Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences

Connie K. So, Catherine T. Best

MARCS Auditory Laboratories, University of Western Sydney, Australia

Key words

lexical tones

non-native speech perception

Perceptual Assimilation Model

phonological and phonetic influences

Abstract

This study examined the perception of the four Mandarin lexical tones by Mandarin-naïve Hong Kong Cantonese, Japanese, and Canadian English listener groups. Their performance on an identification task, following a brief familiarization task, was analyzed in terms of tonal sensitivities (A-prime scores on correct identifications) and tonal errors (confusions). The A-prime results revealed that the English listeners' sensitivity to Tone 4 identifications specifically was significantly lower than that of the other two groups. The analysis of tonal errors revealed that all listener groups showed perceptual confusion of tone pairs with similar phonetic features (T1–T2, T1–T4 and T2–T3 pairs), but not of those with completely dissimilar features (T1–T3, T2–T4, and T3–T4). Language-specific errors were also observed in their performance, which may be explained within the framework of the Perceptual Assimilation Model (PAM: Best, 1995; Best & Tyler, 2007). The findings imply that linguistic experience with native tones does not necessarily facilitate non-native tone perception. Rather, the phonemic status and the phonetic features (similarities or dissimilarities) between the tonal systems of the target language and the listeners' native languages play critical roles in the perception of non-native tones.

1 Introduction

The term *linguistic experience* has been widely used in accounts of how perceptual performance with non-native speech categories is influenced by the listener's native

language (L1). While most previous studies have reported L1 influence on speech perception at the segmental level, its impact on listeners' speech perception at the suprasegmental level is not well understood. In particular, questions about how and to what extent linguistic experience with the features[1] of the prosodic systems (e.g., lexical tones, pitch accents, lexical stress, and intonation) of one's native language affects perception of non-native tone categories have not yet been addressed. To bridge the gap, this study examined the role of differing linguistic experience of native tonality on the perception of non-native Mandarin tones by three groups of Mandarin-naïve listeners: native speakers of Cantonese, Japanese, and English.

Studies on cross-language perception at the segmental level have generally found that adults experience great difficulties producing and perceiving non-native consonant and/or vowel contrasts (Lisker & Abramson, 1970; Polka, 1995; Werker & Tees, 1984) although some non-native segmental contrasts are discriminated quite well (e.g., Best, McRoberts, & Goodell, 2001). The L1 perceptual effects also extend to related difficulties when adults learn a new second language (L2) (Flege, McCutcheon, & Smith, 1987; Guion, Flege, Akahane-Yamada, & Pruitt, 2000; Jamieson & Morosan, 1986; Logan, Lively, & Pisoni, 1991; Polka, 1995). It has been suggested that perception of non-native speech contrasts is constrained by both the *phonological* and the *phonetic properties* of their native language (L1) (Best, 1995; Flege, 1995; Polka, 1991, 1992; Strange, 1995). *Phonological properties* are crucial to the structural transformation of words. They are contrastive and categorical and involve language-specific rules governing the distribution, patterning and context-determined realization of consonants, vowels, and tones. For example, native speakers of Japanese have difficulties in perceiving the English /r/–/l/ distinction, because these phonemes are not contrastive in the Japanese phonological system (Brown, 2000; Hume & Johnson, 2003). In contrast, we use the term *phonetic properties* to refer to characteristics of pronunciation that are not phonologically distinctive in a language (fine-grained, gradient, within-category, non-contrastive details of speech). For instance, when American English listeners categorized the Zulu aspirated voiceless velar stop [kʰ] and ejective [kʼ], they perceived both as the voiceless stop [kʰ], but the non-native sound [kʼ] was perceived as having odd or unusual voice qualities (non-contrastive gradient difference), because English has a [kʰ] but no ejectives (Best et al., 2001).

Effects of linguistic experience have also been reported in some studies on suprasegmental features, such as stress patterns, quantity contrasts (e.g., vowel length, consonant gemination) and sentence-level prosodic patterns, with influences from both phonological and phonetic levels of the L1. Adults' L1 prosodic systems have a profound effect on their perception of the suprasegmental contrasts of non-native languages (Archibald, 1992, 1993; Guion, Harada, & Clark, 2004; Hirata, 2004; McAllister, Flege, & Piske, 2002). For example, both Polish and Spanish learners of English tend to apply their L1 stress assignment rules when performing a stress placement task with spoken English words (Archibald, 1992, 1993; Guion et al., 2004).

---

1 Prosodic features generally include "length, accent and stress, tone, intonation, and potentially a few others" (Fox, 2000, p.1).

Studies on the perception of non-native lexical tones have also found that linguistic experience gained from listeners' native languages guides their perception of non-native tones substantially (Burnham & Francis, 1997; Gandour, 1983; Gandour & Harshman, 1978; Gottfried & Suiter, 1997; Hallé, Chang, & Best, 2004; Leather, 1983; Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004). For example, some found that non-native speakers perceive Mandarin tonal categories differently from native speakers, who can identify subtle differences between tones (Hallé et al., 2004; Leather, 1983; Lee et al., 1996). In addition, in the studies investigating the perception of tonal features (or dimensions) by listeners from different language backgrounds, Gandour (1983, 1984) found that native English listeners tended to focus on pitch height even though English is a non-tone language, while listeners from Chinese languages (e.g., Cantonese and Mandarin) focused on both pitch height and pitch direction when perceiving tones.

Other reports on non-native (or L2) tone perception have suggested that linguistic experience with lexical tones in the native language generally assists listeners in perceiving non-native tones, but in fact its effect on perception of non-native tones is still not very clear. Lee et al. (1996) found that Cantonese listeners perceived Mandarin tones better than did English listeners, but a comparable pattern was not found when Taiwanese Mandarin and English listeners perceived Cantonese tones. (Note: The experiment was conducted in the listeners' own countries—Hong Kong, Taiwan, and the US (New York).) The author claimed that Mandarin tones are much easier to perceive than Cantonese tones; however, that is simply a descriptive statement, it is not the only possible interpretation of the results, and in any case, it does not explain the basis for the proposed difference in ease of perception of the two tone systems. In addition, in a training study with a pair of Thai level tones (mid vs. low), Wayland and Guion (2004) reported that Thai-naïve Mandarin Chinese listeners discriminated the Thai level tones in the pretest better than did naïve English listeners. Both Lee et al. (1996) and Wayland and Guion (2004) concluded that linguistic experience plays a role in adults' perception of non-native tones, specifically suggesting that in general, listeners from a tone language perform better than those from a non-tone language.

That claim implies that adult listeners' linguistic experience in using tones from their native languages facilitates their perception of non-native tones across the board. Presumably, native speakers of tone languages have more experience in using pitch variations (or tones) in their native languages than do speakers of non-tone languages. However, it appears that the previous studies did not systematically control two possible confounding factors. One factor is prior knowledge of the target language. For example, in Lee et al.'s (1996) study, the performance difference could have resulted from the fact that Cantonese speakers in Hong Kong have extensive exposure to Mandarin; many indeed learn Mandarin. Also, none of the prior studies evaluated whether the effects of linguistic experience on non-native tone perception occur at the phonetic and/or at the phonological level. Indeed, it is uncertain whether linguistic experience with tones facilitates non-native tone perception in general, or rather is constrained by the specific tone contrasts in the L1 system, and whether the performance of listeners from language-different backgrounds will differ systematically as a function of how tonality is used in their native languages. The other potential

confounding factor is prior musical training, which was not controlled or assessed in Lee et al. (1996). It has been found that listeners with musical training backgrounds (e.g., piano/violin lessons) generally outperformed non-musically trained listeners on non-native lexical tone perception (Alexander, Wong, & Bradlow, 2005; Burnham & Brooker, 2002; Gottfried & Riester, 2000). Therefore, the present study examined non-native tone perception in listeners of diverse, systematically differing language backgrounds, while controlling for prior experience with the target language, as well as for prior musical training (see the Methods section for details).

