Published in final edited form as:

J Mem Lang. 2012 April 1; 66(3): 438–457. doi:10.1016/j.jml.2011.12.006.

The Perception and Representation of Segmental and Prosodic Mandarin Contrasts in Native Speakers of Cantonese

Xujin Zhang¹, Arthur G. Samuel², and Siyun Liu³

¹Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China and Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, China and Brain and Cognition, School of Psychology, Central China Normal University, Wuhan, China and Psychology Department at Stony Brook University, the United States

²Basque Center on Cognition Brain and Language, Donostia Spain and IKERBASQUE, Basque Foundation for Science, Bilbao Spain and Psychology Department at Stony Brook University, the United States

³Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Wuhan, China and Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, Wuhan, China and Brain and Cognition, School of Psychology, Central China Normal University, Wuhan, China

Abstract

Previous research has found that a speaker's native phonological system has a great influence on perception of another language. In three experiments, we tested the perception and representation of Mandarin phonological contrasts by Guangzhou Cantonese speakers, and compared their performance to that of native Mandarin speakers. Despite their rich experience using Mandarin Chinese, the Cantonese speakers had problems distinguishing specific Mandarin segmental and tonal contrasts that do not exist in Guangzhou Cantonese. However, we found evidence that the subtle differences between two members of a contrast were nonetheless represented in the lexicon. We also found different processing patterns for non-native segmental versus non-native tonal contrasts. The results provide substantial new information about the representation and processing of segmental and prosodic information by individuals listening to a closely-related, very well-learned, but still non-native language.

Keywords

sibling languages; non-native language processing; segmental versus tonal perception; lexical representation of L2 contrasts

Infants are born with an ability to discriminate phonological contrasts in all languages (Streeter, 1976). However, over the course of the first year of life, perception of native speech (L1) greatly improves while that of non-native speech significantly declines (Kuhl et

© 2011 Elsevier Inc. All rights reserved.

Address correspondence to: Siyun Liu, School of Psychology, Central China Normal University, Wuhan 430079, China, liusy@mail.ccnu.edu.cn, Telephone: +86-13886062394.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

al., 2006, Werker & Tees, 1984). This perceptual specialization for native sounds continues to develop into adolescence (Iverson et al. 2003). Related research has focused on the perception of contrasts in a second language (L2) or in a non-native dialect, at both segmental and prosodic levels. Several theoretical models have been proposed to explain why adults experience difficulties in perception and learning of non-native segments, including Flege's Speech Learning Model (Flege, 1995, 1999, 2003) and Best's Perceptual Assimilation Model (Best, McRoberts, & Goodell, 2001; Best & Tyler, 2007).

Flege's Speech Learning Model (SLM) suggests that first and second language sounds exist in a common phonological space and interact with each other. According to SLM, L2 learners can establish new L2 phonetic categories if they detect phonetic differences between an L2 sound and the nearest L1 sound. Thus, the more similar a new L2 category is to the nearest L1 category, the more difficult it is for an L2 learner to establish a new L2 category. For instance, Japanese speakers are well known to have difficulties in learning English /ɹ/ and /l/ (Goto, 1971; Guion et al., 2000; Hattori & Iverson, 2009; Miyawaki et al., 1975). However, Japanese speakers acquire the English phoneme /ɹ/ more successfully than they acquire English /l/ (Aoyama et al., 2004), because English /l/ is more similar than English /ɹ/ to the closest Japanese phoneme (Guion et al., 2000).

Best's Perceptual Assimilation Model (PAM) also assumes that listeners have a strong tendency to assimilate a non-native speech category to the most similar native category. In this model, there are three ways that a given non-native sound could be perceptually assimilated to the native phonological system. It can be perceived as an exemplar of a native category, as an uncategorized segment that is similar to two or more native categories, or as a nonspeech sound that cannot be assimilated. According to PAM, if the two sounds of a non-native contrast can be assimilated to two different native categories, or if one sound of a non-native contrast is categorized while the other is not, discrimination should be fairly good. If the two sounds of a non-native contrast are assimilated to the same native category or fall in between the same native categories, discrimination will generally be poor. If the two sounds are assimilated to a single native category with different goodness of fit, discrimination will be moderate. In support of this view, Japanese speakers were better at distinguishing English /ɹ/-/w/ than English /ɹ/-/l/ (Best & Strange, 1992; Guion et al., 2000) because the former contrast is a goodness difference case, while the latter is a single category case. Although Japanese speakers are inclined to rate English /l/ as being somewhat more similar to the closest Japanese phoneme than English /ɹ/ (Aoyama et al., 2004), they are both poor exemplars of the same Japanese category (Best and Strange, 1992; Hattori & Iverson, 2009). Therefore, even highly experienced Japanese (L1)-English (L2) bilinguals still have difficulty in discriminating this contrast (Guion et al., 2000).

In the current study, we examine the perception of L2 Mandarin contrasts by native speakers of Guangzhou Cantonese. Cantonese and Mandarin have overlapping but different sets of vowels, consonants, and lexical tones, providing a natural setting to compare segmental contrasts to tonal contrasts. In addition, Cantonese speakers get intensive education in Mandarin beginning at the early age of six, which allows an investigation of highly experienced L2 listeners. Our study includes both phonological and lexical level tests. To foreshadow, there are interesting similarities and differences between segmental and tonal processes, and between phonological and lexical levels of representation.

In tone languages, the same syllable with different lexical tones will have diverse meanings. For instance, despite having the same segmental components, the Mandarin words /mɑ(1)/ (“妈”, meaning *mother*), /mɑ(2)/ (“麻”, meaning *flax*), /mɑ(3)/ (“马”, meaning *horse*) and /mɑ(4)/ (“骂”, meaning *scold*) have totally different meanings (the number after each syllable indicates the syllable's lexical tone).

Our experiments took advantage of differences in the segmental and lexical tone inventories of Cantonese and Mandarin. At the segmental level, one of the biggest differences involves fricatives and affricates. Cantonese has only two affricates and one fricative: /ts/, /ts'/ and /s/, while Mandarin consonants include /ts/, /ts'/, /s/, /tʂ/, /tʂ'/, /ʂ/, /tʂ/, /tʂ'/, and /ʂ/. At the prosodic level, the Cantonese tone system is more complex than the Mandarin system: Guangzhou Cantonese has six tone patterns plus shortened versions of three of these (Zhan et al., 2004), while Mandarin has only four tones (Hallé, Chang & Best, 2004; So, 2005a, 2005b). Using a standard 5-level pitch scale (in which 5 represents a high pitch, and 1 a low pitch, so that a 51 tone would start high and end low), Cantonese tones have pitches of 55/53 (there is variability in the ending pitch across speakers), 21/11, 35, 13, 33, 22 and short versions of 55, 33 and 22. Mandarin tones have pitches of 55, 35, 214 and 51, which are usually referred to as Tones 1, 2, 3, and 4, respectively. Due to these differences, there are contrasts in Mandarin that do not exist in Cantonese, at both the segmental and prosodic levels. We will be testing how native Cantonese speakers deal with such contrasts.

In the literature there are very different views on whether Cantonese and Mandarin are two different dialects of Chinese, or are two different languages. Some researchers have argued that they are two different languages (Cai, Pickering, Yan, & Branigan, 2011; Chen et al., 2004; Francis, Ciocca, Ma, & Fenn, 2008) because they are largely (but not entirely) mutually unintelligible. However, others consider them to be two dialects of Chinese because of lexical, syntactic, and orthographic similarities (Lee, Vakoch, & Wurm, 1996; Tang & van Heuven, 2009; Zhan, Li, Huang, & Xu, 2004; Zhang, 1998). In a lecture given in 1945, the Yiddish linguist Max Weinrich said (translated to English) that a language is a dialect with an army and a navy. This description nicely captures the differing opinions with regard to Cantonese and Mandarin, with Mandarin and Cantonese officially considered as two dialects of Chinese by the Chinese Academy of Social Sciences (Xing, 1991).

For our purposes, it is sufficient to note that if they are different languages, as the difficulties in mutual intelligibility would suggest, they are certainly closely related ones. We believe that it is useful to look at studies of L2 processing in terms of whether the native and non-native cases are, like Cantonese and Mandarin, sibling languages. In this spirit, we will refer to closely related languages, including these two, as “siblangs” (sibling languages), to make it clear that the results pertain to pairs that include significant overlap at the lexical and syntactic levels. Results for siblangs may or may not turn out to match those for distant pairs, such as English and Japanese.

Given that we will be testing native Cantonese listeners on non-native (Mandarin) segmental and tonal contrasts, and examining the impact of these factors on lexical processing, we will briefly review prior findings in these domains:

The perception of non-native segmental contrasts

Many researchers have examined the perception of non-native consonantal and vocalic contrasts, with the most widely studied case being the one we have already mentioned: the poor performance of Japanese speakers in discriminating English /ɹ/ and /l/ (Goto, 1971; Guion et al., 2000; Hattori & Iverson, 2009; Miyawaki et al., 1975). Even after several weeks of perceptual training, Japanese speakers typically show only a small (though significant) improvement in distinguishing this contrast (Bradlow et al., 1997; Lively, Logan, & Pisoni, 1993; Lively et al., 1994; Logan, Lively, & Pisoni, 1991). Similarly difficult cases include Italians trying to distinguish some vocalic contrasts (e.g., /e/-/ɛ/, /i/-/ɪ/) in Canadian English (Flege & MacKay, 2004) and American English speakers trying to discriminate French /y/-/ø/ (Gottfried, 1984).

There have been a number of studies of Spanish-Catalan bilinguals that provide a particularly useful context for our study because Catalan and Spanish are also sibilants. Moreover, as with our Guangzhou listeners, the second language is learned early and used widely. A somewhat surprising finding is that even fluent non-native language speakers may have difficulty perceiving segmental contrasts in their L2. Proficient Spanish (L1) – Catalan (L2) bilinguals fail to distinguish certain Catalan phonemes that do not exist in Spanish, such as /e/-/ɛ/ and /s/-/z/ (Pallier, Bosch, & Sebastián-Gallés, 1997; Pallier, Colomé, & Sebastián-Gallés, 2001; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés, Echeverría, & Bosch, 2005; Sebastián-Gallés, Rodríguez-Fornells, de Diego-Balaguer, & Díaz, 2006). Even simultaneous bilinguals, who had been exposed to both Spanish and Catalan from birth, have problems in perceiving segmental contrasts that do not exist in their dominant language (Sebastián-Gallés, Echeverría and Bosch, 2005). Flege and MacKay (2004) have suggested that an early age of L2 acquisition may not guarantee native-like perception of L2 phonemes – only those who learn L2 at an early age and also use their native language less frequently may reach similar perceptual levels as native speakers.

The perception of non-native tonal contrasts

An important difference between the Catalan-Spanish case and ours is the role of lexical tones in Mandarin and Cantonese; Romance languages such as Catalan and Spanish do not use lexical tones. Prior work has shown that speakers whose native language is non-tonal have difficulty perceiving the tones of a tonal language. Some cases that have been examined include the perception of Mandarin words by English speakers (Bent et al., 2006; Gottfried & Suiter, 1997), by Uigur speakers (Gao, 2005), or by French speakers (Hallé et al., 2004), and the perception of Thai words by English speakers (Burnham et al., 1996).

More germane to the current study is the finding that if a speaker's native language is tonal, this can influence the perception of non-native tones. For instance, both Cantonese speakers (Burnham et al., 1996) and Mandarin speakers (Wayland & Guion, 2004) were better at distinguishing Thai tones than English speakers, and training significantly helped the Mandarin speakers but not the English speakers. However, there have been some conflicting findings, suggesting that prior experience with a tonal language may not necessarily aid the acquisition of tones in another language. For example, Hmong speakers, whose native language has seven tones, performed worse than Japanese speakers and English speakers in identifying the four tones in Mandarin (Wang, 2006).

There is a small existing literature on tone perception in sibilants – Mandarin and Hong Kong Cantonese (which has six tones -- 55/53, 35, 33, 21, 22, and 23; Zhan et al., 2004). Francis et al. (2008) examined Mandarin speakers' identification of Hong Kong Cantonese tones and compared it with English speakers. The two groups showed similar performance on identification before training, and English participants actually made greater progress on some of the Cantonese tones after training. So (2005a, 2006a, 2006b) tested Hong Kong Cantonese participants' discrimination of Mandarin tones and found that they were poor at identifying some of the Mandarin tones, despite their native familiarity with lexical tones.

The representation of minimal pairs of words

Compared to the large literature on the perceptual difficulties associated with non-native phonetic contrasts, there has been much less research focused on how these perceptual problems affect word recognition. There have been three approaches to this issue. One approach is to obtain lexical decision judgments for "near words" – tokens in which a phoneme that does not exist in the listener's native language is replaced by a similar one that does exist. For example, for native Dutch speakers listening to English, a word like "task" (which includes the vowel /æ/ that is not in the Dutch inventory) could be presented as

“task” (with the closest Dutch vowel, /ɛ/). Under such conditions, Dutch speakers are much more likely to misidentify English nonwords as words than English speakers (Broersma, 2002). Similarly, Spanish (L1) – Catalan (L2) bilinguals were much more likely than native Catalan participants to erroneously identify (as words) Catalan nonwords that had been produced by substituting the Catalan phoneme /e/ (present in Spanish) for Catalan /ɛ/ (not a Spanish vowel) (Sebastián-Gallés et al., 2005).

The second approach used to study non-native lexical processing employs a priming methodology, and tests whether minimal pairs of words containing ambiguous contrasts are represented as homophones for non-native speakers. For example, Spanish (L1) – Catalan (L2) subjects showed repetition priming for minimal pairs of Catalan words (e.g. /netə/-/netə/) differing in contrasts that do not exist in Spanish (e.g. /e/-/ɛ/), whereas Catalan L1 participants showed repetition priming only when the same word was repeated (Pallier et al., 2001). Because there were no comparable priming effects for nonword stimuli, Pallier et al. demonstrated that the repetition priming paradigm specifically targets lexical effects. The repetition priming for Spanish-dominant bilinguals suggests that the non-native sound (/ɛ/) is represented in the lexicon as the native sound (/e/), making minimal pairs homophonous.

