The effects of perceptual training on speech production of Mandarin sandhi tones by tonal and non-tonal speakers

Si Chen [a,e,f,*], Bei Li [a], Yunjuan He [b], Shuwen Chen [c], Yike Yang [d], Fang Zhou [a]

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, China Hong Kong Special Administrative Region
[b] Department of Modern and Classical Languages, University of North Georgia, USA
[c] Institute of Linguistics, Chinese Academy of Social Sciences, China
[d] Department of Chinese Language and Literature, Hong Kong Shue Yan University, China Hong Kong Special Administrative Region
[e] Research Centre for Language, Cognition, and Neuroscience, China Hong Kong Special Administrative Region
[f] PolyU-PekingU Research Centre on Chinese Linguistics, China Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:
Perceptual training
Tone sandhi
Wug test
Acquisition of sandhi tones

ABSTRACT

This study examined the effects of perceptual training on the acquisition of sandhi tones by both tonal (Cantonese) and non-tonal (English) speakers. In the pre- and post-training tasks, participants were presented with two monosyllables in each trial and were asked to produce them as disyllables. During perceptual training, they heard trials of disyllabic words with and without sandhi rule applications and were trained to identify the two types of disyllabic words. Based on functional data analyses, Cantonese speakers improved in building up more detailed and native-like allotone representations and in turn exhibited more native-like speech production in many cases. However, American learners showed improvement mainly in the offset f0 values or the turning point, indicating that they might have failed to pay attention to the shape of the entire tonal contour during perceptual training. For tonal speakers, the improvement of allotone production occurred on both real and wug words, indicating that the perceptual training may lead to more accurate production of tone sandhi rules regardless of syllable types. American learners, however, might need more training to further modify the allotone representation to perform more accurate sandhi tone application, especially in wug words. Finally, phonetic bias did not correctly predict the success of allotone acquisition since other factors such as the acoustic cues used in processing allotones during perceptual training might have affected the learning results.

1. Introduction

Perceptual training may lead to modifications of phonetic categories of second language learners in both segmentals (Bradlow et al., 1999; McCandliss et al., 2002; Saito and van Poeteren, 2018) and suprasegmentals (Kaan et al., 2007; Lu et al., 2015; Wayland and Guion, 2004; Wayland and Li, 2008). In turn, perception-based training may lead to improvement of speech production (e.g Wang et al. 2003.). Though previous studies reported inaccurate speech production of tones sandhi rules (Chen et al., 2019), the effects of perceptual training on the acquisition of sandhi tones by learners with various linguistic backgrounds remain to be investigated.

1.1. Mandarin and Cantonese tones and tone alternation

There are four tones in Mandarin Chinese: the high-level tone (Tone1), the high-rising tone (Tone2), the middle-dipping tone (Tone3), and the high-falling tone (Tone4). We may use Chao's (1948) five-degree scale of pitch height (from the lowest pitch value1 to the highest pitch value 5) to label the starting and ending points of a pitch contour for each tone: 55 for Tone1 (T1), 35 for Tone 2 (T2), 214 for Tone 3 (T3), and 41 for Tone 4 (T4). Cantonese has a more complex tonal system with six tones: T1 (55/53), T2 (25), T3 (33), T4 (21), T5 (23), T6 (22) in open syllables and three additional tones corresponding to T1, T3 and T6 in the checked syllables (e.g Bauer and Benedict 1997.).

Tone sandhi is a phonological process in tonal languages where the alternations of lexical tones are conditioned by certain phonological environments (Shih, 1997; Chen, 2000; Zhang, 2007) Zhang and Lai

* Corresponding author.
E-mail addresses: sarah.chen@polyu.edu.hk (S. Chen), chensw@cass.org.cn (S. Chen).

https://doi.org/10.1016/j.specom.2022.02.008
Received 17 November 2020; Received in revised form 20 January 2022; Accepted 22 February 2022
Available online 4 March 2022

(2010). concluded that there are two tone sandhi rules in Mandarin: one is the third tone sandhi rule, and the other is the more phonetically motivated half-third sandhi rule. Both rules describe the tonal alternations of T3 triggered by the following tones. There are two views of the third tone sandhi rule (Myers and Tsay, 2003): (1) The "categorical" view considers that the third tone sandhi is similar to the English *a/an* allomorphy, where the original T3 on the first syllable will change to T2 when followed by a T3, as shown in example (1); and 2) The "gradient" view, however, believes that the third tone sandhi does not completely neutralize with T2 and can be considered as an allophonic variation, which is similar to flapping in English. The second gradient view was supported by many recent acoustic studies (e.g Peng 2000, Xu and Prom-on 2014), and hence we will adopt the gradient view in our analyses.

In addition, T3 will become a half T3, where only the rising part of the tonal contour is truncated. This tonal alternation occurs when T3 is followed by T1, T2, or T4, as example (2) presented.

(1) T3 (213) → T2 (35)/_ T3 (213):

ʂʷəi 213 - kʷoo 213 → ʂʷəi 35 - kʷoo 213 "fruit"

(1) T3(213) →21/__{T1(55), T2(35), T4(41)}:

ʂəu 213 - ɕæn 55→ ʂəu 21 - ɕæn 55 "first"
ʂəu 213 - tsʷuu 35→ ʂəu 21 - tsʷuu 35 "brothers"
ʂəu 213 - ʂʷuu 41 → ʂəu 21 - ʂʷuu 41 "operation"

Zhang and Lai (2010) argued that the half-third tone sandhi rule is more phonetically motivated than the third tone sandhi rule since the half-third sandhi rule involves the truncation of the rising part of the pitch contour, which is considered phonetically natural, and it exhibits post-lexical characteristics, which can be applied across the board independent of syntactic brackets. The third tone sandhi rule is instead considered as a historical tone sandhi pattern and less phonetically motivated. This phenomenon is referred to as phonetic bias. Phonetic bias may affect the accuracy of sandhi rule application. We also consider phonetic bias in our design of this study, aiming to examine whether learners may improve better in half sandhi tones than the third sandhi tones. Moreover, different views of the underlying form of T3 have been proposed and the half-third sandhi tone can also be considered as the underlying form, and yet the traditional proposal of the full T3 as the underlying form is well accepted in the field (Zhang, 2017).

Cantonese is reported not to have consistent tone sandhi rules (Matthews and Yip, 1994). Instead, it has tone change, which is different from Mandarin tone sandhi rule application since tone alternation applies in more restricted contexts such as compounds while Mandarin tone sandhi applies once the phonological context is met.

1.2. Acquisition of sandhi tones

The tone sandhi rule applications were investigated among native speakers extensively. Cross-linguistically, native speakers adopt either a lexical mechanism (e.g Zhang et al. 2011.) or a computation mechanism (e.g Chen et al. 2019., Nixon et al. 2015, Politzer-Ahles and Zhang 2021, Zhang et al. 2015, Zhang and Peng 2013) in applying tone sandhi rules. The lexical mechanism refers to the mechanism speakers use in encoding and storing the tone sandhi, where the sandhi tones are stored as part of the lexical morphemes. In contrast, the computation mechanism refers to the mechanism of storing the abstract phonological representations of sandhi and non-sandhi forms and applies the rule whenever the phonological environment is met.

Although Perceptual Assimilation Model (PAM; e.g Best 1995.) and the Speech Learning Model (SLM; e.g Flege 1995.) are designed to account for the acquisition of phonemic categories, the acquisition of phonological processing remains less investigated. The observed results are often interpreted from the perspective of first language (L1) transfer and markedness of sounds (Hancin-Bhatt, 2008). Phonological acquisition by second language (L2) learners was reported in both segmental (Shoemaker, 2014; Spinu et al., 2018; Stemberger et al., 2018) and suprasegmental level (Chen and Kager, 2011; McAndrews, 2019). For example, after training, bilingual French-English-speaking Canadians have an advantage in acquiring the glottal stop as an allophone with a greater increased glottal stop rate compared to their monolingual counterparts, showing that bilingualism may lead to better acquisition of allophones (Spinu et al., 2018). In the suprasegmental level, advanced Dutch learners of Mandarin obtained native-like perceptive ability to discriminate and identify the third sandhi tones in simple tasks. However, their ability was undermined and less native-like if the perception tasks were complicated (Chen and Kager, 2011) Chen et al. (2019). investigated the speech production of tone sandhi by L2 learners with tonal and non-tonal linguistic backgrounds. Their results demonstrated that L2 learners had difficulties in acquiring allotones as their production was acoustically less accurate than that of native speakers. In addition, both tonal and non-tonal L2 learners failed to produce distinct allotones as native speakers. However, Cantonese learners' production of sandhi tones were more accurate than English learners (Chen et al., 2019).

1.3. The effects of perceptual training on L2 acquisition

SLM and the revised model SLM-r have the assumption that learners can add new phonetic categories or modify old ones to the phonetic system in response to new L2 sounds (Flege, 1995, 2007; Fledge and Bohn, 2020). It is also believed that laboratory perceptual training can modify perceptual mechanisms of adults (Wayland and Li, 2008). The perceptual training is effective in improving L2 learners' perception in the segmental level (Bradlow et al., 1997, 1999; Hazan et al, 2005) as well as the suprasegmental level such as tones (Francis et al., 2008; Wang et al., 2003; Wayland & Li, 2008).

