

# Effects of L1 tone on perception of L2 tone - a study of Mandarin tone learning by native Cantonese children\*

XINXIN LI

*The University of Hong Kong*

CAROL KIT SUM TO

*The University of Hong Kong*

MANWA LAWRENCE NG

*The University of Hong Kong*

(Received: May 15, 2015; final revision received: December 14, 2015; accepted: January 26, 2016; first published online 1 March 2016)

*In the present study, the Perceptual Assimilation Model (PAM) was tested on its applicability in child L2 lexical tone acquisition. The possible effect of L1 (Cantonese) lexical tones on L2 (Mandarin) lexical tone learning was explored. Accuracy rate and error patterns were examined with an AX discrimination task and a forced-choice identification task. Forty-nine native Cantonese-speaking students aged 8 years participated in the study. Results revealed that these children exhibited nearly perfect performance in the discrimination of Mandarin tones. However, significant tone differences were detected in the identification task. Tone 4 (T4) was identified with the lowest accuracy, and T1 with the highest. Error analysis revealed that Mandarin T2-T3 was the most confusing pair, followed by the T1-T4 pair. The inherent phonetic similarity between lexical tones in a language and the tone similarities across languages may also have contributed to perception difficulties, which could help to refine and supplement the PAM in the tonal/suprasegmental domain.*

Keywords: Cantonese, Mandarin, second language tone learning, PAM

## Introduction

A foreign accent appears to be prevalent in speakers of a second language (L2) especially in adult foreign language speakers (Flege, 1995). Research has found that the quality and quantity of L2 input, motivation of L2 learning, and/or other psychological reasons such as cognitive and imitation skills can influence outcomes of L2 learning (Flege, 1988, 1992). In addition, superior musical ability has also been related to better L2 speech learning outcomes (Gottfried, 2007; Gottfried & Riester, 2000; Gottfried, Staby & Ziemer, 2004; Lee & Hung, 2008).

Despite the different factors determining individual L2 accents, L2 speakers of the same L1 background share similar accent patterns. This points to the fact that accent is a result of the linguistic influence of the L1 (Best, 1995; Flege, 1995). A large number of studies have examined the perception and production of an L2 in relation to the influence of speakers' L1, and the majority of these studies have focused on segmental learning (Best, 1995; Best & Tyler, 2007). For example, Werker, Frost and McGurk (1992) found that French Canadians tended to substitute English /θ/-/ð/ with French /d/-/t/ when producing English dental fricatives. Goto (1971) reported that Japanese speakers encountered difficulties

in discriminating between English /r/ and /l/ due to the absence of this distinction in Japanese. Best, McRoberts and Sithole (1988) investigated the discrimination of voicing contrast of lateral fricatives in Zulu, a southern African language, by native English speakers. They found that most native English speakers assimilated the non-native sound contrast /ʎ/ and /ʒ/ to English coronal fricatives /z/ and /s/. In order to explain and predict how non-native sounds or sound contrasts are perceived with reference to the degree of similarity between a non-native L2 sound and the L1 sound most similar to it, Best (1995) put forward the Perceptual Assimilation Model (PAM).

## Theoretical framework of PAM

The PAM describes the way in which non-native L2 sounds are perceived by naïve speakers based on how these non-native novel sounds are assimilated to the sounds in the speakers' native language (Best, 1995). It aims not only to provide an explanation for the realizations in the speaker's L2, but also to predict the perception of L2 sounds. In PAM, various categories or assimilation types are proposed to describe the contrasts between L1 and L2 sounds. The model has subsequently been extended to be applicable from naïve speakers without L2 learning experience to L2 learners by Best and Tyler (2007). The non-native to native speech sound assimilations in PAM are categorized into six contrast types: (1) Two-Category Assimilation (TC), (2) Single-Category

\* We would like to thank the principal of the school, all children and their parents for their participation in this study.

Address of correspondence:

Xinxin Li, Division of Speech and Hearing Sciences, Faculty of Education, The University of Hong Kong, Pokfulam, Hong Kong, SAR China  
allylx@hku.hk

Table 1. *The Perceptual Assimilation Model and its application*

Assimilation Types	Explanation	Prediction
Two Category Assimilation (TC)	non-native $T\alpha$ $T\beta$ ↓ ↓	$P(T\alpha \approx T\beta)$ ↓
	native $T_i$ $T_{ii}$	easy
Single-Category Assimilation (SC)	non-native $T\alpha$ $T\beta$ ↘ ↙	$P(T\alpha \approx T\beta)$ ↑
	native $T_i$	difficult
Category-Goodness Difference (CG)	non-native $T\alpha$ $T\beta$ ↘ ↙	$P(T\alpha \approx T\beta)$ ↑ medium to difficult
	native $T_i$	
Uncategorized-Categorized Pair Assimilation (UC)	non-native $T\alpha$ $T\beta$ ↓ ↓	$P(T\alpha \approx T\beta)$ ↓
	native $T_i, T_{ii}$ $T_{iii}$	easy to medium
Uncategorized-Uncategorized Assimilation (UU)	non-native $T\alpha$ $T\beta$ ↓ ↓	$P(T\alpha \approx T\beta)$ ?
	native $T_i, T_{ii}$ $T_{iii}, T_{iv}$	hard to predict
Non-Assimilable (NA)	non-native $T\alpha$ $T\beta$ ↓ ↓	$P(T\alpha \approx T\beta)$ ↓
	native × ×	easy

Notes: 1. Explanation: ↓ ↘ ↙ similar; ↘ less similar; × no similar  
2. Prediction: ≈ confusion; ↓ easy; ↑ difficult; ↑ medium to difficult; ↓ easy to medium

Assimilation (SC), (3) Category-Goodness Difference (CG), (4) Uncategorized-Categorized Pair Assimilation (UC), (5) Uncategorized-Uncategorized Assimilation (UU), and (6) Non-Assimilable (NA) (Best, 1995). All these categories are based on the perceptual similarities between two contrasting sounds in an L2 and how these two L2 sounds are assimilated to sounds in the L1 (see Table 1).

TC means two non-native phones are phonetically similar to two different native phones and can assimilate separately to each other. In this case, it will be relatively easy to discriminate the non-native contrast. Evidence of the TC category comes from perception studies by Best (1990), who investigated the perception of sounds in Ethiopian (an African language) by English speakers. The ejective bilabial stop /pʰ/ and the alveolar stop /tʰ/ were often assimilated to English /p/ and /t/ respectively. English speakers, therefore, achieved an excellent performance in discriminating the L2 sound contrast /pʰ/-/tʰ/ (Best, 1990). Related findings were reported by Best, McRoberts and Goodell (2001), which showed that English speakers could differentiate between Zulu lateral fricatives /h/ and /ɬ/ well due to TC assimilation. While SC means two non-native phones assimilate equally well to a single native phone, which will be hard to discriminate. In the case of the SC type, Best and Strange (1992) posited that the difficulty in discriminating the English sound contrast /r/-/l/ observed in Japanese speakers could be explained by the assimilation of these