However, evidence from the third approach -- eye-tracking -- suggests that there may be differences in the lexical representations for words containing native versus non-native sounds. For example, Japanese participants were more likely to look at a picture of a *locker* when they were asked to click on the picture of a *rocket* than the reverse (Cutler, Weber, & Otake, 2006). Similarly, Weber and Cutler (2004) found competition from words with the native Dutch sound /e/ for targets with the non-native /æ/, but not vice-versa. Due to these asymmetries, the eye-tracking results are difficult to accommodate with the notion of lexical representations that are identical for words with native versus non-native segments. In the current study, we will use a version of the priming methodology that allows us to look for such asymmetries, a test that has not generally been available in priming studies.

The present study

In the current study, three experiments were conducted to test how Guangzhou Cantonese speakers, who learn Mandarin beginning in primary school, perceive and represent consonantal and tonal contrasts that are present in Mandarin Chinese but not in their native Cantonese. We examine how early learners of a sibling language perceive and represent segmental and prosodic contrasts that do not exist in their native language, and what ramifications these differences have for word recognition. Experiment 1 uses a discrimination task to measure how well the Cantonese listeners can discern differences between non-native Mandarin contrasts. Using segmental and tonal contrasts that have been established as difficult, Experiment 2 uses an implicit perceptual processing task to determine whether the members of the non-native contrasts are treated as equivalent sounds when they are not the focus of the task. Finally, in Experiment 3, we use a priming task similar to that of Pallier et al. (2001), but with additional procedures that allow us to look for the kinds of perceptual asymmetries that have been seen using eye-tracking approaches. This addition is the key to examine whether the perceptually difficult non-native contrasts are nonetheless represented differentially in the lexicon. In all three experiments, we test both segmental contrasts and prosodic (tonal) ones.

Experiment 1

The goal of Experiment 1 was to identify specific Mandarin consonants and tones that are difficult for Cantonese native speakers to discriminate, to provide a basis for stimulus selection in the following experiments. Thus we chose 12 pairs of consonantal contrasts formed by 9 Mandarin consonants (i.e., /ts/-/tʂ/, /tsʰ/-/tʂʰ/, /s/-/ʂ/, /ts/-/tʂ/, /tsʰ/-/tʂʰ/, /ts/-/tʂ/

ʃ/, /ts' /-/tʃ/, /ts' /-/tʃ' /, /ts' /-/ʃ/, /s /-/tʃ/, /s /-/tʃ' /, and /s /-/ʃ/) and 6 pairs of tonal contrasts formed by the four Mandarin tones (i.e., T1–T2, T1–T3, T1–T4, T2–T3, T2–T4 and T3–T4). The frequency of each of these Mandarin consonants and tones is provided in Appendix A. We used an oddity discrimination paradigm developed by Flege and MacKay (2004) to determine the most difficult Mandarin consonant and tone contrasts for native speakers of Cantonese.

Method

Participants—Thirteen Guangzhou native Cantonese speakers (mean age of 22 years, ranging from 21 to 25 years) and 13 native Mandarin speakers (mean age of 24 years, ranging from 22 to 26 years) participated in this experiment. The Guangzhou Cantonese speakers were born and raised in Guangzhou and began learning Mandarin in primary school (about six or seven years old). Their average experience using Mandarin was 16 years, ranging from 14 to 19 years. They usually communicate with their teachers and classmates in Mandarin in class, but speak Cantonese outside class. The Mandarin speakers were born and raised in areas where Mandarin is the native dialect (e.g. Beijing and Chengde). During subject recruitment, participants were informed that we were testing native Cantonese or native Mandarin speakers. All participants were tested in the Cognitive Language Lab in Huazhong Normal University in China. No participant reported any hearing problem, and all received small gifts for their participation.

Stimuli—The stimuli for consonant perception were monosyllabic words created by combining the tested consonants with the vowel [ɿ] and Mandarin Tone 1. This procedure yielded 12 minimal pairs of words differing only in their consonants (although the vowel after /tʃ/, /tʃ' /, and /ʃ/ is pronounced slightly differently from that after the other tested consonants). The stimuli for tone perception were monosyllabic words created by combining the syllable /pɑ/ with four Mandarin tones, yielding six minimal pairs of words differing only in tones. Words were recorded by three female Mandarin native speakers. All stimuli were digitized (44.1kHz) and edited with GoldWave 5.22 sound editing software. Appendix B lists the words used in this study.

Procedure—Participants were tested individually in a sound-attenuated booth. On each trial, three words produced by different speakers were presented sequentially. Three-fourths of the trials were *change trials*, containing an odd item that differed in either consonant, such as /ts ɿ (1)/-/tʃ ɿ (1)/-/ts ɿ (1)/, or tone, such as /pʅ(1)/-/pʅ(1)/-/pʅ(2)/. The odd item occurred with equal frequency in all three possible positions in the series, and each contrast was tested in 12 change trials. The remaining 1/4 of the trials were *no-change trials*, containing three identical words that had been produced by different speakers, such as /ts ɿ (1)/-/ts ɿ (1)/-/ts ɿ (1)/ and /pʅ(1)/-/pʅ(1)/-/pʅ(1)/. Change trials were used to investigate participants' sensitivity to different phonological categories, while no-change trials were used to make sure that variation within a single category (speaker-based variation) was ignored by participants.

The experiment was controlled by a Lenovo computer running Windows XP. Stimuli were presented via headphones at a comfortable listening level using E-prime 1.1 (Psychology Software Tools, Inc.) software. On each trial, three items were presented sequentially with an ISI (inter stimulus interval) of 1s. Participants were told to decide which was the odd item and to press the key labeled "1", "2," or "3" corresponding to the position of the odd item in the series. If all items were the same, they were told to press the fourth key, labeled "4". They were asked to make decisions as quickly and as accurately as possible. A subsequent trial began 750ms after the participant had responded. If a participant failed to respond within 3000ms, a new trial would start.

The experimental phase was divided into two blocks. In one block, participants were instructed to focus on consonant changes and only the minimal pairs differing in the consonant were presented. In the other block, participants were instructed to focus on tone changes and only the minimal pairs differing in the tone were presented. Each block was preceded by a practice. Twelve trials were included in the practice before the consonant block started and ten trials were included in the practice before the tone block started. All of the critical contrasts in the experiment were tested in the practice and no feedback was given. The order of the two blocks was counter-balanced across participants, and the order of trials within each block was random.

Results and Discussion

Data from trials in which participants failed to respond within 3000ms were discarded (2.4% of all responses).

Accuracy analysis—An A' score was calculated for each contrast as a measure of how discriminable the two members of the contrast were. The A' score is based on the proportion of hits and false alarms. A hit was defined as the correct selection of the odd item on a change trial, and a false alarm was defined as the report of an odd item on a no-change trial; Appendix C shows the hit and false alarm rates used to compute A' . An A' score of 1.0 reflects perfect discrimination of a contrast, while an A' score of 0.5 corresponds to a chance level of performance (see Flege & MacKay, 2004; Sebastián-Gallés, Echeverría & Bosch, 2005; Sebastian-Gallés et al., 2006; Sebastián-Gallés et al., 2009; Snodgrass, Levy-Berger, & Haydon, 1985).

Consonant discrimination accuracy: Figure 1 shows the two groups' A' scores for each consonantal contrast. A 2 (Group: Guangzhou Cantonese vs. Mandarin) $\times$ 12 (Consonant Type: 12 consonantal contrasts) mixed analysis of variance (ANOVA) was conducted on A' scores for the consonantal contrasts. The main effect of Group was significant, $F(1,24)=5.52$, $p=.027$, $\eta^2=.19$, with the mean A' of Guangzhou Cantonese speakers significantly lower than that of Mandarin speakers (.90 vs. .96). The main effect of Consonant Type was also significant ($F(11,264)=3.67$, $p<.001$, $\eta^2=.13$), ranging from an average score for the two groups of .90 for the /ts-/ɲ/ contrast to an average of .96 for the /s-/tɕ/ contrast, $p<.001$. The interaction between Group and Consonant Type also reached significance, $F(11,264)=1.94$, $p=.035$, $\eta^2=.08$, indicating that some contrasts were more difficult for the Cantonese speakers (compared to the Mandarin speakers) than others. Simple effect tests demonstrated that the Guangzhou Cantonese speakers' scores were lower than the Mandarin speakers' for six consonantal contrasts: /ts-/tɕ/ [$F(1,24)=6.11$, $p=.021$], /ts'-/tɕ'/ [$F(1,24)=9.41$, $p=.005$], /s-/ɲ/ [$F(1,24)=8.54$, $p=.007$], /ts'-/tɕ'/ [$F(1,24)=5.53$, $p=.027$], /ts'-/tɕ'/ [$F(1,24)=6.33$, $p=.019$], and /ts'-/ɲ/ [$F(1,24)=7.92$, $p=.010$]. Of these, the /ts-/tɕ/ contrast produced the largest average difference in absolute A' scores, and the /ts'-/tɕ'/ contrast produced the most reliable difference in performance between the Mandarin and Cantonese speakers, making these two contrasts the best candidates for further study.

Tone discrimination accuracy: Figure 2 shows the two groups' A' scores for each tonal contrast. A 2 (Group: Guangzhou Cantonese vs. Mandarin) $\times$ 6 (Tone Type: 6 tonal contrasts) mixed analysis of variance was conducted on the A' scores for tonal contrasts. The main effect of Group was not significant, $F(1,24)=1.05$, $p=.316$, $\eta^2=.04$. The main effect of Tone was significant, $F(5,120)=2.93$, $p=.016$, $\eta^2=.11$. The average score for the T1–T2 contrast was significantly lower than that for the T1–T3 (.94 vs. .96) and T2–T4 (.94 vs. .96) contrasts, $p<.05$. The average score for the T2–T3 contrast was significantly lower than that for T2–T4 contrast (.94 vs. .96), $p<.05$. The interaction between Group and Tone

Type did not reach significance, $F(5,120)=1.88$, $p=.103$, $\eta^2=.07$, though the T2–T3 contrast clearly showed the largest separation between the two groups.

Reaction time analysis

Consonant discrimination reaction times: Table 1 shows the reaction times for each of the consonant contrast cases. A 2 (Group: Guangzhou Cantonese vs. Mandarin) $\times$ 12 (Consonant Type: 12 consonantal contrasts) mixed analysis of variance was conducted on reaction time for the consonantal contrasts. The main effect of neither Group nor Consonant Type was significant, $F(1,24)=2.36$, $p=.138$, $\eta^2=.09$; $F(11,264)=1.18$, $p=.305$, $\eta^2=.05$. The two-way interaction was also not significant, $F(11,264)=1.37$, $p=.190$, $\eta^2=.05$.

Tone discrimination reaction times: Table 2 shows the mean reaction time for each of the tone contrast cases. A 2 (Group: Cantonese vs. Mandarin) $\times$ 6 (Tone Type: 6 tonal contrasts) mixed analysis of variance on reaction time for tonal contrasts revealed that the main effect of Group was significant, $F(1,24)=4.83$, $p=.038$, $\eta^2=.17$. The response time of the Cantonese speakers was longer than that of the Mandarin speakers (1261ms vs. 1061ms). The main effect of Tone Type also was significant, $F(5,120)=4.36$, $p<.001$, $\eta^2=.15$; the discrimination time for the T2–T3 contrast was significantly longer than that for the T2–T4 contrast (1277ms vs. 1114ms), $p=.001$. No interaction between Group and Tone Type was observed, $F<1$. However, a comparison between the two groups for each tonal contrast revealed that the Cantonese speakers took significantly longer to discriminate the T2–T3 contrast than the Mandarin speakers did, $F(1,24)=5.0$, $p=.035$. Recall that this was the same contrast that produced the largest accuracy difference as well.

Discussion—Although individuals who began learning Mandarin at a young age generally showed good performance on consonantal contrasts, these native Cantonese speakers were significantly poorer than the Mandarin speakers at distinguishing six of the tested consonantal contrasts: /ts/-/tʂ/, /tsʰ/-/tʂʰ/, /s/-/ʂ/, /tsʰ/-/tʂʰ/, /tsʰ/-/tʂʰ/, and /tsʰ/-/tʂʰ/. For two of these contrasts (/ts/-/tʂ/ and /tsʰ/-/tʂʰ/) the difference in performance for the two groups was quite substantial, and the effects primarily affected accuracy.

For the tone contrasts, this pattern reversed -- the accuracy measure did not show robust differences as a function of native language, but there were clear differences in processing speed. The two groups' A' scores did not differ significantly in any of the tested tonal contrasts and were all above .90, which means both the Cantonese and the Mandarin speakers can consistently distinguish Mandarin tonal contrasts. However, we found that the Guangzhou Cantonese group took longer to make tone discrimination judgments than the Mandarin group, particularly for the T2–T3 contrast.

Recall that the primary goal of Experiment 1 was to provide an empirical basis for stimulus selection for the following experiments. For the tonal contrast, T2–T3 is the obvious choice. For the consonantal contrast, both /ts/-/tʂ/ and /tsʰ/-/tʂʰ/ are possible choices. We selected the /tsʰ/-/tʂʰ/ case for three reasons: The statistical reliability of this difference was slightly higher than the /ts/-/tʂ/ case, the reaction time difference was higher, and most importantly, the /tsʰ/-/tʂʰ/ contrast offers a better selection of words to be used as stimuli in Experiment 3.

Experiment 2

The discrimination task in Experiment 1 revealed that Guangzhou Cantonese speakers were poorer than Mandarin speakers at discriminating the /tsʰ/-/tʂʰ/ consonantal contrast and the T2–T3 tonal contrast. However, poor performance on an explicit task does not necessarily mean that an ability does not exist. For example, people may show quite good performance

on an implicit memory task while completely failing on an explicit task (e.g., Warrington & Weiskrantz, 1968, 1970). Thus, we used the Garner speeded classification paradigm (e.g., Tong, Francis, & Gandour, 2008) to obtain an implicit measure of the listeners' sensitivity to non-native contrasts. This paradigm includes two tasks: On a *baseline task*, participants classify stimuli on a target dimension (e.g., consonant) with non-target properties remaining constant on other dimensions (e.g., vowel or tone). On a *filtering task*, participants make the same judgments, but with variation on the non-target dimension (e.g., changed vowels or tones). Any differences in performance between the baseline and the filter tasks are assumed to be due the listeners' sensitivity to the variation of the non-target dimension.