Studies also suggest that the four Mandarin tones are not all perceived and produced equally well by non-native listeners, and that the pattern seems to be language-independent. Some pairs of Mandarin tones are more easily confused than others, apparently because of the similarities in their pitch onset and offset values and in their contour shapes (Gottfried & Suiter, 1997; Kiriloff, 1969; Miracle, 1989; Shen, 1989). The pairs Tone 2–Tone 3 (both have an initial dip in pitch followed by a rising contour) and Tone 1–Tone 4 (both start with a similar pitch height) are frequently found to be difficult for learners of Mandarin from non-tonal native languages (e.g., Dutch and English) to discriminate. This implies that listeners' sensitivity to universal, gradient *phonetic* information was at work during perception, rather than language-specific, contrastive *phonological* information. However, it is not known whether the same perceptual patterns also occur for non-native listeners of other tone languages, since Lee et al. (1996) did not examine discrimination of these tone pairs in their Cantonese listeners. To the extent that phonetic similarities of tone contours constrain non-native tone perception, this should apply to non-native listeners of tone as well as non-tone languages. If the resulting patterns are found irrespective of listeners' native languages, this would imply that perception of tone contrasts is influenced by the tones' phonetic properties in a language-universal way. However, if there are discrepancies in performance patterns among different language groups, this would imply that the use of tonality in listeners' native phonological systems constrains perception. Also, if the effect of L1 linguistic experience on perception of non-native segmental speech categories is shaped by both the phonetics and phonology of the native language, then the influences should also apply to suprasegmental categories such as lexical tones, pitch-accent and stress patterning.

For these reasons, the present study examined phonetic and phonological influences on non-native tone perception by naïve listeners from several languages differing in their use of tonality at the lexical (word) level. The target stimulus language was Mandarin, in which each tone has its unique pitch contour. Mandarin tones are perceived categorically by native speakers (Leather, 1987; Stagray & Downs, 1993; Wang, 1976). Three non-native listener languages were selected in which linguistic use of tonality ranges from extensive to minimal, and only listeners who were Mandarin-naïve were included in the study. Hong Kong Cantonese, a tone language, uses tonal features in every lexical item, and allows evaluation of both phonological and phonetic influences from the L1 on perception of Mandarin tones because its tone system differs systematically from that of Mandarin. Unlike Lee et al. (1996), we limited our Cantonese listeners to those who had never had instruction in Mandarin, and who were not experiencing Mandarin in their daily environment given that they had been living in Canada for several years at the time of testing. Japanese, a pitch-accented
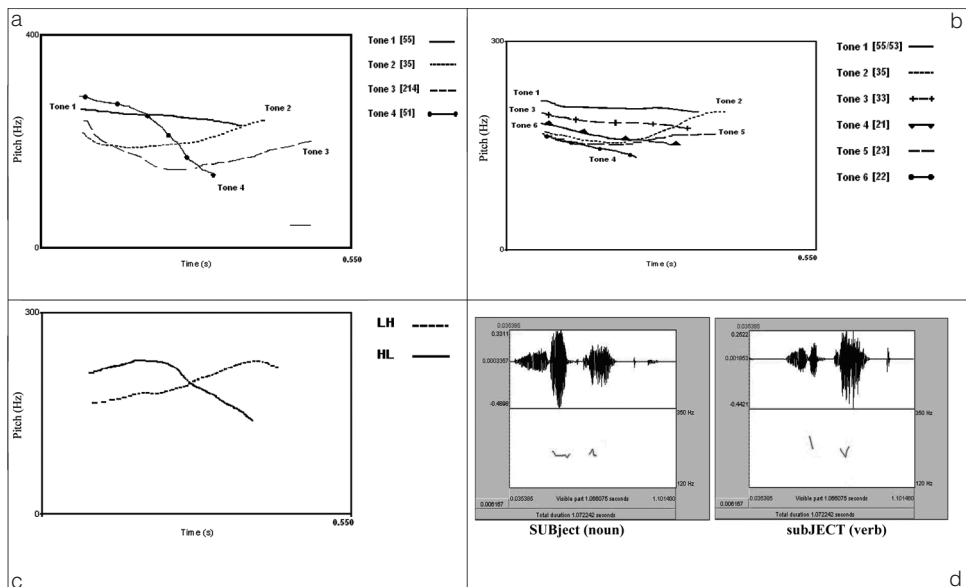
language, was chosen for comparison to the non-native tone language Cantonese, because it uses much more limited pitch variations to differentiate lexical items, and uses them over two timing units (morae) rather than over a single syllable or rhyme as in Cantonese. English, a non-tone language, was chosen because using pitch variations to signal different meanings at the word level (e.g., lexical stress or intonation attached with pragmatic expressions) is very limited.

Mandarin is a "lexical tone language" (Yip, 2002, p.2) with four tones (e.g., Bauer & Benedict, 1997; Duanmu, 2004; Hashimoto, 1972) that are typically described in terms of Chao's (1930) tone letters, which range from 1 (low pitch, or *F0*, within the speaker's range) to 5 (high *F0*): *high level* [55] (Tone 1; hereafter, T1), *mid rising* [35] (Tone 2; T2), *falling rising* [214] (Tone 3; T3), and *high falling* [51] (Tone 4; T4) (see Figure 1(a), top left). For example, /ta55/ 'carry', /ta35/ 'reach', /ta214/ 'hit', and /ta51/ 'large' are all meaningfully different words solely because of their lexical tone differences (minimal tone contrasts).

Cantonese is also a lexical tone language, but with six phonemic tones (see Figure 1(b), top right): three level tones, *high* [55], *mid* [33], and *low* [22], namely, Tone 1 (T1),

Figure 1

Naturally spoken examples of the pitch patterns of the Mandarin (a: top left) and of the Hong Kong Cantonese (b: top right) tone systems based on syllables /ta/ and /si/, respectively; the contrastive pitch patterns of the Japanese pitch-accent system based on the two-mora form /ame/ (c: bottom left), in which H = high (accented) and L = low (unaccented) and the contrastive lexical stress patterns of English two-syllable word "subject" (d: bottom right). Speech samples were spoken by a female native speaker of each language, and target words were embedded in a frame "I say X" in their native language. The *F0* curves were extracted from the speech samples using Praat (Boersma & Weenink, 2005)

Tone 3 (T3), and Tone 6 (T6); two rising tones, *high rising* [25 or 35][2] and *low rising* [23], Tone 2 (T2) and Tone 5 (T5), respectively; and one *low falling* [21] tone (Tone 4: T4). A minimal contrast set for these six Cantonese tones is illustrated by the words /si55/ 'poem', /si35/ 'history', /si33/ 'attempt', /si21/ 'time', /si23/ 'city', /si22/ 'trained person'.

Phonetic similarities and differences can be identified between Cantonese and Mandarin tones. Both Cantonese and Mandarin T1 are high level [55] (e.g., Mandarin: Duanmu, 2000; Howie, 1976; Cantonese: Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002). Similarly, although Cantonese T2 [25 or 35] is labeled as a high rising tone, and Mandarin T2 [35] is labeled as a mid rising tone, they have similar rising pitch and overlapping tone letter values [35] (Mandarin: Duanmu, 2000; Howie, 1976; Cantonese: Bauer & Benedict, 1997; Hashimoto, 1972; Yip, 2002). Further, as displayed in Figure 1(a, b), Cantonese and Mandarin T2 are typically produced with an initial dip followed by a rise (Cantonese: Bauer & Benedict, 1997; So, 1999; Mandarin: Fon & Chiang, 1999; Howie, 1976). Cantonese does not have a direct *phonological* (tonemic) counterpart to Mandarin T3 [214], however, Cantonese low rising T5 [23] does display an initial dip followed by a rise, and thus is *phonetically* similar to Mandarin T3. In addition, Cantonese lacks a high falling toneme corresponding to Mandarin T4 [51], but its T1 [55] does have a high falling allotone [53][3] (Bauer & Benedict 1997; Hashimoto, 1972; Yip, 2002), which is phonetically somewhat similar to Mandarin T4.

