

Discrimination of Cantonese Tones by Speakers of Tone and Non-tone Languages

Zhen Qin and Peggy P. K. Mok
University of Kansas and The Chinese University of Hong Kong

1. Introduction

Listeners often find some non-native sounds difficult to perceive. For example, lexical tones in Chinese languages are difficult for non-native listeners to perceive, although pitch is frequently used in their first language (L1) and is not unfamiliar to them. We conduct a cross-linguistic study to explore the factors contributing to the difficulty of tone perception for non-native listeners.

1.1 Factors in perception of non-native sounds

L1 experience and psychoacoustics are generally considered to be the two main factors in perception of non-native sounds (Polka, 1991).

First of all, it is agreed that the influence from L1 accounts for the difficulty of non-native sounds in cross-linguistic perception. “L1 experience” is used in this paper to represent the general influence from L1. L1 experience has been widely attested in the studies involving naïve listeners and L2 learners. It is well-known that non-native speakers do not perceive and produce sounds in the same way as native speakers because of their different L1 experience (Best, 1995; Flege, 1995; Kuhl & Iverson, 1995).

However, the effects of L1 experience are still controversial regarding phonological and phonetic levels. Early research on language transfer proposed that L1 phonological inventory acts as a kind of “sieve” that filters out acoustic properties which are needed to distinguish new sounds in the L2 (Lado, 1957). Thus, some studies focused on L1 phonological transfer in terms of abstract structure, for instance, features in Feature Competition Model (Hancin-Bhatt & Govindjee, 1999) and in Feature Geometry (Brown, 1998). However, amounts of empirical research demonstrated that it is not enough to consider only abstract phonological structure in L2 speech research. Surface phonetic details can be more important than abstract phonological properties in accounting for the difficulty of L2 speech perception (Best, 1995; Flege, 1995; Kuhl & Iverson, 1995).

L1 experience alone is not sufficient, because it cannot predict some perceptual patterns shared by non-native listeners with different L1 backgrounds. Psychoacoustics is another factor which can predict non-native perception, especially for naïve listeners and learners at initial stages of L2 speech learning. Psychoacoustic factors are language-independent and affect all listeners similarly in speech perception. Burnham (1986) suggested that “robust” sounds which are acoustically salient are easy to distinguish regardless of language experience. Polka (1991) found that the perceptual differences of four Hindi consonant contrasts were attributed in part to different acoustic salience related to voicing in those contrasts. Psychoacoustic factor, for instance, acoustic salience, should be extended to account for perceptual patterns in non-native perception.

Werker and Tees (1984) proposed two processing modes in the perception of non-native sounds. If listeners perceive a non-native sound only based on its psychoacoustic aspects, the listeners hear the sound in an acoustic mode. If listeners perceive a non-native sound relying on their L1 phonological and phonetic knowledge, they hear the sound in a linguistic mode. But the perception of non-native speech could not always be clearly distinguished in the two modes. Later studies showed that three factors, phonemic, phonetic and psychoacoustic factors, combined to exert an influence on perception of non-native sounds (Polka, 1991; Werker & Tees, 1984).

Our primary interest is to examine how different factors interact with each other to affect the perception of non-native tones in specific tasks. Theoretical models, for instance, Best's (1995) Perceptual Assimilation Model (PAM) consider L1 experience at phonetic level as well as psychoacoustic factors. The PAM focuses on surface phonetic details and intends to see whether there is category assimilation between native and non-native sounds. Different cases such as assimilated and non-assimilated cases are predicted on the basis of the articulatory-phonetic similarities between L1 and non-native sounds (Best, 1995). The extended version, PAM-L2 in Best and Tyler (2007), attributes the difficulties of the non-native sounds to both phonological structure and phonetic similarity. Since the factors appear to be incorporated differently in the theoretical models, it is important to investigate how the two main factors affect the perception of non-native sounds in light of the theoretical models.

1.2 Perception of non-native tones

The study aims at investigating how L1 experience and psychoacoustic factors shape the perception of non-native lexical tones.

