JASA ARTICLE

The effects of native prosodic system and segmental context on Cantonese tone perception by Mandarin and Japanese listeners

Min Zhu,[1] Xiaoxiang Chen,[1,a)] and Yuxiao Yang[2,b)]

[1]*College of Foreign Languages, Hunan University, No. 2 Lushan South Road, Yuelu District, Changsha, 410082, China*

[2]*Foreign Studies College, Hunan Normal University, 36 Lushan Road, Yuelu District, Changsha, 410081, China*

ABSTRACT:
This study explores the effects of native prosodic system and segmental context on the perception of Cantonese tones by Mandarin and Japanese listeners. In Experiment 1, 13 Mandarin and 13 Japanese subjects took part in a two-alternative forced-choice discrimination test of Cantonese tones in different segmental contexts (familiar vs unfamiliar). In Experiment 2, 20 Mandarin listeners participated in a perceptual assimilation task that examined the cross-language perceptual similarity between Mandarin and Cantonese tones. Results showed that Mandarin listeners were comparable to Japanese counterparts in discriminability, but the former attended more to pitch contour differences while the latter were more sensitive to pitch height. Moreover, the effect of segmental context was significant exclusively in the Mandarin group, whereas the Japanese group performed stably across syllables in discriminating Cantonese tones. It seemed that unfamiliar context rendered lower perceptual similarity, which further hindered corresponding discrimination by the Mandarin group. In addition, segmental effects were mainly observed in the assimilation patterns of category goodness or uncategorized-categorized. These findings suggested that non-native tone perception could be modulated by listeners' native prosodic structures in a finer way.

© *2021 Acoustical Society of America*. https://doi.org/10.1121/10.0005274

(Received 8 October 2020; revised 19 May 2021; accepted 23 May 2021; published online 15 June 2021)

[Editor: Jody Kreiman] Pages: 4214–4227

I. INTRODUCTION

Tones are employed in more than 60% of the existing languages in the world to distinguish lexical meanings, in particular for most Eastern Asian languages (Yip, 2002). The acquisition of non-native lexical tones could be challenging for second language (L2) learners (Francis *et al.*, 2008; So and Best, 2010, 2014; Tsukada and Kondo, 2019), and it could be even harder than that of segments (Wong and Perrachione, 2007). One of the factors causing this difficulty could stem from the phonological system of the learners' first language (L1). Listeners as early as in infancy could attune their perception to the speech sounds existing in their native languages and ignore the speech contrasts absent from their L1s, forming differential perceptual reorganization for native and non-native speech sounds (Werker and Tees, 1984).

In addition to the influence of native language background, segmental context has been found to affect the perception of tones. Since lexical tones were embedded in monosyllables, it was argued that tone language listeners would integrate tone and syllable perception while non-tone language listeners who lack lexical pitch variations would not (Repp and Lin, 1990). For example, using a speeded classification paradigm, Tong *et al.* (2008) demonstrated that Mandarin listeners showed an asymmetry in processing segmental and suprasegmental dimensions in that tone perception was disturbed more by segmental dimensions than the reverse. In addition, Tong *et al.* (2014) found that syllables exerted greater effects on tonal contrasts of height than those of contour. The current study aims to investigate how these two factors affect the perception of Cantonese tones by Mandarin and Japanese listeners.

In the case of lexical tones, pitch, characterized by height and contour, plays a primary role along with other secondary acoustic correlates such as intensity and duration (Howie, 1976; Moore and Jongman, 1997). Although non-tone languages also employ pitch variations in natural speech, these convey paralinguistic information at the sentence level rather than at the word level as in the case of tone languages (Best, 2019). In the meantime, tone languages also differ from one another. Aside from canonical tone languages like Mandarin and Cantonese, there is another variety of tone languages known as pitch accent languages like Japanese, Swedish, and Norwegian that also have restricted pitch variations at the word level (Yip, 2002). Unlike Mandarin or Cantonese (referred to hereafter as canonical tone languages) where pitch variations occur on individual syllables, pitch accent languages exhibit relative pitch differences between successive syllables. Tone languages can also differ from each other in the numbers of their tones, which has been assumed to affect non-native tone perception in previous studies. Specifically, Lee *et al.* (1996) argued that L1 tonal experience could assist learners to perceive a novel tone language only when L1 is more

[a)]Electronic mail: laxiaoxiangchen@sina.com
[b)]ORCID: 0000-0003-2657-8436.

complicated than the target language through observing an asymmetric tone experience in Mandarin and Cantonese.

In addition, defined by pitch trajectories, lexical tones could be subdivided into level (or static) and contour (or dynamic) tones (Abramson, 1978). For instance, Mandarin has one level tone and three contour tones, while Cantonese has three level tones and three contour tones. It has been argued that L1 pitch realizations could influence the listeners' perceptual weight for different acoustic cues (Francis *et al.*, 2008; Gandour, 1983; Qin and Mok, 2015). Gandour (1983) reported that non-tone language listeners pay more attention to pitch height, while tone language listeners are more influenced by pitch contour in the perception of non-native tones. Furthermore, the author also found that even within tone languages, the weight of these two dimensions could be different in that Mandarin listeners assigned more weight to pitch contour than pitch height, whereas Cantonese listeners relied on both height and contour. Aside from pitch correlates, phonation type (i.e., creaky or breathy) also plays a crucial role in some tone languages, such as Burmese and Vietnamese (Tsukada and Kondo, 2019; Yip, 2002). Brunelle (2009) indicated that creakiness is an important feature of hỏi and ngã tones of Northern Vietnamese. However, the effect of phonation type will not be considered here since voice quality is not a consistent cue in either Cantonese or Mandarin.

Compared with canonical tone languages, there is a paucity of empirical research on the perceptual sensitivity of listeners whose L1 is a pitch accent language. So and Best (2010) studied Cantonese, Japanese and English listeners' discrimination of Mandarin tone pairs. The Japanese group performed systematically better than the Cantonese group for Tone (T) 1-T4 pair. However, what was left unaccounted for was which phonetic cues are crucial for Japanese-speaking listeners. In addition, Tsukada *et al.* (2016) compared Mandarin tonal perception by Japanese listeners with and without learning experience. Results indicated that Mandarin experience assisted learners to outperform naive counterparts for T2-T3, T1-T2 distinctions, and to be more immune to speakers and phonemic variations. Given that the limited literature on the Japanese group was solely based on Mandarin tone perception, which lacks level contrasts, it remains unclear as to how a pitch accent system would influence Japanese listeners' tone perception in the light of the relative weight between height and contour.

A. Theoretical framework of perceptual assimilation model (PAM)

Certain theoretical models have been proposed to predict non-native perceptual difficulties. For instance, PAM (Best, 1995) makes predictions regarding listeners' performance in discriminating non-native sounds based on how they are assimilated to the listeners' native phonological system. Six assimilation patterns were proposed: two category (TC), single category (SC), category goodness (CG), uncategorized-categorized (UC), uncategorized-uncategorized (UU), and non-assimilable (NA). When two non-native sounds are assimilated to two distinct L1 sounds (TC), the discriminability could be optimal, whereas the discriminability could be very poor when the two non-native sounds are equally mapped onto one single L1 sound (SC); when two non-native sounds are mapped onto a single native category but with different degrees of similarity (CG), the discrimination could be moderate to good. The discriminability of these three types of assimilation patterns generally follows the sequence TC > CG > SC. On the other hand, if at least one sound of the non-native phonemic contrast could not be assimilated to certain native categories, but still falls into the phonological space of the L1 (in the cases of UC, UU), listeners would encounter varying degrees of discrimination difficulty depending on the set of L1 categories (Best, 1995; Best and Tyler, 2007). UC and UU types could be further classified as non-overlapping (UC-n; UU-n), partially overlapping (UC-p; UU-p) and completely overlapping (UC-c; UU-c) based on the perceived overlap of L1 categories (Faris *et al.*, 2018; So and Best, 2014). Faris *et al.* (2018) assessed whether naive English listeners' discriminability for Danish monophthongs could be predicted *via* assimilation overlaps, and results revealed that UU-n (/e/-/o/) contrasts were more accurately discriminated than UU-p (/oː/-/uː/) contrasts. Accordingly, UC-n should be discriminated more accurately than UC-p, which in turn could be better than UC-c (Faris *et al.*, 2016, 2018).

PAM has been widely employed to explain the perception of non-native phonemic contrasts based on the mapping relationships of the non-native contrasts onto listeners' L1 categories (Best and Strange, 1992; Best *et al.*, 2001; Hao, 2012). With respect to the segmental level, Best and Strange (1992) argued that the difficulty in discriminating the English /r/-/l/ contrast observed for Japanese speakers could be triggered by the equal assimilation of these two English sounds to the single Japanese category /ɾ/, supporting the case of SC. On the other hand, the near-ceiling performance achieved by English learners in discriminating the Zulu contrast /ɬ/ and /ɮ/ could be due to TC assimilation (Best *et al.*, 2001). Aside from segmental perception, the tenets of PAM were further extended to the realm of the perception of non-native suprasegmental features as PAM-s by So and Best (2014). They explored English and French listeners' discrimination of Mandarin tones, in which the better performance observed in T3-T4 was due to TC assimilation of this pair to French intonation categories. On the other hand, the failure encountered by Cantonese listeners in discriminating Mandarin T1-T4 could be attributed to SC assimilation (Hao, 2012; So and Best, 2010). In general, PAM has provided plausible explanations for non-native perception based on assimilation patterns. However, compared with the segmental level, there is still a lack of empirical research regarding the application of PAM at the suprasegmental level.

