THE INFLUENCE OF L1 PROSODIC BACKGROUND ON THE LEARNING OF MANDARIN TONES: PATTERNS OF TONAL CONFUSION BY CANTONESE AND JAPANESE NAÏVE LISTENERS*

Connie K. So
Simon Fraser University

1. Introduction

Human perception of non-native sound contrasts are strongly influenced by their native (L1) phonological system (Best 1995; Flege 1995). For example, it has been well documented that Japanese speakers have difficulties discriminating English /r/ and /l/ (Goto 1971, Lively, Pisoni, Yamada, Tohkura & Yamada 1994). Likewise, English speakers have difficulty in discerning the difference between the Hindi contrasts: voiceless retroflex stop [ʈ] and voiceless dental stop [t̪] (Werker 1991). Jamieson & Morosan (1989) also found that English consonant contrast /θ/ and /ð/ posed a problem for French learners of English.

Among existing working models of L2 speech perception, such as Speech Learning Model (SLM; Flege 1995) and Perceptual Assimilation Model (PAM; Best 1995), PAM is a phonologically-based model that focuses on the effect of L1 phonological system on perception of non-native sound contrasts. It posits that listeners' native (L1) phonology greatly affects their perception of non-native phones, as listeners perceptually assimilate the non-native phones to their native phonemes (Best 1995; Best, McRoberts & Goodell 2001). A non-native phone may be perceptually assimilated as a ***Categorized*** exemplar for some L1 phonemes, an ***Uncategorized*** segment that falls somewhere in between two native phonemes, or a ***Nonassimilable*** nonspeech sound that has no identifiable similarity to any native phoneme. According to PAM, listeners' abilities to discern the non-native phones are predictable and dependent on how the non-native phones assimilate to native segments. There are a few possible types of assimilations. When two non-native phones assimilate to two phonetically similar native phones separately, ***Two Category assimilation*** **(TC)** occurs and leads to successful perception of the non-native phones. In contrast, when two non-native phones assimilate to a single native phone, ***Single Category assimilation*** **(SC)** takes place and hinders listeners' perception of the non-native phones. In the situation that one of the non-native phones assimilates better than the other, it is termed ***Category Goodness difference*** **(CG)**. In addition, when one non-native sound is not categorized and the other is categorized, it forms an ***Uncategorized-Categorized pair*** **(UC)**. If both are not categorized, then they will be an ***Uncategorized-Uncategorized pair*** **(UU).** However, if the phone is completely dissimilar to any native phonemes, it will be perceived as a ***Non-Assimilable*** **(NA)** nonspeech sound, and native phonology does not have any influence on it. Although PAM originally focuses on speech segments, Hallé, Chang & Best (2004) suggested that PAM can be applicable to suprasegments as well. The present study will explore the possibility of their proposition.

* This research was supported by SSHRC. I would like to thank the audience of the 35th CLA meeting for their comments and discussions. All errors are my own.

Actes du congrès annuel de l'Association canadienne de linguistique 2005.
Proceedings of the 2005 annual conference of the Canadian Linguistic Association.

In addition to the influence of L1 linguistic background, previous studies have shown that laboratory training assists the learning of non-native sound contrasts. Generally, "lengthy periods of auditory training" (Ortega-Llebaria, Faulkner & Hazan 2001:40) will assist non-native learners to discriminate the phonemic contrasts of a second language. For example, French learners of English were trained to improve the abilities to identify the English consonant contrast /θ/ and /ð/ (Jamieson & Morosan 1989). Japanese speakers showed significant improvement in identifying English /l/ and /r/ better after training, and the improvement still presented after six months (Lively et al. 1994).

Similarly, research on lexical tone training has showed that training improves learners' perception, as well as their production, of lexical tones that are not contrastive in their first language (L1). Leather (1990) trained a group of naïve Dutch listeners to perceive and produce Mandarin tones. He found that the naïve listeners benefited more from perceptual training than from production training in learning the tonal system. Wang et al. (1999, 2003) demonstrated that perceptual training effectively helped American learners of Mandarin to perceive and produce lexical tones better than those who did not receive training. In addition, the training effect on learners' performance was still found after six months, and thus the results concurred with the findings of other perceptual studies focusing on segmental level (e.g., Lively et al. 1994). Further, Wayland & Guion (2004) trained both native speakers of English and (Taiwanese) Mandarin to learn a pair of Thai tones (mid vs. low level tones). They found that Mandarin speakers performed better than did the English speakers in discriminating the tone pair, indicating that participants' L1 prosodic background might be at work during lexical tone training.