Using this methodology, Navarra et al. (2005) asked Spanish-dominant Spanish(L1)-Catalan bilinguals and Catalan-dominant Catalan(L1)-Spanish bilinguals to classify the first syllable of disyllabic words. In the baseline condition, the stimuli consisted of two homogeneous lists (in which all of the words contained only /e/, or only /ɛ/, in the second syllable). Two orthogonal lists (in which half the words contained /e/ in the second syllable and half contained /ɛ/ in the second syllable) were used for the filtering task. Although participants were instructed to ignore the second syllable, Catalan-dominant bilinguals responded more slowly in orthogonal lists than in homogeneous lists, indicating that variation of the vowel in the second syllable interfered with processing of the first syllable. In contrast, Spanish-dominant bilinguals did not suffer from this interference. These results suggest that only Catalan-dominant bilinguals were sensitive to the difference between /e/-/ɛ/, with the difference influencing their decisions on supposedly irrelevant syllables. Here we use a similar method to compare the differences between Guangzhou Cantonese native speakers and Mandarin native speakers in processing Mandarin /ts' /- /tɕ' / and T2-T3 contrasts. We test whether task-irrelevant variation of either consonantal or tonal properties will affect classification performance, with any such effect indicating perceptual sensitivity to the varied information.

Method

Participants—Fourteen native Guangzhou Cantonese speakers (mean age of 22 years, ranging from 18 to 25 years) and 14 native Mandarin speakers (mean age of 21 years, ranging from 18 to 25 years) participated in Experiment 2. All of the participants were selected from the same two populations as in the first experiment, but none of them had participated in the previous study. The Guangzhou Cantonese native speakers had been using Mandarin Chinese for an average of 16 years, ranging from 11 to 20 years. No participant reported any hearing problem, and all received small gifts for their participation.

Stimuli—All stimuli were recorded by a native Mandarin speaker. For consonant processing, ten different instances of /ts' α(1)/, ten instances of /tɕ' α(1)/, ten instances of /ts' (1)/ and ten instances of /tɕ' (1)/ were recorded by the speaker. The stimuli were used to construct two homogeneous lists and two orthogonal lists. Each homogeneous list consisted of 20 instances which contained the same consonant [List 1: 10 instances of /ts' ■ (1)/ and 10 instances of /ts' α(1)/; List 2: 10 instances of /tɕ' (1)/ and 10 instances of /tɕ' (1)/]. Each orthogonal list consisted of five different instances of each word [five instances of /ts' (1)/, five instances of /tɕ' (1)/, five instances of /ts' α(1)/ and five instances of /tɕ' α(1)/; the two lists only differed in the particular tokens assigned to each list].

For tone perception, ten different instances of /pi(2)/, /pi(3)/, /pa(2)/, and /pa(3)/ were recorded by the same speaker. Four lists were created as follows: Each homogeneous list consisted of 20 instances which contained the same tone [List 1: 10 instances of /pi(2)/ and 10 instances of /pa(2)/; List 2: 10 instances of /pi(3)/ and 10 instances of /pa(3)/]. Each orthogonal list consisted of five different instances of each word [five instances of /pi(2)/,

five instances of /pi(3)/, five instances of /pa(2)/ and five instances of /pa(3)/; the two lists only differed in the particular tokens assigned to each list]. All stimuli were digitized (44.1kHz) and edited with GoldWave 5.22 sound editing software.

Procedure—Participants were tested in the same lab, with the same equipment, as in Experiment 1. The experiment consisted of two blocks, one for consonant processing and the other for tone processing. Each block contained two homogeneous lists and two orthogonal lists. Each list included 20 tokens (consonants or tones) and every token in each list was repeated twice. Thus, the participants received 160 words in each block (4 lists $\times$ 20 tokens $\times$ 2 repetitions).

The participants were instructed to indicate whether the vowel of each word was /a/ or /ɤ/ by pressing two corresponding keys. They were given written instructions and the two critical phonemes were specified in Pinyin (*a* and *ɤ*), which is the official system to transcribe Chinese characters into the Roman alphabet in China. They were told to ignore the consonants and tones, and to make their decisions as quickly and accurately as possible. Each trial began 750ms after the participant had responded. If a participant failed to respond within 3000ms, a new trial would be presented.

The experiment was preceded by a practice. Eight practice trials were included, and all of the critical contrasts in the experiment were tested in the practice. No feedback was given to the participants. The order of the two blocks and the four lists was counter-balanced across participants. The order of tokens within each list was random.

Results and Discussion

Reaction times were measured from the onset of each word. Reaction times lower than 200 ms or greater than three standard deviations from the mean were discarded (1.0% of all responses).

/ts'/-/tɕ' analysis—The overall accuracy rate of vowel identification was quite high, .97 for Guangzhou Cantonese speakers and .99 for Mandarin speakers. Thus, we focus on the reaction times. The average reaction time for each type of list was calculated for the two groups of participants (Cantonese vs. Mandarin) separately. Figure 3 shows the average reaction times for the consonant contrast. We conducted a 2 (Group: Guangzhou Cantonese vs. Mandarin) $\times$ 2 (List: homogeneous vs. orthogonal) mixed analysis of variance on the reaction times for the /ts' /-/tɕ' / contrast. The main effect of Group was not significant, $F < 1$. The main effect of List was significant, $F(1,26) = 32.22$, $p < .001$, $\eta^2 = .55$, reflecting the faster reaction times for homogeneous lists than for orthogonal lists (574ms vs. 629ms). Critically, we observed a significant interaction between Group and List, $F(1,26) = 16.85$, $p < .001$, $\eta^2 = .39$.

In order to clarify the nature of the critical interaction, we analyzed the simple effect of List for each group. The simple effect of List for the Cantonese speakers was not significant, $F(1,26) = 1.24$, $p = .277$. In contrast, there was a robust effect for the Mandarin speakers, $F(1,26) = 47.83$, $p < .001$, with slower responses in orthogonal lists than homogeneous lists (665ms vs. 569ms). This pattern clearly shows that Mandarin speakers suffered interference from the consonantal variation, while the Cantonese speakers did not.

T2–T3 analysis—Again, the overall accuracy rate of vowel identification was quite high, .97 for Guangzhou Cantonese speakers and .98 for Mandarin speakers. Thus, as with the consonants, we focus on the reaction times. Figure 4 shows the means for the two groups in the two conditions. A 2 (Group: Guangzhou Cantonese vs. Mandarin) $\times$ 2 (List: homogeneous vs. orthogonal) mixed analysis of variance on reaction times showed that

there was no main effect of Group, $F < 1$. The main effect of List was significant, $F(1,26)=67.10$, $p < .001$, $\eta^2=.72$, due to faster reaction times for homogeneous lists than for orthogonal lists (469ms vs. 504ms). For our central theoretical question, the critical result is the interaction between Group and List, and this was significant, $F(1,26)=5.33$, $p=.029$, $\eta^2=.17$.

Simple effect analysis revealed that both the Guangzhou Cantonese speakers [$F(1,26)=17.30$, $p < .001$] and the Mandarin speakers [$F(1,26)=55.13$, $p < .001$] took longer to respond in orthogonal lists than in homogeneous lists. This indicates that both Cantonese and Mandarin speakers suffered interference from the tonal variation when judging the vowels, but the Mandarin speakers (472ms vs. 517ms) were more affected than the Cantonese speakers were (466ms vs. 490ms).

Discussion—Experiment 2 investigated the perception of non-native consonantal and tonal contrasts using an implicit method, and the results nicely complement those obtained in the explicit discrimination task of Experiment 1. When the Mandarin /ts' /-tɕ' / consonant contrast varied in a task-irrelevant way, Cantonese speakers were unaffected, whereas the Mandarin speakers were significantly slower than when the consonant was kept constant. The fact that the Cantonese speakers were not influenced by the /ts' /-tɕ' / variation suggests that they are not sensitive to the difference between Mandarin /ts' / and /tɕ' /. In Experiment 1, this contrast was difficult for them to discriminate, and the reasonably good accuracy that they showed there may have been based on attending to subtle (acoustic) differences that are not usable outside of the simple comparison situation of an explicit discrimination task.

As in Experiment 1, the difference between the two groups was smaller for the tonal contrast than for the consonantal contrast. Both the Guangzhou Cantonese and the Mandarin speakers were slower on the orthogonal lists than on the homogeneous lists. Still, the expected difference was found: Variation of tone was more disruptive for the Mandarin subjects than for the native Cantonese speakers.

Both the explicit task in Experiment 1 and the implicit task in Experiment 2 provide evidence that Guangzhou Cantonese speakers cannot easily distinguish the Mandarin consonant /ts' / from the Mandarin consonant /tɕ' /; both experiments suggest somewhat better but still relatively poor abilities to perceive the Mandarin T2–T3 contrast. Experiment 3 was designed to investigate how the Guangzhou Cantonese speakers lexically represent minimal pairs of words differing only in each of these contrasts.

Experiment 3

The first two experiments focused on the perceptual processing of Mandarin consonantal and tonal contrasts by native Cantonese speakers who had learned Mandarin at an early age. Experiment 3 is designed to examine the lexical representation of these phonological contrasts. Specifically, we wish to determine whether minimal pairs of words containing consonants or tones that are very poorly discriminated by the Guangzhou Cantonese speakers are stored as homophones in their lexical representations.

We use a medium-term auditory repetition priming paradigm (Dufour et al., 2007; Pallier et al., 2001; also see Sumner & Samuel, 2009) to address this question. In this paradigm, participants make lexical decision judgments on a series of words and non-words. Critically, in some cases a word is later followed by either a true repetition, or by an item that only differs by one critical feature. The medium-term repetition priming effect is defined as a reaction time decrease between the first and second occurrences of a repeated item. For the current purposes, this definition also includes the occurrence of an item and its counterpart

in a minimal pair. Pallier et al. and Dufour et al. both found a significant priming effect in the minimal pair condition when the difference between the two items was a non-native segmental distinction (and, Pallier et al. showed that this effect only occurred for words, indicating that the task taps lexical effects). Based on this effect, these authors concluded that minimal pairs of words were stored as homophones in the lexical representations of non-native participants.

In contrast, as discussed in the Introduction, researchers using the eye-tracking methodology (Cutler & Weber, 2004; Cutler, Weber, Otake, 2006; Escudero, Hayes-Harb, Mitterer, 2008) have concluded that in these cases lexical representations include different representations for the two sounds. This conflicting conclusion is based on the presence of asymmetric competition effects: Words with native sounds compete with those having non-native segments more than vice versa. In the Pallier et al. (2001) and Dufour et al. (2007) priming studies, the researchers did not analyze the two opposite priming directions in the minimal pair conditions separately. As such, whether the two members of a minimal pair produce symmetric priming remains uncertain.

Sumner and Samuel (2009) tested speakers of the New York City/Long Island dialect of American English to determine whether their lexical representations included dialectal variants or the forms found in General American English. Using a priming methodology similar to that of Pallier et al. (2001) and Dufour et al. (2007), they found symmetric priming for individuals who overtly produced the local dialect, but found asymmetric priming for individuals in the same linguistic community whose speech did not include the local phonetic variation. These results indicate that the priming technique can reveal differences in the details of lexical representations. Thus, in Experiment 3 we tested and analyzed both orders of presenting the two members of a minimal pair, to see whether there would be an asymmetric pattern of priming. As research using the eye-tracking approach has shown, asymmetries indicate that two forms are not homophonous.

The critical contrasts in Experiment 3 were those that we have shown to be relatively difficult for the Guangzhou Cantonese speakers to discern: /ts' /- /tɕ' / and T2–T3. The results of Experiments 1 and 2 indicate that the /ts' /- /tɕ' / contrast is extremely difficult for the Cantonese speakers, and the T2–T3 contrast is moderately difficult. For comparison purposes, we also tested a consonantal contrast that exists in both Cantonese and Mandarin (/t /- /t' /) and a tonal contrast (T2–T4) that was easily distinguished by both groups in Experiment 1. Both groups of participants should be able to distinguish these phonetic and prosodic contrasts easily, providing a baseline with which we can interpret the two critical cases.

Method

Participants—Thirty-six native Guangzhou Cantonese speakers (mean age of 22 years, ranging from 16 to 29 years) and 36 native Mandarin speakers (mean age of 23 years, ranging from 18 to 26 years) participated in Experiment 3. All were tested in the same lab, with the same equipment, as in the previous experiments. Six Cantonese speakers and four Mandarin speakers had participated in Experiment 1; none had participated in Experiment 2. Participants were recruited in the same way as in the previous experiments. The Guangzhou Cantonese native speakers had been using Mandarin Chinese for an average of 15 years, ranging from 8 to 22 years. No participants reported any hearing problem, and all received small gifts for their participation.

Stimuli—Twenty minimal pairs of disyllabic words based on each difficult contrast (/ts' /- /tɕ' / and T2–T3), and 20 minimal pairs of words based on each control contrast (/t /- /t' / and T2–T4) were initially selected as possible experimental stimuli. For example, the words /ts

ʼi(2)t'ɑŋ(2)/ [“祠堂”, meaning *temple*] and /tɕ'ɿ(2)t'əŋ(2)/ [“池塘”, meaning *pond*] comprise a minimal pair that differs only in the critical consonantal contrast. There was no semantic relationship between the two members in any pair of words. To make sure that participants would accept the experimental stimuli as real words, 22 students who did not take part in the present experiment were asked to make lexical decisions on the 80 words. Only words that were rated as real words by more than 80% of the students were kept. Based on this pilot testing, 16 pairs of words for each contrast were selected as critical stimuli; the tested consonants and tones were present in either the first or second syllable in approximately equal frequency. The median frequencies for the /ts'ɿ/ and /tɕ'ɿ/ words were 195 and 210, respectively. The median frequencies for the /t/ and /t'/ words were 400 and 327, respectively. The median frequencies for the T2 and T3 words were 596 and 552, respectively. The median frequencies for the T2 and T4 words were 511 and 407, respectively (per million, retrieved from CCL corpus: http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai). Appendix D presents the critical stimuli used in Experiment 3.

For the lexical decision task, 64 minimal pairs of nonwords were created in the same way as the critical stimuli, 16 for each contrast. An additional 64 words and 64 nonwords were used as fillers. No critical consonants were included in the fillers. Since there are only four tones in Mandarin and three were tested in the critical stimuli, all Mandarin tones were used in the fillers.

All stimuli were recorded by a native Mandarin speaker. All items were digitized (44.1kHz) and edited with GoldWave 5.22 sound editing software.