two English sounds to the Japanese sound /r/. Japanese speakers, as a result, found it particularly difficult to discriminate the English /r/-/l/. CG means both two non-native phones might assimilate to a single native phone, but the assimilation level is different. One non-native phone may be more similar to the native phone while the other is less similar. An example of the CG pattern is the difficulty encountered by French listeners in discriminating between English /r/-/w/. The two English phonemes /r/ and /w/ tend to be assimilated to French /w/. Although there is an /r/ phoneme in French, perceptually, English /r/ is quite different from the French /r/; instead, it is more similar to the French /w/. But the assimilation levels are different since English /w/ is more similar to French /w/ than English /r/. Therefore, the discrimination accuracy is relatively low (Hallé, Best & Levitt, 1999). It has been suggested that the gradient of perception levels should be TC < CG < SC (Best, 1995; Best, Faber & Levitt, 1996). UC means one non-native phone assimilates to a native phone while the other non-native phone may be uncategorized that falls in between two native phones. The difficulty level to discriminate the non-native phones in UC might be in the middle. UU means both non-native phones might be uncategorized speech segments and it is hard to judge the difficulty level for discrimination. The similarity between L2 sounds and L1-L2 assimilation should be further analyzed. The UC pattern was examined among Japanese speakers by Aoyama (2003) through English final contrasts /m/-/n/ and /m/-/ŋ/. The /n/-/ŋ/

assimilation was categorized as the UU type since /m/ can be assimilated to Japanese /mu/, while /n/ and /ŋ/ fall in between two native phones /N/ and /Ngu/. Relatively good performance was achieved on the UC type, and poor performance was achieved on the UU type, which generally agreed with the prediction of PAM. NA means if the articulatory properties of both non-native phones are dissimilar to any native phone, it can be perceived as non-speech sounds. Hence, good discrimination is supposed to be achieved since there is no influence from native phones when learning these non-native phones. Evidence for the NA pattern has been found in the good perception of Zulu click contrasts, perceived as nonspeech sounds, by native English speakers (Best et al., 1988).

In general, PAM has provided a satisfactory explanation of the segmental learning of various L2s with reference to the sound features of L1s. On the other hand, investigation into the perception of suprasegmental features such as lexical tones by L2 learners is scarce (So & Best, 2010). Lexical tone is a challenging aspect of learning a tone language as an L2. L2 speakers have often demonstrated a disproportionately poorer performance with tones than with segments, or showed persistent problems with lexical tones even after extensive training (Lin, 1985). In theory, the factors pertaining to L2 segmental learning discussed earlier could also be applicable to the learning of L2 lexical tones (Alexander, Wong & Bradlow, 2005). However, the exact interaction among L1 and L2 tones might be different from that among L1 and L2 segments. It is not clear whether the existence of a lexical tone system in the native language would facilitate or hinder the learning of the tone system in the L2. The present study aimed to test the applicability of PAM in L2 lexical tone learning. More specifically, it examined the pattern of L2 lexical tone perception with reference to the native lexical tones present in the L1. The target L2 language under examination was Mandarin, and the L2 learners were native Cantonese-speaking children.

### Lexical tones in Mandarin and Cantonese

Lexical tones are physiologically associated with the rate of vocal fold vibration and are acoustically manifested by fundamental frequency (F0). Different lexical tones can be represented by different F0 heights and/or contours (Howie, 1976; Li & Thompson, 1977). There are four lexical contrastive tones in Mandarin and one neutral tone (Chao, 1948, 1956, 1968). Chao (1956) made use of a five-degree framework to represent the four different Mandarin tones. In this paper, different tones are represented by the capital letter “T” followed by a number, which marks a specific tone. For example, the syllable /ma/ carries different semantic meanings when produced with different lexical tones: /mā/ (T1, high-level: 55) means “mother”;

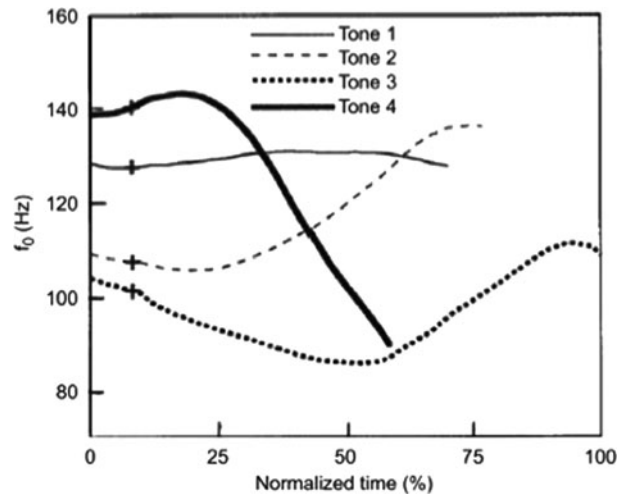


Figure 1. Mean fundamental frequency (F0) contours of four Mandarin tones. Source from Xu (1997).

/má/ (T2, mid-rising: 35) means “hemp”; /mǎ/ (T3, mid-falling-rising: 214) means “horse”; and /mà/ (T4, high falling: 51) means “to scold”.

Perceptual findings from native Mandarin speakers indicate that both F0 height and contour are the most important acoustic cues for identification (Blicher, Diehl & Cohen, 1990). In addition, duration and amplitude also play a role in distinguishing the four tones during perception and production (Moore & Jongman, 1997). In Mandarin, T2 is shorter than T3 but slightly longer than T1, and T4 is the shortest tone among the four. The mean F0 of the four Mandarin lexical tones are represented in Figure 1, which is generalized from 48 tokens with four tones and time is normalized (Xu, 1997).

Generally, Mandarin T1 is a high-level tone with a stable vocal fold vibration with similar starting and ending fundamental frequency (F0). The laryngeal height remains in a steady and highest position. T2 is a rising tone with the starting onset from the mid-level of T1 F0 height and offset ended towards the T1 F0 height. Yet, Ho (1976) pointed out that a slight dip existed in the first 15% area of the whole syllable duration. The overall laryngeal movement is still to go up. T3 is a low falling-rising tone with a tuning point in the lowest F0 height. The low falling contour ends at around 41% of the whole syllable duration then it starts the rising part (Ho, 1976). Besides, the overall contour for T3 shares some similarity with T2 since acoustic analysis revealed that a slight dipping also existed in T2 (Fon & Chiang, 1999; Ho, 1976; So & Best, 2010). T4 is a high falling tone that begins with high F0 similar to T1 F0 height but it directly falls to the lowest F0 height. It is the shortest tone in duration among the four tones with highest starting F0, which means the larynx moves up from its original rest position and then moves down instantly.

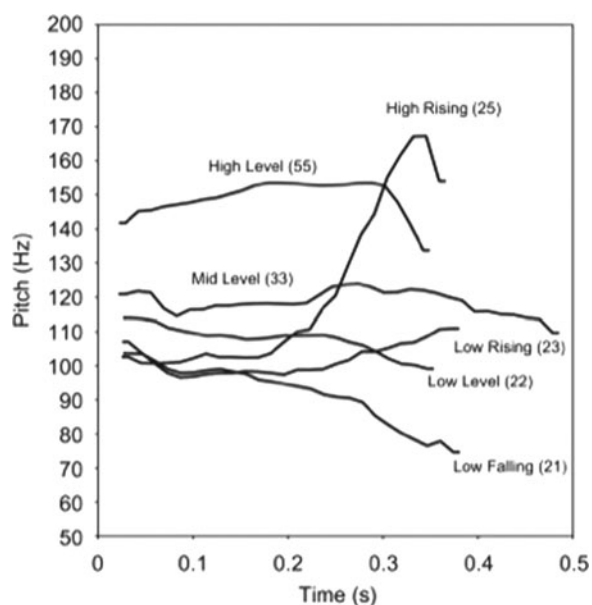


Figure 2. Fundamental frequency (F0) contours for six Cantonese tones. Source from Francis, Ciocca, Ma & Fenn (2008).