On the segmental level, a large number of studies explored the effects of perceptual training on second language acquisition of English such as the acquisition of English phonetic contrasts /ɹ/ and /l/ by native Japanese learners (e.g Bradlow et al. 1999., McCandliss et al. 2002, Saito and van Poeteren 2018). It has been shown that perceptual training facilitates both perception and production of the phonetic contrast /ɹ/ and /l/ by Japanese L2 learners (Shinohara and Iverson, 2018) Lee et al. (2020). compared the effects of perception and production-based training on the acquisition of segmentals and suprasegmentals by Japanese learners of English. Though both groups showed improvement after training, perceptual training was proved to be more effective than production-based training. Similarly, Lu et al. (2015) found no differences between two groups receiving perception-only training and perception-plus-production training in improving the perception of lexical tones. These results suggested that the perceptual training may have an advantage over production training in improving learners' perception.

On the suprasegmental level, previous studies showed inconclusive results regarding whether the tonal background of native language may facilitate L2 learners' ability to acquire a new tonal language after training Wayland and Guion's (2004). training on mid and low Thai tonal contrasts for English speakers and Chinese speakers showed that the latter group of trainees outperformed the former one in various inter-stimulus interval conditions. However, Kaan et. al. (2007) trained Mandarin and English speakers in perceiving Thai mid and low tones, and the training led to a different carryover effect on the perception of untrained high tone deviant with respect to the linguistic background (tonal vs. non-tonal) of the L2 learners, which suggests that English speakers were more sensitive to pitch level than pitch contours (Chandrasekaran et al., 2007; Gandour, 1983; Gandour and Harshman, 1978; Krishnan et al., 2010).

In addition to the training effects on perception (Kaan et al., 2007; Kaan et al., 2008; Song et al., 2008; Wang et al., 1999; Wayland and

Guion, 2004; Wayland and Li, 2008), perception-based training may also help improve speech production of segments and tones by L2 learners (Sakai and Moorman, 2018; Saito, 2011; Saito and Saito, 2017; Wang et al., 2003). SLM hypothesized that accuracy in L2 perception precedes accuracy in L2 speech production. Therefore, the training in perception may lead to an improvement in mental representations and in turn in speech production. Due to the conflicting results regarding the improvement of speech production after perceptual training, Sakai and Moorman (2018) conducted a meta-analysis of cross-linguistic publications on the effects of perceptual training on the acquisition of obstruents, sonorants and vowels. Their results did show improvement in speech production through perceptual training and the improvement in obstruents was bigger than that of vowels or sonorants. It has been proposed that the phonological representation may be associated with an interface module mediating interactions with the articulatory system (Pathi and Mondal, 2021). Since the articulatory system controls the coordination of articulators and the physical realizations of speech production, mental representations can affect speech production through the interface module with articulatory system. These results suggest that speech production and perception are associated and perceptual training of segments may improve mental representations and in turn speech production. However, the effects of perception-based training on acquisition of tone sandhi rules remain unknown.

1.4. The current study

Although perceptual training of segmentals and suprasegmentals has been shown to be effective to help modify the perceptual mechanism (e.g Wayland and Guion 2004., Wayland and Li 2008) and also speech production (e.g Wang et al. 2003.) in adult learners, relatively few studies investigated the acquisition of phonological sandhi rules and the effects of perceptual training on the acquisition of allotones. It has been reported that tonal and non-tonal speakers performed differently in tone sandhi rule acquisition (Chen et al., 2019). However, it is yet unknown whether perceptual training may facilitate learners' acquisition of tone sandhi rules.

The current study aims to answer three research questions: (1) Does perceptual training lead to improvement in the acquisition of sandhi rules by both tonal and non-tonal learners? (2) Does perceptual training lead to a better speech production of the sandhi tones in both real and wug words? (3) Does phonetic bias in rule applications affect the training results?

With regards to the first research question, although perceptual training has been shown to help modify perceptual mechanisms and establish phonetic categories, little is known about how perceptual training may improve the acquisition of allotones. As reported by Chen et al. (2019), Cantonese and American learners have difficulties in learning the dissimilarities among the sandhi tones since their production of allotones were less distinct from each other compared to those produced by native speakers. Also, their findings confirmed that non-native speakers' production of sandhi tones was less detailed and accurate compared to native Mandarin speakers. We hypothesize that through perceptual training, learners will build more detailed and native-like allophonic representations of allotones and they will learn to better differentiate between the allotones. The perceptual training is designed to train learners to differentiate the sandhi forms and the underlying form by letting them identify disyllabic words that have gone through tone sandhi from synthesized disyllables without tone sandhi rule application. This study aims to explore whether perceptual training may lead to richer and more detailed representations of allotones and in turn more accurate speech production. In addition, tonal and non-tonal learners may show differences in the representations of tones after training. Since non-tonal speakers have no existing tonal categories and less experience in using acoustic cues to differentiate tones compared to tonal speakers, it is likely that they might have more difficulties in improving speech production of sandhi tones.

As for the second research question, it has been reported that tonal and non-tonal learners showed no significant differences in the tonal contour between both real and wug morphemes (Chen et al., 2019). We aim to explore if perceptual training help improve sandhi tone production in both real and wug words or if perceptual training only leads to better rule applications in real words. Wug words were included to test whether learners can encode the new sandhi contour as part of the abstract tonal category and whether tone sandhi rules can be better applied without the influence of lexical meanings.

Regarding the third research question, we aim to explore if phonetic bias predicts the success of acquisition after perceptual training. The half-third sandhi rule is reported to be more phonetically motivated as it only involves the truncation of the second half of the pitch contour and is applied across the board (Zhang and Lai, 2010). Thus, it is predicted that perceptual training may lead to more accurate production of the half-third sandhi tone than the third sandhi tone. It is likely that phonetic bias predicts the amount of improvement in rule applications by Cantonese and American learners after training.

2. Methodology

2.1. Subjects

We recruited three groups of participants, twelve native Mandarin Chinese speakers from Beijing (NM) (6M; 6F; age: 26.25 ± 4.61(Mean ± SD); years of living in Beijing: 22.83 ± 3.43), twelve native Cantonese speakers from Hong Kong for the training group (TC) (8F; 4M; age: 22.5 ± 2.15 (Mean ± SD); starting age to learn Mandarin: 17.75 ± 2.49; years of learning: 2.33 ± 0.89)) and twelve native English speakers for the training group (TA) (7F; 5M; age: 22.09 ± 1.87 (Mean ± SD); starting age to learn Mandarin: 18.09 ± 2.34; years of learning: 3.27 ± 1.03). TA subjects were recruited from University of North Georgia, and TC and NM subjects were recruited at the Hong Kong Polytechnic University. TC and TA subjects completed all three phases of the experiment (pre-training, training and post-training), but NM only participated in pre-training recording.

Differential responses in a single group can be used as an experimental design to evaluate the effects of training, which is also known as before-and-after designs (Chambliss and Schutt, 2018; Paulus et al., 2014). Participants are considered as their own control using data before training as the baseline. This type of design has been widely employed in perceptual training studies (e.g Sakai and Moorman, 2018.; Wayland and Li, 2008). The direct comparison of the pre- and post-training performance may inform us about the effectiveness and limitation of the training. We also compared the participants' speech production before and after training. Our whole experiment procedure only took two to three days, and the participants did not receive any other types of training during our experiment, so it is unlikely that they will improve their speech production due to other factors such as other types of training or familiarity with the tasks. To confirm our assumptions that speakers will not improve their speech production due to other factors, we recruited a Cantonese control group to rule out the possibility that participants may make improvements due to other factors such as replication of the tasks. We did not train the Cantonese control group, and they were tested twice on the tone sandhi production. Specifically, we recruited 12 native Cantonese participants (CC) for the control group (7F; 5M; age: 23.17 ± 1.40 (Mean ± SD); starting age to learn Mandarin: 18.58 ± 2.61; years of learning: 2.25 ± 1.34), who participated in the first-time and second-time recording without receiving training to serve as a control baseline mainly to rule out familiarity brought by replicating the tasks. The length of Mandarin Chinese study and the initial exposure age were controlled to be similar among TA, TC, and CC group. All the Cantonese and English speakers were intermediate learners of Mandarin and they have similar proficiency level according to their starting age of learning, years of learning, education backgrounds as reported above. Participants have also provided a score of their

proficiency level in a five-point scale (higher scores stand for higher proficiency) including speech, reading, listening and writing skills. The scores for each groups are reported as follows: Cantonese control group: 2.96 ± 0.62; Cantonese training group: 3.38 ± 0.75; American training group: 3.15 ± 0.42.

2.2. Training

2.2.1. Training stimuli

As noted in the introduction, there are two types of tone sandhi rules: the third tone sandhi and half-third tone sandhi rule. In order to raise the awareness of tone sandhi rule applications in disyllabic tones, as well as the differences among allotones, we used disyllabic words with tone sandhi rules application and those without. The disyllabic words without application of sandhi rules were re-synthesized based on a modification of the disyllabic words with the rule application. Specifically, the first monosyllable in each disyllable was replaced by the pitch contour of the same monosyllable uttered in isolation with the full T3. In this fashion, the segmental information, duration and intensity remain the same for disyllables with and without rule applications. All stimuli were normalized for peak intensity (65dB) and duration. We manipulated the pitch contour using the pitch synchronous overlap add (PSOLA) method (Moulines and Laroche, 1995). All the stimuli used in pre- and post-training tasks and training were processed by Praat (Boersma and Weenink, 2020). The resynthesized stimuli on each syllable were evaluated by two native Chinese speakers and the stimuli were judged to be natural by both speakers.