Procedure—The participants were tested individually in a sound-attenuated booth. The stimuli were presented via headphones at a comfortable listening level. Four counterbalanced lists of the 384 items ([16 pairs of words × four contrasts] + [16 pairs of nonwords × four contrasts] + 64 filler words + 64 filler nonwords) were created and each participant was assigned to one list. In each list, one member of the minimal pair was followed, 8 to 20 items further down in the list, either by itself or by the other member in the pair. The members of a given minimal pair appeared in the same positions in the four lists, but were counterbalanced across the lists. The participants were required to make a lexical decision for each item by pressing corresponding keys. They were told to respond as quickly and accurately as possible. The following trial began 750ms after the participant had responded. If a participant failed to respond within 3500ms, a new trial would be presented. The experiment was preceded by a practice. Five words and five non-words that were not tested in the experiment were used, with no feedback given to the participants.

Each participant was tested in four prime conditions. An example of the four conditions for the /ts'ɿ-/tɕ'ɿ/ contrast is provided in Table 3. As the table shows, the four prime conditions represent the crossing of /ts'ɿ/ and /tɕ'ɿ/ as the first member of a pair with /ts'ɿ/ and /tɕ'ɿ/ as the second member of the pair.

Results and Discussion

Responses to fillers were not included in the analyses. The data of four participants (two Mandarin speakers and two Guangzhou Cantonese speakers) were excluded because of very high error rates (more than 25%). The average error rate for the remaining 68 participants was less than 9%. Reaction times lower than 200 ms or greater than 3 standard deviations from the mean were discarded (3.0% of all responses). The present experiment was a 2 (Group: Mandarin vs. Cantonese) × 4 (Prime Condition: 2 identity conditions vs. 2 minimal pair conditions) × 2 (Occurrence: first vs. second) mixed design. For each contrast, the difference between the first and second occurrence (the priming effect) was assessed within

four prime conditions for each group. Accuracy was consistently very high. In all cases no significant accuracy effects were observed for either words or nonwords, and no significant reaction time effects were observed for nonwords. Thus, we focus here on the reaction times for the critical words, as is typically the case in lexical decision studies. *F* values are reported by participants (*F*₁) and by items (*F*₂). We report analyses of the priming effects here; in Appendix E, we report analyses of the overall reaction times.

Consonant contrasts: /tsʹ/-/tɕʹ/ analysis—Priming effects in each condition are shown in Figure 5. We first consider identity priming -- cases in which the same token occurred twice. Previous studies using medium- and long-term priming indicate that such repetition should lead to faster responses for the second occurrence. As the left side of Figure 5 shows, such repetition priming was robust. A strong priming effect was observed when /tsʹ/ words were repeated (/tsʹ/-/tsʹ/ condition) for both the Guangzhou Cantonese speakers [*F*₁(1,66)=11.46, *p*=.001; *F*₂(1,30)=5.39, *p*=.027] and the Mandarin speakers [*F*₁(1,66)=10.75, *p*=.002; *F*₂(1,30)=7.64, *p*=.010]. There was also a priming effect when /tɕʹ/ words were repeated (/tɕʹ/-/tɕʹ/ condition) for both the Guangzhou Cantonese speakers [*F*₁(1,66)=5.21, *p*=.026; *F*₂(1,30)=14.53, *p*=.001] and the Mandarin speakers [*F*₁(1,66)=4.93, *p*=.030; *F*₂(1,30)=3.49, *p*=.071].

Having established that our testing conditions are sensitive to item repetition, we can see whether minimal pairs were treated as cases of repetition. For the Mandarin speakers, there should not be repetition priming because the first and second members in a pair are different lexical items in Mandarin. The results are consistent with this expectation: No priming effect was found for the minimal pairs when /tɕʹ/ words preceded their /tsʹ/ counterparts for the Mandarin speakers [*F*₁<1; *F*₂<1]; the same null effect obtained when a /tsʹ/ word preceded its /tɕʹ/ minimal pair [*F*₁<1; *F*₂(1,30)=1.87, *p*=.181].

As Figure 5 shows, the results for the native Cantonese speakers were quite different. There was a very large priming effect when /tɕʹ/ words preceded their /tsʹ/ counterparts [*F*₁(1,66)=9.94, *p*=.002; *F*₂(1,30)=5.703, *p*=.023]. This large priming effect indicates that for the Cantonese listeners, the minimal pairs were not distinct. However, if the pairs were stored as homophones (as the large /tɕʹ/-/tsʹ/ priming effect might suggest), then we should see a similarly large priming effect when /tsʹ/ words preceded their /t/ minimal pairs, and the results are not consistent with this symmetric hypothesis: There was a priming effect of approximately 60 ms, but this trend did not approach significance [*F*₁(1,66)=1.63, *p*=.21; *F*₂<1]. Thus, numerically there is a clear asymmetry. However, the 60 ms trend requires caution in our interpretation.

Given the importance of this comparison for our inferences regarding lexical representation of this difficult non-native contrast, and the indeterminate result, we brought in a new group of 24 native Guangzhou participants and had them do the same task as the original group. These participants were drawn from the same population, and were tested on the same equipment in the same lab. Two of these new participants exceeded our cutoff of 25% errors, leaving 22 usable data sets.

The new group showed no hint of priming in the /tsʹ/-/tɕʹ/ condition, actually producing a 13 ms reversal of the priming pattern. We conducted an ANOVA of the priming effects in the four conditions, with Replication (the original group of Cantonese participants vs. the new group of Cantonese participants) as a factor. Figure 6 presents the average priming effects for the four conditions, combining the results of the original and the new Cantonese group. Again, the two identity priming conditions exhibited large priming effects (for the /tsʹ/-/tsʹ/ condition, *F*₁(1, 55)=14.13, *p*<.001; *F*₂(1, 31)=8.91, *p*=.005; for the /tɕʹ/-/tɕʹ/ condition, *F*₁(1, 55)=19.81, *p*<.001; *F*₂(1, 31)=53.81, *p*<.001. As the figure makes clear,

the /tʂʰ/-/tʂʰ/ condition also produced a large priming effect, $F_1(1,55) = 11.81, p = .001$; $F_2(1,31) = 2.81, p = .10$. This priming effect was indistinguishable from both the /tsʰ/-/tsʰ/ identity condition ($F_1 < 1$; $F_2 < 1$) and the /tʂʰ/-/tʂʰ/ identity condition ($F_1 < 1$; $F_2 < 1$). In contrast, the /tsʰ/-/tʂʰ/ condition produced a negligible effect ($F_1 < 1$, $F_2 < 1$), which was reliably different from the average priming effect of the other three conditions, $F_1(1,54) = 5.20, p = .027$, $F_2(1,30) = 6.30, p = .018$.

Thus, using the larger sample, we see clear evidence for asymmetric priming effects, a pattern that is inconsistent with the members of the minimal pairs being stored as homophones. Just as the results of eye-tracking studies have suggested (Cutler & Weber, 2004; Cutler, Weber, Otake, 2006; Escudero, Hayes-Harb, Mitterer, 2008), contrasts that are very difficult to distinguish at the phonetic level are nonetheless represented differentially at the lexical level.

Consonant contrasts: /t/-/tʰ/ analysis—Priming effects in each condition for these consonantal controls are shown in Figure 7. Priming effects were found in the two identity priming conditions for both groups. In the /t/-/t/ condition, the priming effect was significant for both the Guangzhou Cantonese speakers [$F_1(1,66) = 9.02, p = .004$; $F_2(1,30) = 16.47, p < .001$] and the Mandarin speakers [$F_1(1,66) = 12.48, p = .001$; $F_2(1,30) = 13.93, p = .001$]. In the /tʰ/-/tʰ/ condition, the priming effect was also robust for both the Guangzhou Cantonese group [$F_1(1,66) = 19.69, p < .001$; $F_2(1,30) = 18.14, p < .001$] and the Mandarin group [$F_1(1,66) = 5.86, p = .018$; $F_2(1,30) = 4.71, p = .038$].

As the /t/-/tʰ/ pairs are clearly different words, with no phonetic ambiguity even for the Cantonese speakers, there should not be any priming for these minimal pairs. This expectation was confirmed: No priming effect was found in either the /t/-/tʰ/ condition [for the Guangzhou Cantonese speakers: $F_1(1,66) = 1.66, p = .202$; $F_2(1,30) = 1.16, p = .291$; for the Mandarin speakers: $F_1(1,66) = 3.15, p = .081$, $F_2 < 1$] or in the /tʰ/-/t/ condition [for the Guangzhou Cantonese speakers: $F_1(1,66) = 2.44, p = .123$; $F_2 < 1$; for the Mandarin speakers: $F_1 < 1$, $F_2 < 1$].

Tonal contrasts: T2–T3 analysis—Priming effects in each condition are shown in Figure 8. As expected, priming effects were found in the two identity priming conditions for both groups. In the T2–T2 condition, the priming effect was significant for the Guangzhou Cantonese speakers [$F_1(1,66) = 4.96, p = .029$; $F_2(1,30) = 29.52, p < .001$] and the Mandarin speakers [$F_1(1,66) = 10.28, p = .002$; $F_2(1,30) = 5.26, p = .029$]. The same was true in the T3–T3 condition, with a priming effect for both the Guangzhou Cantonese speakers [$F_1(1,66) = 22.82, p < .001$; $F_2(1,30) = 8.52, p = .007$] and the Mandarin speakers [$F_1(1,66) = 10.98, p = .001$; $F_2(1,30) = 13.63, p = .001$].

We saw that with the consonantal minimal pairs there were different results for the native Mandarin speakers (no priming) and the Cantonese speakers (priming, in one direction). The tonal manipulation produced different results: No priming effect was found in either the T2–T3 condition [for the Guangzhou Cantonese speakers: $F_1 < 1$, $F_2 < 1$; for the Mandarin speakers: $F_1(1,66) = 1.54, p = .219$; $F_2(1,30) = 1.35, p = .225$] or in the T3–T2 condition [for the Guangzhou Cantonese speakers: $F_1 < 1$, $F_2 < 1$; for the Mandarin speakers: $F_1 < 1$, $F_2 < 1$]. The results for the Mandarin speakers mirror those for the consonantal case, and reflect the different lexical status of the members of each word pair. It is interesting that the same non-priming result is found for the Cantonese speakers. This pattern is consistent with the results of the first two experiments, which suggested better discrimination of the tonal contrast than the consonantal one.

Tonal contrasts: T2–T4 analysis—Priming effects in each condition for the tonal controls are shown in Figure 9. Priming effects were found in the two identity priming conditions for both groups. In the T2-T2 condition, the priming effect was significant for the Guangzhou Cantonese speakers [$F(1,66)=11.33$, $p=.001$; $F(1,30)=6.05$, $p=.001$] and the Mandarin speakers [$F(1,66)=5.96$, $p=.017$; $F(1,30)=9.58$, $p=.002$]. In the T4-T4 condition, a priming effect was obtained for the Guangzhou Cantonese speakers [$F(1,66)=6.97$, $p=.010$; $F(1,30)=7.457$, $p=.020$] and the Mandarin speakers [$F(1,66)=4.05$, $p=.048$; $F(1,30)=5.25$, $p=.029$]. As we would expect, given the clear discriminability of T2 and T4 for both groups, no priming effect was found in either the T2–T4 condition [for the Guangzhou Cantonese speakers: $F(1,66)=1.90$, $p=.173$; $F(1,30)=2.75$, $p=.108$; for the Mandarin speakers: $F<1$, $F<1$], or in the T4-T2 condition [for the Guangzhou Cantonese speakers: $F(1,66)=1.06$, $p=.307$; $F(1,30)=3.40$, $p=.075$; for the Mandarin speakers: $F<1$, $F<1$].

Discussion—The priming effects varied in systematic ways between the two groups of speakers and among the different minimal contrasts. Both the Cantonese and the Mandarin speakers exhibited strong and consistent repetition priming when the same word was repeated. In contrast, the effects were diverse in the minimal pair conditions across different types of contrasts. As expected, the Mandarin speakers showed no repetition priming for an item and its counterpart in a minimal pair, for all contrasts that were tested. This result confirms that Mandarin speakers were capable of distinguishing these contrasts -- the minimal pairs differing in a consonant or tone were different words for them, and hence there was no lexically-mediated medium term priming.

The Guangzhou Cantonese speakers produced a rich pattern of priming effects for the minimal pairs, providing insight into the lexical representations for a sibling learned at an early age. The /t/-/t'/ control condition confirmed that Mandarin minimal pairs that are based on a good phonetic contrast for the Cantonese subjects produce no priming. Thus, any priming effects reflect shared lexical representations between the first and second members of a pair. In this context, the robust priming effect in the minimal pair condition for the /ts'/-/ts/ contrast indicates that this contrast is not maintained in the lexicon for these early non-native Mandarin learners. However, the lack of priming for the /ts'/-/ts'/ case shows that these minimal pairs are not homophones – if they were, we would have found robust priming for such “repetition”, and we did not. This pattern of asymmetric priming is reminiscent of the asymmetric eyetracking patterns reported by Cutler et al. (2006) and Weber and Cutler (2004). Recall that those authors suggested that such an asymmetry implies that only one member of the pair appears in the lexical representations, one that is closer to the subject’s native language phoneme (in this case, /ts'/). In the General Discussion, we will consider how the different processing requirements for the eyetracking and priming paradigms constrain the theoretical implications for lexical representation and processing.

The results for the tonal conditions were extremely clear. When the tones were kept constant in a pair, robust repetition priming was consistently observed. When tone was varied between the first and second member of a minimal pair (keeping the segmental information constant), neither the Guangzhou Cantonese nor the Mandarin speakers exhibited repetition priming. This was true for the critical T2–T3 case and for the T2–T4 control contrasts. These results indicate that words differing in tones are represented as different lexical items for both Guangzhou Cantonese and Mandarin speakers.

General Discussion

The goal of the present study was to investigate how consonantal and tonal contrasts are perceived and represented by experienced non-native speakers who learn a sibling at an early age. Guangzhou Cantonese and Mandarin participants were tested in three experiments. We first used an explicit discrimination methodology to provide a strong empirical basis for stimulus selection for the subsequent experiments. We then combined the Garner methodology with the medium term priming paradigm to explore the phonetic and lexical processing of non-native contrasts. We believe that this kind of interlocking approach provides a richer picture of non-native language processing than earlier studies have provided.