In Hong Kong Cantonese<sup>1</sup>, there are six contrastive lexical tones (T1 to T6) categorized by pitch height and contour and three short tones (T7 to T9) ending with unreleased stops /-p, -t, -k/. T7, T8, and T9 have the same pitch height as that of T1, T3, and T6 respectively but are shorter in duration (Fok, 1974). The pitch height and contour of the six long Cantonese lexical tones can also be represented by the five-degree framework proposed by Chao (1956) (Zee, 1999). Taking the syllable /ji/ as an example, the six tones are /ji1/ (T1, high-level: 55/53), meaning ‘cure’; /ji2/ (T2, low-mid to high-rising: 25), meaning ‘chair’; /ji3/ (T3, mid-level: 33), meaning ‘opinion’; /ji4/ (T4, low falling: 21), meaning ‘child’; /ji5/ (T5, low rising: 23), meaning ‘ear’; and /ji6/ (T6, low-level: 22), meaning ‘two’ (see Figure 2).

However, there are some concerns about Cantonese T1. Some research has found that both 55 and 53 existed in Cantonese. Kwan (1990) distinguished Cantonese T1 as a high-level tone (55) and a high falling tone (53) separately. Some researchers considered high falling tone (53) as a variant for T1 (Bauer, Cheung & Cheung, 2003; Zee, 1991). But most current Cantonese native speakers seldom tell these differences apart and merge high falling tone (53) with high-level tone (55). Nowadays in Hong Kong, most native speakers use high-level tone (55) as T1 while high falling tone (53) is more frequently adopted as T1 in Guangdong province. Some detectable difference

is that more T1 verbs are pronounced with high falling tone while more nouns are pronounced with high level tone (Bauer & Benedict, 1997). Matthews and Yip (1994) also found that certain high-level tone adverbs in the final position of the utterance were produced with high falling tone by some Hong Kong Cantonese speakers. The current study will focus on the Cantonese spoken in Hong Kong where the use of high falling tone (53) is declining.

### *Native and non-native lexical tone perception*

With respect to production, lexical tones are reported to be acquired earlier than segments by children learning Mandarin as their first language (Clumeck, 1980; Zhu & Dodd, 2000). Zhu and Dodd (2000) examined the phonological development of monolingual Mandarin-speaking children aged from 1;6 to 4;6 through a picture naming task and a story telling task based on raters’ perceptual judgement. They found that tone errors were much less frequent compared to consonants and vowels in all age groups, and the acquisition of tones was completed by almost all the children by the age of two years old. Children as young as two years old showed a good mastery of lexical tones, except for two tone errors, T4 for T2 and T2 for T3. In addition to production, Wong, Schwartz and Jenkins (2005) also investigated the perception of Mandarin tones by Mandarin-speaking children aged from 2;10 to 3;4 in the United States using a monosyllabic word identification task and a naming task. Results revealed a 70%–78% accuracy for the production of T1, T2 and T4 and the most difficult tone was the dipping tone T3, with an accuracy of 44%. In contrast, in the tone perception task, a high accuracy rate of 88% to 95% was observed for all four tones. In terms of tone acquisition, Li and Thompson (1977) reported that T1 and T4 were acquired earlier compared to T2 and T3. Children often confused T2 with T3 in two-to-three-word phrases. Despite the lack of consensus regarding the age and order of native tone acquisition, it is generally accepted that L1 lexical tone acquisition is not challenging for children.

In contrast, the mastery of L2 lexical tones appears to be more demanding than segmental learning. Learners of a non-tone L1 such as English often find it very difficult to learn the L2 lexical tones. This may be due to the absence of tone contrasts in the native sound system. For tonal L2 learners with a tone language L1 background, research has revealed mixed findings (Francis, Ciocca, Ma & Fenn, 2008; Hao, 2012; Lee, Vakoch & Wurm, 1996; So, 2006; Wayland & Guion, 2004; Wayland & Li, 2005). Some studies suggested that L2 learners with a tone language as their L1 can perceive lexical tone contrasts better than L2 learners with a non-tone language (Lee et al., 1996; So, 2006; Wayland & Li, 2005), yet, others reported non-significant differences in non-native tone perception between tone and non-tone language speakers (Francis

<sup>1</sup> There are varieties of Cantonese (Yue dialect), including Hong Kong Cantonese spoken in Hong Kong, Guangzhou Cantonese spoken in Guangdong and Guangxi Province and Taishanese spoken by people in the southern part of Guangdong province, and so forth.

et al., 2008; Hao, 2012). The lack of consistency in the results of previous studies may be due to discrepancies in research design, including small sample size, discrepant amount of tonal L2 exposure (Lee et al., 1996; So, 2006; Wayland & Guion, 2004), and no control over extraneous factors like influence of musical background (Bidelman, Hutka & Moreno, 2013). Besides, all the studies focused only on adults, but not children (Yang, 2010).

**Mandarin lexical tone perception based on PAM**

Two studies have employed PAM to explain the acquisition of suprasegmental features such as lexical tones (Hao, 2012; So, 2006). So (2006) found that the good discrimination of the Mandarin T1-T2 pair by Cantonese speakers corresponded to the TC pattern in PAM since Mandarin T1 and T2 can be closely assimilated to Cantonese T1 and T2. The higher error rate for the Mandarin T1-T4 pair was due to SC assimilation in that Mandarin T1 and T4 are similar to Cantonese T1. With respect to CG, So (2006) found that the T2-T3 pair was the second most confusing pair for Cantonese speakers, who tended to choose Mandarin T2 more frequently when identifying T2 and T3 since Mandarin T2 is more similar to Cantonese T2 than it is to Mandarin T3. The holistic performance was consistent with TC > CG > SC, that is, T1-T2 > T2-T3 > T1-T4, where the T1-T2 pair was the easiest. These findings provided evidence of the feasibility of applying PAM in L2 suprasegmental learning. However, the assimilation of Mandarin tones to Cantonese tones in the study by So (2006) was based on the five-degree framework which just reflects the perceived starting and ending points of F0 contours of the lexical tones. It will be more convincing if the categorization of Mandarin tone pairs into PAM is based on acoustic or perceptual analysis of the similarities and dissimilarities between Mandarin and Cantonese tones.