However, it is still far from clear to what extent listeners' L1 prosodic backgrounds affect their learning of a new tonal system. Most previous studies either recruited native speakers of a single non-tone language (Leather 1990; Wang et al. 1999), or focused on a specific pair of tones (Wayland & Guion 2004) instead of a complete tonal system. A number of issues have remained unresolved. (1) Will ones' L1 prosodic backgrounds always facilitate learning of new non-native tones? (2) Would there be any differences in performance among native speakers of different tone languages? For example, will speakers of a tone language perform differently from those of a pitch accent language, a subtype of tone language (Yip 2002)?

In an attempt to address the above issues, the present research investigated the effect of native language (L1) prosodic system on second language (L2) tonal acquisition. This study examines the perception of Mandarin tones before and after training by native speakers of two tone languages (Cantonese vs. Japanese), which are typologically different (i.e., a tone vs. a pitch accent system). Further, unlike previous studies, this study also attempted to explain listeners' performance differences using PAM. In this case, non-native tones were expected to assimilate to the native tones by the two listener groups.

1.1. Prosodic systems of Mandarin, Cantonese, and Japanese

Mandarin is the target language for this study. As a typical tone language, it uses tones to convey different word meanings. There are four lexical tones in its system: Tone 1 (high level), Tone 2 (mid-rising), Tone 3 (falling-rising), and

Tone 4 (high falling). In terms of Chao's tone letters[1] (1930), they are represented as [55], [35], [214], and [51] respectively. These four tones are distinctive in their pitch contours or tonal patterns. However, the fact that some tones resemble one another causes perceptual difficulties for non-native learners (see Kiriloff 1969; Miracle 1989; Shen 1989). For instance, Tone 1 and Tone 4 start at a high level pitch. For Tone 2 and Tone 3, both have a rising portion of the pitch contours, and also exhibit different degrees of falling-rising patterns. According to Fon & Chiang (1999), (Taiwanese) Mandarin's rising tone (Tone 2) shows a dipping at the initial portion and then a moderately rising pattern in the final portion.[2]

Cantonese is another typical "lexical tone language" (Yip 2002:2). Its tonal system consists of six phonemic tones: Three level tones (Tone 1 [55], Tone 3 [33], and Tone 6 [22]), two rising tones (Tone 2 [35] and Tone 5 [23]) and one falling tone (Tone 4 [21]). As in Mandarin tonal system, Cantonese also has the same high level tone [55] and a high falling tone [53] in its system at the phonetic level, but they are allotones of a Cantonese toneme, Tone 1 [55]. The high falling tone is well documented in a tone sandhi situation[3] (Hashimoto 1972; Bauer & Benedict 1997). Cantonese Tone 2 is also described with the same tone letters [35] as Mandarin Tone 2. In addition, similar to (Taiwanese) Mandarin rising tone (Tone 2), both Cantonese rising tones exhibit a dipping at the initial portion and then a moderately rising pattern in the final portion (Bauer & Benedict 1997; So 1999).

Japanese, a pitch accent language - a subtype of tone languages (Yip 2002) - also has tones in its prosodic system, but with a few contrasting tones only (Yip 2002:4). Japanese's pitch contour is fixed for individual words (Duanmu 2004). It uses pitch (or accentuation) for differentiating lexical items (Ito, Speer & Beckman 2003; McCawley 1978). For example, the syllable /ame/ illustrates a minimal pair based on different pitch accent patterns. It means "rain" when it is accented on the second syllable (i.e., LH), but it means "candy" when it is accented on the first syllable (i.e., HL). Thus, Japanese pitch accent patterns (LH and HL) are similar to those of Mandarin Tone 2 and Tone 4. In fact, generative phonology typically describes Chinese contour tones in a sequence of level tones. Rising and falling tones are described as LH and HL, respectively (Duanmu 2004; also see Woo 1969; Yip 1980). In addition, Chinese CV (full) syllables are typically analyzed with two moras (Duanmu 2004, 2005). This kind of representation matches perfectly with Japanese bisyllabic words (e.g., /ame/), since both Chinese and Japanese words have the same number of moras in this case.

Given the backgrounds of the three languages, it is hypothesized that owing to the influence of their native prosodic systems, Cantonese listeners have greater difficulty in distinguishing Mandarin tones in the pairs, Tone 1-Tone 4 and Tone 2-Tone 3. For the former tone pair, Cantonese speakers may

---

[1] Digits are *tone letters* devised by Chao; the system provides "simplified time-pitch graphs of the voice" (1930:24). Tones are represented by 2 or 3 letter values to indicate the initial and the final or the initial, the medial and the final.

[2] In fact, native Mandarin speaker from Beijing in this study also produced the rising tone (Tone 2) with a "falling and rising" pattern.