Furthermore, regarding the assessment of cross-linguistic phonetic similarity between two tonal categories, numerous studies tended to compare solely the "five-degree" pitch values, without considering the actual perceived similarity between tones from different languages. “Five-degree” values, which only reflect the starting and ending points of pitch contours, could not depict a full picture of pitch processing in the brain. In Hao (2012), Mandarin T2 (35) was perceived as more similar to Cantonese T5 (23) than Cantonese T2 (25) by listeners despite the latter being more similar to Mandarin T2 in terms of the “five-degree” scale. Hence, it would be more convincing to determine perceptual similarities based on a perceptual assimilation task (Strange and Shafer, 2008; Yang *et al.*, 2020). This being the case, a perceptual assimilation task was carried out in this study to assess the perceptual similarity between Mandarin and Cantonese tone systems. Before introducing our study in detail, previous studies concerning the effects of native prosodic system and segmental context will be systematically reviewed below to elaborate on the motivation of the present study.

B. Native prosodic effect on non-native tone perception

A large number of studies have examined the perception of non-native tone contrasts with respect to the influence of speakers’ L1 prosodic system (Burnham *et al.*, 2015; Francis *et al.*, 2008; Hao, 2012; Lee *et al.*, 1996; Qin and Mok, 2015; So and Best, 2010, 2014; Tsukada and Kondo, 2019; Wayland and Guion, 2004; Wang, 2013). It has been widely reported that L2 tone contrasts absent in L1 could be difficult for non-native listeners to discriminate and acquire. A well-documented instance is tone perception by English listeners who displayed much difficulty in discriminating contour tones (Burnham *et al.*, 2015; Lee *et al.*, 1996). Meanwhile, perceiving non-native tones could also be challenging for tone language listeners (Francis *et al.*, 2008; Hao, 2012; Qin and Mok, 2015; So and Best, 2010). For instance, Cantonese level tone contrasts caused much difficulty for Mandarin listeners since they lack level contrasts in their phonological system (Francis *et al.*, 2008; Qin and Mok, 2015).

Generally speaking, the effect of L1 tone experience on the perception of non-native tones has been frequently discussed, yet the results are still mixed. Some research indicated that listeners with tone language backgrounds could perform better than those speaking non-tone languages in perceiving non-native tones (Lee *et al.*, 1996; Qin and Mok, 2015; Wayland and Guion, 2004). Wayland and Guion (2004) demonstrated a positive transfer of L1 tonal experience to L2 tone perception by observing higher accuracy of Mandarin listeners over English peers in distinguishing mid and low tonal contrasts in Thai. However, contradictory results have also suggested that the presence of lexical tones in the native prosodic system does not necessarily assist (Francis *et al.*, 2008; Hao, 2012), and can even hinder non-native tone perception (Tsukada and Kondo, 2019; Wang, 2013). In a cross-language study, Wang (2013) examined Mandarin tone perception by three groups of novices from Hmong, Japanese, and English backgrounds, and found that listeners of Hmong, a complex tone language, performed the worst. Therefore, it might be insufficient to conclude an effect of L1 based solely on whether the language is tonal or not.

Apart from tonal versus non-tonal comparison, Lee *et al.* (1996) reported the role of tonal complexity in non-native tone perception. The study explored the perception of Cantonese and Mandarin tones by Mandarin, Cantonese, and English listeners. Results showed that Cantonese listeners outperformed English listeners in perceiving Mandarin tones, yet the same superiority was not observed from Mandarin listeners in perceiving Cantonese tones compared with English peers. Therefore, Lee *et al.* (1996) claimed that tone language experience in L1 could be positively transferred to L2 only when L1 is more complex than L2. However, in the perception of Mandarin tones, Hmong listeners who have seven tones in L1 appeared to perform less accurately than Japanese and English counterparts (Wang, 2013), which contradicts the viewpoint on tonal complexity. Together, the inconsistent findings from the previous studies might imply that neither typology (tone vs non-tone) nor L1 tonal complexity can make precise predictions on listeners’ performance in non-native tone perception, and an alternative account needs to be explored.

It has been extensively reported that native language experience could shape listeners’ perceptual weight for specific features, which in turn would modulate the unfamiliar language perception (Francis *et al.*, 2008; Gandour, 1983; Yazawa *et al.*, 2020). At the segmental level, Yazawa *et al.* (2020) found that unlike native English speakers who used spectra as a primary cue, Japanese learners of English relied heavily on temporal cues to distinguish high front vowels /iː/ and /ɪ/, which could be triggered by the quantity contrasts in Japanese phonemes. Turning to the suprasegmental level, listeners from different language backgrounds attend to different cues, depending on prosodic features in their L1s (Francis *et al.*, 2008; Gandour, 1983). Tong *et al.* (2014) claimed that Cantonese listeners attended to *F*0 onset for the perception of height tones, and *F*0 direction for the perception of contour tones. Using synthesized tones, Li *et al.* (2016) confirmed that both Thai and Vietnamese listeners were more sensitive to the pitch height dimension due to L1 phonological systems.

The cue-weighting view might account for the findings of Lee *et al.* (1996) reviewed above, in that Cantonese contrasts both level and contour tones in its tonal system, whereas English intonation varies mainly in pitch height. Therefore, Cantonese listeners outperformed English counterparts who lack pitch contour sensitivity in processing Mandarin tones, because the former could directly transfer an L1 perceptual cue (pitch contour) to the Mandarin perception. However, the reverse did not hold true since both Mandarin and English listeners could employ only one dimension (pitch contour or pitch height) in perceiving Cantonese tones. Thus, predicting non-native tone perception from the perspective of L1 pitch height or pitch contour might be more compelling. Another piece of evidence comes from Burmese listeners’ failure in perceiving

Mandarin tones (Tsukada and Kondo, 2019). Although Burmese listeners have tonal experience in general, they tend to employ phonation rather than pitch to convey lexical meanings.

In sum, native experience regarding perceptual cues rather than the presence of L1 tone experience appears to be a better predictor of listeners' perception of novel tones. This proposal was made explicitly by Francis *et al.* (2008), who conducted a training study of Cantonese tone identification by Mandarin and English native speakers. Both groups exhibited comparable performances before and after training, yet differed in terms of specific tonal confusions induced by native perceptual cues. However, the evidence from Francis *et al.* (2008) could be insufficient, since the pitch height variation in English realized as intonation is also applied in Mandarin signaling post-lexical information, e.g., interrogative vs declarative mood (Liu and Xu, 2005), which demonstrated that the employment of pitch height is not "English exclusive" in the context of Francis *et al.* (2008). Furthermore, the pitch height variation in English occurs at the sentence level, which could be intrinsically different from the pitch variations at the lexical level. Therefore, the investigation of Japanese and Mandarin in the study could be more compatible with the issue concerning the effect of L1 perceptual cues on non-native tone perception since both languages employ pitch variations on the lexical level yet with different cue reliance: Japanese, pitch height; Mandarin, pitch contour.

C. Segmental effect on non-native tone perception

In addition to the influence of L1 prosodic system, segmental context has been found to affect tone perception as well (Lee *et al.*, 1996; Repp and Lin, 1990; Tong *et al.*, 2008; Tong *et al.*, 2014). It has been proposed that tone language listeners rather than non-tone language listeners process segmental information and tones in an integral way (Lee *et al.*, 1996; Wayland and Guion, 2004). For instance, Lee *et al.* (1996) found an important role of lexical information in tone perception for tone language listeners through observing higher accuracy achieved on real words over pseudowords. Furthermore, Repp and Lin (1990) and Tong *et al.* (2008) also found dependencies between segmental and suprasegmental features in Mandarin speakers' tone processing. Tong *et al.* (2014) examined the perception of Cantonese tones embedded in various phonetic contexts by native Cantonese children. It was found that children performed remarkably differently across /ji/ and /fu/, which was indicative of interaction between syllables and lexical tones. Moreover, there was a decreasing level of accuracy from the tones carried by the same rime-different syllable onset to different rime-different syllable onset, which further corroborated the claim that tone perception could be context dependent. Despite evidence of segmental effect on tone perception by native speakers, little has been done on non-native speakers. Thus, it is necessary to explore whether Japanese listeners would be more similar to canonical tone language listeners, perceiving syllables and tones integrally, or would be more similar to non-tone language listeners, perceiving them independently.

D. The present study

This study attempts to explore the effects of the native prosodic system and segmental context on Mandarin and Japanese naive listeners' perception of Cantonese tones. The aims of the study are threefold: First, it aims to find out how L1 canonical tone or pitch accent system affects Mandarin and Japanese listeners' perception of Cantonese tones with regard to discriminability and perceptual cue weighting. Second, the study aims to explore how segmental context affects Mandarin and Japanese listeners' perception of Cantonese tones. Third, the study would like to investigate how native and non-native perceptual similarity modulates listeners' discriminability based on PAM-s.

Cantonese is a Chinese dialect mainly spoken in Hong Kong, Macau, and Guangdong Province in Southern China, which has a complex prosodic system with six tones (Bauer and Benedict, 1997). The three level tones [T1 (55), T3 (33), T6 (22)], two rising tones [T2 (25), T5 (23)], and one falling tone [T4 (21)] are unevenly distributed in the acoustic space, with five of them crowding into the lower part of the space and T1 being at the top of the space (Peng, 2006). The $F0$ pattern of Cantonese tones is depicted in Fig. 1. The balanced tone types of Cantonese (three level tones, three contour tones) provide an optimal window to probe into listeners' relative cue weighting. Furthermore, the rich tonal pairs (15 combinations) provide enough opportunities to detect perceptual discrepancies.