Be that as it may, Hao (2012) assessed the perceptual assimilation of Mandarin tones to Cantonese tones by deliberately requested native Cantonese speakers to match a Mandarin tone with a Cantonese one. Results revealed that Mandarin T1 most significantly resembled or was assimilated to Cantonese T1, followed by Cantonese T3. Although Mandarin T2 is similar to Cantonese T2 in terms of F0 height and contour, the perceptual data indicated that it corresponded more to Cantonese T5 (low rising: 23) and followed by T2 (low-mid to high-rising: 25). The internal similarity between Cantonese T2 and T5 might account for the perceptual assimilation finding since Cantonese T2-T5 pair was hard to discriminate even for native Cantonese speakers in Hong Kong (Ciocca & Lui, 2003). In addition, research has revealed that Cantonese T2 and T5 were merged in some Cantonese speakers (Bauer et al., 2003; Mok, Zuo & Wong, 2013). Hence, it is likely that Cantonese children exposed to this merger

Table 2. The assimilation of Mandarin tones to Cantonese tones (Hao, 2012)

Cantonese (Native)	Mandarin (Non-native tones presented)			
	T1 (55)	T2 (35)	T3 (214)	T4 (51)
T1 (55/53)	*81%			*66%
T2 (25)		22%		
T3 (33)	11%	11%		19%
T4 (21)		7%	*78%	
T5 (23)		*52%		
T6 (22)	6%		8%	10%

Notes: 1.\*Stands for the most similar one.  
 2. The percentage means a Cantonese tone was chosen to match the Mandarin tone stimuli. Only more than 5% was listed to make the table easier to read.

will inherit the feature. As for other tones, Hao (2012) found that Mandarin T3 (low falling-rising tone) was more frequently considered similar to Cantonese T4 (low falling: 21), and Mandarin T4 (high falling: 51) was more frequently considered similar to Cantonese T1 (high-level: 55/53)(see Table 2). The perceptual assimilation between Cantonese and Mandarin tones could represent the similarities between the native Cantonese tones and the L2 Mandarin tones. The reported patterns in Hao (2012) were used as a reference to formulate hypotheses when testing PAM in the current study.

Accordingly, Hao (2012) categorized the Mandarin T1-T2 pair into the UC type since Mandarin T2 was found to be most similar to other three Cantonese tones. Besides, the Mandarin T1-T4 pair was considered as the SC/CG pattern since Mandarin T1 and T4 were perceived with the closest similarity to the same Cantonese tone (T1) in the perceptual assimilation test, which well explained its poor perception performance. In addition, Hao (2012) proposed that the Mandarin T2-T3 pair should be categorized into the UC type rather than the CG type since Mandarin T2 was similar to three Cantonese tones, while Mandarin T3 was most similar to one Cantonese tone (i.e., T4) (see Table 1 & Figure 3). In this case, if in agreement with the UC type, the discrimination of Mandarin T2-T3 should not be too hard. However, her perception results showed the worst performance for the Mandarin T2-T3 pair. Hence, Hao (2012) concluded that PAM was not able to predict all confusing tone pairs.

Based on the above analysis, it was hypothesized in the present study that Mandarin T1-T3 and Mandarin T3-T4 could be categorized as the TC type, Mandarin T1-T4 could be classified as the SC type, and the remaining three Mandarin tone pairs, T1-T2, T2-T3, T2-T4, would be the UC type (see Figure 3). In this way, it was predicted that there would be a higher accuracy rate for T1-T3 and T3-T4 than for T1-T4. While the accuracy rate for T1-T2, T2-T3 and T2-T4 should be in the middle.

Assimilation Types	So (2006)	Hao (2012)	Hypothesis
Two Category Assimilation (TC)	M (T1-T2) ↓ ↓ C (T1-T2)		M (T1-T3) & (T3-T4) ↓ ↓ ↓ ↓ C (T1-T4) & (T4-T1)
Single-Category Assimilation (SC)	M (T1-T4) ↘ ↙ C (T1)	M (T1-T4) ↘ ↙ C (T1)	M (T1-T4) ↘ ↙ C (T1)
Category-Goodness Difference (CG)	M (T2-T3) ↘ ↙ C (T2)	M (T1-T4) ↘ ↙ C (T1)	
Uncategorized-Categorized Pair Assimilation (UC)		M (T2-T3) ↓ ↓ C (T2/T3/T5-T4)	M (T1-T2) & (T2-T3) & (T2-T4) ↓ ↓ ↓ ↓ ↓ ↓ C (T1-T2/T5) & (T2/T5-T4) & (T2/T5-T1)
Uncategorized-Uncategorized Assimilation (UU)			
Non-Assimilable (NA)	Mandarin vs. English	Mandarin vs. English	Mandarin (tone) vs. English (non-tone)

Notes: 1. M stands for Mandarin; C stands for Cantonese  
 2. ↓ ↘ ↙ similar; ↘ less similar

Figure 3. The application of the PAM in Mandarin tone studies.

The current study investigated the applicability of the PAM in predicting the patterns of non-native lexical tone perception by children with a tone language as their L1. Two tasks were included, a discrimination test and an identification test. Children’s error patterns in both tasks may reflect the specific influence of the phonemes in their L1.

**Method**

**Participants**

Forty-nine Cantonese-speaking children (27 boys and 22 girls) were recruited from mainstream Chinese primary schools in Hong Kong. They were all Primary Year Two (P2) students with ages from 7;0 to 8;4 years (mean = 7.67 years, SD = 0.29 years). Children at P2 were selected as they would have finished *pinyin* learning including lexical tones.

All participants had no speech and hearing problem according to parents’ and teachers’ reports. To control for influences from other languages and potential confounding factors such as musical background, parents were required to complete a questionnaire that requested information about their home language, the language(s) family members spoke to the child, their language preference when communicating with their child, the language environment of their neighborhood, and the music background of their child.

**Stimuli**

The five different syllables, /bu, di, lu, na, ka/ produced with the four Mandarin tones were used as the stimuli

in both tasks. The stimuli were produced by a native female Mandarin speaker who had attained Level One certificate of the National Putonghua Proficiency Test. All the recordings were carried out in a soundproof booth via a professional grade microphone (SM58A, Shure, USA) connected to an external sound card (PreMobile USB, M-Audio, USA). All stimuli were digitized at 44.1 kHz sampling rate and 16 bits/sample quantization rate, and normalized for RMS amplitude at 70 dB using a speech editing software called Audacity. The speaker read the syllables aloud three times, out of which readings the best pronunciation was selected. Two other native Mandarin speakers were asked to listen to all stimuli to ensure their accuracy. The average duration for each monosyllable was around 0.44 seconds (range = 0.40–0.45 seconds) to control for the effect of duration (Blicher et al., 1990). Acoustic analysis of the four Mandarin lexical tones in five different monosyllables was conducted to explore the pitch contours of the stimuli and illustrated in Figure 4 to Figure 8.

**Procedures**

**Discrimination test**

In the discrimination test, a forced-choice AX discrimination task was used. One tone pair consisting of two identical monosyllables carrying the same or different tones was presented in each trial. The inter-stimulus pause was around 0.44 seconds. Participants were asked to press one of two keys to indicate whether the two lexical tones presented were identical or different. With four lexical tones, a total of 16 tone pairs was formed. All stimuli

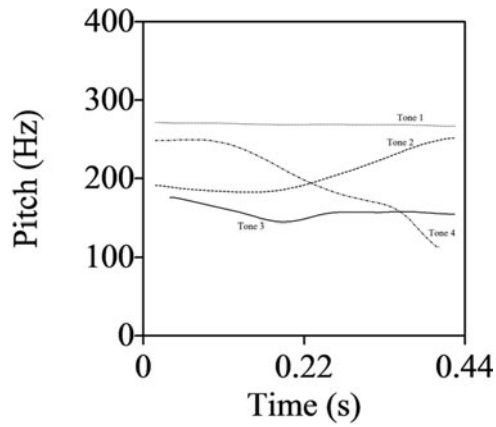


Figure 4. F0 contour of four Mandarin tones in syllable /bu/.