[3] A high falling becomes a high level before a high falling or high level. However, unlike Guangzhou speakers, most Hong Kong Cantonese speakers "have lost the high falling tone, or use it in certain syntactic environments, or use it in free variation with high level" (Bauer & Benedict 1997:167).

perceive them as one single toneme (Tone 1), since the high falling tone is an allotone of high level tone (Tone 1) in Cantonese. For the latter one, they may perceive the tones as Tone 2 most of the time, because Cantonese Tone 2 [35] is similar to Mandarin Tone 2 [35], and Mandarin Tone 2 and Tone 3 share high degree of similarity in their pitch contours.

2. Method

2.1 Participants

Two groups each of 10 listeners, Cantonese (Hong Kong) and Japanese, were recruited as trainees. The Cantonese speakers ranged in age from 18–26 years (*M*=21.7 years), and the Japanese speakers had a mean age of 23.8 years with a range of 18–36 years. They were undergraduate students either at Simon Fraser University or at community colleges in Vancouver, British Columbia. They all passed a pure-tone hearing screen prior to the experiment. All participants had neither learned the target language (Mandarin) nor received formal music training at the time of the study. The latter selection criterion is crucial, since previous studies have shown that listeners with prior musical knowledge/experience outperform those who do not in both production and perception tasks (e.g., Gottfried & Riester 2000). All native speakers of Cantonese and Japanese were born and raised in their homeland and came to Canada after the age of 15 years.

2.2 Stimuli

Four native Mandarin speakers (2 female, 2 male) served as speakers. They were born and raised in Beijing, China. They ranged in age from 22–25 years (*M*=23.5 years). Five instances of each of 18 Mandarin consonant-vowel (CV) syllables with each of the four tones comprising the consonant in the context of a vowel /i/, /u/ or /a/ were recorded from each of the speakers. The target syllable was placed in the final position of a carrier phrase in Chinese. 我 说 X (“I say X”). Individual recordings were made in a sound-treated room in the Phonetics Laboratory at Simon Fraser University using a MB Quart K800 headworn microphone connected to a Marantz CDR300 CD-recorder. All recorded utterances were digitally sampled at 44.1 kHz (16 bits) and converted to audio files. They were used as stimuli for a pretest, training sessions, and two posttests.

The stimuli were judged by two additional native Mandarin speakers (22 and 27 years old, respectively) to evaluate the intelligibility of the stimuli, and they were identified with 100% accuracy. Training stimuli were 288 words (12 root-words x 4 tones x 2 speakers x 3 repetitions) spoken by two of the speakers, and were evenly arranged in six sets of exercises. Testing materials were another 15 root-words with the four tones spoken by the four native Mandarin speakers. The followings are the details of the stimuli used in this experiment:

- Pretest: syllables (*di, da, du*)[4] spoken by one female and one male native Mandarin speakers.

[4] The italicized words are in Mandarin Pinyin, not in IPA.

- Posttest1: the pre-test stimuli were re-used, plus other novel stimuli for the generalization tests (Gen1 and Gen2). Stimuli used during the training sessions were not re-used in the test. The Gen1 used the same stimuli (*di, da, du*) spoken by two novel speakers (another female and male native Mandarin speakers), while the Gen2 employed new stimuli (*chi, cha, chu*) spoken by old speakers.
- Posttest2: the post-test 1 was re-used.

2.3 Procedure

This study employed a pretest-posttest paradigm (cf. Lively et. al. 1994). All experimental sessions were given to individual participants through a custom-made program running in a Dell notebook (Inspiron 600m).

The experiment had four phases. In phase I, all participants were provided with a brief familiarization session for the four lexical tones before they took the pre-test. In phase II, they were trained with either simple auditory training sessions or auditory-visual enhanced training sessions[5]. The numbers of the training sessions were not fixed, which depended on the performance of the listeners. The sooner their scores reached 80% of the training exercises, the fewer training sessions the listeners would take. In phase III, the posttest1 took place 1-2 days after training. Lastly, in phase IV, the posttest2 was given approximately a month after training.

3. Results

The performance between Cantonese and Japanese listeners in the pretest, and the two posttests were compared in three domains: (a) percent correct identification scores, (b) listeners' sensitivity (A') to tonal identification, and (c) their tonal confusions (errors).

3.1 Percent correct identification scores (%)

Mean percent correct identification scores (%) in the pretest and posttests are shown in (1). In general, the scores between the Cantonese and Japanese listeners groups were comparable to each other, though Japanese listeners had slightly higher scores than did the Cantonese listeners in the two posttests.