Each disyllabic stimulus was produced in random order and repeated for three times by 5 female native Mandarin speakers. These five female native Mandarin speakers (age: 23.2 ± 1.79 (Mean ± SD)) did not speak Cantonese. They have lived in Mainland China for most of their lives (22.4 ± 0.9 years (Mean ± SD)) and have lived in Hong Kong for less than one year (month: 8.4 ± 5.37 (Mean ± SD)). The recording took place in the speech lab at the Hong Kong Polytechnic University with a MB Quart K800 C headset. The stimuli were digitized using a PC (44.1 kHz sampling rate). The naturalness and quality were ensured in the chosen training stimuli. In order to examine the generalizability of perceptual training, we only chose a part of the stimuli to train participants. Specifically, the stimuli used in the training sessions were all real words, but only a proportion of stimuli from pre-training recording were used, in order to examine if the training effect is generalizable to all stimuli. Specifically, six real words were used and three of them had Mandarin third tone sandhi rule application (T3+T3: /məi/ /xau/ "美好", /ʂʷəi/ /kʷoo/ "水果" and /ʈʂʷæn/ /liæn/ "转脸"), and the other three had the half-third sandhi rule application (T3+T1: /ʂʷəi/ /tɕəŋ/ "水晶"; T3+T2: /ʈʂʷæn/ /ii/ "转移"; T3+T4: /məi/ / lìi/ "美丽"). Seventy-two tokens of the training words produced by three female native Mandarin speakers with clear pronunciation were chosen as the training stimuli. In this study, we controlled for the variability of stimuli by using only female speakers, but variability in training stimuli may also result in differences in training effects. Early studies suggest that variability in stimuli benefit learners in better generalization to new stimuli and new talkers (Logan et al., 1991; Lively et al., 1993; Wang et al., 1999) since the learners are exposed to acoustic variability and contrasts for better representations of the stimuli. However, more recent studies demonstrate that the subject internal factors may interact with talker variability so that some learners with lower pre-training aptitude (Perrachione et al., 2011) or with less proficiency (Antoniou et al., 2015) may suffer from stimulus variability, which may be due to less efficient employment of cognitive resources by low-aptitude learners (Antoniou and Wong, 2015) Perrachione et al. (2011). also shows that there is an interaction between individual abilities and training paradigm in successful learning of phonological pitch contrasts. In future studies, the role of variability in tone sandhi training can be further tested.

Corresponding to the chosen tokens, we made 72 synthesized tokens, where we used the contour of T3 produced in isolation to replace the tonal contour in the first syllable. For each word with and without tone sandhi application, six utterances of three different talkers were chosen based on the acoustic quality. In total, the participants heard 144 stimuli.

2.2.2. Training procedure

Following the experimental procedure by Wayland and Li (2008), Hanulikova et al. (2012) and Baese-Berk and Samuel (2015), the perceptual training procedure included two phases of identification (ID) training. In the practice session, we first reviewed the concepts of tone sandhi rules and gave six real-word examples of the two tone sandhi rules in Mandarin Chinese. Participants were told to pay attention to the differences between the correct stimuli (with tone sandhi application) and incorrect stimuli (the same two morphemes without tone sandhi application). These stimuli with and without sandhi rule applications were recorded by three native speakers.

In the first phase of identification (ID) training, we presented one single stimulus (a disyllabic word such as mei3 + xɑu3 "goodness") in each trial. The participants were told that the correctly pronounced disyllabic word (tone sandhi rules applied) appears in odd trials, and the resynthesized disyllabic words deemed incorrect (tone sandhi rules not applied) appears in even trials. The Inter-Trial-Interval (ITI) was fixed at 3000 ms. They were told to press the button labelled "A" and "B" on the keyboard to reinforce whether the stimuli have correctly applied the tone sandhi rule or not. The participants were instructed to pay attention to the differences between the stimuli in the odd and even trials. The participants took a short break after the first phase ended.

In the second phase of ID training, each token was presented randomly and then the participants were asked to determine whether the stimulus presented was pronounced correctly (with tone sandhi rule application) or incorrectly (without tone sandhi rule application). Feedback was given immediately after the response. Participants completed all the training sessions within one and a half hours.

2.3. Pre-and post-training

2.3.1. Pre-and post-training stimuli

We used the stimuli in Chen et al. (2019) for the pre- and post-training tasks, which were designed following the studies by Zhang and Lai (2010), Hsieh (1970), (1975), (1976), Zhang and Peng (2013). The first type we used was all real disyllabic words which are named as AO-AO (AO: actual occurring morphemes). The second type include disyllabic sequences consisting of real morphemes (*AO-AO), but the disyllabic sequences do not exist in Mandarin. The third type of stimuli consisted of a real morpheme and a syllable of an accidental gap (AO-AG, where AG stands for an accidental gap, meaning that the morpheme is a legitimate syllable, but the syllable does not bear a specific tone). The fourth type was made up of a syllable of an accidental gap and a real morpheme (AG-AO), and the fifth type used two syllables of accidental gaps (AG-AG). The sixth type used two pseudo-words of non-existing syllables in Mandarin. The first type of stimuli was real words, and the second to sixth types were considered wug words as used in the "wug" tests (Berko, 1958).

The vowels in the first syllable of all types of diagrams were controlled to be the same or similar as much as possible to avoid the intrinsic f0 effect from vowels (Whalen and Levitt, 1995). In addition, pairs with aspirated and unaspirated onsets were not included to avoid consonant perturbation effect on f0 values (Xu and Xu, 2003). Similar to Zhang and Lai (2010), all the chosen diagram and individual characters were highly frequent words selected from character and diagram frequency corpus (Da, 2004). Disyllabic wug words were not real words in any tonal combinations to avoid neighbourhood effects (Zhang and Lai, 2010). A variety of wug words were adopted in order to test the training effects on the applications of tone sandhi rules Table 1. lists some example stimuli to illustrate types of real and wug words.

Filler words consisting of real disyllabic words (AO-AO) and wug

Table 1
Example stimuli of real and wug words.

Tones	Chinese diagram	Transcription	Tones	Chinese diagram	Transcription
T3-T1			T3-T3		
AO-AO	每天	mei tʰiæn	AO-AO	美好	mei xɑu
*AO-AO	每聪	mei tsʰʷoŋ	*AO-AO	美怎	mei tsən
AO+AG	每 nei	mei nei	AO+AG	美shuan	mei ʂʷæn
AG+AO	hei惜	xei ɕii	AG+AO	hei 好	xei xɑu
AG+AG	hei den	xei tən	AG+AG	hei diu	xei tɨəu
Pseudo	chei fai	ʈʂʰei fai	Pseudo	chei bou	ʈʂʰei pəu
T3-T2			T3-T4		
AO-AO	美国	mei kʷoo	AO-AO	美丽	mei lʲii
*AO-AO	美别	mei pʲee	*AO-AO	美特	mei tʰɤɤ
AO+AG	美ka	mei kʰaa	AO+AG	美mang	mei mɑŋ
AG+AO	hei别	xei pʲee	AG+AO	hei怕	xei pʰaa
AG+AG	hei mie	xei mʲee	AG+AG	hei dei	xei tei
Pseudo	Chei diang	ʈʂʰei tʲɑŋ	Pseudo	chei fai	ʈʂʰei fai

disyllabic words (*AO-AO) were also used in order not to reveal the purpose of the experiment. For example, the real word 開發 "develop" /kʰaɪ fa/ and the wug word T1-T1科生/kʰɤ ʂəŋ/ were used as fillers. In total, there were 192 target stimuli and 192 filler words with all possible tonal combinations. All the monosyllabic stimuli were read by a native Mandarin speaker born in Beijing, and Tone 3 was pronounced as a full third tone with a falling-rising f0 contour. All the stimuli were normalized for peak intensity.

2.3.2. *Pre-and post-training procedure*

We used the software 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 800 ms in between the two monosyllables. Using some filler words, participants were given instructions and learned how to put the two monosyllables together to form a disyllabic word in Mandarin. There were six blocks of the stimuli, where the order of presentation was counterbalanced across participants. The stimuli were randomly presented in each block. They were instructed to speak at a normal rate of speech and they were allowed to correct themselves if they made any mistake. Practice sessions and instructions with examples were also offered to participants to ensure they could complete the pre- and post-training tasks.

Similarly, in the experimental session, the participants were asked to produce the disyllabic stimuli in each trial. All the Mandarin and Cantonese participants were recorded by a GMH C 8.100 D headset at the speech lab of the Hong Kong Polytechnic University. All the American participants were recorded by a Sanako SLH07 headset at a quiet room at University of North Georgia. To empirically test the recording conditions and exclude other confounding factors from the stimuli to be recorded, we generated synthesized pure tones varying in amplitude (1pa and 2pa) and frequency (150Hz, 175Hz, 200Hz, 225Hz, 250Hz, 275Hz, 300Hz, 325Hz, 350Hz) using Praat. Then we played the tones and recorded them in both testing conditions. Since our paper relies on f0 measurements, we tested whether the testing condition significantly contributed to f0 measurement in a regression model. The results showed testing conditions did not contribute to f0 measurement ($t(78) = -0.04$, $p = 0.97$). The f0 measurement were also comparable (f0 mean ± sd: lab 1: 263.3 ± 70.99; lab2: 262.7 ± 70.82).Fig. 1 summarizes the whole procedure.

2.3.3. *Statistical analyses*

We used time-normalized pitch contours so that we may better compare the pitch contours (see Xu 2013 for the reasons to use time-normalized pitch contours for pitch contours comparisons). We segmented the recordings and extracted f0 values at 20 equal time points using the Praat script ProsodyPro (Xu, 2013). Following Chen et al. (2017), and we proceeded to apply the Log Z-score transformation (Laplace, 1820) to the f0 values and only the transformed f0 values from the first syllable of each target word were analyzed. We applied the functional data analyses to compare pairs of tonal contours before and after training and the pairs produced by learners and native speakers. In evaluating the training effects, we conducted a comparison of pre- and post-training performance, a comparison of post-training performance by learners and the speech production of native speaker. We included pre-training data from the speech corpus (Chen et al., 2019) for a comparison to the post-training data.