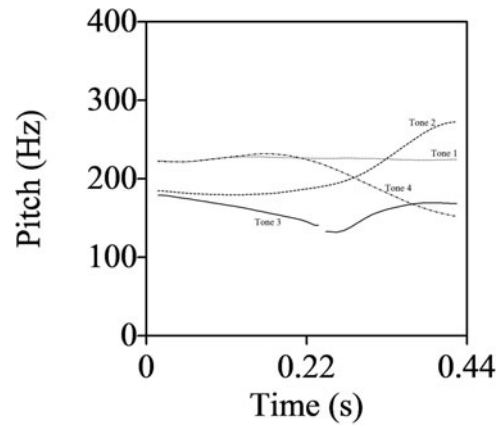


Figure 7. F0 contour of four Mandarin tones in syllable /lu/.

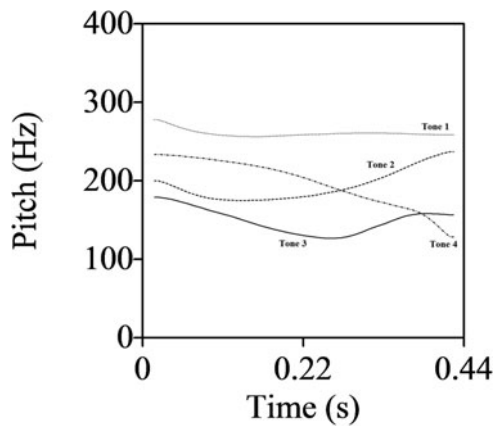


Figure 5. F0 contour of four Mandarin tones in syllable /di/.

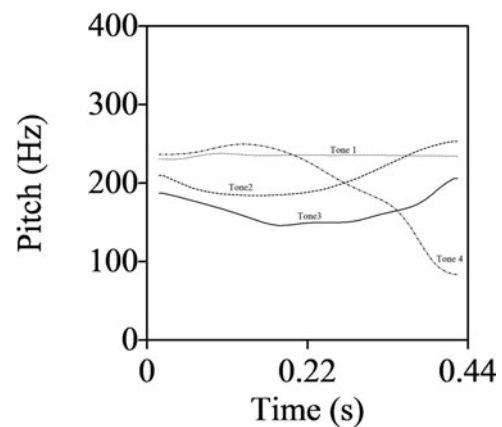


Figure 8. F0 contour of four Mandarin tones in syllable /na/.

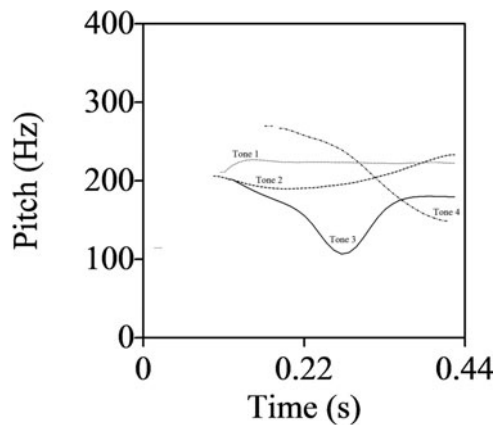


Figure 6. F0 contour of four Mandarin tones in syllable /ka/.

were presented twice in the experiment. This gave a total of 160 presentations (5 syllables  $\times$  16 tone pairs  $\times$  2 times) in the discrimination task, including 40 identical tone pairs and 120 different tone pairs.

All stimuli were presented in a random order by using the E-Prime software (Schneider, Eschman & Zuccolotto, 2002). Two practice trials were provided for students to familiarize themselves with each test and the correct answers were provided upon completion of the practice trial. All experiments were presented through high-quality headphones (HD280 Pro, Sennheiser, USA). The entire listening experiment was carried out in a quiet room in the schools of participating students. Neutral feedback after each test trial was given. The responses were recorded automatically by E-prime software.

**Identification test**

In the identification task, all stimuli were repeated once, yielding a total of 40 stimuli in total (5 syllables  $\times$  4 tones  $\times$  2 times). During the experiment, participants were asked to select the correct tone out of four tones (T1, T2, T3, T4) presented on the screen visually after listening to a target tone. The E-Prime software was used for this experiment (Schneider et al., 2002). Similarly, there were also two practice trials to familiarize participants

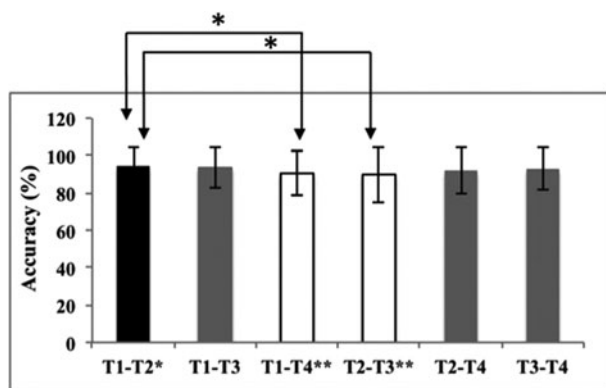


Figure 9. Accuracy rates of tone pairs with ‘different’ stimuli.

with the procedures. A small gift was given to each participant upon completion of all the assessments as an appreciation.

**Results**

**Mandarin lexical tone discrimination test**

**Accuracy**

The average overall accuracy rate in the discrimination test was 91.9% (*SD* = 9.9%). The average accuracy rate was 91.2% (*SD* = 11.4%) for ‘identical’ stimuli, and 92.2% (*SD* = 10.9%) for ‘different’ stimuli.

**Error patterns**

Performance was similar across the tone pairs. For ‘identical’ stimuli, participants committed slightly more errors in discriminating the T4-T4 pair, with an error rate of 10%, followed by T1-T1 (9.0%), T2-T2 (8.8%) and T3-T3 (7.4%). Friedman’s ANOVA showed no significant difference between the four ‘identical’ tone pairs ( $\chi^2(3) = 2.056, p > .05$ ), implying that the four ‘identical’ tone pairs were at a similar level of difficulty. For the ‘different’ stimuli pairs (see Figure 9), participants made slightly more errors in discriminating the T2-T3 (error rate of 10.3%) and T1-T4 (error rate of 9.7%) pairs. A relatively higher accuracy rate was found for the T1-T2 pair and the T1-T3 pair, with an error rate of only 5.6% and 6.5% respectively. Friedman’s ANOVA test revealed that there were significant differences between six ‘different’ tone pairs ( $\chi^2(5) = 18.732, p < .05$ ). Further *post hoc* comparisons using Wilcoxon signed-rank tests with Bonferroni correction ( $p = .003$ ) revealed that the T1-T2 pair had a significantly higher accuracy rate than that of the T1-T4 pair ( $t = 66.50, r = -.334$ ) and the T2-T3 pair ( $t = 45, r = -.338$ ).