(1)

Groups	Pretest	Posttest1			Posttest2		
	Test	Test	Gen1	Gen2	Test	Gen1	Gen2
Can	59%	86%	86%	86%	83%	84%	81%
Jap	58%	90%	93%	90%	90%	93%	88%

The data were submitted to a mixed ANOVA analysis with Group (Can vs. Jap) as the between-subjects factor, and Test (pretest, posttest1, posttest2) as the within-subjects factor. The analysis revealed that there was a significant effect of Test [$F(2,36)=119.597$, $p<0.0001$]. However, neither the effect of Group nor the interaction between Group and Test was significant ($ps>0.05$). Two separate ANOVAs were carried out to further investigate the effect of Test

[5] The performance differences between the two training approaches will not be discussed in this study, as they will be reported elsewhere.

on listener groups' performance. The analyses found that the effect of Test was significant for each group [*Fs*(9,18)=47.336 and 74.682, *ps*<0.0001]. Tukey pair-wire comparisons further indicated that the scores for the two posttests were significantly higher than those of the pretest (*ps* <0.05).

With respect to the two Generalization tests (Gen1 & Gen2), individual t-tests were used to examine the difference in mean % correct between the two groups for each of the posttests. It has been found that Cantonese listeners' mean % correct scores were comparable to those of Japanese listeners in all Gen tests (*ps*>0.05), except the Gen2 in posttest 1 [*t*(9)=-2.565, *p*<0.05].

In sum, the results suggested that listeners of both groups showed improvements after the training, but their performance was consistently similar to each other. For the Gen tests, only minor differences were found in the mean % correct scores between the two listener groups.

3.2 Sensitivity to tonal identification (A' scores)

Listeners' mean A' scores for the four lexical tones in the identification tests (pretest, posttest1, and posttest2) are listed in (2), and the ones for the Gen1 and Gen2 are shown in (3) and (4), respectively. Overall, Japanese listeners' mean A' scores for the tests were greater than those of Cantonese listeners.

(2)

	Pretest				**Posttest1**				**Posttest2**			
	T1	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
Can	.85	.78	.80	.86	.96	.96	.95	.94	.96	.94	.91	.95
Jap	.86	.76	.82	.83	.98	.98	.95	.96	.98	.97	.95	.98

(3)

	Posttest1 (Gen1)				**Posttest2 (Gen1)**			
	T1	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
Can	.97	.96	.96	.95	.96	.96	.93	.94
Jap	.98	.98	.97	.98	.99	.97	.96	.98

(4)

	Posttest1 (Gen2)				**Posttest2 (Gen2)**			
	T1	**T2**	**T3**	**T4**	**T1**	**T2**	**T3**	**T4**
Can	.96	.96	.94	.95	.95	.93	.89	.95
Jap	.97	.97	.95	.97	.97	.93	.93	.97

To determine how sensitive listeners were to correctly identify the target tone from the other three tones, a 3-way mixed ANOVA was used with Group as the between-subjects factor, and Test and Tone (T1, T2, T3, &T4) as the two within-subjects factors. The analysis indicated significant effects on listeners' mean A' scores of Test [*F*(2,36)=109.163, *p*<0.0001] and Tone [*F*(3,54)=4.314, *p*<0.001]. The interaction between Test and Tone was also significant [*F*(6,108)=5.733, *p*<0.0001]. However, no significant effects of Group (*p*>0.05), and interaction among Group, Test, and Tone (*p*>0.05) were obtained. The findings suggested that the four tones are not equally identified in the pretest and the two posttests. However, since this was beyond the focus of this study, the interaction was not explored further. Regarding the performance of the two groups, Japanese mean A' scores were similar to those of Cantonese listeners.

For the Generalization tests in the two posttests, individual mixed ANOVAs were performed with Group as the between-subjects factor and Tones as the within-subjects factor. The results revealed that the effect of Tones was only found in the Gen1 [$F(3,54)=3.707$, $p<0.05$] and Gen2 [$F(3,54)=5.856$, $p<0.01$] in the posttest2. The other ones (Gen2 in the posttest1 and Gen1 in the posttest2) were not significant (*ps*>0.05). The effects of Group, and Group x Tones interaction did not approach significance ($p>0.05$).