Comparison of non-native segmental and prosodic processing

The results of Experiment 1 revealed that the Guangzhou Cantonese speakers had difficulty distinguishing specific Mandarin consonantal contrasts, particularly the /ts/-/tʂ/ and /ts' /-tʂ' / contrasts. These results are consistent with the results for another pair of siblings, Catalan and Spanish (Pallier et al., 1997; Pallier et al., 2001; Sebastián-Gallés & Soto-Faraco, 1999; Sebastián-Gallés et al., 2005; Sebastián-Gallés et al., 2006): Non-native speakers are poor at distinguishing certain segmental contrasts despite their otherwise high proficiency. The Guangzhou Cantonese speakers who participated in our research have a good command of Mandarin Chinese. They started learning it at an early age -- six or seven -- which is within the age range usually assumed to allow mastering a second language. Nonetheless, their perception of certain Mandarin consonantal contrasts was significantly worse than their native counterparts. These explicit discrimination difficulties were reflected in both the implicit perceptual measure of phonological processing in Experiment 2 and the lexical priming effects in Experiment 3.

Because the same individuals were tested on both segmental and tonal contrasts, the present research provides the first direct comparison of dialect effects for tonal versus segmental contrasts. The three experiments were designed to provide this comparison at increasingly abstract levels of processing: The discrimination test in Experiment 1 can be done on an acoustic basis, the Garner interference of Experiment 2 can be generated at a phonological (or for tones, suprasegmental) level, and the priming effects in Experiment 3 are lexically driven (Pallier et al., 2001).

Generally speaking, the results for the tonal contrasts in the three experiments could be characterized as an attenuated version of what was found for the consonantal cases, but with important variation as a function of the tested level. On the explicit discrimination test, the Guangzhou Cantonese speakers demonstrated better tone perception than segment perception; in both cases their perceptual ability was worse than that of the native Mandarin speakers, but the difference was smaller for tones than segments. The pattern of Garner interference also showed stronger differences for segmental contrasts than for tonal, in an interesting way. Unlike the native Mandarin speakers, the non-native speakers suffered no interference from variation of /ts' /-tʂ' /, indicating that during perception they are not sensitive to this phonological contrast. When the variation was based on tone, both native and non-native speakers were impaired by the variation, with native speakers showing significantly more interference than the non-native ones. At the lexical level, we again saw evidence that non-native speakers do not represent the segmental contrast well, leading to significant priming from words (the /tʂ' /-ts' / case) that did not produce such priming for the native speakers. Again, for the non-native speakers the tonal contrast was better represented than the segmental one, with the two groups of speakers showing similar patterns of lexical activation for words differing in tones.

The Guangzhou Cantonese speakers might be better at processing Mandarin tones than consonants because of particular aspects of their language experience with Mandarin Chinese. This suggestion is grounded in research indicating that non-native tones seem to be easier to learn than non-native segments. English speakers' identification accuracies of Mandarin tones increased from 69% to 90% after several weeks of training (Wang, Spence, Jongman & Sereno, 1999), and identification of non-native tones increased from a baseline of 50–59% to at least 81–88% after intensive training of only a matter of hours (So, 2006a). With certain training methods, performance was as high as 91%–97% (So, 2006a). Importantly, these learning effects were found to be long lasting.

These large training effects for tones contrast with the much smaller changes observed in attempts to improve segment perception. As we have discussed, for contrasts such as English /ɹ/-/l/, the effect of training can be impressively weak. Three weeks of training for Japanese speakers trying to identify English /ɹ/ and /l/ only yielded an increase from 78% to 86% (Logan et al., 1991). Other training studies also found only a 12% (Lively et al., 1994) or a 16% (Bradlow et al., 1997) improvement in non-native consonant learning. The Guangzhou Cantonese speakers in the current study received comprehensive education in Mandarin Chinese beginning at an early age. Given the tone training results, we would expect them to be able to process Mandarin tones relatively well, and the results are consistent with this.

In this context, it is worth noting that our Cantonese listeners showed little of the perceptual difficulties experienced by the Cantonese participants in So's studies (2005a; 2005b; 2006a; 2006b). A critical difference is that So tested Hong Kong Cantonese speakers, whereas our listeners were Guangzhou Cantonese speakers. There are some differences between the Hong Kong Cantonese and Guangzhou Cantonese tone systems, but more importantly, there are differences between the language experiences of the two populations. The Hong Kong Cantonese speakers who participated in So's studies were naïve listeners of Mandarin Chinese — Hong Kong speakers do not have much education or training in Mandarin. In contrast, our participants were quite proficient in Mandarin, despite its being non-native.

We noted in the Introduction that Best et al.'s PAM model has the potential to be applied to the case we are studying. Overall, the model does seem to be able to offer at least a qualitative account of how our Cantonese listeners perceived the non-native Mandarin contrasts. Cantonese does not have the consonants /tʂ/, /tʂ'/, /ʂ/, /tʂ/, /tʂ' / and /ʂ/, and although /ts/, /ts' / and /s/ exist in both Mandarin and Cantonese, they are pronounced somewhat differently in the two sibilants (Bauer & Benedict, 1997). In Experiment 1 the Guangzhou Cantonese speakers were significantly worse than native Mandarin speakers at discerning the /ts/-/tʂ/, /ts' /-/'tʂ' /, /s/-/'ʂ/, /ts' /-/'tʂ/, /ts' /-/'tʂ' /, and /ts' /-/'ʂ/ contrasts. They may have assimilated Mandarin /ts/ and /tʂ/ to Cantonese /ts/, Mandarin /ts' / and /tʂ' / to Cantonese /ts' /, and Mandarin /s/ and /ʂ/ to Cantonese /s/. However, the two members of each contrast would likely differ in their goodness of fit to the Cantonese consonants. Mandarin /ts/, /ts' /, and /s/ should be relatively good exemplars of the Cantonese categories, while Mandarin /tʂ/, /tʂ' / and /ʂ/ should be poorer exemplars. Similarly, Mandarin /tʂ/, /tʂ' / and /ʂ/ could be assimilated to Cantonese /ts' /, but they are all poor instances of this category. PAM would predict that single category assimilation contrasts would be relatively difficult for the Cantonese speakers to perceive. In contrast, the Cantonese speakers were relatively good at discriminating the other six Mandarin consonantal contrasts in Experiment 1 (/ts/-/'tʂ/, /ts' /-/'tʂ' /, /s/-/'ʂ/, /ts' /-/'tʂ/, /ts' /-/'tʂ' /, and /ts' /-/'ʂ/). One reason may be that although we used the same syllable-final vowel (i.e., *i*) with all of the consonants, this vowel sounds slightly different when following /ts/, /ts' /, and /s/ versus /tʂ/, /tʂ' /, and /ʂ/, with the former more like [ɿ] and the latter more like [i]. This difference might have helped participants distinguish the contrasts, but it is interesting to note that the discrimination

between /ts'/ and /tɕ/, /tɕ'/, /ɕ/ was rather poor. PAM provides a better prediction for this result. It is plausible that Mandarin /tɕ/, /tɕ'/ and /ɕ/ would all be assimilated to Cantonese /ts'/ while Mandarin /ts/ and /s/ are assimilated to Cantonese /ts/ and /s/, respectively. According to PAM, if the two members of a contrast can be assimilated to different native categories, they should be well discriminated, consistent with our findings.

PAM can also provide a reasonable explanation for our tone results. The Guangzhou Cantonese speakers experienced less difficulty in discriminating Mandarin tonal contrasts than segmental ones. The largest effect of non-native experience was that the Cantonese speakers took longer to discriminate Mandarin Tones 2 and 3 than the native Mandarin subjects. Thus, the native dialect had some effect, but the constraint is much smaller than that for the consonant system. Guangzhou Cantonese has a similar tone (35) to that of Mandarin Tone 2 (35), but not to Mandarin Tone 3 (214). Therefore, while Mandarin Tone 2 could have been assimilated to Cantonese 35, the situation for Tone 3 is more difficult. So (2006b) has suggested that Mandarin Tone 3 (214) might also be assimilated to Cantonese 35, or to Cantonese 23; others have suggested it might be more similar to Cantonese 33 (Francis et al., 2008). Given this unsettled analysis, in PAM's taxonomy of possible relationships of the two members of a contrast, Mandarin Tones 2 and 3 might best be considered a "Categorized (Tone 2) – Uncategorized (Tone 3)" contrast because Tone 3 could be assimilated to more than one Cantonese tone category. PAM predicts that contrasts of this type should be distinguished moderately well, which is what we observed.

It is also possible that Cantonese speakers have difficulty distinguishing this tonal contrast because they are more sensitive to the average height of the pitch of a tone than the direction of the pitch, while Mandarin speakers weigh direction more than height (Francis, et al., 2008; Gandour, 1983; Gandour & Harshman, 1978; Guion & Pederson, 2007). Since Mandarin Tone 2 (35) and Tone 3 (214) start and end with similar f0 heights, but differ in the direction of f0 change, it is reasonable to expect that Cantonese speakers will find this contrast difficult.

The lexical representation of non-native contrasts

There is a small existing literature that has examined how non-native contrasts are stored in the lexicon. Experiment 3 was designed to provide new evidence to address this issue, using medium term repetition priming. The task's sensitivity was confirmed by the robust identity priming effects in every condition, for both native and non-native listeners. In contrast, when a lexical tone was varied between the first and second member of a minimal pair, neither the Guangzhou Cantonese nor the Mandarin speakers displayed a priming effect. These results indicate that words differing in tones are represented as different lexical items for both early learners and native speakers. Given that training studies show relatively good learning of tones compared to learning of non-native segmental contrasts, it makes sense that native Cantonese speakers who have had early and extensive experience with Mandarin would show evidence of incorporating the proper tones in their lexical representations.

For the segmental /ts' /- /tɕ' / contrast, the native Mandarin speakers produced a pattern of priming that is exactly as one would expect: Because two tokens that differ in this segmental contrast are different Mandarin words, prior presentation of one does not prime later presentation of the other. The question posed in Experiment 3 was whether Guangzhou Cantonese speakers would show priming effects for these minimal pairs, given that Experiments 1 and 2 showed /ts' /- /tɕ' / to be perceptually almost indistinguishable for them. Previous priming studies of the Catalan-Spanish sibilants (Pallier et al., 2001) and of two dialects of French (Dufour et al., 2007) found such priming for minimal pairs. Here, when the member of the minimal pair with /tɕ' / preceded the member with /ts' /, we found robust priming, with effects that were comparable to those produced by identity priming (see

Figure 6). These results converge with those of the prior priming studies and appear to suggest that even early and highly proficient non-native speakers treat the non-native sound (e.g., /tʂʰ/) as though it were their closest-matching native sound (e.g., /tsʰ/).

However, such a conclusion is challenged by the lack of priming when the prime-target roles were reversed: Prior presentation of the minimal pair token with /tsʰ/ did not prime the corresponding target token with /tʂʰ/. Experiment 3 provided a richer pattern of priming effects than was available in the Dufour et al. (2007) and Pallier et al. (2001) studies because our experiment allowed us to look for possible asymmetrical priming. This design was motivated by the asymmetrical competition effects found in eye-tracking studies (Cutler, Weber, & Otake, 2006; Weber and Cutler, 2004). In the eye-tracking studies, words that contain a segment that is similar to the subject's native segment (e.g., /e/ for a Dutch subject listening to L2 English) compete against words that contain a non-native segment (e.g., /æ/ for the Dutch subject), but not vice-versa. An asymmetry of this type rules out any strictly homophonous view of minimal pairs of this sort.

Cutler, Weber, and Otake (2006) suggested that the asymmetric competition pattern was consistent with the L1 phonetic category (for our listeners, /tsʰ/) being dominant relative to the L2 phonetic category (in this case, /tʂʰ/). The phonetic dominance in turn gives words with the L1 segment an advantage in contacting lexical representations, producing the asymmetric competition. This hypothesis is generally consistent with the pattern of medium term priming effects we observed for the Cantonese listeners hearing Mandarin words, though certain details need to be specified. We suggest the following three hypotheses: (1) The “good” native-like sound (here, /tsʰ/) increases the activation of the (strong) phonetic representation for /tsʰ/, and has little or no effect on the (weak) phonetic representation for /tʂʰ/. (2) The “poor” non-native sound (here, /tʂʰ/) activates both the most similar L1 phonetic representation (/tsʰ/) and the weaker L2 phonetic representation (/tʂʰ/). One way to instantiate the idea of strong/weak phonetic representations could be to have higher baseline activation levels for the strong native sounds than for the weak L2 sounds. (3) Because the L1 phonetic representation is stronger, there will initially be more activation of lexical candidates compatible with the L1 interpretation, but over the course of a few hundred milliseconds the acoustic mismatch may eventually produce stronger activation of the L2 word. This will depend on the acoustic differences between the two sounds, and the level of expertise that the listener has achieved in L2.

These hypotheses are consistent with Cutler et al.'s (2006) model, and thus are consistent with the eyetracking results (e.g., there is some evidence that an L1 competitor may initially be more active than the actual L2 target). In the medium-term priming paradigm these hypotheses also fit the observed asymmetric priming: When the prime contains the L1 sound, it will only produce substantial activation of the L1 phonetic representation and the corresponding lexical item. As a result, when a later target word occurs, there will be priming if it also has the L1 sound, but not if the target is the L2 minimal pair item; because there had been no activation of the L2 word when the prime occurred, there is no priming later. In contrast, when the prime has the weaker L2 segment in it, our hypothesis says that this will activate both the L1 and L2 phonetic representations, and the corresponding L1 and L2 words. Therefore, when a later target word occurs, there will be priming regardless of which member of the minimal pair it is, because both were activated previously. Note that this model posits different lexical representations for the two members of the minimal pair. The dominance of the L1 phonetic representation presumably underlies the very poor performance on both explicit and implicit discrimination tasks (Experiments 1 and 2).

It remains to be determined what the nature is of the L2 phonetic representation, and thus the nature of the lexical code for L2 words, but a plausible possibility is that they contain all of

the “clear” features of the segment, with the difficult feature remaining unspecified. For example, for the /tsʰ/-/tʃ/ pair, the listener presumably can tell that the L2 segment is a voiceless fricative with some kind of middle place of articulation, but without the specifics of the place.

