity of Mandarin third tone sandhi: a wug test. In G. Peng, & F. Shi (Eds.), *Eastward flows the great river: Festschrift in honor of Prof. William S-Y. Wang on his 80th birthday* (pp. 256-282). Hong Kong: City University of Hong Kong Press.

Zhang, C. C., Xia, Q. S., & Peng, G. (2015). Mandarin third tone sandhi requires more effortful phonological encoding in speech production: Evidence from an ERP study. *Journal of Neurolinguistics, 33,* 149-162.

Zhang, H. (2017). The effect of theoretical assumptions on pedagogical methods: a case study of second language Chinese tones. *International Journal of Applied Linguistics*, 27(2), 363-382.

Zhang, J., & Lai, Y. (2010). Testing the role of phonetic knowledge in Mandarin tone sandhi. *Phonology,* 27(1),153-201.

Zhang, J., Lai, Y., & Sailor, C. (2011). Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. *Lingua*, *121*(2), 181-206.

Zhang, X., Samuel, A. G., & Liu, S. (2012). The perception and representation of segmental and prosodic Mandarin contrasts in native speakers of Cantonese. *Journal of Memory and Language*, *66*(3), 438-457.

Zuraw, K. (2000). Patterned Exceptions in Phonology (Ph.D. dissertation). UCLA.

Zuraw, K. (2007). The role of phonetic knowledge in phonological patterning: Corpus and survey evidence from Tagalog. *Language, 83,* 277-316.