Japanese is a "pitch-accent language," which is a "subtype of tone language" (Yip, 2002, p.4). Although Japanese words can be accented or unaccented in their underlying forms, the words' pitch-accent patterns are predictable from the position of the accented (high tone: H) mora (the basic speech timing unit) due to the *pitch-accent rules*[4] (McCawley, 1978; Tsujimura, 1996). In some cases, pitch-accent patterns (e.g., LH vs. HL) can differentiate otherwise segmentally identical words with different lexical meanings (Haraguchi, 1999; Ito, Speer, & Beckman, 2003; McCawley, 1978; Tsujimura, 1996). For example, the Japanese two-mora sequence /ame/ illustrates a minimal pitch-accent pair (see Figure 1(c), bottom left) meaning "candy" when accented on the second mora (i.e., LH: [amé]), but "rain" when accented on the first (i.e., HL: [áme]) (Tsujimura, 1996, p.76).

Although the basic prosodic units for Japanese pitch accents and Mandarin tones are different (mora vs. syllable or rhyme: see Chao, 1968; McCawley, 1978), both Mandarin lexical tones and Japanese pitch accents have phonemic status in their languages, because they are minimally contrastive and mark lexical distinctions. Moreover, the *F0* patterns for Japanese LH and HL pitch accents are phonetically similar to those for Mandarin T2 [35] and T4 [51], with the exception that in Mandarin

---

2   Early literature generally described Cantonese Tone 2 as [35] (e.g., Hashimoto, 1972), but recently it has been described as [25], due to the fact that it shares a similar tonal onset as that of the Cantonese T5, low rising tone [23] (e.g., Bauer & Benedict, 1997; So, 1999; also see Figure 1(b)).

3   However, most Hong Kong speakers (our subjects' origin) "have lost the high falling tone, or use it in certain syntactic environments, or use it in free variation with high level" Tone 1 (Bauer & Benedict, 1997, p.167).

4   All morae that precede the accent are assigned high tones, and morae that follow the accent are assigned low tones. The first mora of the word is assigned a low tone when it is unaccented.

the pattern is realized over a single timing unit (syllable or rhyme) while in Japanese is it realized over two timing units (two morae). Indeed, some phonologists (Duanmu, 2004; Woo, 1969; Yip, 1980) describe Mandarin contour tones as a sequence of level tones. For example, rising and falling tones are described as LH and HL, respectively. Lastly, in terms of syllable weight, both Japanese (C)VV, (C)VC, and (C)VCV sequences (e.g., *ame*) and Mandarin CV: and CVC syllables are represented with two morae (Cutler & Otake, 1999; Duanmu, 1993, 2005; McCawley, 1978; Sugito, 2003; Yip, 2002).

English is a non-tone language, because it uses neither lexical tones *nor* pitch accents. It has been characterized as a "stress-accent language" (Beckman, 1986). Its use of distinctive pitch at the word level is very restricted. Even for lexical stress, for example, pitch is just one of several acoustic components (along with loudness, duration, and vowel realization/reduction) used to indicate stress in English homophonous pairs, such as *SUBject* (noun) and *subJECT* (verb) (Fear, Cutler, & Butterfield, 1995; Pennington & Ellis, 2004; see Figure 1(d), bottom right). Stressed versus unstressed syllables are typically produced with vowel quality and length differences (Beckman, 1986; Cutler & Otake, 1999; Fox, 2000; Gussenhoven, 2004) that are more consistent and salient than $F0$ differences. At the phrasal level, English intonation generally uses different pitch patterns to loosely associate with pragmatic functions, such as rising pitch for yes/no questions, and falling pitch patterns for statements. Sometimes, even the same word, "yes" for example, can be pronounced as a statement as well as a question depending on the pitch pattern (a rising or a falling one) when it is pronounced. However, akin to the inconsistent role of $F0$ in stress patterning in English, the linguistic meaning between specific pitch patterns and specific pragmatic functions is not fixed (Fox, 2000; Gussenhoven, 2004; Ladd, 1996). This is much different from lexical tones and pitch accents that can signal different lexical meanings, and have phonological status.

In the present study, we hypothesized that the *phonological and phonetic properties* of listeners' L1 *prosodic systems* (e.g., tone, pitch accent, and intonation) affect their perception of non-native tones. That is, we posit that the main influence is not due simply to the amount of experience with the use of tonality at the word level (i.e., the use of pitch alone to change lexical meaning),[5] but rather to the influence of the phonological status and/or phonetic features of tonal categories and contrasts in the listener's native prosodic system. Specifically, we expected that given their native system of lexical tone contrasts, Mandarin-naïve Cantonese listeners would have greater difficulty in distinguishing Mandarin tone pairs T1–T4 and T2–T3, relative to the other two groups. They could be expected to perceive both Mandarin T1 [55] and T4 [51] as exemplars of Cantonese T1 [55], since the high falling tone is an allotone of Cantonese T1. Cantonese listeners may tend to perceive both Mandarin T2 [35] and T3 [214] as their T2 [25], which is quite similar to Mandarin T2 [35] in that both have a slight dip followed by a rise (something like [325] if given a narrow numeric

---

5 It is clear that the experience of using pitch variation at the word level for the three languages ranges from extensive (Cantonese) to minimal (English). Cantonese as a tonal language will provide its speakers with substantial experience using pitch variations at the word level, more so than Japanese, and English will in turn provide the least systematic use of pure tonal information for lexical distinctions.

transcription), as well as to Mandarin T3 [214]. That is, to a naïve Cantonese listener these two Mandarin tones both share a high degree of phonetic similarity to a single Cantonese tone, T2 [25]. In contrast, neither Japanese nor English listeners were expected to experience such interference from their native prosodic systems, which lack such dipping pitch contours. Further, because Japanese, but not English, has both rising and falling pitch-accent patterns that are phonetically similar to Mandarin T2 [35] and T4 [51], Japanese listeners should more easily distinguish those Mandarin tones than should English listeners.

If Mandarin-naïve Cantonese listeners perform worse on tone pairs T1–T4 and T2–T3 than do the other listener groups, this will indicate that learners' phonological systems play an important role in the perception of non-native Mandarin tones. However, if Cantonese listeners instead perform better than the other two groups on all tone pairs, this will imply that general experience with the use of lexical tone, rather than specific phonological properties of the native tone system, underlie the perception of non-native tones.

2   Methods

2.1 Participants

Thirty adults were paid to participate in this study. They were recruited based on their native languages: (Hong Kong) Cantonese, Japanese, and (Canadian) English. The Cantonese listeners ($n = 10$) ranged in age from 18 to 26 years ($M = 21.7$ years), and the Japanese listeners ($n = 10$) ranged in age from 18 to 36 years ($M = 23.8$ years). All native speakers of Cantonese and of Japanese were born and raised in their home countries (Hong Kong and Japan, respectively), and came to Canada after the age of 15 years. The Canadian English listeners ($n = 10$) ranged in age from 18 to 35 years ($M = 21.7$ years), and were all born and raised in Canada with English as their only fluent language (functional monolinguals). All of the participants were undergraduate students either at Simon Fraser University or at other universities in Vancouver, British Columbia. They all passed a pure-tone hearing screening (250, 500, 1000, 2000, 4000, and 8000 Hz at 25 dB HL) prior to the experiment.

The selection of all the participants was based on two criteria: they had neither learned the target language (Mandarin) nor received formal musical training prior to or during the time of the study.[6] The former criterion ensured that experience with the target L2 tones was not present to affect the trainees' performance. The latter criterion was also crucial, because as mentioned earlier, previous studies have shown that listeners with musical training outperformed those without such training in both production and perception tasks with non-native tones (Alexander et al., 2005; Burnham & Brooker, 2002; Gottfried & Riester, 2000).