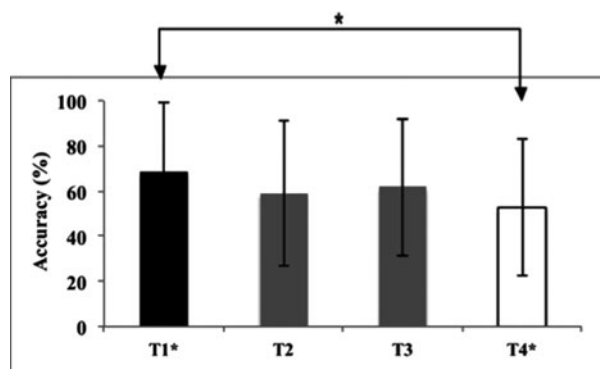


Figure 10. Overall accuracy rates in identification of the four tones.

Table 3. Confusion matrix of tone identification

Target	Response			
	T1	T2	T3	T4
T1	<b>68.6%</b>	11.6%	7.1%	12.7%
T2	10%	<b>59.2%</b>	*23.9%	6.9%
T3	8.6%	*24.3%	<b>61.8%</b>	5.3%
T4	18.4%	15.1%	13.9%	<b>52.6%</b>

**Mandarin lexical tone identification test**

**Accuracy**

The confusion matrix of tone identification is shown in Figure 10. The mean overall accuracy rate in the identification test was 60.6% (*SD* = 21.8%). The accuracy rate for Mandarin T1, T2, T3, and T4 was 68.6% (*SD* = 30.8%), 59.2% (*SD* = 32.3%), 61.8% (*SD* = 30.4%), and 52.6% (*SD* = 30.1%) respectively. Repeated-measures ANOVAs showed significant differences between the four lexical tones [ $F(2.45, 117.61) = 3.3, p < .05$ ]. *Post hoc* analyses revealed that T1 had a significantly higher accuracy rate than that of T4 ( $p = .024$ ).

**Error patterns**

Participants made more errors in identifying Mandarin T4 and T2, with error rates of 47.4% and 40.8% respectively. As shown in Table 3 and Figure 10, for Mandarin T1 errors, 11.6% of the responses consisted of mistakenly identifying T1 as Mandarin T2, 7.1% as Mandarin T3, and 12.7% as Mandarin T4.

The Cantonese-speaking children were most likely to mix up Mandarin T2 and T3 (see Figure 11). Most of their errors in identifying Mandarin T2 consisted of misidentifying the tone as Mandarin T3, with an error rate of 23.9%. Friedman’s ANOVA test on T2 error distribution showed significant differences between three other tones ( $\chi^2(2) = 25.406, p < .05$ ). *Post hoc* Wilcoxon signed-rank tests with corrected significance level  $p = .0167$



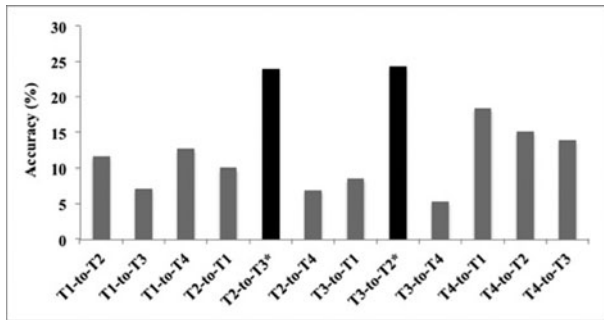


Figure 11. Mean error rates for each tone in identification of the four tones.

showed that the children misidentified Mandarin T2 as Mandarin T3 significantly more than they misidentified it as Mandarin T1 ( $T = 133.50$ ,  $r = -.302$ ), and more than they misidentified it as Mandarin T4 ( $T = 60.50$ ,  $r = -.448$ ).

Similarly, most Mandarin T3 errors consisted of misidentifying this tone as T2. Friedman's ANOVA test also revealed a significant difference in Mandarin T3 error distribution ( $\chi^2(2) = 26.275$ ,  $p < .05$ ). Wilcoxon signed-rank tests showed significantly more errors for Mandarin T2 than for Mandarin T1 ( $T = 116.50$ ,  $r = -.345$ ), and for Mandarin T4 ( $T = 85.00$ ,  $r = -.433$ ).

The errors in the identification of Mandarin T4 were evenly distributed across the other three tones, with 18.4% identifying it as Mandarin T1, 15.1% as Mandarin T4, and 13.9% as Mandarin T3 (see Table 3). Friedman's ANOVA showed no significant difference between these three tones ( $\chi^2(2) = 1.000$ ,  $p > .05$ ).

## Discussion

The applicability of the PAM in Mandarin lexical tone acquisition was explored in a discrimination task and an identification task in terms of accuracy rates. Error patterns were examined with reference to confusing tone pairs. The possible effect of L1 lexical tones on L2 lexical tone learning in children is discussed.

### Discrimination

The nearly perfect performance in the discrimination task implied that the Cantonese-speaking children did not encounter much difficulty in discriminating non-native Mandarin lexical tone contrasts. The existence of lexical tone contrasts in their native language may have accounted for this. When the analyses were broken down with reference to the 'identical' and 'different' pairs, no significant differences were found in the discrimination of the 'identical' tone pairs (T1-T1, T2-T2, T3-T3, T4-T4), but statistically significant differences were detected in the discrimination of the 'different' tone pairs (T1-T2,

T1-T3, T1-T4, T2-T3, T2-T4, T3-T4). A higher accuracy rate was found for T1-T2 and a lower accuracy rate for T2-T3 and T1-T4. Only some of these discrimination patterns were predicted by PAM but not all.

For the Mandarin T1-T2 pair, predicted as a UC type, results indicated that the Cantonese children performed well, which was consistent with the hypothesis. Similar findings were reported in So (2010) showing that Cantonese adult speakers were more successful in telling the difference between Mandarin T1 and T2 than were English and Japanese speakers. This may imply that their assimilation to L1 Cantonese tones facilitated the discrimination of Mandarin T1-T2. Two other tone pairs, Mandarin T1-T3 and T3-T4, were classified into the TC category and manifested a similar accuracy to that of Mandarin T1-T2, which was generally consistent with the prediction.

As was stated in the hypotheses, Mandarin T1-T4 could be categorized into the SC Type, and the Cantonese-speaking children performed poorly on this tone pair since both Mandarin T1 and T4 would be assimilated to Cantonese T1. The result of the present study was in agreement with the prediction, in that the degree of accuracy in discriminating the Mandarin T1-T4 pair was lower than the degree of accuracy in discriminating other tone pairs, and especially significantly lower than that in discriminating the T1-T2 pair. The Mandarin tone identification tests by So (2006) and Hao (2012) also revealed similar difficulties for Cantonese adult speakers in distinguishing between Mandarin T1 and T4. Hence, it turned out that for native Cantonese speakers regardless of children or adults, Mandarin T1 (55) and T4 (51) were very confusing. With regard to the similarities between Mandarin T1-T4 and Cantonese T1, it can be explained by the allotonic features of Cantonese T1 (55/53) (Bauer et al., 2003). That is the high-level tone (55) and the high falling tone (53) are two variants of Cantonese T1. Besides, the F0 contour of Cantonese T1 consists of a stable and high-level F0 and a little dropping towards the end of the contour (Francis et al., 2008). The high level contour of Cantonese T1 that is similar to Mandarin T1 and the falling part that is similar to Mandarin T4 may make Cantonese speakers hard to discriminate between Mandarin T1 and T4. Yet, it will be more convincing if there is cross-language acoustic analysis on the similarities between Cantonese T1 and its two similar Mandarin tones.