1. Mandarin and Japanese prosodic systems

Mandarin has four lexical tones, with one level tone and three contour tones (Chao, 1968). Each lexical tone is carried by a monosyllable and is used to differentiate lexical meanings such as T1 mā (55): 妈 "mother"; T2 má (35): 麻 "hemp"; T3 mǎ (214): 马 "horse"; T4 mà (51): 骂 "to scold" (see Fig. 1). It is worth mentioning that T3, apart from citation forms, is usually produced as a low falling tone in different phonological contexts. Four lexical tones are distributed evenly in the acoustic space and are distinct from each other in $F0$ contour (Peng *et al.*, 2012).

Japanese is a pitch accent language, which can be considered a subtype of tone languages (Yip, 2002). It relies on the position of the accented mora (H) to differentiate lexical meanings. Taking "あめ/ame/" as an example, if the accent occurs on the first mora, it means "rain"; if the accent occurs on the second mora, it means "malt." However, Japanese uses pitch variations in a very restrictive way, over two timing units (morae) rather than one single syllable. Moreover, Japanese pitch accents are sparsely distributed or even absent on some words and dialects in contrast to the abundant tonal employment in Mandarin. Accordingly, pitch variations in Japanese over two morae can be realized as high-high (HH), high-low (HL), and low-high (LH), taking "fuu" as an example (see Fig. 1).

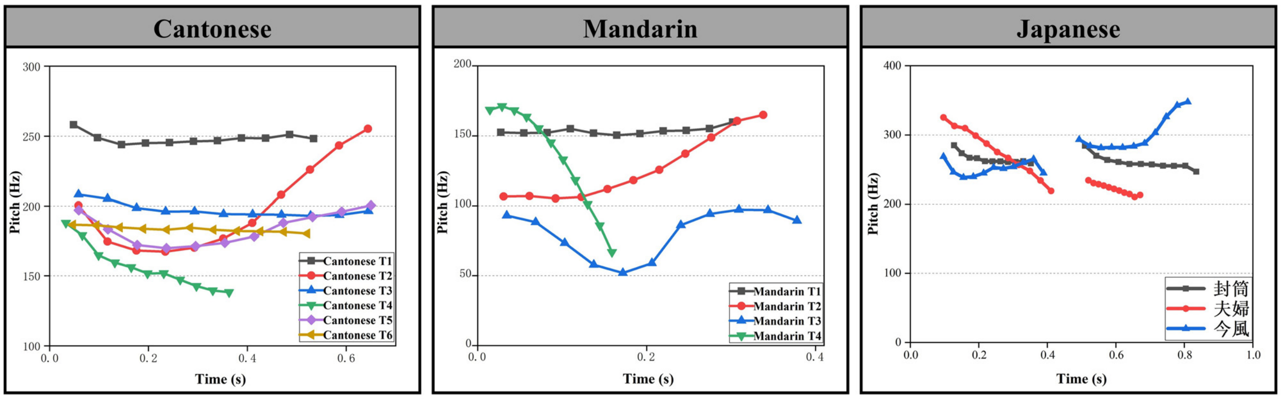

FIG. 1. (Color online) $F0$ patterns of Cantonese tones (left panel), Mandarin tones (mid panel), and Japanese pitch accents (right panel). The Cantonese tones were carried by the syllable /ji/; the Mandarin tones were carried by the syllable /ma/; the Japanese pitch accents were carried by the word /fuu/ adapted from So (2010).

2. *Research questions and hypotheses of the study*

The present study addresses the following three specific research questions. Two experiments were designed; Experiment 1 (a discrimination test) aimed to tackle the first two questions; Experiment 2 (a perceptual assimilation task), the third question.

1. *How does canonical tone and pitch accent L1s affect the perception of Cantonese tones by Mandarin and Japanese learners?*
   Drawing upon previous findings, it was hypothesized that Mandarin and Japanese listeners exhibit different patterns in discriminating Cantonese tones in that the former attend more to pitch contour, while the latter rely more on pitch height.
2. *How do familiar and unfamiliar segmental contexts affect the perception of Cantonese tones by Mandarin and Japanese listeners?*
   Based on previous findings, we hypothesized that the effect of segmental context on Cantonese tone perception is language specific, so that Mandarin listeners process syllables and tones integrally; Japanese listeners, on the other hand, perceive syllables and tones integrally if patterned with canonical tone language listeners, otherwise they perceive them more independently as English listeners do.
3. *How does cross language perceptual similarity shape the discrimination of Cantonese tone contrasts?*
   It was hypothesized that cross language perceptual similarity of tones is susceptible to segmental contexts, and further affects corresponding discrimination as posited by PAM-s.

II. EXPERIMENT 1

A. Method

1. *Participants*

Thirteen native Mandarin speakers (NM) (six male and seven female; average age = 22.2 years, standard deviation, SD = 1.23), 13 native Japanese speakers (NJ) (six male and seven female; average age = 20.7 years, SD = 1.32) and 15 native Cantonese speakers (NC) as the control group (eight male and seven female; average age = 20.1 years, SD = 1.29) were recruited in the experiment. All of them were undergraduate and graduate students of Hunan University and Hunan Normal University in Changsha, China. None of the participants had been exposed to Cantonese or to formal musical training outside the classroom before.

All NM were born and grew up in Northern China, speaking standard Mandarin. NJ were exchange students in Changsha, residing in China for two to six months at the time of testing. Before coming to China, they had never been exposed to canonical tone languages before. All NC were natives of Guangzhou and Foshan cities, Guangdong Province. Moreover, their Cantonese proficiency was verified by four native Cantonese speakers from their reading *The North Wind and the Sun* (IPA, 1999).

All participants were confirmed as having no speaking and hearing disorders *via* a pure-tone hearing screening (250–8000 Hz at 25 dB hearing level, HL). Informed consent was signed by each participant in compliance with a protocol approved by Human Research Ethics Committee of Hunan University, and they were rewarded monetarily for their participation.

2. *Materials*

Monosyllabic words were used. Two target syllables /ji/ and /ts^{h}ɐm/ together with two fillers /fu/ and /jɐu/ were embedded in a carrier sentence: ŋɔ kɔŋ x̲ (“I say x̲”) with six tones (So and Best, 2010). Each sentence was recorded five times by two native Cantonese speakers (one female and one male, both 20 years old) from Guangzhou in a sound-treated room, yielding a total of 240 sentences (4 syllables × 6 tones × 5 repetitions × 2 speakers).

Syllables /ji/ and /ts^{h}ɐm/ were used for two reasons. First, they could be affixed to any of the six tones to form real words in Cantonese.[1] Furthermore the syllable /ji/ has a counterpart in both Mandarin and Japanese, while /ts^{h}ɐm/ is new for both groups, so that the effect of familiar vs unfamiliar segmental context could be examined.

All the recordings were carried out individually using a microphone (Shure Beta 58a, Niles, IL) with an external sound card (Avid Mbox 3, Burlington, MA) at 44.1 kHz sampling rate and 16-bit resolution. The recorded words are listed in Table I. Before recording, the two speakers were asked to read an article in Cantonese for five minutes to activate their Cantonese speaking mode. Then they read carrier sentences presented randomly on the computer screen at a natural speed.

All target words were extracted using Praat (Boersma and Weenink, 2019) for the inspection of spectrograms and waveforms. Two tokens with similar duration and high quality ($F0$ curve clarity) were selected for each word to constitute the final test materials. To control duration and intensity, 96 stimuli (2 tokens × 2 speakers × 4 syllables × 6 tones) were normalized to 75 dB intensity and 600 ms duration to ensure that naive listeners could hear the stimuli clearly. All stimuli were verified as correct by the four native Cantonese speakers mentioned above.

3. Procedures

An AX (two-alternative) forced-choice discrimination test was performed by three groups of listeners independently using a laptop and a head-mounted microphone *via* Experiment MFC 7 in Praat. The participants were told that they would hear pairs of sounds from an unfamiliar language and would not receive any feedback. The whole experiment lasted about 30 min, including short breaks between blocks.

The test composed of 288 trials was divided into four blocks by syllable and speaker. Each block was made up of 36 tonal pairs (pairwise combinations with six tones) repeated twice, thus constituting totally 60 different pairs and 12 same pairs per block.[2] The presentation order of the syllable and the speaker was counterbalanced across participants. Within each trial, listeners heard two words consecutively with an inter-stimulus interval (ISI) of 500 ms and then were asked to determine whether the two sounds were the same or not by clicking the box representing "same" or "different" on the screen. Instructions were written in the participants' native languages. It is worth mentioning that the two sounds used in the "same" pairs in speech stimuli were not totally identical in acoustics, but two tokens of one word, so listeners had to make a decision based on "words" rather than "sounds." Once participants had made their choice, the next trial would appear 500 ms later automatically. Before the formal test, a familiarization session using fillers was completed by each participant.

Since the first aim of the current study was to test how different prosodic systems affect Mandarin and Japanese listeners' cue weighting in discriminating non-native tones, the 15 Cantonese tones were divided into two types: contrast by height (T1-T3, T1-T6, T2-T5, T3-T6) and contrast by contour (T1-T2, T1-T4, T1-T5, T2-T3, T2-T4, T2-T6, T3-T4, T3-T5, T4-T5, T4-T6, T5-T6). The discriminability of Mandarin and Japanese groups was assessed *via* sensitivity (hit rate: correct responses for different pairs) following Qin and Mok (2015).