Given the dominance of the L1 segments over the L2 segments at the phonetic level, why is the contrast represented at the lexical level? One possibility is that early learners receive repeated instructions about the distinction between minimal pairs of words when learning Mandarin in class. For example, Cantonese students may be told to keep in mind that the Mandarin word /tsʰɿ(2)tʰɑŋ(2)/ (“祠堂”, meaning *temple*) starts with the consonant /tsʰ/ and means a building devoted to the worship of ancestors, while /tʃɿ(2)tʰɑŋ(2)/ (“池塘”, meaning *pond*) starts with the consonant /tʃ/ and means a small area of still water. In addition, young Guangzhou Cantonese speakers usually use the Chinese Pinyin entry system to chat with each other when surfing on the Internet. The word /tsʰɿ(2)tʰɑŋ(2)/ is transcribed as *ci(2)tang(2)* in Pinyin and /tʃɿ(2)tʰɑŋ(2)/ is transcribed as *chi(2)tang(2)*. The different pinyin codes (*c* vs. *ch*) require them to make the distinction between minimal pairs in order to avoid typing wrong words. This type of analysis led Cutler et al. (2006) to suggest that it may be necessary to test people who have not received formal education of a non-native language to get uncontaminated results.

Conclusion

The findings in the present study indicate that the constraints of a native phonological system on perception of contrasts in a non-native dialect vary across segmental and prosodic systems. There were substantial limitations imposed by the native dialect on Cantonese listeners' ability to discriminate Mandarin consonantal contrasts, with weaker constraints at work for tonal contrasts. It is important to remember that these results were found for individuals who learned Mandarin at an early age, and who had been using it frequently for over a decade. Thus, early and extensive language experience with a non-native dialect does not guarantee native-like perception. At the same time, language experience seems to have a more substantial impact on lexical representations. Minimal pairs of words that differed in contrasts that were very poorly discriminated (Experiment 1) and that failed to generate Garner interference (Experiment 2) were nonetheless not represented as homophones at the lexical level. This suggests that extensive experience hearing non-native words can produce lexical representations that provide more information about the contrast than might be expected from assessments of the contrast at a phonological level. This pattern is most consistent with a view of lexical representations in which repeated exposure to a word builds up a lexical representation that conforms to the average phonological properties of the tokens that have been heard. This type of lexicon could allow non-native speakers to achieve rather good comprehension, even when their ability to map the lexical items onto appropriate motor patterns is limited. This is, of course, consistent with the well-established pattern of comprehension significantly leading production, for both first and second language (or dialect) acquisition.

Acknowledgments

Funding for the research reported here was provided by the Fundamental Research Funds for the Central Universities (CCNU11A02018), the funds for Human Development and Mental Health Hubei Key Laboratory (Central China Normal University) (Grant No.201007), and by NIH grant HD R01 059787.

References

- Aoyama K, Flege JE, Guion SG, Akahane-Yamada R, Yamada T. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics*. 2004; 32(2):233–250.
- Bauer, RS.; Benedict, PK. *Modern Cantonese Phonology*. Berlin: Mouton de Gruyter; 1997.
- Bent T, Bradlow AR, Wright BA. The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance*. 2006; 32(1):97–103. [PubMed: 16478329]
- Best CT, McRoberts GW, Goodell E. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Acoustical Society of America*. 2001; 109(2):775–794.
- Best CT, Strange W. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*. 1992; 20(3):305–330.
- Best, CT.; Tyler, MD. Nonnative and second-language speech perception: Commonalities and complementarities. In: Munro, MJ.; Bohn, OS., editors. *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins; 2007. p. 13–34.
- Bradlow AR, Pisoni DB, Akahane-Yamada R, Tohkura Y. Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*. 1997; 101(4):2299–2310. [PubMed: 9104031]
- Broersma, M. Comprehension of non-native speech: Inaccurate phoneme processing and activation of lexical competitors. *Proceedings of the Seventh International Conference on Spoken Language Processing*; Denver, CO. 2002. p. 261–264.
- Burnham, D.; Francis, E.; Webster, D.; Luksaneeyanawin, S.; Attapaiboon, C.; Lacerda, F., et al. Perception of lexical tone across languages: evidence for a linguistic mode of processing. In: Bunnell, HT.; Idsardi, W., editors. *Proceedings of the Fourth International Conference on Spoken Language Processing*. Vol. 1. Wilmington, DE: Applied Science and Engineering Laboratories; 1996. p. 2514–2517.
- Cai ZG, Pickering MJ, Yan H, Branigan HP. Lexical and syntactic representations in closely related languages: Evidence from Cantonese–Mandarin bilinguals. *Journal of Memory and Language*. 2011; 65(4):421–445.
- Chen X, Anderson RC, Li W, Hao M, Wu X, Shu H. Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology*. 2004; 96(1):142–151.
- Cutler A, Weber A, Otake T. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*. 2006; 34(2):269–284.
- Dufour S, Nguyen N, Frauenfelder UH. The perception of phonemic contrasts in a non-native dialect. *Journal of the Acoustical Society of America*. 2007; 121(4):EL131–136. [PubMed: 17471757]
- Escudero P, Hayes-Harb R, Mitterer H. Novel second-language words and asymmetric lexical access. *Journal of Phonetics*. 2008; 36(2):345–360.
- Flege, JE. Second language speech learning: Theory, findings, and problems. In: Strange, W., editor. *Speech perception and linguistic experience*. Baltimore: York Press; 1995. p. 233–277.
- Flege, JE. Age of learning and second-language speech. In: Birdsong, D., editor. *New perspectives on the critical period hypothesis for second language acquisition*. Mahwah, NJ: Erlbaum; 1999. p. 101–132.
- Flege, JE. Assessing constraints on second-language segmental production and perception. In: Meyer, A.; Schiller, Niels, editors. *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Mouton de Gruyter; Berlin: 2003. p. 319–355.
- Flege JE, MacKay IRA. Perceiving vowels in a second language. *Studies in Second Language Acquisition*. 2004; 26:1–34.
- Francis AL, Ciocca V, Ma L, Fenn K. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*. 2008; 36(2):268–294.
- Gandour JT. Tone perception in far Eastern languages. *Journal of Phonetics*. 1983; 11:149–175.

- Gandour JT, Harshman RA. Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech*. 1978; 21:1–33. [PubMed: 692240]
- Gao J. Identification of Lexical Tones in Mandarin Chinese by Uygur Students of Different Proficiency Levels. *Journal of Xinjiang Normal University (Social Sciences)*. 2005; 26(4):174–178.
- Goto H. Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*. 1971; 9:317–323. [PubMed: 5149302]
- Gottfried TL. Effects of Consonant context on the perception of French vowels. *Journal of Phonetics*. 1984; 12:91–114.
- Gottfried TL, Suiter TL. Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*. 1997; 25(2):207–231.
- Guion SG, Flege JE, Akahane-Yamada R, Pruitt JC. An investigation of current models of second language speech perception: The case of Japanese adults’ perception of English consonants. *Journal of the Acoustical Society of America*. 2000; 107:2711. [PubMed: 10830393]
- Hallé PA, Chang YC, Best CT. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*. 2004; 32(3):395–421.
- Hattori K, Iverson P. English /r/-/l/ category assimilation by Japanese adults: individual differences and the link to identification accuracy. *Journal of the Acoustical Society of America*. 2009; 125(1): 469–479. [PubMed: 19173432]
- Iverson P, Kuhl PK, Akahane-Yamada R, Diesch E, Tohkura Y, Kettermann A, et al. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*. 2003; 87(1):B47–B57. [PubMed: 12499111]
- Kuhl PK, Steven E, Hayashi A, Deguchi T, Kiritani S, Iverson P. Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*. 2006; 9:13–21.
- Lee YS, Vokoch DA, Wurm LH. Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*. 1996; 25(5):527–542. [PubMed: 8865624]
- Lively SE, Logan JS, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*. 1993; 94:1242–1255. [PubMed: 8408964]
- Lively SE, Pisoni DB, Yamada RA, Tohkura Y, Yamada T. Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *Journal of the Acoustical Society of America*. 1994; 96(4):2076–2087. [PubMed: 7963022]
- Logan JS, Lively SE, Pisoni DB. Training Japanese listeners to identify English /r/ and /l/: a first report. *Journal of the Acoustical Society of America*. 1991; 89(2):874–886. [PubMed: 2016438]
- Miyawaki K, Strange W, Verbrugge R, Liberman A, Jenkins J, Fujimura O. An effect of language experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*. 1975; 18:331–340.
- Navarra J, Sebastián-Gallés N, Soto-Faraco S. The perception of second language sounds in early bilinguals: New evidence from an implicit measure. *Journal of Experimental Psychology: Human Perception and Performance*. 2005; 31(5):912–918. [PubMed: 16262488]
- Pallier C, Bosch L, Sebastián-Gallés N. A limit on behavioral plasticity in speech perception. *Cognition*. 1997; 64:B9–B17. [PubMed: 9426508]
- Pallier C, Colomé A, Sebastián-Gallés N. The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*. 2001; 12(6):445–449. [PubMed: 11760129]
- Sebastián-Gallés N, Soto-Faraco S. Online processing of native and non-native phonemic contrasts in early bilinguals. *Cognition*. 1999; 72:111–123. [PubMed: 10553668]
- Sebastián-Gallés N, Echeverría S, Bosch L. The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*. 2005; 52:240–255.
- Sebastián-Gallés N, Rodríguez-Fornells A, de Diego-Balaguer R, Díaz B. First- and Second-language Phonological Representations in the Mental Lexicon. *Journal of Cognitive Neuroscience*. 2006; 18(8):1277–1291. [PubMed: 16859414]

- Sebastián-Gallés N, Vera-Constán F, Larsson J, Costa A, Deco G. Lexical Plasticity in Early Bilinguals Does Not Alter Phoneme Categories: II. Experimental Evidence. *Journal of Cognitive Neuroscience*. 2009; 21(12):2343–2357. [PubMed: 18855552]
- Snodgrass, J.; Levy-Berger, G.; Haydon, M. *Human experimental psychology*. Oxford: Oxford University Press; 1985.
- So CK. The effect of L1 prosodic backgrounds of Cantonese and Japanese speakers on the perception of Mandarin tones after training. *Journal of the Acoustical Society of America*. 2005a; 117(4): 2427.
- So, CK. The influence of L1 prosodic background on the learning of Mandarin tones: Patterns of tonal confusion by Cantonese and Japanese naïve listeners. In: Gurski, C., editor. *Proceedings of the 2005 CLA Annual Conference*; Paper presented at the 2005 annual conference of the Canadian Linguistic Association; Canada: 2005b.
- So, CK. Doctoral dissertation. Simon Fraser University; 2006a. Effect of L1 prosodic background and AV training on learning Mandarin tones by speakers of Cantonese, Japanese, and English.
- So, CK. Perception of non-native tonal contrasts: Effects of native phonological and phonetic influences. In: Warren, P.; Watson, CL., editors. *Proceedings of the 11th Australian International Conference on Speech Science & Technology*; Presented at the 11th Australian International Conference on Speech Science & Technology; Auckland, New Zealand: 2006b. p. 438-443.
- Streeter LA. Language perception of 2-month old infants shows effects of both innate mechanisms and experience. *Nature*. 1976; 259:39–41. [PubMed: 1256541]
- Sumner M, Samuel AG. The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*. 2009; 60:487–501.
- Tang C, van Heuven VJ. Mutual intelligibility of Chinese dialects experimentally tested. *Lingua*. 2009; 119:709–732.
- Tong YX, Francis AL, Gandour JT. Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*. 2008; 23(5):689–708.
- Wang, X. Perception of L2 Tones: L1 Lexical Tone Experience May Not Help. *Speech Prosody 2006*; Presented at the Speech Prosody; 2006; Dresden, Germany. 2006.
- Wang Y, Spence MM, Jongman A, Sereno JA. Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*. 1999; 106(6):3649–3658. [PubMed: 10615703]
- Warrington EK, Weiskrantz L. New method of testing long-term retention with special reference to amnesic patients. *Nature*. 1968; 217:972–974. [PubMed: 5642857]
- Warrington EK, Weiskrantz L. Amnesic syndrome: Consolidation or retrieval? *Nature*. 1970; 228:629–30.
- Wayland RP, Guion SG. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*. 2004; 54(4):681–712.
- Weber A, Cutler A. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*. 2004; 50(1):1–25.
- Werker JF, Tees RC. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*. 1984; 7:49–63.
- Xing, F. *Xiandai Hanyu [Modern Chinese]* (in Chinese). Beijing: Chinese High Education Press; 1991.
- Zhan, PH.; Li, RL.; Huang, JJ.; Xu, BH. *Chinese Dialects and Survey* (in Chinese). Hubei: Hubei Education Press; 2004.
- Zhang, X. Dialect MT: A case study between Cantonese and Mandarin. Paper presented at the COLING-ACL'98 (17th International Conference on Computational Linguistics (COLING'98) and 36th Annual Meeting of the Association for Computational Linguistics (ACL'98)); August 10–14, 1998; Montreal, Canada: University de Montreal; 1998. p. 1460-1464.

Appendix A

Consonant frequencies in Mandarin Chinese (retrieved from CCL corpus:
http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp?dir=xiandai).