On the other hand, the other confusing tone pair, Mandarin T2-T3, could not be predicted by the UC pattern in PAM. It was originally hypothesized that participants would demonstrate a relatively good to fair performance for T2-T3, but the actual discrimination result indicated that Mandarin T2-T3 was comparatively harder than other tone pairs, and especially more difficult than T1-T2. Factors other than the similarity between L1 and L2 sounds may have contributed to this pattern. First, the

assimilated L1 Cantonese tones with Mandarin T2-T3, that is Cantonese T2, T5 and T4, are internally ambiguous since all started with similar pitch level (Mok & Wong, 2010). With acoustic analysis on the speech of a female Cantonese speaker, Mok and Wong (2010) found that Cantonese tones with lower pitch range shared more tonal similarities, including Cantonese T2, T4, and T5. These tones shared the same starting pitch level and maintained it throughout the first half of the syllable duration. Similarly, the inherent similarities between Mandarin T2 and T3 may make it difficult to differentiate. The turning point of pitch contour is crucial for differentiating Mandarin T2 and T3 and may result in the perceived similarity between these two tones. The alike concave shapes of these two tones, Mandarin T2 and T3, may lead to confusion for native Mandarin speakers and also for non-native speakers (Shen & Lin, 1991). Kiriloff (1969) reported the results of two perception tests on Mandarin tones for 10 Australian students and claimed that for the English speaking learners, Mandarin T2-T3 was more difficult to acquire, followed by T1-T4. Similar research results were found in Wang (1995) examining the Mandarin tone learning by English speakers. Apart from the inherent similarity, it may be possible that the tone sandhi rule for Mandarin T3 may contribute to the confusion. According to the Mandarin tone sandhi rule, T3 will be changed to T2 if it is followed by T3 in a phrase, that means, a T3 T3 series will be changed to T2 T3. With this rule, Cantonese speakers may perceive Mandarin T2 and T3 as phonetic variants rather than contrastive phonemes. In addition, even though the overall F0 contour for Cantonese T2 and T5 is low rising, there is a dipping feature on phonetic perspective (Bauer & Benedict, 1997; So & Best, 2010). But the turning point of pitch contour for Cantonese T2 and T5 is not as distinctive as it is in Mandarin T3. Since Mandarin T2 is assimilated to Cantonese T2 and T5, the alike dipping feature between Cantonese T2, T5 and Mandarin T3 may contribute to the confusion of Mandarin T2-T3 pair for Cantonese speakers. In general, the Cantonese children in the present study encountered the same T2-T3 problem as that encountered by Cantonese and English adults in discriminating Mandarin tone pairs.

### **Identification**

The Cantonese-speaking children achieved the highest degree of identification accuracy for Mandarin T1, followed by T3 and T2, and finally T4. For the error distribution of Mandarin T2 and T3, significantly more errors consisted of misidentifying T2 as T3, and vice versa, which were consistent with the T2-T3 difficulty in the discrimination test.

The poor performance in differentiating the Mandarin T2-T3 pair was not applicable to the UC type hypothesis in identification. Based on Hao's (2012) assimilation test,

Mandarin T2 could be mostly assimilated to Cantonese T5 and then T2, while Mandarin T3 was similar to Cantonese T4. This pattern might suggest a facilitating factor in discriminating Mandarin T2 and T3 and so Cantonese children would not make too many errors in misidentifying Mandarin T2 to T3 and T3 to T2. Nevertheless, results showed that this pair posed great problem for the Cantonese L1 children. The same pattern also holds for Cantonese and English L1 adult speakers (Hao, 2012; So & Best, 2010). Hence, as discussed in the discrimination section, the internal similarity of Mandarin T2 and T3 and the assumption of phonetic variants as a result of the tone sandhi rules of T3 may also hinder identification of Mandarin T2 and T3. Also, the assimilation level of Mandarin T2 with Cantonese T2 and T5, and the similarity on the phonetic dipping feature of Mandarin T3 with Cantonese T2 and T5 might account for the misidentification of Mandarin T2 and T3 to each other. This may suggest that modifications are needed for PAM to be applied to L2 lexical tone learning. Factors other than native to non-native sounds assimilation, the inherent phonetic similarity of L2 sounds and language-internal variations need to be taken into account. Also, it is crucial to explore the inner assimilation of L1 sounds, especially those L1 sounds that share a higher degree of similarity with the L2 sound pair, which could help to explain the perception of L2 sounds or sound contrasts.

With respect to the most problematic tone, Mandarin T4, tone errors were generally evenly distributed across three tones, with a few more cases of participants misinterpreting it as T1. A similar case was found with Mandarin T1, with relatively more cases of participants misinterpreting it as T4. This pattern was consistent with the SC type for the Mandarin T1-T4 contrast. However, Mandarin T1-T4 has been found to be more challenging for Cantonese speakers than for English speakers (Hao, 2012; So & Best, 2010). Compared to Cantonese speakers, there is no L1 tone influence for English speakers. The presence of Cantonese T1 in the L1 may be considered to be interfering for the identification of Mandarin T1 and T4.

As for the order of Mandarin tone identification, the current results (T1>T3>T2>T4 where T1 is the easiest) were different from the findings for adult Cantonese speakers learning Mandarin reported in previous studies (Alexander et al., 2005; Hao, 2012; Li & Thompson, 1977; So & Best, 2010; Wong et al., 2005). The studies by Hao (2012) and So (2006) showed that adult Cantonese speakers performed significantly better in identifying Mandarin T4 and T1 compared to T3 and T2. The differences in the orders between children and adult L2 learners may suggest the need for a further study on the influence of age of learning on Mandarin lexical tone learning. Cantonese-speaking children in this study were at the initial stage of learning Mandarin. Tracking future

developmental changes in Mandarin lexical tones can further improve our understanding of L2 Mandarin tone perception by Cantonese speakers.

To conclude, Mandarin lexical tone perception by children with an L1 tone language background was explored in this study, and part of the results from both tests were in accordance with the hypothesis generalized from the PAM. PAM mainly proposes that non-native sounds are perceived based on the similarities and dissimilarities with native speech sounds. These various assimilation types are categorized through the contrasts between sounds in the native language and those in the L2. However, factors other than this, such as the inherent phonetic similarity among lexical tones in a language and the language variations, may also contribute to the perceptual difficulties. Hence, to provide more comprehensive theoretical support for L2 suprasegmental acquisition, the language internal similarities, variations and cross-language similarities could be used for reference to refine the PAM in the suprasegmental/tonal domain.