We focus on the perception of tones, because it is of growing importance to extend the theoretical models developed primarily based on segmental observations to the suprasegmental domain. Evidence of the various models on non-native perception mainly comes from studies of consonants (Flege & Hillenbrand, 1986; Iverson et al., 2003) and vowels (Flege, 1987). Insufficient studies investigated non-native perception in the suprasegmental domain. Thus, it is interesting to examine how non-native tones were perceived in the framework of the theoretical models.

1.2.1 The phonemic status of pitch

The phonemic status of pitch in L1, whether pitch is lexically decoded, plays an important role in determining the performance of tonal perception.

Categorical perception of tones is linked closely with tonal categorization. Prior studies found that speakers of tone languages perceived tones in a categorical manner, e.g. Mandarin (Hallé et al., 2004; Peng et al., 2010; Wang, 1976; Xu, Gandour, & Francis, 2006) and Cantonese (Francis, Ciocca, & Ng, 2003; Peng et al., 2010). In order to understand what role the phonemic status of pitch plays in the perception of non-native tones, it is important to compare how tones are perceived by different L1 speakers.

For instance, Wang (1976), using an 11 step continuum of the syllable [i], found a typical pattern of categorical perception for Mandarin Tone 1 (T1) and Tone 2 (T2). A sharp category boundary was found between steps 6 and 8 in an identification task, and a corresponding accuracy peak was revealed in the same area in a discrimination task. Likewise, Hallé, Chang

and Best (2004) reported that Taiwan Mandarin speakers showed a “quasi-categorical” manner in the perception of tones.

In contrast to speakers of tone languages, speakers of non-tone languages do not hear tones categorically. Wang (1976) illustrated that English speakers showed a flatter category boundary and a lower discrimination accuracy compared with Mandarin speakers. While Mandarin speakers exhibited a linguistic boundary in perceiving pitch contours, English speakers appeared to make judgments on the basis of psychoacoustic properties of the stimuli. Similar to English speakers, French speakers in Hallé et al. (2004) and German speakers in Peng et al. (2010) were reported to hear tones in a non-categorical manner and differ from Mandarin speakers in discrimination and identification patterns.

Although speakers of non-tone languages do not hear tones categorically, they may have a psychoacoustic advantage over speakers of tone languages. While Mandarin speakers were not sensitive to minor pitch changes in their tonal categorization, English speakers were able to detect subtle pitch variations, that is, within-category differences (Leather, 1987; Stagger & Downs, 1993). For instance, Hallé et al. (2004) found that French speakers did not hear tones categorically, but they were able to discriminate tonal variations by relying on the acoustic properties of the stimuli.

In short, most studies showed that speakers of tone languages differ from those of non-tone languages in tonal perception.

1.2.2. Perceptual cues of tones

Pitch is the primary cue in the perception of Mandarin tones (Howie, 1976; Lin & Wang, 1984) and Cantonese tones (Fok, 1974; Khouw & Ciocca, 2007).

Although pitch is the primary cue in tonal perception, speakers of tone languages and non-tone languages often attend to different dimensions of pitch. Previous studies agreed that English speakers attended to pitch height rather than pitch direction, whereas Mandarin and Cantonese speakers were sensitive to pitch direction (Francis et al., 2008; Gandour, 1983).

Difference perceptual cues were found not only between speakers of tone and non-tone languages, but also amongst speakers of different tone languages. In Gandour (1983), differences regarding pitch dimensions were found between Taiwan-Mandarin and Cantonese speakers. While Taiwan-Mandarin speakers assigned more weight to pitch direction than pitch height, Cantonese speakers were sensitive to both pitch direction and pitch height.

In brief, although pitch is the primary cue in tonal perception, listeners with different L1 backgrounds often attend to different dimensions of tones.

1.2.3 Psychoacoustic factors

While L1 experience contributes to listeners’ different performance in phonemic and phonetic levels, psychoacoustic factors, which are language-independent, affect all listeners in a similar way.