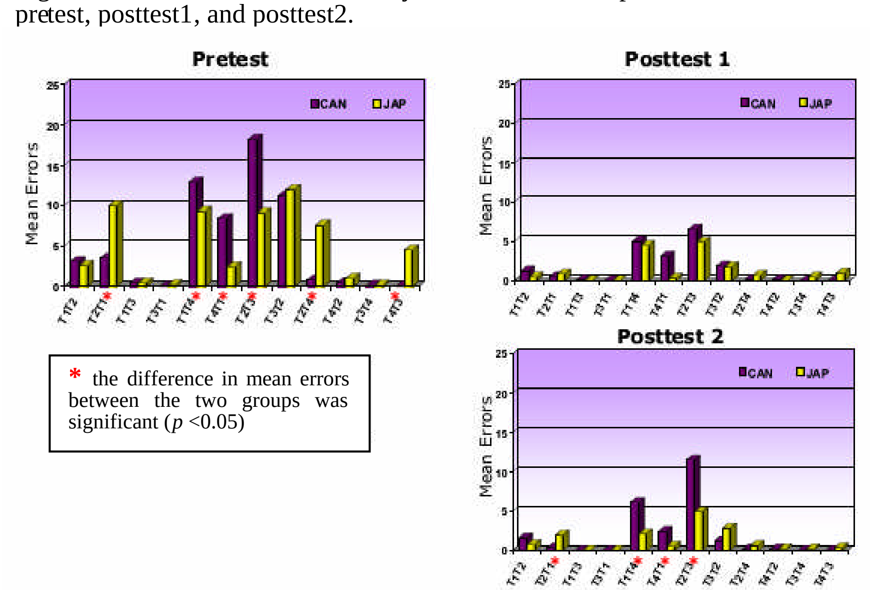
Overall, as expected, Cantonese and Japanese listeners showed improvement after training but no significant differences in their sensitivities of tones were found. This pattern was also observed in the generalization tests.

3.3 Tonal confusion (errors)

Listener groups' mean errors for 12 tone pairs[6] in the pretest and posttests are graphed in Figure 1, and the ones in the Gen1 and Gen2 are presented in Figures 2-3, respectively. Generally, a few patterns were observed:

(1) As expected, listeners' errors in the pretest are greater than those in the posttest1 and posttest2.
(2) Errors in the posttest1 are fewer than those in the other two tests.
(3) Errors in all the tests were mostly found in T1-T4 and T2-T3 pairs (bidirectional), except that Japanese listeners showed more error types in the pretest. However, Cantonese listeners seemed to have substantially more errors in these two tone pairs.

Figure 1. Mean tonal confusions by Cantonese and Japanese listeners in the pretest, posttest1, and posttest2.



[6] In this study, each tone pair indicates a relationship between a response tone and a target tone. For example, in the T1T2 pair, Tone 1 is listener group's response (error) when the target tone is Tone 2.

Figure 2. Mean tonal confusions by Cantonese and Japanese listeners in Gen1 in posttest1 and posttest2.

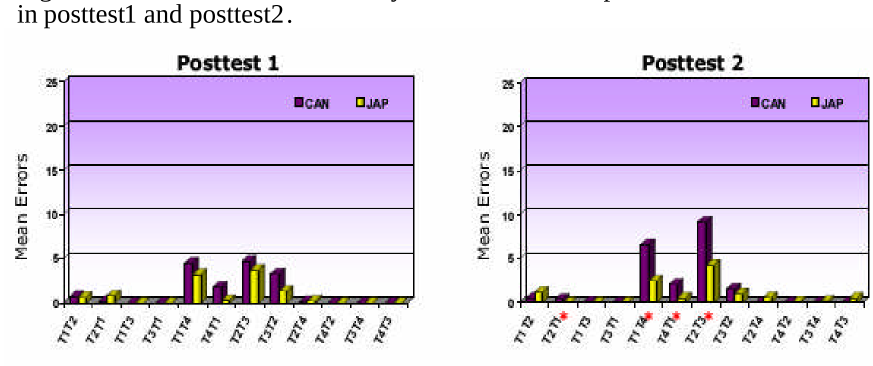
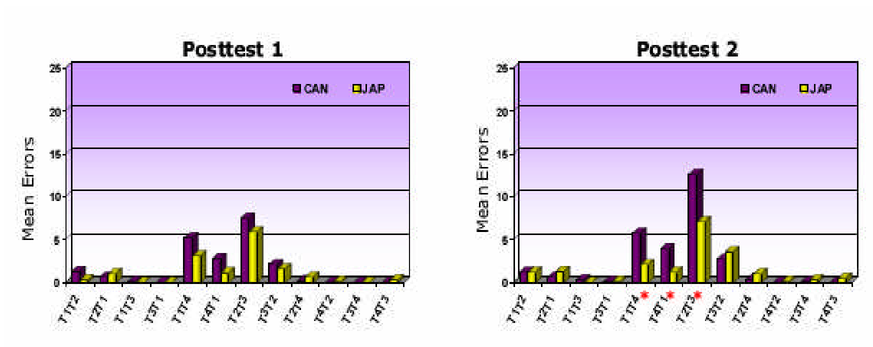


Figure 3. Mean tonal confusions by Cantonese and Japanese listeners in Gen2 in posttest1 and posttest2.