Specifically, we fit pairs of normalized f0 contours with the following model:

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

where $y_i(t_j)$ stands for the f0 value at time point t_j of utterance i by each subject, and $i = 1, \ldots, n$, and $j = 1, \ldots, m$. The error term, ε_{ij} follows a normal distribution, $N(0, \sigma^2)$. To compare a pair of tonal contours, from each vowel, we chose 20 break points and fit these contours using four B-spline basis functions. The basis function expansion for $f_i(t_j)$ we used was in the following form:

$$f_i(t) = \sum_{k=1}^{K} c_{ki}\, \varphi_k(t) \tag{2}$$

where c_{ki} is the coefficient for the k^{th} basis function that we used to model the ith utterance. Following the procedure described by Chen et al. (2017), for each pair of f0 contours to test, 20 break points were chosen first. Then we used four B-spline basis functions to fit f0 contours, where we determined the optimal smoothing value, λ, by the generalized cross-validation measure. After all f0 curves were fitted, we then proceeded to conduct a functional t-test to compare the differences

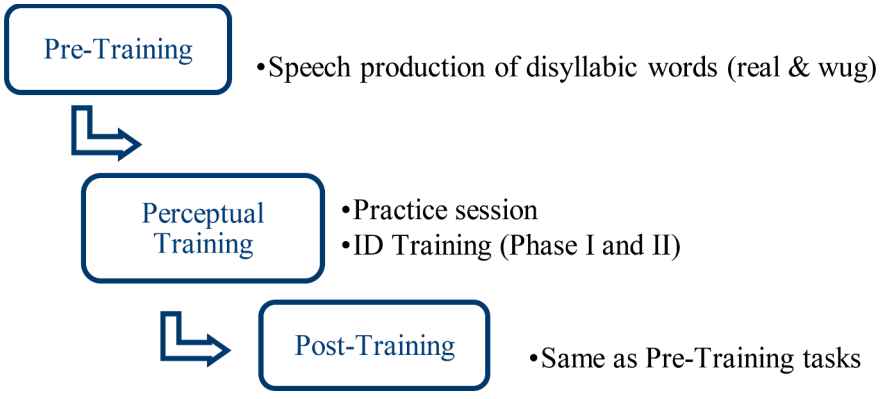

Fig. 1. A flowchart of the whole experimental procedure.

between the two f0 curves in each pair. A functional t-test constructs a null distribution by randomly shuffling the labels of the two curves. We used 200 random samples, and calculated observed t-statistic, point-wise 0.05 critical value and maximum 0.05 critical value. Statistical significance is reached when observed t-statistics exceed the critical value.

3. Results

3.1. Tone sandhi production by Cantonese speakers

3.1.1. The third sandhi tone produced by Cantonese speakers in the training group

From Fig. 2, the tonal contours of sandhi T3 changed significantly after training in both real and wug words, where sandhi T3 produced after training had higher pitch values and closer to that of native speakers. In addition, for real words, the rising slope of the tonal contours also became much steeper and more similar to the shape of tonal contours produced by native speakers.

In real words, before training, Cantonese speakers produced the sandhi T3 that showed significant differences in the entire tonal contour (region: 0–100%) compared to native speakers. From Fig. 3, the red solid line represents observed statistics, the blue dashed line represents pointwise critical value. All the observed test statistics are above the maximum 0.05 critical value (1.97) for the entire region. Since statistical significance was reached when observed t-statistics exceeded critical values, the two curves were significantly different for the whole contour.

Similar analyses were applied to other pairs of tonal contours. After training, Cantonese speakers' production was not significantly different from native speakers in any region. Similarly, in wug words, before training, Cantonese speakers produced the sandhi T3 differently from

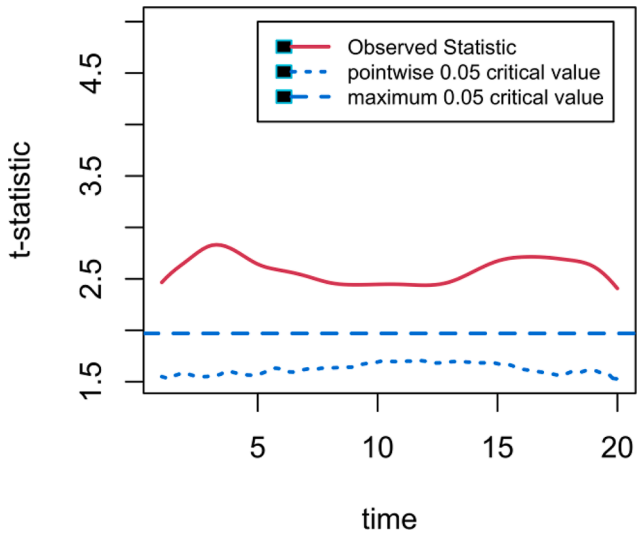

Fig. 3. Functional data analysis of tonal contours produced by Cantonese and Mandarin speakers on real words before training.

the native speakers in the entire tonal contour (region: 0–100%). After training, the region with significant differences also shrunk, where only the regions 0–17 and 74–100% of the contour were still different from the native speakers. We also compared the f0 contours on real vs. wug words before and after training. The results showed no significant differences in the entire contour for real words. For wug words, the contour differed from the region 12–94% before and after training. In addition, a comparison between trained and unseen items was conducted. While

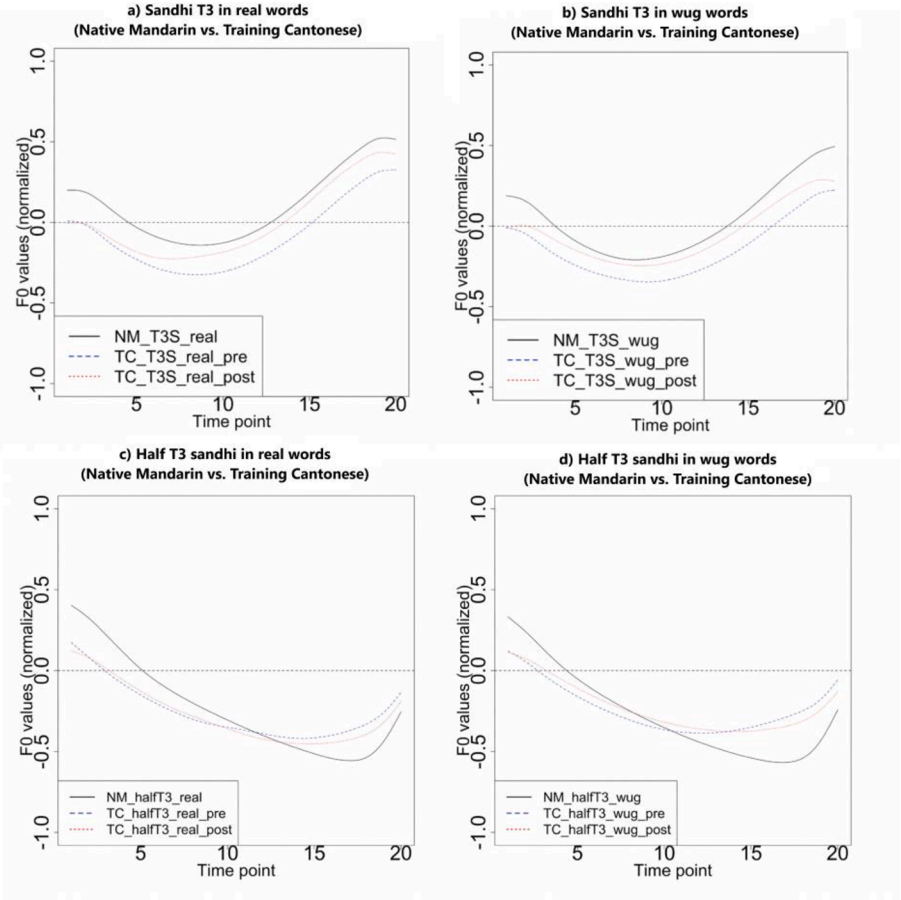

Fig. 2. (a) A comparison of native speakers' and Cantonese speakers' production of sandhi T3 in real words before and after training; (b) A comparison of native speakers' and Cantonese speakers' production of sandhi T3 in wug words before and after training; (c) A comparison of native speakers' and Cantonese speakers' production of half T3 sandhi in real words before and after training; (d) A comparison of native speakers' and Cantonese speakers' production of half T3 sandhi in wug words before and after training.

there was no significant differences before training, the trained and unseen items showed slight difference toward the end of the contour (94–98%).

3.1.2. The half-third sandhi tone produced by Cantonese speakers in the training group

In real words, before training, Cantonese speakers produced the half-third sandhi tone that differed from native speakers in the region 0–19%. After training, they were significantly different from native speakers in the region 0–16%. In wug words, before training, Cantonese speakers' half-third sandhi differed from the native speakers in the regions 0–28% and 63–100%. After training, the region of significantly different regions of the contour slightly shrunk, where only the regions of 0–17 and 60–100% were still different from the native speakers.

As shown in Fig. 2, even though the post-training half T3 tonal contour in real and wug words showed closer f0 values to the native norm toward the end of the contour, they still demonstrated significant differences from the native norm. It was also observed that the pitch range of half T3 produced by Cantonese speakers was narrower than the native half T3 regardless of the training. The native norm showed steeper falling slopes and less rising proportion toward the end of the contours. We also compared the f0 contours on real vs. wug words before and after training. The results showed no significant differences in the entire contour. In addition, there was no significant differences between trained and unseen items before and after training.