It should be noted that the Hong Kong Cantonese listeners recruited in this study were totally naïve to Mandarin for the following reasons. In Hong Kong (HK),

---

6   These criteria were quite stringent, and only about 10% of potential subjects met them. Originally, around 300 potential participants were recruited and screened but only the 30 included participants met both criteria.

there are actually people who are "gap people"—those who have never received formal training in Mandarin throughout their education in HK, but there are not many of these people. Because of this, the HK Cantonese speakers in this study were carefully selected and all fulfill the strict requirements (see footnote 6). Unlike the HK Cantonese speakers in the previous report (Lee et al., 1996), those in the present study left HK to move to Canada at the age of 15 years, on average. At the time of their move to Canada, Mandarin was not a compulsory course for HK high school students, and there were still quite a number of high school students who did not take any Mandarin courses. Our HK Cantonese speakers were restricted to include only native speakers from this specific group. Also, they had been living in Canada generally for an average of six years before they participated in the present study, thus further limiting their exposure to Mandarin. In the language background questionnaire, they all reported that they had never taken any Mandarin language courses, and that their daily languages were English and Cantonese only. Thus, their exposure to the Mandarin language was minimal, especially during their years in Canada.

2.2 Materials
A total of 120 tokens of six Mandarin syllables, *di*, *da*, *du*, *chi*, *cha*, and *chu* (in IPA: /ti/, /ta/, /tu/, /tʂʰi/, /tʂʰa/, and /tʂʰu/),[7] each produced 5 times with all four Mandarin tones (see Gottfried & Suiter, 1997) were recorded by two native Mandarin speakers from Beijing (a female and a male; $M = 23.5$ years). The first three syllables were used during the identification (ID) test, while the last three syllables were used during the familiarization session which was given to participants prior to the ID test. Note that these /tV/ syllables were used as the test stimuli because of the following reasons: (i) they were used in a previous study (Gottfried & Suiter, 1997); (ii) they consist of the three "point" vowels of the vowel space plus a simple short-lag unaspirated stop consonant (in IPA: /t/ + /i/, /u/ or /a/); (iii) these segments are found in virtually all languages. Since the listeners were all naïve to the target language, Mandarin, we limited the variation contributed by the consonants, so that listeners' performance should be restricted to tone perception. However, although we controlled variations caused by different syllable structures and consonant onsets or codas, we also used stimuli involving speech token variation. This was done by using two tokens per syllable and tone, since each token will involve slightly different acoustic properties (e.g., duration and *F0* values).

The target word was placed in the final position of a carrier sentence in Chinese [我说 X̲ ("I say X̲")]. Individual recordings were made in a sound-treated room in the Phonetics Laboratory at Simon Fraser University using a high quality microphone (Sennheiser MD46) connected to a CD-Recorder (Marantz CDR300). All digital files from the CDs (44.1 kHz) were extracted using a speech editing program (Goldwave, v.5.07), and saved as audio sound files in WAV format on a PC laptop computer (Dell Inspiron 600m). All target words were excised from the sentence frame, and normalized to the mean peak intensity. The 144 final tokens of *di*, *da*, *du* (3 syllables

---

7 For the word /tʂʰi/, the vowel /i/ will change to [ɹ], when it follows a retroflex consonant (Howie, 1976).

× 4 tones × 2 samples per tone × 2 speakers × 3 repetitions) and the 24 tokens of *chi*, *cha*, and *chu* (3 syllables × 4 tones × 2 speakers) were presented three times each in random order to four additional native Mandarin speakers (M = 24.5 years) who evaluated the intelligibility of the tones of the stimuli (see Guion et al., 2000; Wang, Spence, Jongman, & Sereno, 1999). All final stimuli were 100% correctly identified by each of these native Mandarin speakers in a four-alternative forced-choice Mandarin tone identification task.

2.3 Procedure

Prior to the experimental task, i.e., an ID test, there was a familiarization block consisting of 24 speech samples of the four Mandarin tones on three syllables (*chi*, *cha*, *chu*) spoken by the female and the male native Mandarin speakers. The goal of familiarization was for listeners to learn the tone labels for the stimuli: high level (Tone 1), mid rising (Tone 2), falling rising (Tone 3), and high falling (Tone 4). On the computer screen, 24 buttons were displayed according to the four Mandarin tones and speakers, along with their Mandarin tone labels. Each button was linked to a speech sample. Once the listener clicked on a button, an audio tone sample was presented. Participants were encouraged to listen to each of the speech samples at least once. This process was self-paced; however, participants were not allowed to spend more than two minutes in this familiarization block.

The participants then proceeded to the four-alternative forced-choice identification task. The test consisted of 144 trials in two blocks (a female set and a male set) which were presented in a counterbalanced order across participants in each group, and stimuli were randomized and presented individually within each block. Similar to the familiarization session, the process was self-paced. For each trial, participants were given five buttons on a computer screen. A *stimulus button* (colored white) was located at the top of the screen. Underneath the stimulus button, there were four *response buttons* (colored yellow) labeled with the four Mandarin tones. Participants were instructed to first click on the stimulus button once and listen to the stimulus over the headphones. Then, they were asked to give an answer by selecting one of the four labeled buttons. No feedback was given during the test. Once the listeners selected a response button, the program automatically provided a prompt to the next trial. The decision times for responses were generally less than 4 seconds.
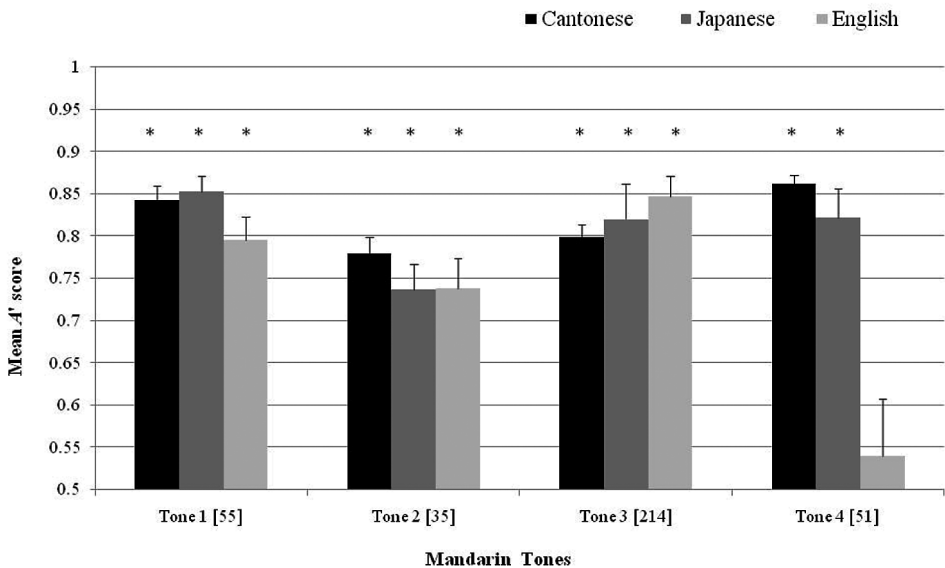
3 Results

Listeners' performance on the identification test was statistically evaluated in two different analyses. First, listeners' sensitivity to correct identification of each lexical tone was examined in terms of A-prime (A') scores (Snodgrass, Levy-Berger, & Haydon, 1985). The greater the accuracy of the listeners' tonal identifications, the closer the A' scores are to 1.0. In contrast, if the listeners are no better than chance in identifying a target tone, the A' score would be around 0.5. Second, in order to determine the types and the frequencies of the listeners' tonal confusions, the listeners' tonal identification errors were also examined. Confusion matrices were constructed for each L1 group.

3.1 Tone sensitivity: A-prime (*A'*) scores

Listeners' *A'* scores ranged from 0.54 to 0.86 (see Figure 2). Generally, the listeners' *A'* scores for T2 were lower than for the other three tones (T1, T3, and T4), except for the English listeners' *A'* scores for T4 (*A'* scores ≈ 0.54, the lowest score of all).