A 3-way mixed ANOVA was performed on the mean tonal errors of the listeners in both Cantonese and Japanese groups, with Group as the between-subjects factor, and Test and TonePair[7] (T1T2, T2T1, T1T3, T3T1, T1T4, T4T1, T2T3, T3T2, T2T4, T4T2, T3T4, T4T3) as the two within-subjects factors. The results showed that the effect of Group was not found (p>0.05), but there were significant effects of Test [F(2,36)=119.831, p<0.0001], and TonePair [F(11,198)=35.805, p<0.0001]. The Group x Test interaction was not significant (p>0.05). However, the interactions of Group x TonePair, and Test x TonePair were significant [F(11,198)=6.746, p<0.0001, and F(22,396)=9.951, p<0.0001, respectively]. In addition, a Group x Test x TonePair interaction was also significant [F(22,396)=3.217, p<0.0001].

To further explore the 3-way interaction among Group, Test, and TonePair, three separate mixed ANOVAs were performed for the tests (pretest, posttest1 and posttest2). Consistently, no effect of Group was found in all three tests (ps>0.05). But the effect of TonePair was found in all tests [Fs(11,198)=29.841, 10.603, and 18.876, ps<0.0001]. However, significant

[7] Huang (2001) also used tone pair as a factor to investigate the differences in reaction time of her listeners' perception responses.

effect of Group x TonePair was found only in the pretest and posttest2 *Fs*(11,198)=6.958, and 2.230, *ps*<0.0001], but not in the posttest1 (*p*>0.05), suggesting that the differences in mean errors between Cantonese and Japanese listeners in each test (pretest and posttest2) were greater in some, but not all, tone pairs. Individual t-tests were used to compare the mean errors between the two groups in the pretest and posttest2. The results indicated that group differences in mean errors were mainly significant in the tone pairs (bidirectional): T1-T4, T2-T3, and sometimes T1-T2. The detailed patterns are shown in Figure 1. The red asterisk (*) in the figure indicates the mean difference between the groups was significant (*p*<0.05).

For the Generalization tests, individual mixed ANOVAs were performed for the Gen1 and Gen2 in the posttest1 and posttest2. Consistently, no effect of Group was found [*Fs*(1,18)=0.884, 0.864, 1.866 and 1.507, *ps*>0.05], but the effect of TonePair was found in all tests [*Fs*(11,198)=7.447, 12.780, 11.939, and 21.804, *ps*<0.0001]. However, significant effect of Group x TonePair was found only in the Gen1 and Gen2 in the posttest2 [*Fs*(11,198)=2.23, and 2.862, *ps*<0.05 and 0.01, respectively], but not those in the posttest1 (*ps*>0.05). These suggest that the differences in mean errors between Cantonese and Japanese listeners are greater in some tone pairs in the posttest2, but not in the posttest1. Individual t-tests were used to compare the mean errors between the two groups for the Gen tests in the posttest1 and posttest2. The results indicated that group differences in mean errors were significant mainly in the tone pairs (bidirectional): Tone 1-Tone 4, and Tone 2-Tone 3, and sometimes Tone 1-Tone 2. The detailed patterns are shown in Figures 2 & 3.

In all, Cantonese listeners showed significantly more mean errors than did the Japanese listeners for the pairs, Tone 1-Tone 4, and Tone 2-Tone 3, in the pretest and posttest2.

4. Discussions

Concerning the hypothesis proposed earlier, as expected, Japanese learners performed better than did the Cantonese learners in all of the identification tasks in the posttests. However, the differences in performance between the two groups were not that noticeable in the identification scores and the tonal sensitivities (A' scores) for each lexical tone, but their differences became more obvious and evident in the analysis of their tonal confusions.

For the tonal confusion patterns, both Cantonese and Japanese had more confusion for the two tone pairs (Tone 1-Tone 4 and Tone 2-Tone 3). It is possible that these tones in each pair share a high degree of similarities. As for the Tone 1-Tone 4 pair, both tones begin with a similar high pitch level, and for the Tone 2-Tone 3 pair, both have a dip and a rising patterns. In fact, previous studies (Kiriloff 1969; Miracle 1989; Shen 1989) have indicated that non-native learners have great difficulties in producing and perceiving different lexical tones. In particular, the tone pairs, Tone 2 - Tone 3 and Tone 1 - Tone 4, are the most problematic ones. However, the remaining question is why Cantonese trainees showed significantly more confusions than Japanese trainees did.