Consonant	(/ts/) (ts' /)	(/s/) (ts' /)	(/tʃ/) (tʃ' /)	(/ʃ/) (tʃ' /)	(/tʃ' /)	(/tʃ' /)	(/tʃ' /)	(/tʃ' /)	(/tʃ' /)
Frequency	0.034	0.013	0.016	0.065	0.030	0.075	0.075	0.030	0.053

Tone frequencies in Mandarin Chinese (Tang, 1995)

Tone	T1	T2	T3	T4
Frequency	0.187	0.194	0.175	0.358

Appendix B: Stimuli used in Experiment 1 and 2

Stimuli for consonant perception in Experiment 1:

- /ts■(1)/ (*zi*(1): “资”, meaning *money*),
- /ts' ■(1)/ (*ci*(1): “疵”, meaning *flaw*),
- /s■(1)/ (*si*(1): “思”, meaning *think*),
- /tʃ■(1)/ (*zhi*(1): “知”, meaning *know*),
- /tʃ' ■(1)/ (*chi*(1): “吃”, meaning *eat*),
- /ʃ■(1)/ (*shi*(1): “诗”, meaning *poem*),
- /tʃi(1)/ (*ji*(1): “鸡”, meaning *chicken*),
- /tʃ' i(1)/ (*qi*(1): “妻”, meaning *wife*),
- /ʃi(1)/ (*xi*(1): “西”, meaning *west*)

Stimuli for tone perception in Experiment 1:

- /pɑ(1)/ (*ba*(1): “八”, meaning *eight*),
- /pɑ(2)/ (*ba*(2): “拔”, meaning *pull out*),
- /pɑ(3)/ (*ba*(3): “把”, meaning *handle*),
- /pɑ(4)/ (*ba*(4): “爸”, meaning *father*)

Stimuli for consonant perception in Experiment 2:

- /ts' ■(1)/ (*ci*(1): “疵”, meaning *flaw*),
- /ts' ɑ(1)/ (*ca*(1): “擦”, meaning *clean*),
- /tʃ' ■(1)/ (*chi*(1): “吃”, meaning *eat*),
- /tʃ' ɑ(1)/ (*cha*(1): “叉”, meaning *cross*)

Stimuli for tone perception in Experiment 2:

- /pi(2)/ (*bi*(2): “鼻”, meaning *nose*),
- /pɑ(2)/ (*ba*(2): “拔”, meaning *pull out*),
- /pi(3)/ (*bi*(3): “笔”, meaning *pen*),
- /pɑ(3)/ (*ba*(3): “靶”, meaning *target*)

Appendix C: The average hit and false alarm rates used to compute A' in Experiment 1

Consonantal contrasts:			
	Consonantal Contrasts	Cantonese Speakers	Mandarin Speakers
FA		0.17	0.08
	/ts/-/tʂ/	0.76	0.94
	/ts' /-/tʂ' /	0.9	0.99
	/s/-/ʂ/	0.97	0.96
	/ts/-/tʂ/	0.88	0.95
	/ts/-/tʂ' /	0.76	0.95
	/ts/-/ʂ/	0.87	0.97
HIT			
	/ts' /-/tʂ/	0.88	0.95
	/ts' /-/tʂ' /	0.9	0.96
	/ts' /-/ʂ/	0.81	0.79
	/s/-/tʂ/	0.92	0.98
	/s/-/tʂ' /	0.87	0.92
	/s/-/ʂ/	0.9	0.99
Tonal contrasts:			
	Tonal Contrasts	Cantonese Speakers	Mandarin Speakers
FA		0.08	0.05
	T1–T2	0.87	0.89
	T1–T3	0.94	0.94
	T1–T4	0.90	0.93
HIT	T2–T3	0.82	0.92
	T2–T4	0.93	0.95
	T3–T4	0.91	0.94

Appendix D: Critical stimuli used in Experiment 3

/ts' /-/tʂ' /			
/ts' /			
Chinese transcript	pinyin	IPA	meaning
擦嘴	ca(1)zui(3)	ts' a(1)tsuei(3)	wipe the mouth
残余	can(2)yu(2)	ts' an(2)y(2)	left
操守	cao(1)shou(3)	ts' du(1)ʂou(3)	ethics
祠堂	ci(2)tang(2)	ts' i(2)t' aŋ(2)	temple
葱郁	cong(1)yu(4)	ts' uŋ(1)y(4)	verdant

<i>/ts' /- /tɕ' /</i>			
<i>/ts' /</i>			
Chinese transcript	pinyin	IPA	meaning
丛生	<i>cong(2)sheng(1)</i>	ts' uŋ(2)ʂəŋ(1)	<i>clusters</i>
粗布	<i>cu(1)bu(4)</i>	ts' u(1)pu(4)	<i>coarse cloth</i>
粗口	<i>cu(1)kou(3)</i>	ts' u(1)k' ou(3)	<i>harsh words</i>
粗重	<i>cu(1)zhong(4)</i>	ts' u(1)tɕuŋ(4)	<i>heavy</i>
电磁	<i>dian(4)ci(2)</i>	tian(4)ts' ɿ(2)	<i>electromagnetic</i>
短促	<i>duan(3)cu(4)</i>	tuan(3)ts' u(4)	<i>brief</i>
推辞	<i>tui(1)ci(2)</i>	t' uei(1)ts' ɿ(2)	<i>decline</i>
小草	<i>xiao(3)cao(3)</i>	ɕiɕu(3)ts' ɕu(3)	<i>grass</i>
言辞	<i>yan(2)ci(2)</i>	ian(2)ts' ɿ(2)	<i>words</i>
阶层	<i>jie(1)ceng(2)</i>	tɕie(1)ts' əŋ(2)	<i>hierarchy</i>
深藏	<i>shen(1)cang(2)</i>	ʂən(1)ts' ɕŋ(2)	<i>deep</i>
<i>/tɕ' /</i>			
插嘴	<i>cha(1)zui(3)</i>	tɕ' a(1)tsuei(3)	<i>interrupt</i>
单于	<i>chan(2)yu(2)</i>	tɕ' an(2)y(2)	<i>leader of the Huns</i>
抄手	<i>chao(1)shou(3)</i>	tɕ' ɕu(1)ʂou(3)	<i>wanton soup</i>
池塘	<i>chi(2)tang(2)</i>	tɕ' ɿ(2)t' ɕŋ(2)	<i>pond</i>
充裕	<i>chong(1)yu(4)</i>	tɕ' uŋ(1)y(4)	<i>ample</i>
重生	<i>chong(2)sheng(1)</i>	tɕ' uŋ(2)ʂəŋ(1)	<i>rebirth</i>
初步	<i>chu(1)bu(4)</i>	tɕ' u(1)pu(4)	<i>initial</i>
出口	<i>chu(1)kou(3)</i>	tɕ' u(1)k' ou(3)	<i>exit</i>
出众	<i>chu(1)zhong(4)</i>	tɕ' u(1)tɕuŋ(4)	<i>superior</i>
电池	<i>dian(4)chi(2)</i>	tian(4)tɕ' ɿ(2)	<i>battery</i>
短处	<i>duan(3)chu(4)</i>	tuan(3)tɕ' u(4)	<i>disadvantage</i>
推迟	<i>tui(1)chi(2)</i>	t' uei(1)tɕ' ɿ(2)	<i>postpone</i>
小炒	<i>xiao(3)chao(3)</i>	ɕiɕu(3)tɕ' ɕu(3)	<i>dishes</i>
延迟	<i>yan(2)chi(2)</i>	ian(2)tɕ' ɿ(2)	<i>delay</i>
阶乘	<i>jie(1)cheng(2)</i>	tɕie(1)tɕ' əŋ(2)	<i>factorial</i>

<i>/ts' /- /tɕ' /</i>			
<i>/ts' /</i>			
Chinese transcript	pinyin	IPA	meaning
伸长	<i>shen(1)chang(2)</i>	ʂən(1)tɕ' əŋ(2)	<i>extend</i>
<i>/t' /- /t' /</i>			
<i>/t' /</i>			
担心	<i>dan(1)xin(1)</i>	tan(1)ɕin(1)	<i>concern</i>
胆子	<i>dan(3)zi0</i>	tan(3)tsɿ(0)	<i>courage</i>
但是	<i>dan(4)shi(4)</i>	tan(4)ʂɿ(4)	<i>but</i>
倒装	<i>dao(4)zhuang(1)</i>	tɕu(4)tɕu;ɕŋ(1)	<i>inversion</i>
抵御	<i>di(3)yu(4)</i>	ti(3).y(4)	<i>resist</i>
地带	<i>di(4)dai(4)</i>	ti(4)tai(4)	<i>zone</i>
颠覆	<i>dian(1)fu(4)</i>	tian(1)fu(4)	<i>subversion</i>
调动	<i>diao(4)dong(4)</i>	tiɕu(4)tu;ŋ(4)	<i>transfer</i>
冬天	<i>dong(1)tian(1)</i>	tuŋ(1)t' ian(1)	<i>winter</i>
动心	<i>dong(4)xin(1)</i>	tu;ŋ(4)ɕin(1)	<i>tempted</i>
洞悉	<i>dong(4)xi(1)</i>	tu;ŋ(4)ɕi(1)	<i>insight</i>
赌注	<i>du(3)zhu(4)</i>	tu(3)tɕu(4)	<i>bet</i>
肚子	<i>du(4)zi0</i>	tu(4)tsɿ(0)	<i>belly</i>
对话	<i>dui(4)hua(4)</i>	tuei(4)xua(4)	<i>dialogue</i>
男丁	<i>nan(2)ding(1)</i>	nan(2)ti;ŋ(1)	<i>male</i>
跑调	<i>pao(3)diao(4)</i>	p' ɕu(3)tiɕu(4)	<i>out of tune</i>
<i>/t' /</i>			
贪心	<i>tan(1)xin(1)</i>	t' an(1) ɕin(1)	<i>greedy</i>
毯子	<i>tan(3)zi0</i>	t' an(3)tsɿ(0)	<i>blanket</i>
探视	<i>tan(4)shi(4)</i>	t' an(4)ʂɿ(4)	<i>visit</i>
套装	<i>tao(4)zhuang(1)</i>	t' ɕu(4)tɕuɕŋ(1)	<i>suit</i>
体育	<i>ti(3)yu(4)</i>	t' i(3) y(4)	<i>sport</i>
替代	<i>ti(4)dai(4)</i>	t' i(4)tai(4)	<i>substitute</i>
天赋	<i>tian(1)fu(4)</i>	t' ian(1)fu(4)	<i>talent</i>

<i>/ts' /- /tɕ' /</i>			
<i>/ts' /</i>			
Chinese transcript	pinyin	IPA	meaning
跳动	<i>tiao(4)dong(4)</i>	t' iɕu(4)tu;ŋ(4)	<i>beat</i>
通天	<i>tong(1)tian(1)</i>	t' u;ŋ(1)t' ian(1)	<i>Babel</i>
痛心	<i>tong(4)xin(1)</i>	t' u;ŋ(4)ɕin(1)	<i>sad</i>
痛惜	<i>tong(4)xi(1)</i>	t' u;ŋ(4)ɕi(1)	<i>deplore</i>
土著	<i>tu(3)zhu(4)</i>	t' u(3)tɕu(4)	<i>indigenous</i>
兔子	<i>tu(4)zi0</i>	t' u(4)tɕi(0)	<i>rabbit</i>
退化	<i>tui(4)hua(4)</i>	t' uei(4)xua(4)	<i>degeneration</i>
难听	<i>nan(2)ting(1)</i>	nan(2)t' i;ŋ(1)	<i>not nice</i>
跑跳	<i>pao(3)tiao(4)</i>	p' u(3)t' iɕu(4)	<i>jump</i>
T2–T3			
T2			
白布	<i>bai(2)bu(4)</i>	pai(2)pu(4)	<i>white cloth</i>
磨砂	<i>mo(2)sha(1)</i>	mo(2)ɕa(1)	<i>scrub</i>
防线	<i>fang(2)xian(4)</i>	fɕŋ(2)ɕian(4)	<i>defense</i>
徒弟	<i>tu(2)di(4)</i>	t' u(2)ti(4)	<i>apprentice</i>
活力	<i>huo(2)li(4)</i>	xuo(2)li(4)	<i>vitality</i>
皮肤	<i>pi(2)fu(1)</i>	p' i(2)fu(1)	<i>skin</i>
媒人	<i>mei(2)ren(2)</i>	mei(2)ʒən(2)	<i>matchmaker</i>
蟾蜍	<i>chan(2)chu(2)</i>	tɕ' an(2)tɕ' u(2)	<i>toad</i>
安详	<i>an(1)xiang(2)</i>	an(1) ɕiɕŋ(2)	<i>serene</i>
边缘	<i>bian(2)yuan(2)</i>	pian(2)yan(2)	<i>edge</i>
步伐	<i>bu(4)fa(2)</i>	pu(4)fa(2)	<i>pace</i>
大学	<i>da(4)xue(2)</i>	ta(4)ɕye(2)	<i>university</i>
单薄	<i>dan(1)bao(2)</i>	tan(1)p u(2)	<i>thin</i>
生活	<i>sheng(1)huo(2)</i>	ɕəŋ(1)xuo(2)	<i>life</i>
清除	<i>qing(1)chu(2)</i>	tɕ' iŋ(1)tɕ' u(2)	<i>remove</i>
历时	<i>li(4)shi(2)</i>	li(4)ɕi(2)	<i>last</i>

<i>/ts' /- /tɕ' /</i>			
<i>/ts' /</i>			
Chinese transcript	pinyin	IPA	meaning
T3			
摆布	<i>bai(3)bu(4)</i>	<i>pai(3)pu(4)</i>	<i>manipulate</i>
抹杀	<i>mo(3)sha(1)</i>	<i>mo(3) ɕa(1)</i>	<i>obliterate</i>
纺线	<i>fang(3)xian(4)</i>	<i>fɕɿŋ(3) ɕian(4)</i>	<i>spin</i>
土地	<i>tu(3)di(4)</i>	<i>t' u(3)ti(4)</i>	<i>land</i>
火力	<i>huo(3)li(4)</i>	<i>xuo(3)li(4)</i>	<i>fire</i>
匹夫	<i>pi(3)fu(1)</i>	<i>p' i(3)fu(1)</i>	<i>ordinary man</i>
美人	<i>mei(3)ren(2)</i>	<i>mei(3)ʐən(2)</i>	<i>beauty</i>
铲除	<i>chan(3)chu(2)</i>	<i>tɕ' an(3)tɕ' u(2)</i>	<i>eradicate</i>
安享	<i>an(1)xiang(3)</i>	<i>an(1)ɕiɕɿŋ(3)</i>	<i>enjoy</i>
边远	<i>bian(1)yuan(3)</i>	<i>pian(1)yan(3)</i>	<i>remote</i>
不法	<i>bu(4)fa(3)</i>	<i>pu(4)fa(3)</i>	<i>illegal</i>
大雪	<i>da(4)xue(3)</i>	<i>ta(4)ɕye(3)</i>	<i>snow</i>
担保	<i>dan(1)bao(3)</i>	<i>tan(1)pɕu(3)</i>	<i>guarantee</i>
生火	<i>sheng(1)huo(3)</i>	<i>ɕəŋ(1)xuo(3)</i>	<i>light a fire</i>
清楚	<i>qing(1)chu(3)</i>	<i>tɕ' i:ŋ(1)tɕ' u(3)</i>	<i>clear</i>
历史	<i>li(4)shi(3)</i>	<i>li(4)ʂi(3)</i>	<i>history</i>
T2–T4			
T2			
人性	<i>ren(2)xing(4)</i>	<i>ʐən(2)ɕiŋ(4)</i>	<i>humanity</i>
防治	<i>fang(2)zhi(4)</i>	<i>fɕɿŋ(2)tɕi(4)</i>	<i>prevention</i>
符合	<i>fu(2)he(2)</i>	<i>fu(2)xɤ(2)</i>	<i>conform to</i>
福利	<i>fu(2)li(4)</i>	<i>fu(2)li(4)</i>	<i>welfare</i>
工人	<i>gong(1)ren(2)</i>	<i>ku:ŋ(1) ʐən(2)</i>	<i>worker</i>
假使	<i>jia(2)shi(3)</i>	<i>tɕia(2) ɕi(3)</i>	<i>if</i>
弧形	<i>hu(2)xing(2)</i>	<i>xu(2) ɕi:ŋ(2)</i>	<i>arc</i>
回报	<i>hui(2)bao(4)</i>	<i>xuei(2)pɕu(4)</i>	<i>reward</i>