Figure 2

Mean *A'* scores for the four Mandarin tones by native speakers of Cantonese, Japanese, and English. The asterisk (*) indicates the *A'* score differs significantly from the 0.5 level (chance level performance)

Listeners' mean *A'* scores were submitted to a two-way mixed-design ANOVA with L1 Group (Cantonese, Japanese, English) as the between-subjects factor, and Tone (T1, T2, T3, T4) as the within-subjects factor. Significant effects of L1 group, $F(2, 27) = 6.130$, $p < .001$, and Tone, $F(3, 81) = 7.556$, $p < .001$, were found. Post-hoc Tukey's tests revealed that the listeners' mean *A'* scores for T2 were significantly lower than their mean scores for T1 and T3 ($ps < .001$).

The interaction of Group × Tone was also significant, $F(6, 81) = 9.102$, $p < .001$. To explore this interaction, four separate 1-way ANOVAs were carried out to evaluate the effect of Group on the mean *A'* scores for the four tones. The analyses indicated a significant Group effect for T4 only, $F(2, 27) = 15.847$, $p < .001$, but not for the other three tones ($ps > .05$). Post-hoc Tukey's tests revealed that the English listeners' mean *A'* score for T4 (mean *A'* ≈ 0.54) was significantly lower than that of the Cantonese (mean *A'* ≈ 0.86) and the Japanese (mean *A'* ≈ 0.82) listeners ($ps < .001$).

3.2 Tonal confusions

Table 1 summarizes listeners' tonal confusions according to their language backgrounds. Some patterns are found across all four listener language groups, while

Table 1

Confusion matrices for responses in the ID test for each listener group.

CANTONESE

Target	Response (%)				
	T1	T2	T3	T4	Total
T1	**66.94%**	9.72%	0%	23.33%	100%
T2	8.61%	**59.17%**	31.11%	1.11%	100%
T3	1.11%	50.56%	**48.33%**	0%	100%
T4	35.83%	2.22%	0.28%	**61.67%**	100%

JAPANESE

Target	Response (%)				
	T1	T2	T3	T4	Total
T1	**65%**	27.78%	0.56%	6.67%	100%
T2	7.22%	**56.67%**	33.33%	2.78%	100%
T3	1.11%	25.28%	**61.11%**	12.50%	100%
T4	25.83%	21.11%	0.28%	**52.78%**	100%

ENGLISH

Target	Response (%)				
	T1	T2	T3	T4	Total
T1	**69.17%**	17.22%	5.83%	7.78%	100%
T2	20%	**52.22%**	19.17%	8.61%	100%
T3	1.94%	22.22%	**60.28%**	15.56%	100%
T4	57.50%	20.28%	3.06%	**19.17%**	100%

Boldface values along the diagonal indicate percentages of correctly identified tones

others are language-particular (see Figure 3). First, tonal confusions were mainly found in three pairs of tones: T1–T4 ([55]–[51]), T2–T3 ([35]–[214]), and T1–T2 ([55]–[35]), bi-directionally.[8] Confusions appear to be proportional to the phonetic feature similarities between the tones in the pairs, since the tones in each of the three pairs do share some phonetic similarities at any of the three reference points along the contours—*pitch onset*, *mid-contour*, or *pitch offset* (see Figure 1). In contrast, errors for the T1–T3 ([55]–[214]), T2–T4 ([35]–[51]), and T3–T4 ([214]–[51]) pairs were apparently fewer. This appears to relate to the extreme dissimilarity of the phonetic features at the three reference points (see Figure 1).

The listeners' mean tonal errors were submitted to a three-way mixed-design ANOVA with Group (×3) as the between-subjects factor, and Feature similarity (Similar-features (SF) vs. Dissimilar-features (DF)) and Tone Pair (six levels for each feature group: T1→T2, T2→T1, T1→T4, T4→T1, T2→T3, and T3→T2 for SF; T2→T4,

---

8 For each tone pair, the tonal misidentifications could go in two directions. In the case of the T1–T4 pair, for example, T1 could be misidentified as T4, or vice versa.

T4→T2, T3→T4, T4→T3, T1→T3, and T3→T1 for DF) as the within-subjects factors.[9] The main effects of Group, $F(1, 27) = 3.510$, $p < .05$, Feature, $F(1, 27) = 168.383$, $p < .001$, and Tone Pair, $F(5, 135) = 3.243$, $p < .001$, were significant. The two-way interactions of Tone Pair × Group, $F(10, 135) = 5.463$, $p < .001$, and Feature × Group, $F(2, 27) = 6.753$, $p < .001$, as well as the three-way interaction (Group × Features × Tone Pair), $F(10, 135) = 5.307$, $p < .001$, were all significant.

To explore the 3-way interaction, two separate ANOVAs (Group × Tone Pair) were carried out for the dissimilar-feature and the similar-feature group, respectively. As for the dissimilar-feature group, the analysis found that the effects of Tone Pair, $F(5,135) = 17.855$, $p < .001$, and Group, $F(2, 27) = 3.738$, $p < .05$, were significant, as well as their interaction Tone Pair × Group, $F(10, 135) = 6.767$, $p < .001$. Individual 1-way ANOVAs (× 6) found that the Group effect was significant in all six pairs, T1→T2, $F(2, 27) = 4.206$, $p < .05$, T2→T1, $F(2, 27) = 5.272$, $p < .01$, T1→T4, $F(2, 27) = 11.113$, $p < .01$, T4→T1, $F(2, 27) = 8.721$, $p < .01$, T2→T3, $F(2, 27) = 5.334$, $p < .01$, and T3→T2, $F(2, 27) = 3.979$, $p < .01$. Individual post-hoc Tukey's tests were performed to examine the Group effect for each pair. The results are listed below.

> T1→T2: the English listeners incorrectly identified T1 as Target T2 significantly more than the Japanese listeners ($p < .05$) but not the Cantonese listeners, *ns*. The mean error difference between the Cantonese and the Japanese groups was not significant, *ns*.
>
> T2→T1: the Japanese listeners incorrectly identified T2 as Target T1 significantly more than the Cantonese listeners ($p < .05$) but not the English listeners, *ns*. The mean error difference between the Cantonese and the English groups was not significant, *ns*.
>
> T1→T4: the English listeners incorrectly identified T1 as Target T4 significantly more often than both the Japanese and the Cantonese listeners ($ps < .05$). The mean error difference between the Cantonese and the Japanese groups was not significant ($p > .05$).
>
> T4→T1: the Cantonese listeners incorrectly identified T4 as Target T1 significantly more frequently than both the Japanese and the English listeners ($ps < .05$). The mean error difference between the English and the Japanese groups was not significant ($p > .05$).
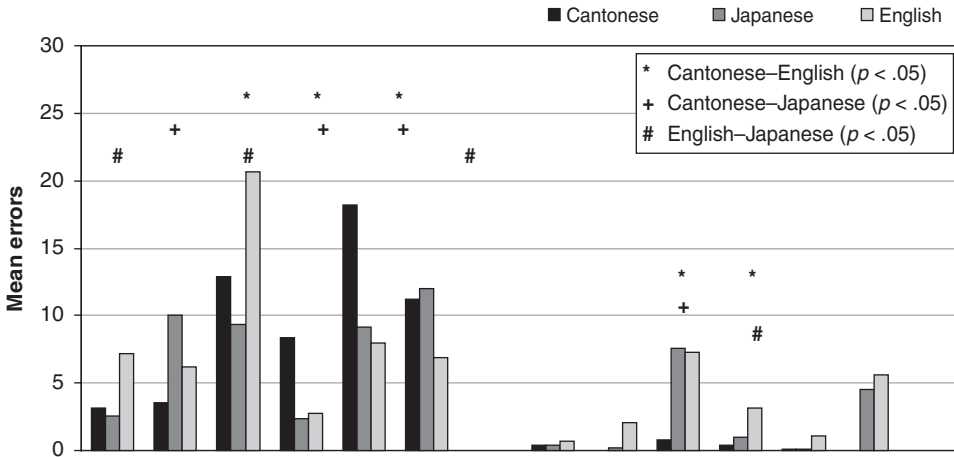>
> T2→T3: the Cantonese listeners incorrectly identified T2 as Target T3 significantly more often than both the Japanese and the English listeners. The mean error difference between the English and the Japanese groups was not significant, *ns*.
>
> T3→T2: the English listeners incorrectly identified T3 as Target T2 significantly less frequently than the Japanese ($p < .05$), but not the Cantonese listeners, *ns*. The mean error difference between the Cantonese and the English groups was not significant, *ns*.