3.1.3. The sandhi tone production by Cantonese speakers in the control group

For both real and wug words, the third sandhi tone was produced with significantly different tonal contours compared to that of native speakers in both the first and second time. As shown in Fig. 4, in the first test, the region 55–100% in real words and the regions 0–3 and

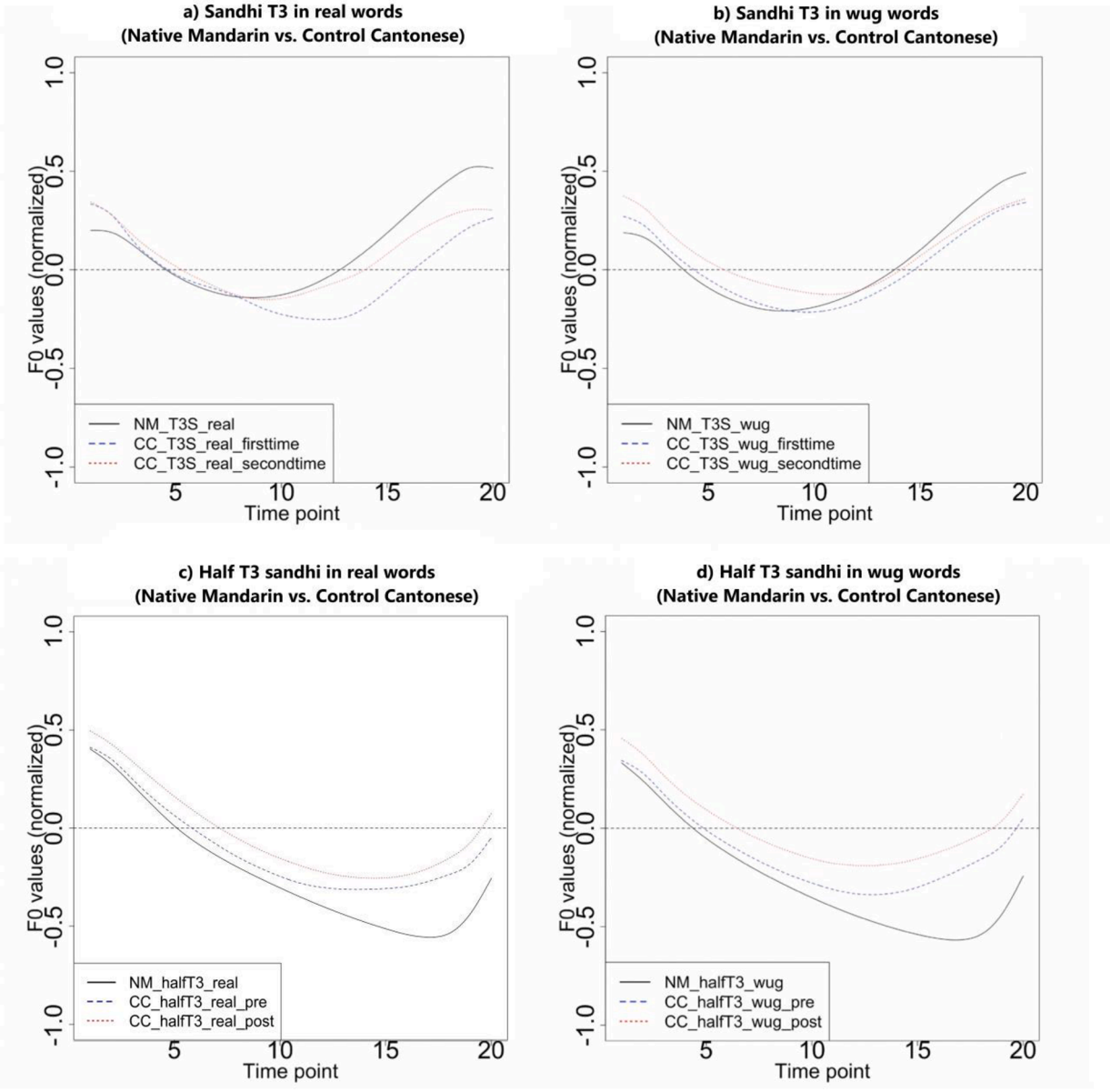

Fig. 4. (a) A comparison of native speakers' and Cantonese speakers' (control group) production of the third sandhi tone in real words during the first and second time of production; (b) A comparison of native speakers' and Cantonese speakers' (control group) production of the third sandhi tone in wug words during the first and second time of production; (c) A comparison of native speakers' and Cantonese speakers' (control group) production of half T3 sandhi in real words during the first and second time of production; (d) A comparison of native speakers' and Cantonese speakers' (control group) production of half T3 sandhi in wug words during the first and second time of production.

72–100% in wug words showed significant differences from the native norm. In the second time of the production task, the region 91–100% in real words and the regions 0–48 and 83–100% in wug words reached significance compared to the native norm. Although some improvement was shown in real words for the second time of production, the improvement was not consistent, showing larger significantly different regions for wug words in the second time of production.

Moreover, the half-third sandhi tone was also produced with tonal contours significantly different from those produced by native speakers in the first and second time in both real and wug words. As shown in Fig. 4, in the first test, the region 76–100% in real words as well as the region 51–100% in wug words were significantly different from the native norm. In the second test, the region 59–100% in real words and the regions 0–100% in wug words reached significance compared to the native norm. From Fig. 4 and the functional t-test results, there was no consistent improvement in the second time of the speech production both in real and wug words. Therefore, we confirmed that other factors such as task familiarity could hardly lead to consistent improvements in tone sandhi production and continued to employ the before-and-after experimental design for English speakers without recruiting a control group.

3.2. *Tone sandhi production by English speakers*

3.2.1. *The third sandhi tone produced by English speakers*

For both before and after training, the sandhi T3 was produced with flatter tonal contours by American learners, compared to those produced by native speakers in both real and wug words. Before training, the region 87–100% in real words as well as the regions 25–46 and 88–100% in wug words were significantly different from the native norm. After training, compared to the native norm, only 0–2% reached significance in real words and the regions 0–7 and 16–80% reached significance in wug words. As shown in Fig. 5, after the perceptual training, English speakers produced the sandhi T3 (both in real and wug words) with significantly higher pitch values across the entire contours. The offset values became closer to the native speakers after the perceptual training, and yet the entire contour was much flatter than native speakers. We also compared the f0 contours on real vs. wug words before and after training. The results showed significant differences in 1–6 and 31–100% in the contour. The results indicated that English speakers may need more training to improve the shape of tonal contours. During the current perceptual training, they might have paid more attention to the offset values instead of the shape of the tonal contour, hence they learned to produce more accurate offset values, but the shape of the tonal contours did not improve much. In addition, there was no significant different regions between trained and unseen items before and after training.

3.2.2. *Half T3 sandhi produced by English speakers*

Fig. 5 (c, d) showed the half-third sandhi contours produced before and after training in real and wug words, respectively. Comparing tonal contours in Fig. 5 (c, d), the rising proportion in real words and wug words shrunk. Also, the turning point of the contour became closer to native speakers. However, the initial part of the tonal contour suffered from the adjustment of the turning point. Specifically, before training, the region 47–100% was significantly different from native Mandarin speakers in real words, and the region 18–100% reached significance in wug words. After training, the significantly different region expanded to 23–100% for real words and 2–100% for wug words. We also compared the f0 contours on real vs. wug words before and after training. The results showed significant differences in the region 46–100%. In addition, there was no significant differences between trained and unseen items before and only a slight region 87–100% after training.

3.3. *Speech production of T3 variants before and after training*

In order to examine whether non-native speakers learned to better

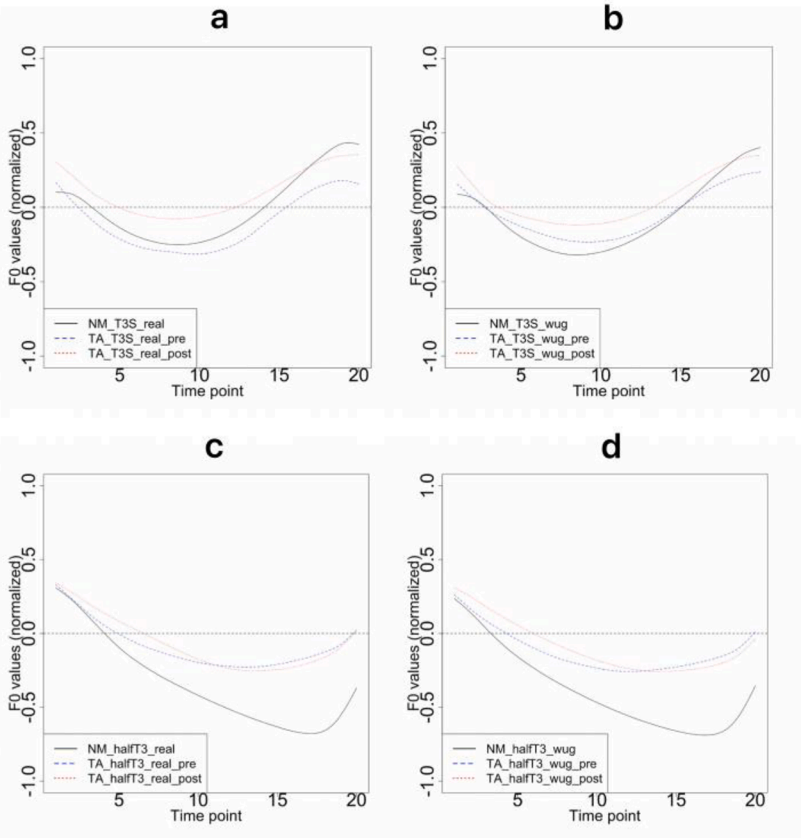

Fig. 5. (a) A comparison of native speakers' and English speakers' production of the third sandhi tone in real words before and after training; (b) A comparison of native speakers' and English speakers' production of the third sandhi tone in wug words before and after training; (c) A comparison of native speakers' and English speakers' production of the half-third sandhi tone in real words before and after training; (d) A comparison of native speakers' and English speakers' production of the half-third sandhi tone in wug words before and after training.

differentiate between T3 variants after training in both real and wug words, we plotted the speech production of T3 variants in one graph for both native and non-native speakers. As shown in Fig. 6 (b) and (f), before training, non-native speakers tended to produce tonal contours less discernible compared to native speakers. Native Mandarin speakers produced much sharper rising third sandhi tones than those produced by English speakers, and native speakers showed higher offsets in the third sandhi tones than both Cantonese and English speakers. What makes these tonal contours of T3 variants more discernible for native speakers is that they produced a sharper falling tonal contour for the half-third sandhi tone with little rising portion at the end of the contour, compared to both Cantonese and English speakers.