Previous studies found that some Mandarin tone pairs, T2-T3 and T1-T4, were difficult for both native and non-native speakers because of their acoustical similarity (Bent, 2005; Huang, 2004). While T2 and T3 both have a rising pitch contour, T1 and T4 both start with a high pitch (Howie, 1976). So (2010) found an asymmetrical pattern, which was shared by speakers of tone language (Cantonese) and non-tone languages (Japanese and English). All the listeners found the

T2-T3 and T1-T4 pairs more confusable than the other pairs in an identification task. Although the confusion of T2-T3 for native speakers was found to be partially attributed to the neutralization of the two tones in tone sandhi in Huang (2004), the difficulty of the two pairs for non-native listeners demonstrated clearly the effect of acoustic similarity. Briefly, not all tones are perceived equally. Some tones, which are acoustically similar, are more confusable than others.

Psychoacoustic factors have an impact on categorical perception of tones as well. Francis et al. (2003) found that the regions of natural auditory sensitivity, together with language-specific factors, determined the boundary of Cantonese tonal categories, because some regions in the perceptual space were acoustically more salient than other regions. Xu et al. (2006) showed that contour tones were more categorically perceived than level tones because pitch movement (rising or falling), which was acoustically salient, was present in contour tones.

1.3 The present study

Both speakers of tone language (i.e. Mandarin) and non-tone languages (i.e. English and French) were included in the study. First of all, the phonemic status of pitch in L1, whether pitch is lexically decoded, plays an important role in determining the performance of tonal perception. Most previous studies showed that speakers of tone languages performed better than speakers of non-tone languages in perceiving and processing lexical tones (Lee, Vakoch, & Wurm, 1996; Wayland & Guion, 2004; Wang, Jongman, & Sereno, 2001; Wang, Behne, Jongman, & Sereno, 2004). Burnham and Mattock (2007) suggested that speakers of non-tone languages (e.g. English speakers) processed lexical tones primarily in an acoustic mode by depending mainly on psychoacoustic similarity of tonal stimuli; speakers of tone languages relied more on their native tonal system and perceived tones primarily in a linguistic mode. Mandarin speakers were included in our study to contrast with non-tone language speakers for examining the phonemic status of pitch.

Second, the possible perceptual differences of speakers of two non-tone languages have not been systematically compared. Extensive studies explored the perception of tones, but most of them focused on the differences between speakers of tone and non-tone language (e.g. Francis, Ciocca, Ma & Fenn, 2008). A literature search suggests that speakers of two non-tone languages have not been compared empirically. While French does not have any lexical prosody, English has lexical stress (Beckman, 1986). Given such differences, Hallé et al. (2004) predicted that French speakers would outperform English speakers in perceiving non-native lexical tones, because pitch in French is not lexically constrained and is used freely. However, they did not provide any concrete evidence to substantiate their prediction. Hence, the current study aims to investigate whether pitch in English and French prosody would result in perceptual differences.

Moreover, Cantonese tones were chosen as test stimuli because of the complicated Cantonese tonal system: both level and contour tones are found, and the number of level and contour tones is balanced. Pitch height and pitch direction tend to be equally used in the Cantonese tonal system. Since test stimuli may bias specific dimensions, the potential bias can be minimized by using Cantonese tones as stimuli (Francis et al, 2008; Gandour, 1983). However, native speakers of Cantonese were not included in our study, because native speakers have been tested in detail using the same paradigm and similar stimuli in Mok, Zuo and Wong (2013).

2. Pitch in Cantonese, Mandarin and English and French

2.1 Lexical tones in Cantonese

Cantonese encodes pitch in tones for lexical meanings. As is illustrated in Figure 1, there are six lexical tones in Cantonese (Chao, 1947; Bauer & Benedict, 1997). The six tones are labeled according to Bauer and Benedict (1997) as follows: T1, a High Level (55) tone; T2, a High Rising (25) tone; T3, a Mid-Level (33) tone; T4, a Low Falling (21) tone; T5, a Low Rising (23) tone; T6, a Low Level (22). The six tones are contrastive in meaning. For example, /si 55/ ‘silk’, /si 25/ ‘history’, /si 33/ ‘to try’, /si 21/ ‘time’, /si 23/ ‘city’, /si 22/ ‘matter’. There are also three checked tones: T7, a High level (5); a Mid-Level T8, (3); T9, a Low Level (2). The three checked tones are shorter in duration, but they do not differ from the three corresponding level tones in pitch patterns. Only the six unchecked tones are included in our experiments.