9 In this study, each tone pair indicates a relationship between a response tone and a target tone. For example, in the T1→T2 pair, Tone 1 is the listener group's response (error) when the target tone is Tone 2.

Figure 3

Mean errors for the 12 tone pairs by native speakers of Cantonese, Japanese, and English. Each of the tone pairs on the x-axis represents a response to → target category. Symbols indicate significant pairwise group differences

As for the similar-feature set, the effects of Tone Pair, $F(5, 135) = 9.359$, $p < .001$, and Group, $F(2, 27) = 7.097$, $p < .01$, were significant, as well as their interaction, Tone Pair × Group, $F(10, 135) = 2.155$, $p < .05$. Individual 1-way ANOVAs (×6) showed a significant Group effect for only two pairs, T2→T4, Fs(2, 27) = 4.933, $p < .01$, and T4→T2, Fs(2, 27) = 5.817, $p < .001$. For the T1→T3, T3→T1, T3→T4, and T4→T3 pairs, the performance of all groups was comparable. No single group difference among the listener groups was found (ps > .05). Individual post-hoc Tukey's tests were performed for each of the two pairs for which the Groups differed significantly:

> T2→T4: the Cantonese listeners incorrectly identified T2 as Target T4 significantly less often than both the Japanese and the English listeners (ps < .05). The mean error difference between the English and the Japanese groups was not significant. T4→T2: the English listeners incorrectly identified T4 as Target T2 significantly more frequently than both the Japanese and the Cantonese listeners (ps < .05). The mean error difference between the Cantonese and Japanese groups was not significant (p > .05).

4 Discussion

4.1 L1 prosodic effects on perception of non-native tones

In this study, it was hypothesized that listeners' native (L1) prosodic systems would influence their performance when perceiving non-native tones (i.e., Mandarin, in this study). Specifically, due to the differences between their L1 prosodic systems, it was expected that the Cantonese listeners would have greater difficulty differentiating

the two pairs of Mandarin tones (T1–T4 and T2–T3), while the Japanese listeners would have less difficulty in learning the four Mandarin tones. The English listeners' performance was predicted to fall between those of the other two language groups, because English does not have a system for lexical tone and does not use pitch variations at the word level, and therefore, English speakers may be less sensitive to the pitch patterns at the word level than are the Japanese speakers.

The results indicate somewhat complicated effects of linguistic experience on non-native lexical tone perception. At first glance, the listeners' overall performance in the analyses of tonal sensitivities (A' scores) did not fully support the hypothesis. In particular, both the Cantonese and the Japanese listeners (native speakers of tone languages) outperformed the English listeners (native speakers of a non-tone language), but the Cantonese listeners did not perform better than the Japanese listeners. This indicates that listeners' L1 systems affect the perception of non-native tonal contrasts. In particular, native speakers of tone languages (Cantonese and Japanese) outperformed those of a non-tone language (English). Thus, linguistic experience with native tones plays a role in listeners' identification of non-native tones, as reported in Lee et al. (1996) and Wayland and Guion (2004).

However, the analysis of listeners' tone errors provides a new perspective on non-native tone perception. Tone pairs that share some similar features (T1–T2, T2–T3, and T1–T4) are more difficult to identify than other pairs that have only dissimilar features (T1–T3, T2–T4, and T3–T4). We will come back to this in Section 4.3. Interestingly, among the three problematic "similar" tone pairs, the Cantonese listeners consistently and significantly misidentified T4 [51] as Target T1 [55] and T2 [35] as Target T3 [214], but they made many fewer errors for the T1–T2 pair than did the English and the Japanese listeners. This exactly supports the hypothesis that Cantonese listeners would have problems in perceiving the tonal contrast between Mandarin T1 [55] and T4 [51] because they are similar to the allotones of Cantonese T1 [55], and between Mandarin T2 [35] and T3 [214] because they are phonetically similar to Cantonese T2 [25]. The same consistent patterns were not found in the Japanese and English listeners.

4.2 Lexical tone assimilations?

The results indicated that the Cantonese listeners had more difficulties in perceiving the T1–T4 and T2–T3 pairs ([55]–[51] and [35]–[214]) but not the T1–T2 pair ([55]–[35]). Perhaps, Perceptual Assimilation Model (PAM: Best, 1995; Best & Tyler, 2007) predictions about assimilations may be used to explain the patterns. It is possible that the Cantonese listeners assimilated Mandarin T1 [55] and T4 [51] to the Cantonese T1 (high level), which has two allotones, high level [55] and high falling [53]. This would correspond to a Single Category (SC) assimilation according to PAM. SC occurs when two non-native phones assimilate equally well or poorly to a single native phoneme. Poor discrimination is predicted. The results of this study were consistent with this suggestion, because the Cantonese listeners frequently misidentified Mandarin T4 as T1. The T2–T3 pair can be interpreted as a possible Category Goodness (CG) assimilation pair. CG occurs when both non-native phones assimilate to a single category; one often assimilates better than the other. Discrimination is moderate to excellent. In this, Mandarin T2 (mid rising [35]) and T3 (falling rising [214]), were

assimilated to Cantonese T2 (high rising [25]). Since Mandarin T2, rather than T3, is phonetically similar to Cantonese T2, Cantonese listeners should have a tendency to select Mandarin T2 as the better match most of the time. Our finding that the Cantonese listeners misidentified T2 as Target T3 was consistent with this assumption. There are two possible reasons for this tendency. First, the Cantonese high rising tone (T2 [25]) is phonetically similar to Mandarin T2 [35] in terms of the $F0$ patterns ($F0$ height and shape). Both are described with the same tone letters [35] in the literature (Cantonese: Hashimoto, 1972; Yip, 2002; Mandarin: Howie, 1976). The Cantonese T2 is produced with a falling and rising pattern (Bauer & Benedict, 1997; So, 1999), and so is the Mandarin T2 (Fon & Chiang, 1999). Second, the Cantonese tonal system does not have a tone that corresponds phonologically to Mandarin T3 [214]. When a Cantonese speaker listens to a Mandarin T3, the best candidate (the closest L1 tone) will be a Cantonese T2, because Mandarin T2 and T3 share considerable similarities in their pitch contours (e.g., the dip and the rising portions). Lastly, the T1–T2 pair may have been a Two Category (TC) Assimilation pair for Cantonese listeners. TC occurs when two non-native phones assimilate to two separate native phonemes. Discrimination between the non-native contrasts in these cases is expected to be excellent. Mandarin T1 [55] and T2 [35] may be perfectly assimilated to the Cantonese T1 [55] and T2 [35], respectively. Since the two Mandarin tones perfectly matched two separate Cantonese tones (in terms of the tone letters), as evidenced by the results of the present study, the Cantonese listeners did not experience difficulty in identifying Mandarin T1 and T2 when compared to the English and the Japanese listeners.

Our findings raise an interesting issue: PAM predicts that listeners should have more difficulties in perceiving the contrasts in the SC pair rather than the CG pair, but, interestingly, the Cantonese listeners in this study exhibited more problems in perceiving the tone contrast in the CG pair (T2–T3) than the SC pair (T1–T4). This may be related to the facts that Cantonese does have some durational differences among tones at the phonetic level, such as the entering tones (T7, T8, and T9), which are traditionally considered to be the short variations (or allotones) of the three level tones (T1, T3, and T6), respectively. Also, Cantonese T4 is the shortest in the system and easily distinguished from the other five phonemic tones. Thus, native speakers of Cantonese may still be sensitive to the vowel length difference (a phonetic feature), and use it as a perceptual cue for distinguishing Mandarin T1 and T4.