As shown in Fig. 6 (c) and (g), after training, both Cantonese and English speakers showed more discernible tonal contours of T3 variants. Specifically, Cantonese speakers learned to produce the third sandhi tone with a steeper rising slope and higher offset values in both real and wug words after training. For the half-third sandhi tone, Cantonese speakers produced the tonal contours with a turning point closer to the end of the contour, resulting in lower offset values in both real and wug words after training. The English speakers produced third sandhi tone with higher f0 values overall in both real and wug words after training. For the half-third sandhi tone, similar to Cantonese speakers, English speakers also learned to produce the tonal contours with a turning point closer to the end of the contour in both real and wug words after training. Compared to the changes in the two training groups (TA and TC), the speech production of the two allotones produced by Cantonese speakers in the control group did not become more discernible from each other in the first vs. second speech production. Perceptual training did lead to more discernible allotones in the two training groups.

4. Discussion

4.1. The effects of perceptual training in improving tone sandhi rule applications in tonal and non-tonal speakers

The acquisition of phonemic categories by second language learners have been extensively investigated and models such as PAM (e.g Best 1995.) and SLM (e.g Fledge 1995.) have been proposed to predict the learning outcome. However, the acquisition of phonological processing still remains to be investigated. Difficulties in acquiring phonological rules have been noted in previous studies, where L1 transfer and markedness usually play a role in the observed learning patterns (e.g Hancin-Bhatt 2008.). For example, Major (2008) summarized the effect of L1 transfer in second language acquisition of phonology, including sound substitution using L1, overdifferentiation (L1 has phonemic distinctions while the L2 does not) and underdifferentiation (L2 has phonemic distinction, but L1 only has allophonic distinction). It has been reported that Cantonese speakers may have positive transfer from the Cantonese rising tones in producing the Mandarin sandhi T3. In producing the half-third sandhi tone, Cantonese speakers also showed more rising at the end of the sandhi tone, indicating a possible transfer effect from the Cantonese rising tones or a low falling tone Cantonese T4 (21), similar to the low-falling contour of the Mandarin half-third sandhi tone T3 (Chen et al., 2019).

Perception-based training has been shown to be effective in improving both speech perception of lexical tones (e.g Kaan et al. 2007., Wayland and Guion 2004, Wayland and Li 2008) and speech production of tones (Wang et al., 2003). SLM proposed that when learners are exposed to new L2 sounds, they are able to establish new phonetic categories or modify their existing L1 phonetic categories (Flege, 1995). Also, perceptual training may help adult learners modify their perceptual mechanisms and help them establish tonal categories though few studies examined the effect of perceptual training in helping acquire variants of tones.

Our results showed that perceptual training significantly improved the speech production of the sandhi T3 produced in real and wug words

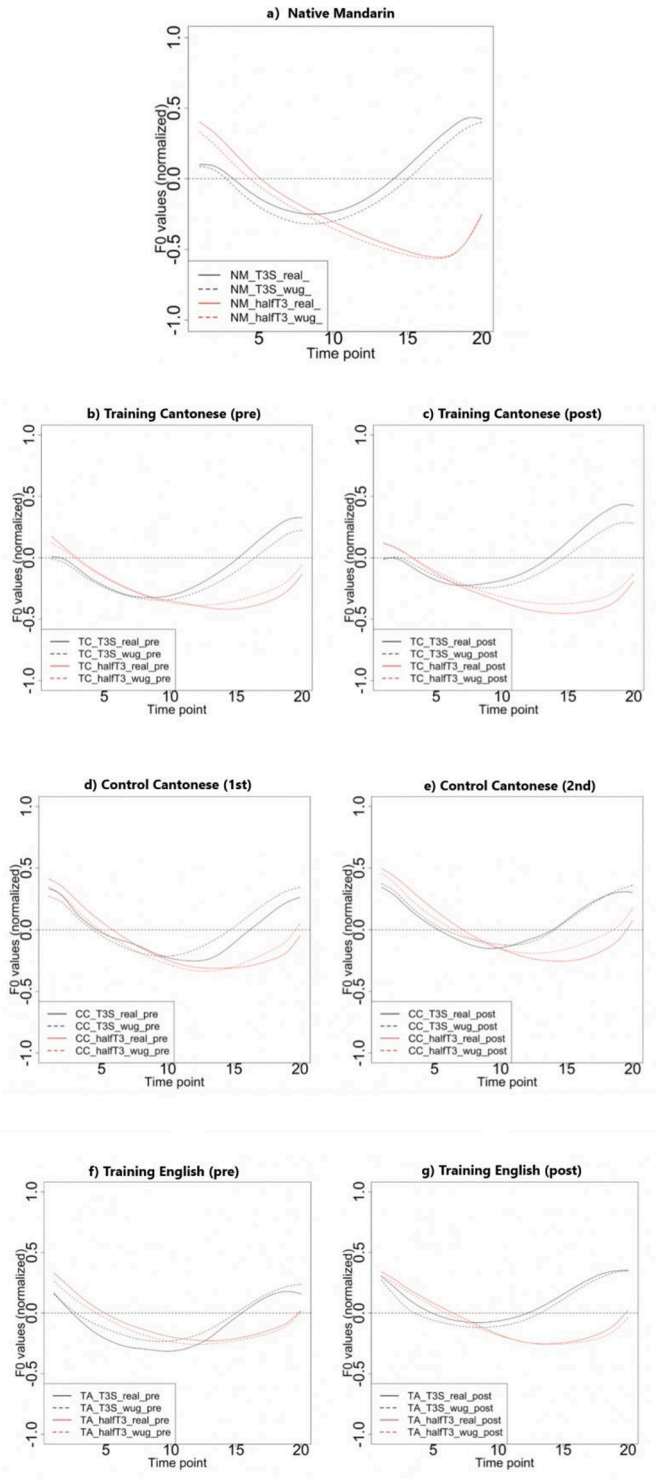

Fig. 6. (a) Speech production of T3 variants by native Mandarin speakers; (b) Speech production of T3 variants by Cantonese speakers before training; (c) Speech production of T3 variants by Cantonese speakers after training; (d) Speech production of T3 variants by Cantonese speakers in the control group for the first time; (e) Speech production of T3 variants by Cantonese speakers in the control group for the second time; (f) Speech production of T3 variants by English speakers before training; (g) Speech production of T3 variants by English speakers after training.

by native Cantonese speakers. Specifically, the training helped raise f0 values and made the rising slope steeper so that the f0 values became much closer to the native speakers. However, perceptual training only led to slight improvement in the tonal contours of the half-third sandhi tone. After training, Cantonese learners showed less steep rising at the end of the tonal contour in both real and wug words, indicating that they modified the half T3 representation with more truncation of the rising part toward the end of the tonal contour though the significant regions based on a comparison to the native speakers only decreased 3% for real word and 8% for wug words.

In addition, our results showed that English speakers did improve in the speech production of the third sandhi tone after training in terms of reduced significant regions. However, for both real and wug words, American learners may have paid more attention to the tonal offset rather than the shape of the entire tonal contour. This finding was consistent with early studies (e.g Gandour and Harshman 1978.), where non-tonal language speakers were reported to focus more on either the onset or the offset values rather than the shape of the tonal contour in tone processing. In contrast, tonal speakers usually focus more on the shape of the tonal contour (Lee and Nusbaum, 1993) in processing tones. Though perceptual training of tones may help shift non-tonal speakers' attention to the acoustic cues and make them focus more on the shape of the tonal contour (Kaan et al., 2008), our results showed that English speakers attended more to the offset values rather than the shape of the tonal contour during perceptual training of allotones.

Similarly, for the half-third sandhi rule, our results indicate that after training, English speakers tended to produce the turning point of the half-third sandhi tonal contour closer to native speakers in both real and wug words. However, the initial part of the tonal contour suffered from this adjustment.

Unlike Cantonese speakers, English speakers have no existing tonal categories, so they need to establish tonal categories as well as representations of tonal variants after being exposed to new L2 sounds. In addition, they need to go through the stage of learning the similarities and dissimilarities of the allotones. Although perceptual training in non-tonal speakers helped them learn the dissimilarities of allotones and in turn better differentiate the tonal variants, the training led to more improvement in the offset values and turning point instead of the shape of the tonal contour. The results indicate that the way non-tonal speakers process acoustic cues of allotones during training may result in the limitations of training effects. Non-tonal learners thus need more training to change their processing of acoustic cues and shift their attention from the offset f0 values to the shape of the entire tonal contours in processing tonal variants. When the learners have more experience with tone categories and attend more to the tonal shape, they may achieve a better sandhi tone production.