B. Results of Experiment 1

In the discrimination task, only a few errors of the same pairs were found for each group. Besides, this study paid attention to the results of different pairs. Therefore, only results for the different pairs are reported below, which were evaluated by hit rate.

1. Overall performance in discrimination

In the discrimination test, the mean hit rates were 0.77, 0.76, 0.92 for NM, NJ, and NC, respectively. For the syllable /ji/, the mean hit rates of NM, NJ, and NC groups were 0.81, 0.77, 0.94, while they were 0.72, 0.76, and 0.91 for the syllable /ts^{h}ɐm/. Table II displays the mean hit rates of the three groups for each contrast type as a function of different syllables. For statistical analyses, Generalized Linear Mixed Effect models (GLMM) from the R package *lme4* were computed. "Response" was calculated as the dependent variable with correct responses coded as "1," incorrect responses coded as "0." "Subject" with "Syllable" and "Item" with "Group" were computed as random slopes after model comparisons. "Group (NC vs NJ vs NC)," "Syllable (/ji/ vs /ts^{h}ɐm/)," and "Contrast type (height vs contour)" were computed as fixed effects. Main and interaction effects were calculated *via* likelihood ratio tests using the package *car*. *Post hoc* Tukey tests were realized by the package *multcomp*, simple main effects were observed by the package *emmeans* with the adjustment of Tukey. The GLMMs on the fixed effects showed that the hit rates for the syllable /ji/ were significantly higher than those for the syllable /ts^{h}ɐm/ [$\beta = 0.52$, standard error, SE $= 0.07$, $z = 7.21$, $p < 0.001$], and they were significantly higher in contour contrasts than height contrasts [$\beta = 2.54$, SE $= 0.63$, $z = 4.03$, $p < 0.001$]. Moreover, the *post hoc* test on "Group" showed no significant difference between NM and NJ in hit rates

TABLE I. The wordlist of target words /ji/ and /ts^{h}ɐm/ carrying six tones (including fillers).

	T1 (55)	T2 (25)	T3 (33)	T4 (21)	T5 (23)	T6 (22)
/ji/	醫 Doctor	椅 Chair	意 Meaning	兒 Son	耳 Ear	二 Two
/ts^{h}ɐm/	侵 Invasion	寢 Dormitory	摻 Mingle	尋 Find	蕈Mushroom	譖 Slander
/fu/	夫 Husband	斧 Axe	富 Rich	符 Symbol	婦 Woman	父 Father
/jɐu/	休 Rest	柚Grapefruit	幼 Young	油 Oil	友 Friend	右 Right

TABLE II. Mean hit rates of the three groups for each contrast type across syllables in discriminating Cantonese tones.

Group	Syllable	Contrast type	Hit rate.mean	Hit rate.sd
Cantonese	/ts^{h}ɐm/	Contour	0.98	0.13
Cantonese	/ji/	Contour	0.99	0.11
Cantonese	/ts^{h}ɐm/	Height	0.69	0.46
Cantonese	/ji/	Height	0.8	0.4
Japanese	/ts^{h}ɐm/	Contour	0.82	0.38
Japanese	/ji/	Contour	0.81	0.39
Japanese	/ts^{h}ɐm/	Height	0.56	0.5
Japanese	/ji/	Height	0.65	0.48
Mandarin	/ts^{h}ɐm/	Contour	0.89	0.31
Mandarin	/ji/	Contour	0.95	0.22
Mandarin	/ts^{h}ɐm/	Height	0.26	0.44
Mandarin	/ji/	Height	0.43	0.5

[$\beta = 0.47$, SE $= 0.27$, $z = 1.78$, $p = 0.18$], while both groups were significantly lower than NC ($ps < 0.001$). The GLMM with all three fixed effects revealed significant main effects of "Group" [χ^2 (2) $= 121.54$, $p < 0.001$], "Syllable" [χ^2 (1) $= 51.5$, $p < 0.001$] and "Contrast type" [χ^2 (1) $= 33.95$, $p < 0.001$]. Moreover, significant interactions were observed between "Group" and "Syllable" [χ^2 (2) $= 24.82$, $p < 0.001$], "Group" and "Contrast type" [χ^2 (2) $= 75.43$, $p < 0.001$], and "Syllable" and "Contrast type" [χ^2 (1) $= 8.87$, $p < 0.01$]. There was no significant three-way interaction [χ^2 (2) $= 2.09$, $p = 0.35$]. Simple main effect tests on "Group" × "Syllable" interaction revealed that NC significantly outperformed NM and NJ for both syllables /ji/ and /ts^{h}ɐm/ ($ps < 0.001$), but the two experimental groups were not significantly different for the syllable /ji/ [$\beta = 0.21$, SE $= 0.18$, $z = 1.19$, $p = 0.14$], whereas NJ outperformed NM for the syllable /ts^{h}ɐm/ [$\beta = 0.53$, SE $= 0.18$, $z = 3$, $p < 0.05$].

2. Performance for specific contrast type

The discrimination performances of the three groups for the two contrast types are presented in Fig. 2. Simple main effect tests on "Group" × "Contrast type" interaction revealed a significantly better performance of NC than the two experimental groups for both contrast types ($ps < 0.01$). Additionally, in the comparisons between the two experimental groups, results showed that NM performed significantly better than NJ for contour contrasts [$\beta = 1.2$, SE $= 0.181$, $z = 6.642$, $p < 0.001$], whereas NJ significantly outperformed NM for height contrasts [$\beta = 1.51$, SE $= 0.261$, $z = 5.808$, $p < 0.001$]. Moreover, NM, analogous to NC, discriminated contour contrasts more accurately than those differing by height ($ps < 0.001$). NJ, on the other hand, exhibited a comparable performance for both types of tone pairs [$\beta = 0.24$, SE $= 0.11$, $z = 2.28$, $p = 0.14$].

Besides discrepancies, the three groups also shared universal patterns in the perception of Cantonese tones. Simple main effect tests for "Contrast type" × "Syllable" interaction showed that all three groups performed worse for height pairs than contour pairs regardless of syllables ($ps < 0.001$). It was possibly due to the intrinsic high degree of acoustic similarity (Francis *et al.*, 2008; Qin and Mok, 2015). Height contrasts were confirmed as less dynamic and categorical than contour ones (Abramson, 1978; Francis *et al.*, 2003). Besides, T2-T5, T3-T6 have been found prone to merge among younger generations (Mok *et al.*, 2013). It is likely that some Cantonese speakers might inherit the feature after long-term exposure of this merger, which contributes to high ambiguity among these pairs in the absence of contexts.

3. Performance across syllables

As mentioned above, there was a significant interaction between "Group" and "Syllable." For the purpose of further examining the effect of segmental context on tonal discrimination, simple main effect tests were performed for each group's performance across the two syllables. Results revealed that the effect of "Syllable" was found to be significant only for NC and NM [NC: $\beta = 0.651$, SE $= 0.189$,

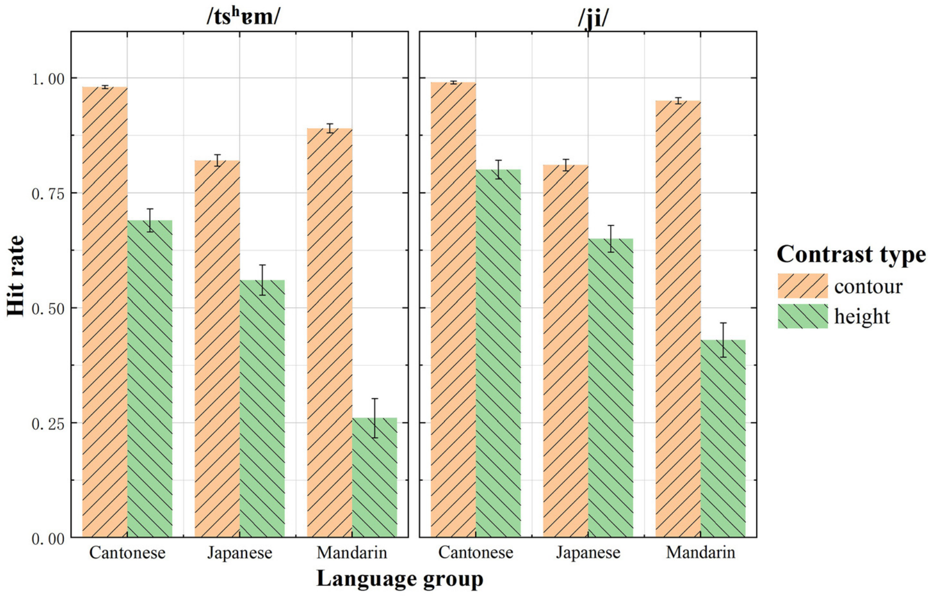

FIG. 2. (Color online) The distribution of hit rate (±SE) for specific contrast types as a function of group (Mandarin, Japanese, and Cantonese) in discriminating Cantonese tones.

$z = 3.447$, $p < 0.01$; NM: $\beta = 0.98$, SE $= 0.125$, $z = 7.861$, $p < 0.001$], in that higher scores were achieved for syllable /ji/ than syllable /ts^{h}ɐm/ for both groups. Conversely, NJ performed stably across two syllables [$\beta = 0.24$, SE $= 0.105$, $z = 2.281$, $p = 0.14$]. These findings might indicate a language-specific effect of syllable in tone perception. That is to say, speakers from canonical tone languages might be more susceptible to syllables in non-native tone perception compared with pitch-accent speakers; however, NC's lower performance for syllable /ts^{h}ɐm/ might be attributed to the fact that the embedded T5 and T6 were less used in natural communication. As for NM, only Cantonese T1 (55) produced with syllable /ji/ could form a real word in Mandarin, it was reasonable to infer that NM's superior performance in /ji/ was due to integral processing of tone and syllable rather than lexical effect.