## References

- Alexander, J., Wong, P. C. M., & Bradlow, A. (2005). *Lexical tone perception in musicians and nonmusicians*. Paper presented at the Interspeech 2005-Eurospeech—9th European Conference on Speech Communication and Technology, Lisbon, Portugal.
- Aoyama, K. (2003). Perception of syllable-initial and syllable-final nasals in English by Korean and Japanese speakers. [Article]. *Second Language Research*, 19(3), 251.
- Bauer, R. S., & Benedict, P. K. (Eds.). (1997). *Modern Cantonese phonology*. Berlin and New York: Mouton de Gruyter.
- Bauer, R. S., Cheung, K.-H., & Cheung, P.-M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change* 15, 211–225.
- Best, C. T. (1990). Adult Perception of Nonnative Contrasts Differing in Assimilation to Native Phonological Categories. *Journal of the Acoustical Society of America*, 88, 177.
- Best, C. T. (1995). *A direct realist perspective on cross-language speech perception*. Timonium, MD: York Press.
- Best, C. T., Faber, A., & Levitt, A. (1996). Assimilation of non-native vowel contrasts to the American English vowel system. *The Journal of the Acoustical Society of America*, 99(4), 2602–2603.
- Best, C. T., McRoberts, G. W., & Goodell, N. M. (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, 775–794.
- Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology*, 14, 345–360.
- Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on crosslanguage perception of approximants. *Journal of Phonetics*, 20, 305–330.
- Best, C. T., & Tyler, M. D. (2007). *Nonnative and second-language speech perception: Commonalities and complementarities*. Amsterdam: John Benjamins.
- Bidelman, G. M., Hutka, S., & Moreno, S. (2013). Tone language speakers and musicians share enhanced perceptual and cognitive abilities for musical pitch: Evidence for bidirectionality between the domains of language and music. *PloS One*, 8(4).
- Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*, 18(1), 37–49.
- Chao, Y. R. (1948). *Mandarin primer*. Cambridge: Harvard University Press.
- Chao, Y. R. (1956). *Tone, intonation, singsong, chanting, recitatives, tone composition, and a tone composition in Chinese*. Mouton: The Hague.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*, 1(2), 141–147.
- Clumeck, H. V. (1980). The acquisition of tone. In G. Yeni-Komshian, J. Kavanagh & C. A. Ferguson (Eds.), *Child Phonology: Production (Vol. 1)*. New York: Academic Press.
- Flege, J. E. (1988). The development of skill in producing word-final English stops: Kinematic parameters. *Journal of the Acoustical Society of America*, 84(5).
- Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological Development, Models, Research, and Applications*. Timonium, MD: York Press.
- 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*. Timonium, MD: York Press.
- Fok, C. (1974). A perceptual study of tones in Cantonese. *HongKong: Centre of Asian Studies, University of HongKong*.
- Fon, J., & Chiang, W. Y. (1999). What does Chao have to say about tones. *Journal of Chinese Linguistics* 27, 15–37.
- Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, 36(2).
- Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “L” and “R”. *Neuropsychologia*, 9, 317–323.
- Gottfried, T. L. (2007). Music and language learning. In M. O. S. Bohn and M. J. (Ed.), *Language Experience in Second Language Speech learning: In Honor of James Emil Flege* (pp. 221–258). Amsterdam: John Benjamins.
- Gottfried, T. L., & Riester, D. (2000). Relation of pitch glide perception and Mandarin tone identification. *Journal of Acoustical Society of America*, 108(2604).

- Gottfried, T. L., Staby, A. M., & Ziemer, C. J. (2004). Musical experience and Mandarin tone discrimination and imitation. *Journal of Acoustical Society of America*, 115(2545).
- Hallé, P. A., Best, C. T., & Levitt, A. (1999). Phonetic vs. phonological influences on French listeners' perception of American English approximants. *Journal of Phonetics*, 27(3), 281–306.
- Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non-tonal language speakers. *Journal of Phonetics*, 40(2), 269–279.
- Ho, A. T. (1976). The acoustic variation of Mandarin tones. *Phonetica*, 33, 353–367.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*: Cambridge University Press.
- Kiriloff, C. (1969). On the auditory discrimination of tones in Mandarin. *Phonetica*, 20, 63–67.
- Kwan, C. W. (1990). *The right word in Cantonese*. Hong Kong: Commercial Press.
- Lee, C. Y., & Hung, T. H. (2008). Identification of Mandarin tones by English-speaking musicians and non-musicians. *Journal of Acoustical Society of America*, 124(3235–3248).
- Lee, Y. S., Vakoch, D., & Wurm, L. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25(5), 527–544.
- Li, C. N., & Thompson, S. A. (1977). The acquisition of tone in Mandarin-speaking children. *Journal of Child Language*, 4(2), 185–199.
- Lin, W. C. J. (1985). Teaching Mandarin tones to adult English speakers: analysis of difficulties with suggested remedies. *RELC Journal (Regional Language Centre of Southeast Asian Ministers of Education Organisation)*, 16(2), 31–46.
- Matthews, S., & Yip, Y. (1994). *Cantonese: a comprehensive grammar*. London: Routledge.
- Mok, P. M. K., & Wong, P. W. Y. (2010). *Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables*. Paper presented at the In Speech prosody 2010, 5th, international conference., Chicago.
- Mok, P. M. K., Zuo, D., & Wong, P. W. Y. (2013). Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese. *Language Variation and Change*, 25, 341–370.
- Moore, C. B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864–1877.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime User's Guide*. Pittsburgh: Psychology Software Tools Inc.
- Shen, X. S., & Lin, M. (1991). A perceptual study of Mandarin tones 2 and 3. *Language & Speech*, 34(2), 145–156.
- So, C. K. (2006). *Effects of L1 prosodic background and AV training on learning Mandarin tones by speakers of Cantonese, English and Japanese*. Doctoral dissertation, Simon Fraser University.
- So, C. K., & Best, C. T. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273–293.
- Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In E. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech Perception, Production, and linguistic Structure*. Tokyo: Ohmsha.
- Wang, Y. (1995). American learners' tone acquisition. *Language Teaching and Research*, 2, 126–140.
- Wayland, R., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54(4), 681–712.
- Wayland, R., & Li, B. (2005). *Training native Chinese and native English listeners to perceive Thai tones*. Paper presented at the ISCA Workshop on Plasticity in Speech Perception (PSP 2005).
- Werker, J. F., Frost, P. E., & McGurk, H. (1992). Cross-language influences on bimodal speech perception. *Canadian Journal of Psychology*, 46, 551–568.
- Wong, P., Schwartz, R. G., & Jenkins, J. J. (2005). Perception and production of lexical tones by 3-Year-Old Mandarin-Speaking Children. *Journal of Speech Language Hearing Research*, 48, 1065–1079.
- Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics*, 25, 61–83.
- Yang, B. (2010). *A Model of Mandarin Tone Categories— A study of perception and production*. Doctoral dissertation, The University of Iowa.
- Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*, 21(1), 46–48.
- Zee, E. (1999). Chinese (Hong Kong Cantonese). In I. P. Association (Ed.), *Handbook of the International Phonetic Association* (pp. 58–60). Cambridge: Cambridge University Press.
- Zhu, H., & Dodd, B. (2000). The phonological acquisition of Putonghua (Modern Standard Chinese). *Journal of Child Language*, 27(1), 3–42.