Based on the results of perceptual training on both tonal (Cantonese) and non-tonal (American) learners, we may answer our first research questions. Our results confirmed that perceptual training led to great improvement in the acquisition of the third tone sandhi rule by tonal learners. However, perceptual training only slightly modified the tonal contour of the half-third sandhi rule by tonal learners to show less rising portion and lower f0 values toward the end of the contour, which may be due to a ceiling effect as their speech production was already quite close to native speakers before training, especially in real words. After training, tonal learners produced more discernible contours of tonal variants than those produced before training. Perceptual training thus may not only help modify phonetic categories as reported in the literature, but also help tonal learners better distinguish between allotones. Therefore, our results confirmed that for tonal speakers, perception and production are associated not only for segments but for phonological rule applications. Perceptual training thus may lead to the modification of mental representations of allotones and in turn the speech production of them.

For non-tonal speakers, perceptual training helped them produce offset f0 values and turning points closer to those produced by native speakers in many cases, but they failed to modify the entire tonal contour. Due to modification of offset values and turning points, American learners did produce more discernible tonal contours of these T3 variants. Perceptual training thus led to better differentiation between T3 variants by American learners, but the training effects demonstrated limitations due to different processing of acoustic cues between American learners and native speakers.

Shea and Curtin (2006) argued that learners need to learn the dissimilarities among the variants in acquiring allophonic variants Chen et al. (2019). reported the difficulties of learners to discern the differences between allotones. After perceptual training, our results showed that the Cantonese learners improved their production of allophonic variants to be more native-like, and made allophonic variants, namely the sandhi T3 and the half-third sandhi tone, more distinct from each other. Thus perceptual training may facilitate tonal learners' establishment of allotone representations and help them learn the dissimilarities among allophonic variants.

4.2. The effects of perceptual training in improving tone sandhi rule applications in real and wug words

We include wug words to test whether learners can encode the new sandhi contour as part of the abstract tonal category instead of merely encode and store the sandhi contours lexically. Using wug words, we can examine how tone sandhi rules are applied without the influence of lexical meanings. We may exclude the possibility that learners have learned these wug words and stored sandhi contours lexically before. Also, wug tests have been used in testing if there is a phonetic bias in sandhi rule application by native Mandarin speakers (Zhang and Lai, 2010). It has been hypothesized that the half-third sandhi rule is more phonetically motivated than the third tone sandhi rule since the half-third sandhi rule involves the truncation of the rising part of the pitch contour, which is considered phonetically natural and motivated. Their results showed that speakers applied the half-third sandhi rule with greater accuracy compared to the third tone sandhi rule. We also consider phonetic bias in our design of this study, aiming to examine whether learners may improve better in half sandhi tones than the third sandhi tone. Using wug words may help us test the effect of phonetic bias on learning.

As for the second research question, perceptual training facilitated tone sandhi rule applications in both real and wug words by tonal learners for the third tone sandhi rule. Our results demonstrated that tonal learners were effective in producing the third sandhi tones in both real and wug words. The improvement in both real and wug words indicates that Cantonese learners may have encoded the new sandhi contour as part of the abstract tonal category instead of merely encode and store the sandhi contours lexically. For the half-third sandhi rule, the improvement was more significant in wug words than in real words, which may be due to the fact that f0 values in real words were already quite close to the native norm even before training, hence the improvement was smaller than that in wug words. The results indicate that for tonal learners, perceptual training may lead to more native-like allotone representations so that the speech production of sandhi tones became more accurate once the phonological environment was met.

For non-tonal learners, the speech production of the third sandhi tone in real words improved more than in wug words. In addition, American learners produced the half-third sandhi tone with better offset f0 values, but the initial contours suffered from the adjustment for both real and wug words so that the region with significant difference from native speaker did not decrease. The results thus indicate that for non-tonal learners, perceptual training may help them produce real words better than wug words. American learners may be in the process of modifying and internalizing allotone representations as well as encoding the new sandhi contour as part of the abstract tonal category in order to achieve more accurate speech production in wug words.

4.3. The effects of phonetic bias in learning allophonic variants

The third research question we explored is whether the third sandhi tone and the half-third sandhi tone could be produced equally accurately by Cantonese and English speakers. For native Mandarin speakers, it has been proposed that there is a phonetic bias in tone sandhi rule application, where the half-third sandhi rule is considered to be more natural phonetically compared to the third-tone sandhi rule (Zhang and Lai, 2010). It is proposed that the more phonetically motivated half-third sandhi rule can be realized with more accuracy. Therefore, it is hypothesized that after perceptual training, a more phonetically motivated tone sandhi rule can be better learned than a rule less phonetically motivated.

For tonal speakers, our results showed that perceptual training significantly improved the speech production of the sandhi T3 in both real and wug words. However, perceptual training only led to slight improvement in the tonal contours of the half-third sandhi tone. Contrary to the prediction that phonetically more motivated tone sandhi rule shows more improvement after training, our results indicated that the improvement was limited, probably due to a ceiling effect as Cantonese speakers already produced the half-third sandhi tone with f0 values close to native speakers.

Moreover, in terms of the significantly different regions from native speakers, American learners improved more for the third sandhi rule than the half-third sandhi rule, which is also against the prediction concerning phonetic bias. In producing the half-third sandhi tone, due to their adjustment of turning points without modification of the entire tonal contours, American learners produced initial f0 values more differently from native speakers after training. These results thus suggest that phonetically more motivated rules may not necessarily lead to more improvement after training. Factors such as perceptual processing of acoustic cues and the ceiling effect may affect the final training results.

5. Conclusions

The current study investigated the effects of perceptual training on the acquisition of two Mandarin tone sandhi rules by tonal and non-tonal speakers. First, perceptual training may improve the production of some tone sandhi rules by tonal learners, and yet non-tonal learners may need more training to help them process acoustic cues of tones. Second, perceptual training may help tonal learners establish more accurate representations of allotones and help them produce more distinct allotones in both real and wug words in general. However, non-tonal learners may be in the process of modifying allotone representations and have not yet encoded and stored the modified sandhi contours as an abstract tonal category in order to apply sandhi rules more effectively on wug words. Third, phonetic bias may not necessarily predict the success in acquiring tone sandhi rules by non-native speakers. In future studies, we may design a study to examine the improvement of perception during the training phases and the corresponding development in the speech production to examine the relationship between the improvement in perception and production Eqs. (1) and (2).

CRediT authorship contribution statement

Si Chen: Conceptualization, Methodology, Writing – original draft. **Bei Li:** Data curation, Writing – review & editing. **Yunjuan He:** Data curation, Writing – review & editing. **Shuwen Chen:** Formal analysis. **Yike Yang:** Data curation, Writing – review & editing. **Fang Zhou:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study is supported by grants from and from Department of Chinese and Bilingual Studies [88C3] [88DW] and Faculty of Humanities [1- ZVHH], [1-ZVHJ], [ZVNV] the Hong Kong Polytechnic University. We thank the editor and anonymous reviewers for their valuable comments and suggestions.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.specom.2022.02.008.

References

Antoniou, M., Wong, P.C.M., 2015. Poor phonetic perceivers are affected by cognitive load when resolving talker variability. J. Acoust. Soc. Am. 138 (2), 571–574.
Baese-Berk, M.M., Samuel, A.G., 2015. Listeners beware: speech production may be bad for learning speech sounds. J. Mem. Lang. 89, 23–36.
Bauer, R.S., Benedict, P.K., 1997. Modern Cantonese Phonology. Mouton de Gruyter, Berlin.
Berko, J., 1958. The child's learning of English morphology. Word (14), 150–177.
Best, C.T., 1995. A direct realist view of cross-language speech perception. In: Strange, W. (Ed.), Speech Perception and Linguistic Experience: Issues in Cross-Language Research. York Press, Timonium, MD, pp. 171–204.
Boersma, P. & Weenink, D. (1992–2022):Praat: doing phonetics by computer [Computer program].Version 6.1.16, retrieved 6 June 2020 from http://www.praat.org/.
Bradlow, A.R., Pisoni, D.B., Yamada, R.A., Tohkura, Y., 1997. Training Japanese listeners to identify English /r/and /l/: IV. Some effects of perceptual learning on speech production. J. Acoust. Soc. Am. 101, 2299–2310.
Bradlow, A.R., Akahane-Yamada, R., Pisoni, D.B., Tohkura, Y.I., 1999. Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. Percept. Psychophys. 61 (5), 977–985.
Chao, Y.R., 1948. Mandarin Primer. Harvard University Press, Cambridge.
Chambliss, D.F., Schutt, R.K., 2018. Making sense of the social world: Methods of investigation. Sage Publications.
Chandrasekaran, B., Krishnan, A., Gandour, J.T., 2007. Mismatch negativity to pitch contours is influenced by language experience. Brain Res. 1128, 148–156.
Chen, A., Kager, R., 2011. The perception of lexical tones and tone sandhi in L2: success or failure?. In: Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII), pp. 444–447.
Chen, M.Y., 2000. Tone Sandhi: Patterns Across Chinese Dialects, 92. Cambridge University Press.
Chen, S., He, Y., Wayland, R., Yang, Y., Li, B., Yuen, C.W., 2019. Mechanisms of tone sandhi rule application by tonal and non-tonal non-native speakers. Speech Commun. 115 (October), 67–77. https://doi.org/10.1016/j.specom.2019.10.008.
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 Commun. 88, 17–38.
Da, J. (2004). Chinese text computing: syllable frequencies with tones. Available at http://lingua.mtsu.edu/chinese-computing/phonology/syllabletone.php.
Flege, J.E., 1995. Second language speech learning: theory, findings, and problems. In: Strange, W. (Ed.), Speech Perception and Linguistic experience: Issues in Cross-Language Research. York Press, Timonium, MD, pp. 233–277.
ed. by Flege, J.E., Bohn, O.S., 2020. The revised speech learning model (SLM-r). In: Wayland, R. (Ed.), Second Language Speech Learning: Theoretical and Empirical Progres. Cambridge University Press. ed. by.
Francis, A.L., Ciocca, V., Ma, L., Fenn, K., 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. J. Phon. 36 (2), 268–294.
Gandour, J., 1983. Tone perception in Far Eastern languages. J. Phon. 11 (2), 149–175.
Gandour, J., Harshman, R., 1978. Cross-language difference in tone perception: a multidimensional scaling investigation. Lang. Speech 21, 1–33.
Hancin-Bhatt, B., 2008. Second language phonology in optimality theory. In: Hansen Edwards, J.G., Zampini, M.L. (Eds.), Phonology and second language acquisition. John Benjamins, Philadelphia.
Hanulikova, A., Dediu, D., Fang, Z., Bašnaková, J., Huettig, F., 2012. Individual differences in the acquisition of a complex L2 phonology: a training study. Lang. Learn. 62 (2), 79–109.
Hazan, V., Sennema, A., Iba, M., Faulkner, A., 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. Speech Commun. 47 (3), 360–378.
Hsieh, H., 1970. The psychological reality of tone sandhi rules in Taiwanese. Chic. Linguist. Soc. 6, 489–503.
ed. by Hsieh, H., Koerner, E.F.K., 1975. How generative is phonology?". The Transformational-Generative Paradigm and Modern Linguistic Theory. Benjamins, Amsterdam, pp. 109–144. ed. by.
Hsieh, H., 1976. On the unreality of some phonological rules. Lingua 38, 1–19.
Kaan, E., Barkley, C.M., Bao, M., Wayland, R., 2008. Thai lexical tone perception in native speakers of Thai, English and Mandarin Chinese: an event-related potentials training study. BMC Neurosci. 9 (1), 53.
Kaan, E., Wayland, R., Bao, M., Barkley, C., 2007. Effects of native language and training on lexical tone perception: an ERP study. Brain Res. 1148, 113–122.