PAM (Best 1995, Best et al. 2001) provides a theoretical account for their discrepancies in tonal confusions. With respect to tonal assimilation from Mandarin to Cantonese, both ***Single Category*** and ***Category Goodness Assimilations*** **(i.e., SC** and **CC)** were found. The former one (SC) was seen in the case that Mandarin Tone 1 (high level) and Tone 4 (high falling) assimilated to Cantonese Tone 1 (high level) which has two allotones (high level [55] and high falling [53]). The consequence was that Cantonese listeners frequently confused Mandarin Tone 4 with Tone 1. The latter one (CC) was found in the case that Mandarin Tone 2 (mid-rising) and Tone 3 (falling-rising) assimilated to Cantonese Tone 2 (high rising). In this case, Mandarin Tone 2 assimilated to Cantonese Tone 2 better than its Tone 3 counterpart, and therefore Cantonese listeners had a tendency to select Tone 2 as responses most of the time. There are two possible reasons for this tendency: First, Cantonese high rising tone (Tone 2) is phonetically similar to Mandarin Tone 2: (1) Both are described with the same tone letters [35] in the literature (Hashimoto 1972; Yip 2002 for Cantonese, and Howie 1976 for Mandarin), and (2) Cantonese Tone 2 is also produced with a falling and rising pattern (Bauer & Benedict 1997; So 1999). Second, Cantonese tonal system does not have a tone that is similar to Mandarin Tone 3 [214]. When Cantonese speakers listen to Tone 3, the best candidate will be Cantonese Tone 2, because Mandarin Tone 2 and Tone 3 do show considerable similarities in their pitch contours (e.g., the dip and the rising portions). Therefore, it can be inferred that Cantonese tonal system hinders the learning of Mandarin tones.

With respect to tonal assimilation from Mandarin to Japanese, ***Two Category Assimilation*** **(TC)*,*** ***Uncategorized-Uncategorized pair*** **(UU),** and ***Uncategorized-Categorized pair*** **(UC)** were observed. The TC one was seen in the case that Mandarin Tone 2 (mid-rising) and Tone 4 (high falling) assimilated to Japanese LH and HL pitch accent patterns, respectively. For Mandarin Tone 1 and Tone 3, they could be interpreted as an ***Uncategorized-Uncategorized pair***, because they did not assimilate to any "tone" or "pitch accent pattern" in the Japanese prosodic system. According to PAM, listeners' perception of uncategorized sounds (i.e., tones in this study) is less influenced by their L1 systems, but it depends on how much listeners perceive the similarities of the non-native contrasts. Therefore, that Japanese listeners perceived Tone 1 and Tone 3 quite well may be due to the fact that Tone 1 and Tone 3 are more perceptually salient. Tone 1 involves high pitch production and Tone 3 involves creaky voice quality near the center portion (e.g., Belotel-Grenié & Grenié 2004). As a result, the pairs, Tone 1-Tone 4 and Tone 2-Tone 3, actually formed two ***Uncategorized-Categorized*** (UC) pairs. Since PAM predicts that listeners should be able to discriminate the non-native sounds of a UC pair quite well, the result that Japanese listeners showed significantly fewer tonal errors in the two tone pairs in the posttest2 was in agreement with the prediction. Thus, Japanese pitch accent system actually facilitates the perception of Mandarin tones.

5. Conclusion

This study showed that laboratory training improved both trainee groups' perceptual ability to discern the Mandarin tones in general, but Japanese listeners showed a tendency to learn to perceive the Mandarin tones better than Cantonese listeners in the posttests. Analyses revealed that Cantonese and Japanese listeners' percent correct identification scores and A' scores are comparable to each other, but not in their tonal confusion. Although the errors of both groups were mainly found in the two tone pairs, Tone 1-Tone 4 and Tone 2-Tone 3, Cantonese listeners indeed significantly made more errors in the two pairs than did the Japanese listeners in the pretest and posttest2. The differences could be attributed to the fact that their L1 prosodic backgrounds play an important role in the process of learning a new tonal system. While Cantonese tonal system hinders their tone learning, Japanese pitch accent system facilitates the learning process. Thus, the findings suggest that trainees' tonal experience does not necessarily facilitate their learning of new tones. The phonemic status of their existing L1 tones can affect the learning process or the establishment of a new tonal system. In addition, this study provides supporting evidence for the assumption that PAM can be extended to "suprasegmental tiers" (Hallé et al. 2004).

References

Bauer, Robert and Paul Benedict. 1997. *Modern Cantonese Phonology.* Berlin and New York: Mouton de Gruyter.