4. Performance across tone pairs in NM

As stated above, syllables had a significant effect on the discrimination of Cantonese tones by NM. In order to further explore the segmental effect and address the third research question of the current study, namely, how perceptual similarity affects the discrimination of Cantonese tone contrasts, different pairs were compared across syllables for NM.

The performance of NM for 15 tone pairs across /ji/ and /ts^{h}ɐm/ is shown in Fig. 3. A "Tone contrast" (15 pairs) × "Syllable" analysis of variance (ANOVA) was performed.[3] Results revealed significant effects of "Syllable" [$F(1, 3090) = 59.37$, $p < 0.001$] and "Tone contrast" [$F(14, 3090) = 172.94$, $p < 0.001$]; there was also a significant interaction between the two factors [$F(14, 3090) = 6.13$, $p < 0.001$]. Simple main effect tests for the interaction showed that the syllable effect was significant only for T1-T6 ($p < 0.05$), T3-T4 ($p < 0.001$), and T3-T6 ($p < 0.001$). Other tone pairs were comparable with respect to hit rates over the two syllables ($ps > 0.05$). In addition, *post hoc* pairwise comparisons between tone contrasts were conducted within each syllable. For /ji/, the best performances of NM were observed with T1-T2, T1-T4, and T1-T5, which were significantly higher than T1-T3, T1-T6, T2-T5, T3-T6, T4-T6 in hit rates ($ps < 0.05$); the worst performances were found with T1-T3, T2-T5, T3-T6, which were significantly lower than other pairs ($ps < 0.05$). The discrimination of T1-T6 was moderate, which was found to be significantly higher than T1-T3, T2-T5 but lower than the pairs like T1-T2, T1-T4 ($ps < 0.05$). For syllable /ts^{h}ɐm/, T1-T2, T1-T5, T2-T3 were three tone pairs that NM found to be the easiest to discriminate, the hit rates of which were significantly higher than height pairs ($ps < 0.05$). Similarly, T1-T3, T2-T5, and T3-T6 were likewise significantly worse than other pairs in hit rates ($ps < 0.05$). It is worth noting that the discriminability of pairs involving T4 was unstable. For instance, T1-T4 was the second best in /ji/, while it was only moderate in /ts^{h}ɐm/. Moreover, T3-T4 was comparable to T1-T2 and T1-T5 for the familiar syllable /ji/ ($ps > 0.05$), yet it was significantly lower than T1-T2 and T1-T5 in the context of the unfamiliar syllable /ts^{h}ɐm/ ($ps < 0.01$).

C. Interim discussion

Experiment 1 explored the effects of native prosodic system and segmental context on listeners' non-native tone perception. Through a discrimination test, it was shown that the NM and NJ were comparable in overall performance, but showed significantly different patterns. In addition, NM were susceptible to the influence of syllables, performing more accurately for the familiar syllable than the unfamiliar one, while NJ demonstrated a stable pattern irrespective of syllables. This implied a language-specific effect of segmental context on tone perception, in line with previous studies (Repp and Lin, 1990; Tong *et al.*, 2008; Tong *et al.*, 2014).

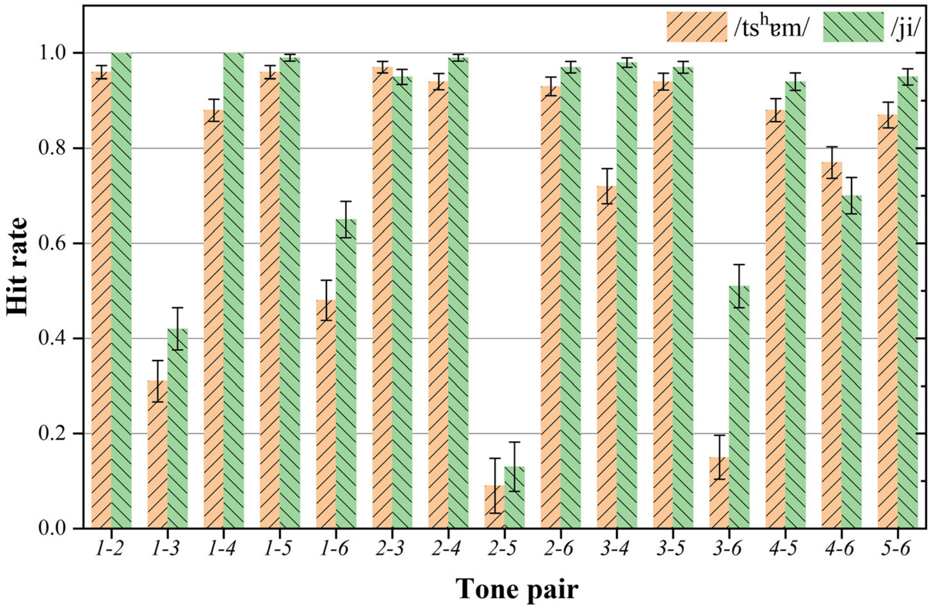

FIG. 3. (Color online) Mean hit rate (±SE) for 15 tone pairs by the Mandarin listeners across two syllables.

1. *The effect of native prosodic system*

The findings in Experiment 1 supported the hypothesis that language-dependent cue weighting rather than general experience with tones might determine non-native tone perception, conforming to the studies of Francis *et al*. (2008). In this study, although Mandarin is a canonical tone language, NM did not significantly outperform NJ in overall performance. However, the two groups of listeners diverged in perceptual patterns for specific tone pairs. NM outperformed NJ for tone pairs which are contrasted by pitch contour, whereas NJ performed better than NM for those distinguished by pitch height. These findings provided evidence for the transfer from the native prosodic system to the perception of non-native tones, in that NM might primarily rely on pitch contour to distinguish tones, while NJ would utilize pitch height in the differentiation of pitch accents at the word level. Therefore, Japanese and Mandarin listeners have different cue weightings in perceiving tones. In addition, the cue-weighting view could explain the previous findings. For instance, Wang (2013) found that L1 tone experience did not provide much aid for Hmong listeners in discriminating Mandarin tones compared with Australian English listeners. From the perspective of cue weighting, the results could be accounted for by the fact that Hmong tone system is mainly characterized by pitch height, while Mandarin relies on pitch contour. The mismatch of the perceptual cues might lead to invalid transfer from the Hmong tonal system to the perception of Mandarin tones at the initial stage.

Moreover, the experiment also displayed some universal patterns of both groups on their processing of Cantonese tones. Contour pairs were found easier to discriminate than height pairs which were difficult even for NC in citation form, as suggested by previous findings (Mok *et al*., 2013; Peng *et al*., 2012; Qin and Mok, 2015). The common mode in the perception of Cantonese tones could plausibly be triggered by the psychoacoustic factors mentioned in previous studies (Hao, 2012; So and Best, 2010). For example, Cantonese T2 and T5 share the same contour with minor discrepancies in height, so do T3 and T6, while contour tonal contrasts, like T1-T2, diverge not only in pitch height, but also in direction. Besides, previous studies also indicated that height tone pairs were perceived more continuously than contour ones, and the latter were found to be less susceptible to other factors such as speaker variability (Peng *et al*., 2012).

2. *The effect of segmental context*

Nonparallel segmental effect on the two experimental groups corroborated the hypothesis that the segmental effect might be language-specific. Results showed that NM exhibited fluctuations for syllable /ji/ (familiar) and /ts^{h}ɐm/ (unfamiliar), which was consistent with previous studies reporting that tone language speakers would process segmental and tonal information in an integral way (Lee *et al*., 1996; Repp and Lin, 1990; Tong *et al*., 2008; Tong *et al*., 2014). However, for NJ, the situation was quite the reverse; they were found to be stable across the two syllables, patterning closer to English listeners in some sense. The independent processing of NJ might be attributed to the pitch realization in their prosodic system where monosyllabic words do not carry tones, and pitch varies over two syllables, in contrast to the monosyllabic tones in Mandarin and Cantonese. That being said, it remains unclear as to how the familiarity of the syllable influences Mandarin listeners' discriminating Cantonese tones, as well as the reason for the decline of discriminability in specific tone pairs, T1-T6, T3-T4, and T3-T6 carried by the unfamiliar syllable.

The above-mentioned perceptual disparities might be related to the perceptual similarities between Mandarin and Cantonese tones. According to PAM-s, listeners' performances could be subject to the perceptual similarity between the suprasegmental categories in two languages. The discrimination might be excellent if a non-native tonal contrast was assimilated to two categories (TC), but the discrimination could be struggling to varying degrees if a non-native tonal contrast was assimilated to one category or if one tone could not be categorized (SC, CG, UC). It is likely that NM's decline of discriminability in the unfamiliar context could be caused by the perceptual mapping between Mandarin and Cantonese tones. Since the discriminability between different tonal pairs for NM had been explored above, Experiment 2 was conducted to unveil how listeners would assimilate Cantonese tones to their native tone categories in different contexts, seeking plausible accounts for the observed discrimination patterns (segmental effect) from the standpoint of PAM-s. Since Japanese lacks overt tonal categories, the NJ group was not included in this experiment since they could not be expected to consistently establish the mappings of the pitch patterns between Japanese and Cantonese.