<i>/ts' /-tɕ' /</i>			
<i>/ts' /</i>			
Chinese transcript	pinyin	IPA	meaning
集结	<i>ji(2)jie(2)</i>	tɕi(2)tɕie(2)	<i>assembly</i>
同时	<i>tong(2)shi(2)</i>	t' u;ŋ(2) ɕi(2)	<i>at the same time</i>
年级	<i>nian(2)ji(2)</i>	nian(2)tɕi(2)	<i>grade</i>
务实	<i>wu(4)shi(2)</i>	u(4)ɕi(2)	<i>pragmatic</i>
风俗	<i>feng(1)su(2)</i>	fəŋ(1)su(2)	<i>custom</i>
风靡	<i>feng(1)mi(2)</i>	fəŋ(1)mi(2)	<i>popular</i>
过时	<i>guo(4)shi(2)</i>	kuo(4) ɕi(2)	<i>out of date</i>
年华	<i>nian(2)hua(2)</i>	nian(2)xua(2)	<i>years</i>
T4			
任性	<i>ren(4)xing(4)</i>	ʒən(4) ɕi;ŋ(4)	<i>capricious</i>
放置	<i>fang(4)zhi(4)</i>	fʊŋ(4)tɕi(4)	<i>place</i>
复合	<i>fu(4)he(2)</i>	fu(4)xv(2)	<i>compound</i>
富丽	<i>fu(4)li(4)</i>	fu(4)li(4)	<i>magnificent</i>
公认	<i>gong(1)ren(4)</i>	ku;ŋ(1)ʒən(4)	<i>accepted</i>
驾驶	<i>jia(4)shi(3)</i>	tɕia(4)ɕi(3)	<i>drive</i>
户型	<i>hu(4)xing(2)</i>	xu(4)ɕi;ŋ(2)	<i>unit</i>
汇报	<i>hui(4)bao(4)</i>	xuei(4)pʊu(4)	<i>report</i>
季节	<i>ji(4)jie(2)</i>	tɕi(4)tɕie(2)	<i>season</i>
同事	<i>tong(2)shi(4)</i>	t' u;ŋ(2)ɕi(4)	<i>colleague</i>
年纪	<i>nian(2)ji(4)</i>	nian(2)tɕi(4)	<i>age</i>
误事	<i>wu(4)shi(4)</i>	u(4)ɕi(4)	<i>make trouble</i>
风速	<i>feng(1)su(4)</i>	fəŋ(1)su(4)	<i>wind velocity</i>
蜂蜜	<i>feng(1)mi(4)</i>	fəŋ(1)mi(4)	<i>honey</i>
过世	<i>guo(2)shi(4)</i>	kuo(2)ɕi(4)	<i>pass away</i>
年画	<i>nian(2)hua(4)</i>	nian(2)xua(4)	<i>new year picture</i>

Appendix E: Analyses of overall reaction times in Experiment 3

Consonant contrasts: *c-ch analysis*

We conducted a 2 (Group: Mandarin vs. Cantonese) $\times$ 4 (Prime Condition: 2 identity conditions vs. 2 minimal pair conditions) $\times$ 2 (Occurrence: first vs. second) mixed analysis of variance on absolute reaction times for the /ts' /-/tʂ' / contrast. The main effect of Group was not significant, $F(1, 30)=1.98$, $p=.17$, $\eta^2=.06$. The main effect of Prime Condition was significant, $F(3, 198)=9.12$, $p<.001$, $\eta^2=.12$; $F(3, 90)=8.46$, $p<.001$, $\eta^2=.22$: the reaction times of the two identity conditions (1336ms for the /ts' /-/tʂ' / condition and 1387ms for the /tʂ' /-/tʂ' / condition) were faster than those of the two minimal pair conditions (1430ms for the /ts' /-/tʂ' / condition and 1465ms for the /tʂ' /-/tʂ' / condition), $ps<.01$. The main effect of Occurrence was significant, $F(1, 66)=24.71$, $p<.001$, $\eta^2=.27$; $F(1, 30)=23.24$, $p<.001$, $\eta^2=.44$: the reaction time of the first occurrence (1331ms) was significantly longer than that of the second occurrence (1286ms). The interaction was not significant, $ps>.05$.

Consonant contrasts: */t/-t' / analysis*

The same analysis was conducted on absolute reaction times for the *d-t* contrast. The main effect of Group was not significant, $F(1, 66)=1.94$, $p=.168$, $\eta^2=.03$; $F(1, 30)=3.08$, $p=.09$, $\eta^2=.09$. The main effect of Prime Condition was significant, $F(3, 198)=8.14$, $p<.001$, $\eta^2=.11$; $F(3, 90)=2.77$, $p=.046$, $\eta^2=.08$: the reaction times of the two identity conditions (1263ms for the /t/-t' / condition and 1286ms for the /t' /-t' / condition) were faster than those of the two minimal pair conditions (1353ms for the /t/-t' / condition and 1330ms for the /t' /-t' / condition), $ps<.01$. The main effect of Occurrence was significant by participants, $F(1, 66)=9.56$, $p=.003$, $\eta^2=.13$, with the reaction time of the first occurrence (1331ms) significantly longer than that of the second occurrence (1285ms). It was not significant by items, $F(1, 30)=1.83$, $p=.186$, $\eta^2=.06$. The interaction of Prime Condition and Occurrence was significant, $F(3, 198)=12.73$, $p<.001$, $\eta^2=.16$; $F(3, 90)=3.08$, $p=.031$, $\eta^2=.09$. The simple effect of Occurrence for each Prime condition revealed that the reaction time of the first occurrence was significantly longer than that of the second occurrence in the /t/-t' / condition, $F(1, 67)=21.63$, $p<.001$; $F(1, 31)=22.95$, $p<.001$. The effect of Occurrence was significant in the /t' /-t' / condition by participants, $F(1, 67)=23.16$, $p<.001$; but not significant by items, $F(1, 30)=1.00$, $p=.32$. None of the other interactions was significant, $ps>.05$.

Tonal contrasts: *T2–T3 analysis*

The same analysis was conducted on absolute reaction times for the T2–T3 contrast. The main effect of Group was not significant by participants, $F(1, 66)=3.70$, $p=.059$, $\eta^2=.05$. It was significant by items, $F(1, 30)=11.31$, $p=.002$, $\eta^2=.27$, with Mandarin speakers (1325ms) faster than Cantonese speakers (1438ms). The main effect of Prime Condition was significant, $F(3, 198)=5.08$, $p=.002$, $\eta^2=.07$; $F(3, 90)=2.93$, $p=.038$, $\eta^2=.04$: the reaction times of the two identity conditions (1305ms for the T2–T2 condition and 1349ms for the T3–T3 condition) were faster than those of the two minimal pair conditions (1385ms for the T2–T3 condition and 1398ms for the T3–T2 condition), $ps<.01$. The main effect of Occurrence was significant, $F(1, 66)=14.46$, $p<.001$, $\eta^2=.18$; $F(1, 30)=9.58$, $p=.004$, $\eta^2=.24$: the reaction time of the first occurrence (1388ms) was longer than that of the second occurrence (1330ms). The interaction of Prime Condition and Occurrence was significant by participants, $F(3, 198)=7.69$, $p<.001$, $\eta^2=.10$; not by items, $F(3, 30)=1.00$, $p=.32$. For the analysis by participants, the simple effect of Occurrence for each Prime condition revealed that the reaction time of the first occurrence was significantly longer than that of the second

occurrence in the T2-T2 condition, $F(1, 67)=14.87$, $p<.001$, and in the T3-T3 condition, $F(1, 67)=32.70$, $p<.001$. None of the other interactions was significant, $ps>.05$.

Tonal contrasts: T2–T4 analysis

The same analysis was conducted on reaction times for the T2–T4 contrast. The main effect of Group was not significant by participants, $F(1, 66)=1.59$, $p=.212$, $\eta^2=.02$; marginally significant by items, $F(1, 30)=4.12$, $p=.051$, $\eta^2=.12$, with Mandarin speakers (1339ms) responding faster than Cantonese speakers (1417ms). The main effect of Prime Condition was significant, $F(3, 198)=9.89$, $p<.001$, $\eta^2=.13$; $F(3, 90)=3.93$, $p=.011$, $\eta^2=.12$: the reaction times of the two identity conditions (1300ms for the T2-T2 condition and 1351ms for the T4-T4 condition) were faster than those of the two minimal pair conditions (1379ms for the T2-T4 condition and 1403ms for the T4-T2 condition), $ps<.05$. The main effect of Occurrence was significant, $F(1, 66)=6.95$, $p=.01$, $\eta^2=.10$; $F(1, 30)=9.68$, $p=.004$, $\eta^2=.24$: the reaction time of the first occurrence (1376ms) was longer than that of the second occurrence (1340ms). The interaction of Prime Condition and Occurrence was significant by participants, $F(3, 198)=4.68$, $p=.004$, $\eta^2=.07$; not by items, $F(3, 90)=2.41$, $p=.072$, $\eta^2=.07$. The simple effect of Occurrence for each Prime condition revealed that the reaction time of the first occurrence was significantly longer than that of the second occurrence in the T2-T2 condition, $F(1, 67)=10.95$, $p=.002$, and in the T4-T4 condition, $F(1, 67)=17.00$, $p<.001$. None of the other interactions was significant, $ps>.05$.

We studied perceptual processing of Mandarin by native Cantonese speakers who learn Mandarin when they are young.

Despite early and extensive L2 experience, lexical representation was not equivalent to that of native speakers.

The perception and representation of tones are better than that of segments for non-native tone language speakers.

L2 words containing L1-like segments compete strongly against L2 words with segments that match L1 poorly.

Figure 1. Mean A' scores (with error bars representing the standard error of the mean) for 12 Mandarin consonantal contrasts by the Cantonese speakers and Mandarin speakers in Experiment 1.

Figure 2. Mean A' scores (with error bars representing the standard error of the mean) for the six Mandarin tonal contrasts by the Cantonese speakers and Mandarin speakers in Experiment 1.

Figure 3. Mean reaction times (with error bars representing the standard error of the mean) for /ts'/ words and /tɕ'/ words by the Cantonese speakers and Mandarin speakers in homogeneous lists and orthogonal lists in Experiment 2.

Figure 4. Mean reaction times (with error bars representing the standard error of the mean) for T2 words and T3 words by the Cantonese speakers and Mandarin speakers in homogeneous lists and orthogonal lists in Experiment 2.

Figure 5.

Priming effects (with error bars representing the standard error of the mean) for the Cantonese speakers and Mandarin speakers across the four prime conditions for the /ts' /-/tɕ' / contrast in Experiment 3.

Figure 6. Priming effects (with error bars representing the standard error of the mean) for the Cantonese speakers (collapsed across the two testing groups) across the four prime conditions for the /ts'/-/tɕ'/' contrast in Experiment 3.

Figure 7. Priming effects (with error bars representing the standard error of the mean) for the Cantonese speakers and Mandarin speakers across the four prime conditions for the /t/-/t'/ contrast in Experiment 3.

Figure 8. Priming effects (with error bars representing the standard error of the mean) for the Cantonese speakers and Mandarin speakers across the four prime conditions for the T2–T3 contrast in Experiment 3.

Figure 9. Priming effects (with error bars representing the standard error of the mean) for the Cantonese speakers and Mandarin speakers across the four prime conditions for the T2–T4 contrast in Experiment 3.

Table 1

Mean reaction time (ms) for 12 Mandarin consonantal contrasts for the Cantonese speakers and Mandarin speakers in Experiment 1 (with standard deviations in parenthesis)

	Cantonese speakers	Mandarin speakers
/ts/-/tʂ/	1108 (60)	1021 (60)
/ts'/-/tʂ'/	1157 (56)	1029 (56)
/s/-/ʂ/	1132 (60)	1068 (60)
/ts/-/tʂ/	1080 (59)	1052 (60)
/ts/-/tʂ'/	1061 (67)	991 (67)
/ts/-/ʂ/	1006 (65)	968 (65)
/ts'/-/tʂ/	1086 (71)	936 (71)
/ts'/-/tʂ'/	1215 (57)	945 (57)
/ts'/-/ʂ/	1029 (59)	1017 (59)
/s/-/tʂ/	1055 (62)	1000 (62)
/s/-/tʂ'/	1200 (68)	994 (68)
/s/-/ʂ/	1050 (74)	1036 (74)

Table 2

Mean reaction time (ms) for six Mandarin tonal contrasts for the Cantonese speakers and Mandarin speakers in Experiment 1 (with standard deviations in parenthesis)

	Cantonese speakers	Mandarin speakers
T1–T2	1307 (317)	1096 (210)
T1–T3	1261 (280)	1086 (258)
T1–T4	1231 (340)	1025 (184)
T2–T3 *	1411 (359)	1143 (243)
T2–T4	1190 (207)	1038 (273)
T3–T4	1232 (257)	1030 (287)

*
 $p < .05$

Table 3

Examples of the four conditions for the /ts' /-/tɕ' /contrast

Prime condition	First occurrence	Second occurrence
/ts' /-/ts' /	/ts' ɿ (2)t' ɔŋ (2)/	/ts' ɿ(2)t' ɔŋ(2)/
/tɕ' /-/tɕ' /	/tɕ' ɿ(2)t' ɔŋ(2)/	/tɕ' ɿ(2)t' ɔŋ(2)/
/ts' /-/tɕ' /	/ts' ' (2)t' ɔŋ(2)/	/tɕ' ɿ(2)t' ɔŋ(2)/
/tɕ' /-/ts' /	/tɕ' ɿ(2)t' ɔŋ(2)/	/ts' ɿ(2)t' ɔŋ(2)/

Here, /ts' ɿ(2)t' ɔŋ(2)/[“ 祠堂”] means *temple* and /tɕ' ɿ(2)t' ɔŋ(2)/[“ 池塘”] means *pond*.