Japanese listeners, on the other hand, might assimilate Mandarin T2 [35] and T4 [51] to the Japanese LH and HL pitch-accent patterns, respectively, because their pitch contours are similar. If so, these tones may be assimilated as a TC pair, and listeners' perception of the pair is supposed to be excellent. However, in this study, the Japanese listeners had difficulties in identifying Mandarin T2 (mid rising) and T4 (high falling). Their difficulties may be partly due to their existing pitch-accent patterns. Although their pitch contours are similar to those of T2 and T4, it is possible that the Japanese listeners had not yet established the mappings of the pitch patterns between the two languages in the brief laboratory training task. In addition, the Japanese listeners also had a tendency to use the label T2 (mid rising) as answers. This can be explained by the fact that two-mora words in Japanese (e.g., *ame*) are likely to be pronounced with a rising pattern. A two-mora word may be produced with either a rising or a falling pitch-accent pattern (i.e., LH or HL). However, when the same word is unaccented,

Japanese speakers will also produce it with a rising pattern (see Cutler & Otake, 1999; Fujisaki, Ohno, & Tomita, 1996; Nagano-Madsen, 2003). Thus, it is not surprising that the Japanese listeners frequently selected Mandarin T2 [35] as their response.

If Mandarin T2 [35] and T4 [51] were perceptually assimilated to the Japanese LH and HL pitch-accent patterns, respectively, then Mandarin T1 [55] and T3 [214] might be a UU pair, failing to assimilate to any tone or pitch-accent pattern in the Japanese prosodic system. UU takes place if both non-native phones are uncategorized. Discrimination is still affected by L1 phones for uncategorized assimilations, but less so than for categorized ones. Moreover, the L1 effects are spread across the several L1 phonemes that are perceived as similar to the non-native phone. Discrimination should range between fair to good, depending on how similar the non-native phones are perceived to be relative to each other and to the closest L1 phonemes. According to PAM, listeners' perceptions of uncategorized sounds are less influenced by their L1 systems, but this depends on how well those listeners perceive the similarities of the uncategorized non-native contrasts. Perhaps, T1 and T3 have some phonetic properties (e.g., vowel duration and $F0$ patterns) that are relatively easy to perceive. T1 involves high pitch with limited pitch movement; T3 involves low pitch in the centre portion, and is produced with longer vowel duration (e.g., Ho, 1976; Howie, 1976). Therefore, the pairs, T1–T4, T2–T3, and T1–T2, may have been assimilated by Japanese listeners as three UC pairs. According to PAM predictions in UC cases, then, listeners should be able to discriminate the non-native sounds of a UC pair quite well.

For English, a stress-accent language (Beckman, 1986), the issue of tonal assimilation is more complicated. Hallé et al. (2004) suggested that there are two possible interpretations: lexical tones could be perceived either as uncategorized speech categories or as nonspeech. On the one hand, English employs tone/pitch contours at the sentential level (as intonation) to indicate discourse characteristics of utterances. For example, a falling pitch pattern signals a statement, and a rising pattern indicates a yes/no question. Thus, English listeners may perceive Mandarin tones as uncategorized speech categories, because English does have tone contours at the phrasal and sentential levels. On the other hand, English listeners may perceive lexical tones as "nonlinguistic melodic variations" (Hallé et al., 2004, p.416). Lexical tones are not part of the phonological system of English, and are not perceived as phonemic categories (Hallé et al., 2004). From this perspective, tones are nonspeech melodies that are *Non-Assimilable* (NA) to the listeners' phonological system, in the framework of the PAM.

However, if lexical tones can be perceived as certain kinds of prosodic categories, they could instead be phonologically categorized in PAM. Listeners may assimilate tones in terms of phrasal or sentential (e.g., T2 [35] to question intonation pattern), or even emotion intonation categories. They may also assimilate to English stress patterns (e.g., T4 [51] to the SW or trochee pattern, and T2 [35] to the WS or iambic pattern). Alternatively, it is also possible that tones may be assimilated to nonspeech, or musical melodies. That is, English listeners might indeed perceive tone contours as (nonspeech) melodic contours similar to the ones used in their prosodic systems (including intonation system). In any of these cases, the pitch patterns do not bear any linguistic significance similar to those of lexical tones or pitch-accent patterns, and do not have any phonemic (or tonemic) status in English.

4.3 Asymmetrical (differential) patterns in the perception of lexical tones

The results also confirmed that there was one consistent asymmetrical perceptual pattern among Mandarin tones by the three native listener groups, and the pattern seemed to be language-independent. Specifically, the tones in the T1–T4, T2–T3, and T1–T2 pairs ([55]–[51], [35]–[214], and [55]–[35]), sharing some phonetic similarities (e.g., pitch contours, and pitch height for the tonal onset and/or the offset), were easily confused with their counterparts. This is consistent with Polka's suggestions (1991, 1992) that a high degree of phonetic similarity between two non-native segments could increase perceptual difficulty for the listener. For the T1–T4 pair, both tones begin with a similar high pitch level. For the T2–T3 pair, both have a dip and a rising pattern. For the T1–T2 pair, both pitch contours end at a high pitch level. In fact, previous studies (Kiriloff, 1969; Miracle, 1989; Shen, 1989) have reported that non-native language learners have great difficulties in producing and perceiving different lexical tones. In particular, the tone pairs T2–T3 and T1–T4 are the most problematic.

In contrast, the tones in the T1–T3, T2–T4, and T3–T4 pairs ([55]–[214], [35]–[51], and [214]–[51]), which share no similarities in their tone contours, appeared to be less confusable. This may relate to the fact that the phonetic properties of the tones in each pair are dissimilar. For example, the $F0$ patterns for T1 (high level) and T3 (falling rising) are very different, at least in terms of the $F0$ patterns (level vs. falling rising) and duration (T3 is longer than T1; e.g., Howie, 1976). Thus, the results of the present study confirm that the phonetic characteristics (similar vs. dissimilar) of lexical tones also exert an effect on listeners' perception of Mandarin tones. Tones with more dissimilar features will be easier to discern and maybe to be learned, whereas tones that share similar features are likely to cause more perceptual and learning difficulties for listeners.

5 Conclusion

In sum, the effect of L1 prosodic backgrounds on categorization of non-native tones was supported in the present study. The results demonstrated that Mandarin-naïve Cantonese listeners' performance was constrained by their phonological system (e.g., the phonemic status and the $F0$ patterns of certain tones in the system). Similar L1 constraints were not observed in either the Japanese or the English listeners. Therefore, the effect of linguistic experience is more related to the constraints of the phonological systems of listeners' native languages than the degree of tonality use. This is consistent with PAM's assumptions and predictions. In addition, this study also found that there is a consistent asymmetrical (differential) perceptual pattern among Mandarin tones by the three native listener groups, (HK) Cantonese, Japanese, and English, and the pattern seems to be language-independent.

References

ALEXANDER, J. A., WONG, P. C. M., & BRADLOW, A. R. (2005). Lexical tone perception by musicians and non-musicians. In *Proceedings of Interspeech 2005*. Lisbon, Portugal.

ARCHIBALD, J. (1992). Transfer of L1 parameter settings: Some empirical evidence from Polish metrics. *Canadian Journal of Linguistics*, **37**, 301–339.

ARCHIBALD, J. (1993). The learnability of English metrical parameters by adult Spanish speakers. *International Review of Applied Linguistics*, **31**, 129–142.

BAUER, R. S., & BENEDICT, P. K. (1997). *Modern Cantonese Phonology*. Berlin and New York: Mouton de Gruyter.

BECKMAN, M. E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris.

BEST, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp.171–204). Timonium, MD: York Press.

BEST, C. T., McROBERTS, G. W., & GOODELL, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, **109**(2), 775–794.

BEST, C. T., & TYLER, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. Munro (Eds.), *Second-language Speech Learning: The Role of Language Experience in Speech Perception and Production. A Festschrift in Honour of James E. Flege* (pp.13–34). Amsterdam: John Benjamins.

BOERSMA, P., & WEENINK, D. (2005). Praat: Doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/

BROWN, C. (2000). The interrelation between speech perception and phonological acquisition from infant to adult. In J. Archibald (Ed.), *Second Language Acquisition and Linguistic Theory* (pp.4–63). Oxford: Blackwell.