III. EXPERIMENT 2

A. Method

1. *Participants*

Twenty Mandarin listeners (ten male and ten female; average age = 21.9 years, SD = 1.55) were involved in Experiment 2, including thirteen subjects in Experiment 1. The criteria for screening participants were the same as those in Experiment 1.

2. *Materials*

The stimuli in Experiment 1 continued to be applied in Experiment 2.

3. *Procedures*

NM took part in the perceptual assimilation task independently in a quiet classroom. A total of 48 tokens (2 speakers × 2 tokens × 2 syllables × 6 tones) were grouped into two blocks by syllable (/ji/ and /ts^{h}ɐm/). Within each block, the stimuli were randomly presented with the program Experiment MFC 7 mentioned above. Participants were asked to assimilate each tone into the Mandarin tonal system. Five choices were displayed on the screen, including the four tones in Mandarin and a "无" (none) button. Listeners were allowed to select "无" only when they could not assimilate the tone to any of the Mandarin tones (So and Best, 2014; Yang and Chen, 2019). In this experiment, the participants were able to listen to the stimuli as many times as they wished by clicking the replay button. After they selected the category, they were instructed to rate the similarity between the tone they heard and the corresponding Mandarin counterpart based on a 7-point Likert scale (1 represents "least similar" while 7 represents "very similar"). If the listeners chose "无" in the last step, the similarity rating would be abandoned. Similarly, before the formal experiment, a familiarization session with 12 samples was completed by each participant.

4. Data analysis

In line with previous studies, data of the perceptual assimilation task were analyzed through assimilation percentages and similarity ratings. In addition, degrees of response diversity were calculated, following Wu *et al.* (2014) to measure assimilation consistency *via* Eq. (1),

$$K' = \frac{1}{\sum_{i=1}^{R} P_i^2}. \qquad (1)$$

R represents the total number of L1 tone categories and Pi represents the percentage of responses that a non-native tone is assimilated to a particular L1 tone category. The minimum diversity ($K' = 1$) indicates that the non-native tone category has been consistently assimilated to a single L1 tone category, while the maximum diversity (K' = the number of L1 tone categories) indicates that the non-native tone has been marginally mapped to all given choices in an equal manner. Degree of response diversity serves as a useful parameter in revealing the degree of assimilation between two phonetic inventories. In the current study, the maximum diversity is 5 for Mandarin listeners (four tone categories and one "none" option). If the K' is large (close to 5), the assimilation between Cantonese and Mandarin tones could be weak, since the mapping is dispersed to multiple targets with low degrees of similarity. On the contrary, if the K' is small (close to 1), the Cantonese Mandarin perceptual assimilation could be robust since there is a clear Mandarin mapping for the Cantonese tone.

B. Results of Experiment 2

1. Perceptual assimilation task

The left panel in Fig. 4 demonstrates the assimilation between Cantonese and Mandarin tones over /ts^{h}ɐm/ and /ji/ by NM. As shown, the assimilation patterns are similar across the two syllables except for T4. For the statistical analysis, the assimilation percentages of Cantonese tones were analyzed *via* GLMMs. The multi-level categorical variable of "Mandarin choice" was transferred into binomial distribution using the package *mlogit*. "1" (the corresponding Mandarin tone was chosen) and "0" (the corresponding Mandarin tone was not chosen) were added as dependent variables with "Mandarin choice" as the fixed effect. "Subject" and "Item" with "Mandarin choice" were computed as random slopes. The assimilation criteria used here were consistent with So and Best (2014) in that the frequency of the assimilated item must be significantly higher than both chance level and that of any other choices. In the present study, the chance level is 20% for each category (five options). Results showed that "Mandarin choice" had significant main effects for all Cantonese tones (*ps* < 0.001) except T4 for the syllable /ts^{h}ɐm/ [χ^2 (4) = 3.91, p = 0.42].

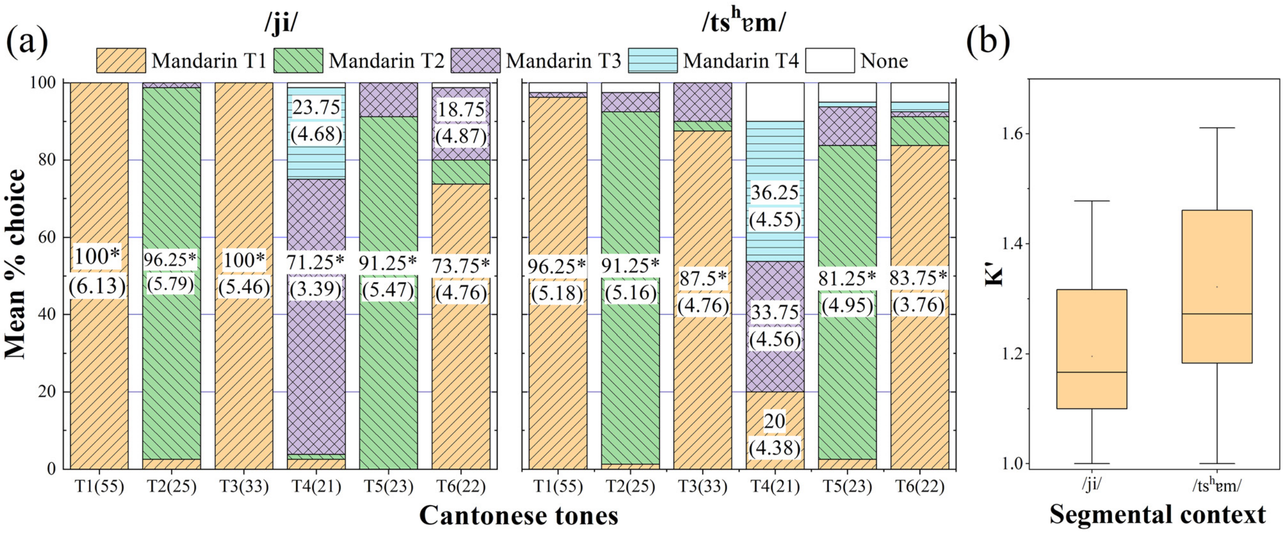

FIG. 4. (Color online) Assimilation percentages of NM and corresponding mean goodness-of-fit ratings of Mandarin categories for each Cantonese tone (left panel) and the degrees of assimilation diversity (K') for two syllables (right panel). Asterisks * indicate the tone could be assimilated. Categories that were chosen less than 10% are not labeled.

Post hoc pairwise comparisons showed that the following mappings were all significantly higher than chance level (20%) and then other mappings for both syllables: Cantonese (C) T1 to Mandarin (M) T1; CT2 to MT2; CT3 to MT1; CT5 to MT2, and CT6 to MT1 ($ps < 0.001$), suggestive of the corresponding assimilation patterns. CT4, on the other hand, was unassimilated to any Mandarin categories in the syllable /tsʰɐm/ since there was no main effect of Mandarin choice as mentioned above. In sum, most Cantonese tones were assimilated to the Mandarin tonal system on the basis of the assimilation criteria, except CT4 for the syllable /tsʰɐm/. Specifically, CT1, CT3, CT6 were all assimilated to MT1, then both CT2 and CT5 were assimilated to MT2, and CT4 was assimilated to MT3 for the syllable /ji/. Due to the fact that more than one Cantonese tone was assimilated to MT1 and MT2, goodness-of-fit rating scores were submitted to Linear Mixed Effects models to determine the assimilation type of SC or CG. "Similarity ratings" and "Cantonese tones" were calculated as the dependent variable and the fixed effect, respectively. "Subject" and "Item" with "Cantonese tone" were computed as random slopes. The visual inspection of Q-Q plots and plots of residuals revealed no obvious deviations from homoskedasticity. For MT1, results showed that goodness-of-fit ratings for CT1 did not significantly differ from those for CT3 [$\beta = 0.55$, SE $= 0.28$, $t = 1.96$, $p = 0.07$], yet a significant discrepancy in ratings was observed in CT1-CT6 [$\beta = 1.4$, SE $= 0.31$, $t = 4.46$, $p < 0.001$] and in CT3-CT6 [$\beta = 0.87$, SE $= 0.28$, $t = 3.1$, $p < 0.01$]. Hence, CT1-CT3 fits SC while CT1-CT6 and CT3-CT6 fit CG. Similarly, no significant difference was found between CT2 and CT5 in goodness-of-fit scores in the assimilation to MT2 [$\beta = 0.26$, SE $= 0.22$, $t = 1.17$, $p = 0.26$], indicating that both two rising tones were equally assimilated to MT2 as SC. Taken together, according to PAM-s, the discriminability of CT1-CT3, CT1-CT6, CT2-CT5, CT3-CT6, being SC or CG, could be poor or moderate to good, which could be worse than other pairs considered as TC.

Comparing the assimilation patterns across two syllables, it was found that the assimilation percentages of the categorized tones declined when the carrying syllable was unfamiliar, except for CT6. In addition, it was also accompanied by a decline of goodness-of-fit ratings. Furthermore, the most obvious divergence between the two syllables lay in CT4. Specifically, CT4 was assimilated to MT3 by NM in syllable /ji/, yet it was uncategorized in syllable /tsʰɐm/. As such, any tone pairs including CT4 in syllable /tsʰɐm/ would form non-overlapping contrasts of UC, since no L1 category for CT4 was above chance level. According to So and Best (2014), the discriminability of UC-n is predicted to be good, modestly lower than the excellent discriminability of TC. Therefore, it was hypothesized that tone pairs involving CT4 would be discriminated slightly worse in syllable /tsʰɐm/ (UC-n) than /ji/ (TC) for NM according to PAM-s.