Belotel-Grenié, Agnès and Michel Grenié. 2004. The Creaky Voice Phonation And The Organisation Of Chinese Discourse. Paper presented at *International Symposium on Tonal Aspects of Languages: With Emphasis on Tone Languages March 28-31, 2004*, Beijing, China.

Best, Catherine. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in cross-language research*, ed. Winifred Strange, 171-204. New York.

Best, Catherine, Gerald McRoberts, and Elizabeth Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *The Journal of the Acoustical Society of America* 109:775-794.

Chao, Yuan-Ren. 1930. A System of Tone Letters. *La Maître phonétique* 45:24-27.

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

Duanmu, San. 2005. Chinese (Mandarin), phonology of, in Encyclopedia of Language and Linguistics. 2nd edition. Elsevier Publishing House.

Flege, James. 1995. Second language speech learning: Theory, findings, and problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed. Winifred Strange, 233-277. Timonium, Maryland: York Press.

Fon, Janice and Wen-Yu Chiang. 1999. What Does Chao Have to Say about Tones. *Journal of Chinese Linguistics* 27:15–37.

Goto, Hiromu. 1971. Auditory perception by normal Japanese adults of the sounds "L" and "R". *Neuropsychologia* 9:317-323.

Gottfried, Terry and Devon Riester. 2000. Preliminary results from study of pitch perception and Mandarin tone identification. *Journal of the Acoustical Society of America.* 108:2604.

Hallé, Pierre, Yueh-Chin Chang, and Catherine Best. 2004. Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics* 32:395-421.

Hashimoto, Anne O.-K. 1972. *Phonology of Cantonese.* Cambridge: Cambridge University Press.

Howie, John. 1976. *Acoustical Studies of Mandarin Vowels and Tones*. Cambridge: Cambridge University Press.

Huang, Tsan. 2001. The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In *OSU Working Papers in Linguistics*, 23-42.

Ito, Kiwako, Shari Speer, and Mary Beckman. 2003, March. The influence of given-new status and lexical accent on intonation in Japanese spontaneous speech. Paper presented at *Presentation to the Annual CUNY Conference on Human Sentence Processing*, Boston, MA.

Jamieson, Donald and David Morosan. 1989. Training new, nonnative speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology* 43:88-96.

Kiriloff, C. 1969. On the auditory discrimination of tones in Mandarin. *Phonetica* 20:63-67.

Leather, Johnathan. 1990. Perceptual and productive learning of Chinese lexical tone by Dutch and English speakers. Paper presented at *New Sounds: proceedings the 1990 Amsterdam Symposium on the Acquisition of Second-language Speech*, University of Amsterdam.

Lively, Scott, David Pisoni, Reiko Yamada, Yoh'chi Tohkura, and Tsuneo Yamada. 1994. Training Japanese listeners to identify English /r/ and /l/ III: Long-term retention of new phonetic categories. *Journal of Acoustical Society of American* 89:874-886.

Miracle, Charles. 1989. Tone production of American students of Chinese: A preliminary acoustic study. *Journal of Chinese Language Teachers Association* 24:49-65.

McCawley, James D. 1978. What is a tone language? In *Tone: A linguistics survey*, ed. Victoria A. Fromkin, 113-131. New York: Academic Press.

Ortega-Llebaria, Marta, Andrew Faulkner, and Valerie Hazan. 2001. Auditory-visual L2 speech perception: Effects of visual cues and acoustic-phonetic context for Spanish learners of English. *Speech, Hearing and Language: work in progress* 13:39-51.

Shen, Xiaonan S. 1989. Toward a register approach in teaching Mandarin tones. *Journal of Chinese Language Teachers Association* 24:27-47.

So, Connie K. 1999. An acoustic analysis of Cantonese rising tones. *Journal of the Acoustical Society of America* 106:2153.

Wang, Yue, Michelle Spence, Allard Jongman and Joan Sereno. 1999. Training American listeners to perceive Mandarin tones. *The Journal of Acoustical Society of America* 106:3649-3658.

Wang, Yue, Allard Jongman, and Joan Sereno. 2003. Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of Acoustical Society of America* 113:1033-1043.

Wayland, Ratree and Susan Guion. 2004. Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning* 54:681-712.

Werker, Janet. 1991. The ontogeny of speech perception. In *Modularity andthe motor theory of speech perception*, eds. I.G.Mattingly and M. Studdert-Kennedy. Erlbaum, NJ: Hillsdale.

Woo, Nancy. 1969. *Prosody and Phonology*: Doctoral dissertation. MIT.

Yip, Moira. 1980. The tonal phonology of Chinese. Doctoral dissertation, MIT.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.