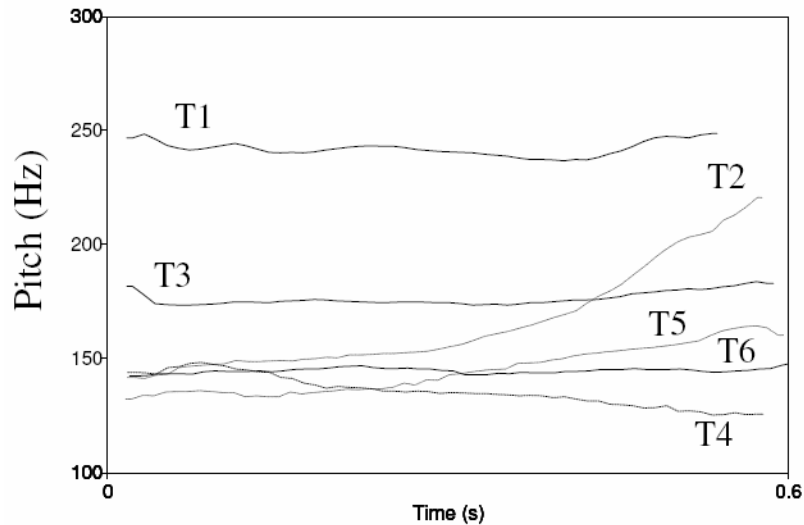


Figure 1: Pitch track of the six Cantonese tones (*figure from Mok, Zuo & Wong 2013*)

The relative distribution of the six tones in an acoustic space has been studied (Peng, 2006). As shown in Figure 1, while T1 was separated clearly from the other tones, the two tones in the T2-T5 pair and the T3-T6 pair overlapped a lot in acoustic space. In addition, it is worth noting that the acoustic distribution is crowded at the pitch onset and most of the tones (e.g T2, T4, T5, T6) differ only at the offset. Studies have shown that pitch offset may be an essential cue for native speakers to categorize the tones (Khouw & Ciocca, 2007). Due to the acoustic similarities between some Cantonese tones, several pairs are harder to discriminate than others. For instance, the pairs of the two rising tones (T2-T5) and the two level tones (T3-T6) were found to be difficult for native speakers to distinguish in perception and for children to learn in their phonological acquisition (Ciocca & Lui, 2003; Khouw & Ciocca, 2007; Mok, Zuo & Wong, 2013). Similarly, the psychoacoustic similarities of T2-T5 and T3-T6 should contribute to the confusion for non-native listeners too.

2.2 Lexical tones in Mandarin

As illustrated in Figure 2, Mandarin has four lexical tones: T1, a High Level (55) tone; T2, a Rising (35) tone; T3, a Dipping (214) tone; and T4, a Falling (51) tone (Chao, 1968; Li & Thompson, 1989). For example, /pa 55/ ‘eight’, /pa 35/ ‘to pull out’, /pa 214/ ‘to hold’, /pa 51/ ‘father’. The four Mandarin tones are clearly different in both pitch patterns and durations.

Moreover, Mandarin has a tone sandhi rule: T3 (214) becomes T2 (35) before another T3 and becomes a low falling tone (21) when preceding other tones. So T2 and the low falling tone are the allotones of T3. There is also a “neutral tone” in Mandarin, which is an unstressed syllable. Since the neutral tone has a smaller pitch range and a shorter duration than the four canonical tones, it can be described as a mid-level pitch target (Chen & Xu, 2006).