BURNHAM, D., & BROOKER, R. (2002). Absolute pitch and lexical tones: Tone perception by non-musician, musician, and absolute pitch non-tonal language speakers. In J. Hansen & B. Pellom (Eds.), *The 7th International Conference on Spoken Language Processing* (pp.257–260). Denver, USA.

BURNHAM, D., & FRANCIS, E. (1997). The role of linguistic experience in the perception of Thai tones. In A. Abramson (Ed.), *Southeast Asian Linguistic Studies in Honour of Vichin Panupong* (pp.29–47). Bangkok: Chulalongkorn University Press.

CHAO, Y.-R. (1930). A system of tone letters. *La Maître phonétique*, **45**, 24–27.

CHAO, Y. R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

CUTLER, A., & OTAKE, T. (1999). Pitch accent in spoken-word recognition in Japanese. *Journal of the Acoustical Society of America*, **105**(3), 1877–1888.

DUANMU, S. (1993). Rime length, stress, and association domains. *Journal of East Asian Linguistics*, **2**(1), 1–44.

DUANMU, S. (2000). *The Phonology of Standard Chinese*. Oxford: Oxford University Press.

DUANMU, S. (2004). Tone and non-tone languages: An alternative to language typology and parameters. *Language and Linguistics*, **5**(4), 891–923.

DUANMU, S. (2005). Chinese (Mandarin), phonology of. In *Encyclopaedia of Language and Linguistics* (2nd ed.) (pp.351–355). Oxford: Elsevier.

FEAR, B. D., CUTLER, A., & BUTTERFIELD, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, **97**(3), 1893–1904.

FLEGE, J. E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Research* (pp.233–277). Timonium, MD: York Press.

FLEGE, J. E., McCUTCHEON, M. J., & SMITH, S. C. (1987). The development of skill in producing word-final English stops. *Journal of the Acoustical Society of America*, **82**, 433–447.

FON, J., & CHIANG, W.-Y. (1999). What does Chao have to say about tones. *Journal of Chinese Linguistics*, **27**(1), 15–37.

FOX, A. (2000). *Prosodic Features and Prosodic Structure: The Phonology of Suprasegmentals*. New York: Oxford University Press.

FUJISAKI, H., OHNO, S., & TOMITA, O. (1996). *On the levels of accentuation in spoken Japanese*. Paper presented at the Fourth International Conference on Spoken Language Processing.

GANDOUR, J. T. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics*, **11**, 149–175.

GANDOUR, J. T. (1984). Tone dissimilarity judgments by Chinese listeners. *Journal of Chinese Linguistics*, **12**, 235–261.

GANDOUR, J. T., & HARSHMAN, R. A. (1978). Cross-language difference in tone perception: A multidimensional scaling investigation. *Language and Speech*, **21**, 1–33.

GOTTFRIED, T. L., & RIESTER, D. N. (2000). Preliminary results from study of pitch perception and Mandarin tone identification. *Journal of the Acoustical Society of America*, **108**, 2604.

GOTTFRIED, T. L., & SUITER, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics*, **25**(2), 207–231.

GUION, S. G., FLEGE, J. E., AKAHANE-YAMADA, R., & PRUITT, J. C. (2000). 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*, **107**(5), 2711–2724.

GUION, S. G., HARADA, T., & CLARK, J. J. (2004). Early and late Spanish–English bilinguals' acquisition of English word stress patterns. *Bilingualism: Language and Cognition*, **7**, 207–226.

GUSSENHOVEN, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.

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

HARAGUCHI, S. (1999). Accent. In N. Tsujimura (Ed.), *The Handbook of Japanese Linguistics* (pp.1–30). Oxford: Blackwell.

HASHIMOTO, A. O.-K. Y. (1972). *Phonology of Cantonese*. Cambridge: Cambridge University Press.

HIRATA, Y. (2004). Training native English speakers to perceive Japanese length contrasts in word versus sentence contexts. *Journal of the Acoustical Society of America*, **116**(4), 2384–2394.

HO, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, **33**, 353–367.

HOWIE, J. (1976). *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

HUME, E., & JOHNSON, K. (2003). The impact of partial phonological contrast on speech perception. In M. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp.2385–2388). Barcelona, Spain.

ITO, K., SPEER, S. R., & BECKMAN, M. E. (2003, March). *The influence of given-new status and lexical accent on intonation in Japanese spontaneous speech*. Paper presented at the Presentation to the Annual CUNY Conference on Human Sentence Processing, Boston, MA.

JAMIESON, D. G., & MOROSAN, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /D/-/T/ contrast by francophones. *Perception & Psychophysics*, **40**(4), 205–215.

KIRILOFF, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica*, **20**, 63–67.

LADD, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

LEATHER, J. (1983). Speaker normalization in perception of lexical tones. *Journal of Phonetics*, **11**, 373–382.

LEATHER, J. (1987). Interrelation of perceptual and productive learning in the initial acquisition of second-language tone. In A. James & J. Leather (Eds.), *Second Language Speech: Structure and Process* (pp.77–101). Berlin: Mouton de Gruyter.

LEE, Y.-S., VAKOCH, D. A., & WURM, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, **25**(5), 527–542.

LISKER, L., & ABRAMSON, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In B. Hala, M. Romportl, & P. Janota (Eds.), *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp.563–567). Prague: Academia.

LOGAN, J. S., LIVELY, S. E., & PISONI, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, **89**(2), 874–886.

McALLISTER, R., FLEGE, J. E., & PISKE, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics*, **30**(2), 229–258.

McCAWLEY, J. D. (1978). What is a tone language? In V. A. Fromkin (Ed.), *Tone: A Linguistics Survey* (pp.113–131). New York: Academic Press.

MIRACLE, C. (1989). Tone production of American students of Chinese: A preliminary acoustic study. *Journal of the Chinese Language Teachers Association*, **24**, 49–65.

NAGANO-MADSEN, Y. (2003). Phonetic realization of the HL and LH accents in Japanese. In S. Kaji (Ed.), *Proceedings of the Symposium "Cross-linguistic Studies of Tonal Phenomena"*. Tokyo: Tokyo Press.

PENNINGTON, M. C., & ELLIS, N. C. (2004). Cantonese speakers' memory for English sentences with prosodic cues. *The Modern Language Journal*, **84**(3), 372–389.

POLKA, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America*, **89**(6), 2961–2977.

POLKA, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, **52**(1), 37–52.

POLKA, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *Journal of the Acoustical Society of America*, **97**(2), 1286–1296.

SHEN, X. S. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, **24**(3), 27–47.

SNODGRASS, J. G., LEVY-BERGER, G., & HAYDON, M. (1985). *Human Experimental Psychology*. New York: Oxford University Press Inc.

SO, C. K. (1999). An acoustic analysis of Cantonese rising tones. *Journal of the Acoustical Society of America*, **106**(4), 2153.

STAGRAY, J. R., & DOWNS, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, **21**, 143–163.

STRANGE, W. (Ed.). (1995). *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Timonium, MD: York Press.

SUGITO, M. (2003). Timing relationships between prosodic and segmental control in Osaka Japanese word accent. *Phonetica*, **60**(1), 1–16.

TSUJIMURA, N. (1996). *An Introduction to Japanese Linguistics*. Oxford: Blackwell.

WANG, W. S.-Y. (1976). Language change. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences* (Vol. 280, pp.61–72).

WANG, Y., SPENCE, M. M., JONGMAN, A., & SERENO, J. A. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, **106**(6), 3649–3658.

WAYLAND, R. P., & GUION, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, **54**(4), 681–712.

WERKER, J. F., & TEES, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, **75**, 1866–1878.

WOO, N. (1969). *Prosody and Phonology*. Cambridge, MA: The MIT Press.

YIP, M. (2002). *Tone*. Cambridge: Cambridge University Press.

YIP, M. (1980). *The Tonal Phonology of Chinese*. Cambridge, MA: The MIT Press.