Krishnan, A., Gandour, J.T., Bidelman, G.M., 2010. The effects of tone language experience on pitch processing in the brainstem. J. Neurolinguist. 23 (1), 81–95.
Laplace, P.S., 1820. Théorie Analytique des Probabilités. Mme Ve Courcier, Paris.
Lee, L., Nusbaum, H.C., 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. Percep. Psychophys. 53 (2), 157–165.
Lee, B., Plonsky, L., Saito, K., 2020. The effects of perception- vs. production-based pronunciation instruction. System 88, 102185. https://doi.org/10.1016/j.system.2019.102185.
Lively, S.E., Logan, J.S., Pisoni, D.B., 1993. Training Japanese listeners to identify English/r/and/l/. II: the role of phonetic environment and talker variability in learning new perceptual categories. J. Acoust. Soc. Am. 94 (3), 1242–1255.
Logan, J.S., Lively, S.E., Pisoni, D.B., 1991. Training Japanese listeners to identify English/r/and/l: a first report. J. Acoust. Soc. Am. 89 (2), 874–886.
Lu, S., Wayland, R., Kaan, E., 2015. Effects of production training and perceptual training on lexical tone perception-a behavioral and ERP study. Brain Res. 1624, 28–44.
Major, R.C., 2008. Transfer in second language phonology. Phon. Second Lang. Acquis. 36, 63–94.
Matthews, S., Yip, V., 1994. Cantonese: A comprehensive Grammar. Routledge.
McAndrews, M., 2019. Short periods of instruction improve learners' phonological categories for L2 suprasegmental features. System 82, 151–160. https://doi.org/10.1016/j.system.2019.04.007.
McCandliss, B.D., Fiez, J.A., Protopapas, A., Conway, M., McClelland, J.L., 2002. Success and failure in teaching the [r]-[l] contrast to Japanese adults: tests of a Hebbian model of plasticity and stabilization in spoken language perception. Cogn. Affect. Behav. Neurosci. 2 (2), 89–108.
Moulines, E., Laroche, J., 1995. Nonparametric techniques for pitch-scale and time-scale modification of speech. Speech Commun. 16, 175–205.
Myers, J., Tsay, J., 2003. Investigating the phonetics of Mandarin tone sandhi. Taiwan J. Linguist. 1 (1), 29–68.
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. Cogn. Neurosci. 30 (5), 491–505.
Pathi, S., Mondal, P., 2021. The mental representation of sounds in speech sound disorders. Humanit. Soc. Sci. Commun. 8 (1), 1–12.
Paulus, J.K., Dahabreh, I.J., Balk, E.M., Avendano, E.E., Lau, J., Ip, S., 2014. Opportunities and challenges in using studies without a control group in comparative effectiveness reviews. Res. Synth. Methods 5 (2), 152–161.
Peng, S.H., 2000. Lexical versus 'phonological'representations of mandarin sandhi tones. In: Broe, M.B., Pierrehumbert, J.B. (Eds.), Papers in Laboratory Phonology V: Acquisition and the Lexicon. Cambridge University Press, Cambridge, pp. 152–167.
Perrachione, T.K., Lee, J., Ha, L.Y., Wong, P.C., 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. J. Acoust. Soc. Am. 130 (1), 461–472.
Politzer-Ahles, S., Zhang, J., 2021. Evidence for the role of tone sandhi in Mandarin speech production. Journal of Chinese Linguistics [Special Issue: Studies on Tonal Aspect of Languages].
Saito, K., 2011. Examining the role of explicit phonetic instruction in native-like and comprehensible pronunciation development: an instructed SLA approach to L2 phonology. Lang. Aware. 20 (1), 45–59.
Saito, K., van Poeteren, K., 2018. The perception–production link revisited: the case of Japanese learners' English/ɹ/performance. Int. J. Appl. Linguist. 28 (1), 3–17.
Saito, Y., Saito, K., 2017. Differential effects of instruction on the development of second language comprehensibility, word stress, rhythm, and intonation: the case of inexperienced Japanese EFL learners. Lang. Teach. Res. 21 (5), 589–608. https://doi.org/10.1177/1362168816643111.
Sakai, M., Moorman, C., 2018. Can perceptual training improve the production of second language phonemes? A meta-analytic review of 25 years of perceptual training research. Appl. Psycholinguist. 39 (1), 187–224.
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.
Shih, C., 1997. Mandarin third tone sandhi and prosodic structure. Linguist. Model. 20, 81–124.
Shinohara, Y., Iverson, P., 2018. High variability identification and discrimination training for Japanese speakers learning English /r/–/l/. J. Phon. 66, 242–251. https://doi.org/10.1016/j.wocn.2017.11.002.
Shoemaker, E., 2014. The exploitation of subphonemic acoustic detail in L2 speech segmentation. Stud. Second Lang. Acquis. 36 (4), 709–731.
Song, J.H., Skoe, E., Wong, P.C., Kraus, N., 2008. Plasticity in the adult human auditory brainstem following short-term linguistic training. J. Cogn. Neurosci. 20 (10), 1892–1902.
Spinu, L.E., Hwang, J., Lohmann, R., 2018. Is there a bilingual advantage in phonetic and phonological acquisition? The initial learning of word-final coronal stop realization in a novel accent of English. Int. J. Biling. 22 (3), 350–370. https://doi.org/10.1177/1367006916681080.
Stemberger, J.P., Bernhardt, B.M., 2018. Tap and trill clusters in typical and protracted phonological development: challenging segments in complex phonological environments. Introduction to the special issue. Clin. Linguist. Phon. 32 (5-6), 411–423.
Wang, Y., Spence, M.M., Jongman, A., Sereno, J.A., 1999. Training American listeners to perceive mandarin tone. J. Acoust. Soc. Am. 106, 3649–3658.
Wang, Y., Jongman, A., Sereno, J.A., 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. J. Acoust. Soc. Am. 113, 1033–1043.
Wayland, R., Guion, S., 2004. Training native English and native Chinese speakers to perceive Thai tones. Lang. Learn. 54, 681–712.
Wayland, R., Li, B., 2008. Effects of two training procedures in cross-language perception of tones. J. Phon. 36 (2), 250–267.
Whalen, D.H., Levitt, A.G., 1995. The universality of intrinsic F0 of vowels. J. Phon. 23, 349–366.
Xu, C.X., Xu, Y., 2003. Effects of consonant aspiration on Mandarin tones. J. Int. Phon. Assoc. 33, 165–181.
Xu, Y., 2013. ProsodyPro — A tool for large-scale systematic prosody analysis. In: Tools and Resources for the Analysis of Speech Prosody (TRASP 2013), pp. 7–10 (Aix-enProvence, France 2013).
Xu, Y., Prom-On, S., 2014. Toward invariant functional representations of variable surface fundamental frequency contours: synthesizing speech melody via model-based stochastic learning. Speech Commun. 57, 181–208.
Zhang, C., Peng, G., 2013. Productivity of Mandarin third tone sandhi: a wug test. In: Peng, G., Shi, F. (Eds.), Eastward Flows the Great River: Festschrift in Honor of Prof. William S-Y. Wang on his 80th Birthday. City University of Hong Kong Press, Hong Kong, pp. 256–282.
Zhang, C., Xia, Q., Peng, G., 2015. Mandarin third tone sandhi requires more effortful phonological encoding in speech production: evidence from an ERP study. J. Neurolinguist. 33, 149–162.
Zhang, H., 2017. The effect of theoretical assumptions on pedagogical methods: a case study of second language Chinese tones. Int. J. Appl. Linguist. 27 (2), 363–382.
Zhang, J., 2007. A directional asymmetry in Chinese tone sandhi systems. J. East Asian Linguist. 16 https://doi.org/10.1007/s10831-007-9016-2.
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.

[This paper was published at Speech Communication 139 (2022) 10–21]