2. Degree of diversity

To further explore the segmental effects on perceptual assimilation, NM's degrees of mapping diversity (K') for Cantonese tones in the context of syllable /ji/ and /tsʰɐm/ were compared using a Linear Mixed Effects model. "K'" and "Syllable" were calculated as the dependent variable and the fixed effect, respectively. "Subject" with "Syllable" was calculated as the random slope. The visual inspection of Q-Q plots and plots of residuals revealed no obvious deviations from homoskedasticity. Conspicuous disparities of K' in two contexts are shown in the right panel of Fig. 4. Statistical results showed that the K' for the syllable /ji/ was significantly lower than that for the syllable /tsʰɐm/ [$\beta = -0.13$, SE $= 0.05$, $t = -2.6$, $p < 0.05$], which was indicative of a higher degree of assimilation for /ji/ than for /tsʰɐm/. The above findings imply that in contrast to the familiar syllable /ji/, NM tend to perceive Cantonese tones as less similar to Mandarin counterparts in the context of the unfamiliar syllable /tsʰɐm/. Combined with the assimilation patterns shown earlier, it is suggested that segmental context could have a large impact on the perceived similarity of Cantonese tones by Mandarin listeners, which in turn influenced the discrimination performance.

C. Interim discussion

First, results of the perceptual assimilation patterns from Experiment 2 could explain NM's discriminability for individual tone pairs in Experiment 1. In general, NM performed significantly better for CT1-CT2, CT1-CT5 than CT2-CT5, CT1-CT3 with CT1-CT6 falling in between. It followed the sequence TC > CG > SC, supporting PAM-s proposal. Additionally, Experiment 2 confirmed and extended the findings from Experiment 1 regarding the role of syllables in NM's perception of Cantonese tones, and shed light on the intrinsic reasons behind this phenomenon. Turning to the assimilation patterns and K' values in Experiment 2, it was found that NM could not certainly assimilate Cantonese tones into their native tonal system with the syllable /tsʰɐm/ as they did with /ji/. In other words, familiar context renders significantly higher perceptual similarity. More specifically, the most prominent difference was observed with CT4. In /ji/, a familiar syllable for NM, CT4 was assimilated to MT3, whereas NM could not assimilate CT4 to any of the Mandarin categories when the tone was carried by the unfamiliar syllable /tsʰɐm/. These results might imply the interference of the unfamiliar syllable on listeners' assimilation, which could influence discriminability. Within the framework of PAM-s, the discrimination of TC category is better than that of UC category. As demonstrated earlier, tone pairs including CT4 fitted TC in syllable /ji/, while they followed UC-n by the syllable /tsʰɐm/. Therefore, the discriminability of tone pairs containing CT4 would decrease when the context changed from /ji/ to /tsʰɐm/, which explains the significant decline of hit rate in CT3-CT4 for /tsʰɐm/ in Experiment 1. On the other hand, CT1-CT6, CT3-CT6, both being assimilated to the single Mandarin category in the two syllabic contexts, were discriminated less sensitively in /tsʰɐm/. This asymmetry seems unaccountable by PAM-s. However, an alternative

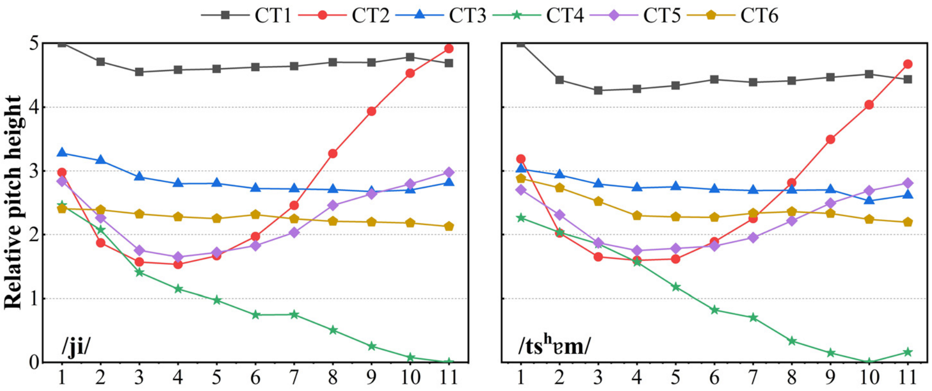

FIG. 5. (Color online) Pitch tracks of the Cantonese tones produced with syllables /ji/ and /ts^{h}ɐm/ by the female Cantonese speaker in the present study.

explanation could stem from cognitive views. Given that the unfamiliar syllable /ts^{h}ɐm/ would reduce the perceptual similarity of non-native tones, it is likely that listeners' attention would be distracted from pitch processing to segment processing, whereas in familiar contexts, they could concentrate more on subtle pitch distinctions since familiar syllables do not require much cognitive processing. Besides, pitch tracks of Cantonese tones produced on two syllables could provide an explanation from the acoustic perspective. It can be seen in Fig. 5 that pitch tracks are slightly more crowded when produced with /ts^{h}ɐm/ than /ji/, especially for CT1-CT6, CT3-CT6. Hence, NM might have encountered greater difficulty in the context of /ts^{h}ɐm/. Generally speaking, perceptual similarity between two tonal systems would play an important role in discriminating tone pairs. Although unfamiliar syllables could cause more confusion for tone perception, the impact was mainly observed in CG and UC pairs.

IV. GENERAL DISCUSSION

This study revealed important roles of the native prosodic system and segmental context on tone perception. First, comparable accuracy but distinct confusions of NM and NJ confirmed that even though pitch accent L1 speakers had more limitations in pitch realizations, this might not induce a disadvantage in perceiving L2 tones, in line with So and Best (2010). It would seem to be that in the initial stage of non-native tone perception, perceptual cues in L1 rather than L1 status (tone vs non-tone) could directly modulate listeners' performance. If perceptual cues between L1 and the target language are compatible, the L1 perceptual cue could be deployed to aid tone perception in a novel language; otherwise listeners could not benefit from their native prosodic systems irrespective of tonality. It can also find support from Wang (2013), who found that Hmong listeners with poor performance initially made great progress in perceiving Mandarin tones after three-to-four week training, shifting more attention from native features to the previously ignored non-native acoustic cues. On the other hand, findings of NM and NJ in the current study might contradict the Lee *et al.* (1996) proposal on tonal complexity. With totally different pitch numbers and scope, NM and NJ performed equally well in perceiving novel tones. In addition, Lee *et al.* (1996) did not exclude the extraneous factors like musical experience and dialects which might influence the results.

Second, segmental contexts had an asymmetric impact on the native speakers of two languages. Although Mandarin and Japanese belong to the same typology in general (Yip, 2002), NM performed better when tones were carried by the familiar syllable than the unfamiliar one, whereas NJ performed stably across both syllables. It seems that NM integrates tones with syllables, whereas NJ, being influenced by their pitch accent system, does not. These findings suggest that segmental effect could be language specific, supporting previous research (Repp and Lin, 1990; Tong *et al.*, 2008). However, results of NJ appeared to be inexplicable. Apart from different pitch realizations mentioned above, microprosodic factors might play a role. Given that different types of vowels and consonants bear different acoustic properties, it is likely that $F0$ perturbations induced by vowel intrinsic $F0$ and obstruents contrasts (voicing or aspiration) would affect tone perception (Cao and Zhang, 2019; Hombert, 1978; Zheng, 2014). Hombert (1978) and Zheng (2014) demonstrated that vowel intrinsic $F0$ was negatively associated with its openness so that close vowels would have a higher $F0$ than open ones. As for obstruents, Cao and Zhang (2019) examined Mandarin and Japanese listeners' perception of tone continua and indicated that a tone carried by aspirated affricates was more easily perceived as the tone which has relatively lower onset $F0$ than unaspirated affricates. Since syllables /ji/ and /ts^{h}ɐm/ differ from each other in both onset obstruents and vowels, microprosodic $F0$ perturbations might partially affect the results found for segmental effect. Future research could be better designed to control the segments or to investigate the effect of phonotactics on novel lexical tone perception.

V. CONCLUSION

The present study investigated the effects of L1 prosodic system and segmental context on the perception of Cantonese tones by Mandarin and Japanese listeners. NM and NJ differed in perceptual patterns but not in overall performance, suggesting that L1 perceptual cues instead of tonal experience modulate non-native tone perception, supporting Francis *et al.* (2008). Besides, the effect of segmental context could be language specific. NM performed worse when tones were carried by the unfamiliar syllable, whereas NJ demonstrated a stable pattern regardless of segmental contexts. The discrepancy between the two groups was possibly due to the different pitch realizations in their prosodic systems (Mandarin: monosyllable; Japanese: cross syllables). These findings unanimously indicate that even within the same typology, different L1 perceptual cues and syllables could also lead to largely different performances in non-native tone perception, contributing new findings to the extant literature from a fine-grained perspective. In future research, disyllabic words or sentence contexts could also be examined to obtain more holistic understandings of the effect of segmental context on the perception of non-native tones by listeners with pitch accent L1s.

ACKNOWLEDGMENTS

This project was supported by Hunan Provincial Social Science Funds (Grant Nos. 18YBA071 and 19YBQ112). We are very grateful to the anonymous reviewers for their professional comments and suggestions on our paper. We would like to thank Dr. Daniel Hirst for his suggestions and for the careful proofreading of this paper.

[1] Although T5 and T6 in the context of the syllable /tsʰɐm/ are rarely used in natural speech, it could not affect the results found for Mandarin and Japanese listeners as an unfamiliar syllable. Moreover, it seems relatively difficult to find another Cantonese syllable meeting the two requirements mentioned in the article. For this reason, /tsʰɐm/ is still an eligible syllable. And to balance the experiment, we choose only one familiar syllable and one unfamiliar syllable. Therefore, fillers /fu/ and /jɐu/ would not be used in the formal experiment though /fu/ as a familar syllable also satisfies our criteria.

[2] The number of different pairs was larger than that of same pairs in the current study, which, intuitively, might induce bias for "different" responses for the same trials, rendering more errors. The results, in effect, revealed only very few errors for the same pairs. Additionally, this study focuses on the results of different pairs. Therefore, the unbalanced design did not appear to have affected the results.

[3] We adopted ANOVA to explore Tonal contrasts × Syllable interaction since GLMM could not be used for convergence issues.

Abramson, A. S. (**1978**). "Static and dynamic acoustic cues in distinctive tones," Lang. Speech **21**(4), 319–325.

Bauer, R. S., and Benedict, P. K. (**1997**). "Cantonese tones," in *Modern Cantonese Phonology*, edited by W. Winter (Mouton de Gruyter, New York), pp. 109–143.

Best, C. T. (**1995**). "A direct realist perspective on cross-language speech perception," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-Language Speech Research*, edited by W. Strange (York Press, Timonium, MD), pp. 167–200.

Best, C. T. (**2019**). "The diversity of tone languages and the roles of pitch variation in non-tone languages: Considerations for tone perception research," Front. Psychol. **10**, 364.

Best, C. T., McRoberts, G. W., and Goodell, E. (**2001**). "Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system," J. Acoust. Soc. Am. **109**(2), 775–794.

Best, C. T., and Strange, W. (**1992**). "Effects of phonological and phonetic factors on cross-language perception of approximants," J. Phon. **20**(3), 305–330.

Best, C. T., and Tyler, M. D. (**2007**). "Nonnative and second-language speech perception: Commonalities and complementarities," in *Language Experience in Second Language Speech Learning: In Honor of James Flege*, edited by M. J. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Boersma, P., and Weenink, D. (**2019**). "Praat: Doing phonetics by computer (version 6.0.49) [computer program]," http://www.praat.org/ (Last viewed September 25, 2019).

Brunelle, M. (**2009**). "Tone perception in Northern and Southern Vietnamese," J. Phon. **37**(1), 79–96.

Burnham, D., Kasisopa, B., Reid, A., Luksaneeyanawin, S., Lacerda, F., Attina, V., Rattansone, N. X., Schwarz, I.-C., and Webster, D. (**2015**). "Universality and language-specific experience in the perception of lexical tone and pitch," Appl. Psycholinguist. **36**(6), 1459–1491.

Cao, C., and Zhang, J. (**2019**). "Effects of affricate's aspiration on realization and perception of lexical tones in Standard Chinese," J. Acoust. Soc. Am. **146**(2), 1036–1044.

Chao, Y. R. (**1968**). "Phonology," in *A Grammar of Spoken Chinese* (The Commercial Press, Beijing), pp. 52–56.

Faris, M. M., Best, C. T., and Tyler, M. D. (**2016**). "An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized," J. Acoust. Soc. Am. **139**(1), EL1–EL5.

Faris, M. M., Best, C. T., and Tyler, M. D. (**2018**). "Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories," J. Phon. **70**, 1–19.

Francis, A. L., Ciocca, V., Ma, L., and Fenn, K. (**2008**). "Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers," J. Phon. **36**(2), 268–294.

Francis, A. L., Ciocca, V., and Ng, B. (**2003**). "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**(7), 1029–1044.

Gandour, J. T. (**1983**). "Tone perception in far Eastern languages," J. Phon. **11**(2), 149–175.

Hao, Y.-C. (**2012**). "Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers," J. Phon. **40**(2), 269–279.

Hombert, J. (**1978**). "Consonant types, vowel quality, and tone," in *Tone: A Linguistic Survey*, edited by V. Fromkin (Academic Press, New York), pp. 77–111.

Howie, J. M. (**1976**). "Fundamental frequency of the tones," in *Acoustical Studies of Mandarin Vowels and Tones*, edited by J. M. Howie (Cambridge University Press, New York), pp. 147–200.

IPA (**1999**). "Handbook of the international phonetic association," in *A Guide to the Use of the International Phonetic Alphabet* (Cambridge University Press, New York).

Lee, Y. S., Vakoch, D. A., and Wurm, L. H. (**1996**). "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," J. Psycholinguist. Res. **25**(5), 527–542.

Li, B., Jing, S., and Bao, M. (**2016**). "Effects of phonetic similarity in the identification of Mandarin tones," J. Psycholinguist. Res. **46**(1), 107–124.

Liu, F., and Xu, Y. (**2005**). "Parallel encoding of focus and interrogative meaning in Mandarin intonation," Phonetica **62**(2–4), 70–87.

Mok, P. P. K., Zuo, D., and Wong, P. W. Y. (**2013**). "Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese," Lang. Var. Change **25**(3), 341–370.

Moore, C. B., and Jongman, A. (**1997**). "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**(3), 1864–1877.

Peng, G. (**2006**). "Temporal and tonal aspects of Chinese syllables: A syllabus-based comparative study of Mandarin and Cantonese," J. Chin. Linguist. **34**(1), 135–154.

Peng, G., Zhang, C., Zheng, H. Y., Minett, J. W., and Wang, W. S. Y. (**2012**). "The effect of intertalker variations on acoustic-perceptual mapping in Cantonese and Mandarin tone systems," J. Speech Lang. Hear. Res. **55**(2), 579–595.

Qin, Z., and Mok, P. (**2015**). "Discrimination of Cantonese tones by speakers of tone and non-tone languages," Kansas Work. Papers Linguist. **34**, 1–25.

Repp, B. H., and Lin, H. (**1990**). "Integration of segmental and tonal information in speech perception: A cross-linguistic study," J. Acoust. Soc. Am. **87**(S1), S46–S46.

So, C. K. (**2010**). "Categorizing Mandarin tones into Japanese pitch-accent categories: The role of phonetic properties," in *Proceedings of Interspeech 2010 Satellite Workshop on Second Language Studies*, September 26–30, Tokyo, Japan.

So, C. K., and Best, C. T. (**2010**). "Cross-language perception of non-native tonal contrasts: Effects of native phonological and phonetic influences," Lang. Speech **53**(2), 273–293.

So, C. K., and Best, C. T. (**2014**). "Phonetic influences on English and French listeners' assimilation of Mandarin tones to native prosodic categories," Stud. Second Lang. Acquis. **36**(2), 195–221.

Strange, W., and Shafer, V. L. (**2008**). "Speech perception in second language learners: The re-education of selective perception," in *Phonology and Second Language Acquisition*, edited by J. H. Edwards and M. L. Zampini (John Benjamins, Amsterdam), pp. 153–192.

Tong, Y., Francis, A., and Gandour, J. (**2008**). "Processing dependencies between segmental and suprasegmental features in Mandarin Chinese," Lang. Cogn. Process. **23**(5), 689–708.

Tong, X., McBride, C., and Burnham, D. (**2014**). "Cues for lexical tone perception in children: Acoustic correlates and phonetic context effects," J. Speech Lang. Hear. Res. **57**(5), 1589–1605.

Tsukada, K., and Kondo, M. (**2019**). "The perception of Mandarin lexical tones by native speakers of Burmese," Lang. Speech **62**(4), 625–640.

Tsukada, K., Kondo, M., and Sunaoka, K. (**2016**). "The perception of Mandarin lexical tones by native Japanese adult listeners with and without Mandarin learning experience," J. Second Lang. Pronun. **2**(2), 225–252.

Wang, X. (**2013**). "Perception of Mandarin tones: The effect of L1 background and training," Mod. Lang. J. **97**(1), 144–160.

Wayland, R. P., and Guion, S. G. (**2004**). "Training English and Chinese listeners to perceive Thai tones: A preliminary report," Lang. Learn. **54**(4), 681–712.

Werker, J. F., and Tees, R. C. (**1984**). "Phonemic and phonetic factors in adult cross-language speech perception," J. Acoust. Soc. Am. **75**(6), 1866–1878.

Wong, P. C. M., and Perrachione, T. K. (**2007**). "Learning pitch patterns in lexical identification by native English-speaking adults," Appl. Psycholinguist. **28**(4), 565–585.

Wu, X., Munro, M. J., and Wang, Y. (**2014**). "Tone assimilation by Mandarin and Thai listeners with and without L2 experience," J. Phon. **46**, 86–100.

Yang, Y., and Chen, X. (**2019**). "Within-organ contrast in second language perception: The perception of Russian initial /r-l/ contrast by Chinese learners," J. Acoust. Soc. Am. **146**(2), EL117–EL123.

Yang, Y., Chen, X., and Xiao, Q. (**2020**). "Cross-linguistic similarity in L2 speech learning: Evidence from the acquisition of russian stop contrasts by mandarin speakers," Second Lang. Res. (published online).

Yazawa, K., Whang, J., Kondo, M., and Escudero, P. (**2020**). "Language-dependent cue weighting: An investigation of perception modes in L2 learning," Second Lang. Res. **36**(4), 557–581.

Yip, M. (**2002**). *Tone* (Cambridge University Press, Cambridge, UK), pp. 1–5.

Zheng, Q. (**2014**). "Effects of vowels on Mandarin tone categorical perception," Acta Psychol. Sin. **46**(9), 1223–1231.