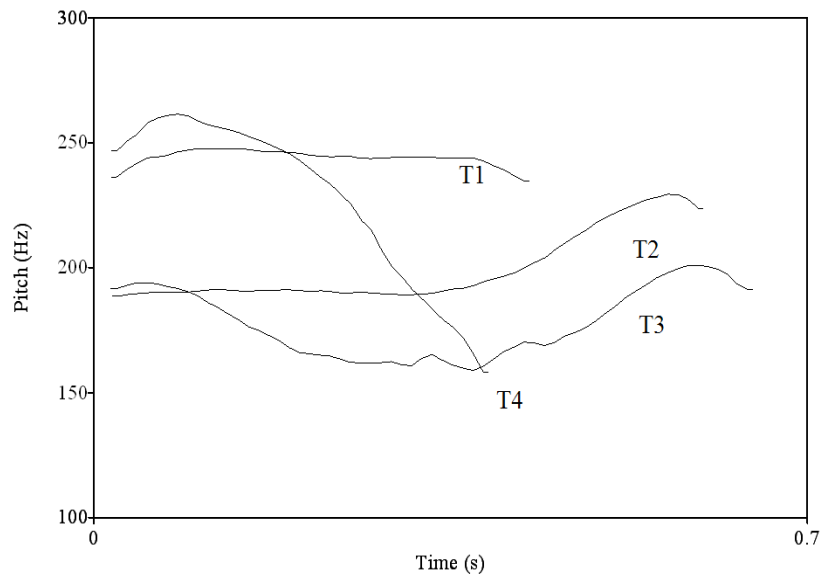


Figure 2: Pitch track of the four Mandarin tones¹

2.3 Pitch in English and French

Although neither English nor French has lexical tones, the two languages differ in how pitch is used in stress and intonation. While English speakers are sensitive to pitch variations, which are used contrastively in lexical stress, French speakers do not use stress in a contrastive way and the cue of pitch is less important for them. Thus, French speakers are regarded as likely candidates with a low sensitivity to pitch variations at the syllable level in our study.

Regarding lexical stress, English has a stress placement system that results in different words being stressed on different syllables (Cutler, 2005). All English words have stress, and English speakers use stress in a contrastive way to recognize some words. For instance, record (noun) and record (verb) form a minimal pair with respect to stress placement. An overall higher pitch was found in stressed syllable than in unstressed syllable (Beckman, 1986; Fry, 1955).

¹ The pitch tracks are extracted from a recording of /yi/ with four Mandarin tones by a female native Beijing Mandarin speaker.

On the contrary, French does not have lexical stress (Cutler, 2005; Vaissière, 1991). While the final syllable in a word or in a phrase is always accented with a rising contour and an increase in duration (“the primary stress”), the first syllable is optionally accented with a rising contour (“the secondary stress”) (Di Cristo, 1998; Welby, 2006). The so-called “stress” functions as a boundary marker and as a demarcative cue to the edges of a word or phrase (Vaissière, 1991; Welby, 2006). However, stress is not contrastively used in French. Some studies suggested that French does not have stress at all in word domain (Dupoux, Pallier, Sebastian & Mehler, 1997). Psycholinguistic research found that French listeners had difficulty distinguishing words which differ in the position of stress, and a “stress deafness” of French listeners was shown in listening nonsense words and L2 words (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). Briefly, English and French have a functional difference with respect to the contrastive use of lexical stress.

Regarding intonation, English and French intonation contours differ in pitch accent, which appear to be correlated with the difference of stress. While French intonation is often characterized by a sequence of rising pitch movements which function as segmentation cues and boundary markers (Delattre, 1951; Welby, 2006), the presence of lexical stress and its dynamic acoustic realization in English allow for more pitch accents and intonation contour varieties (Blum, 1999; Vaissière, 2002). For example, Gussenhoven (2004) illustrated that the tonal grammar of English may derive more types of pitch accents (4 types in English) than that of French (2 types in French). Different from English which has the mobile characteristics of pitch accent² and many potential accents in the pre-nuclear position, French has pitch accent in a fixed position. French has smaller number of words in each intonation group and appears to involve many short intonation groups rather than contour varieties. Therefore, the intonations of English and French differ in detailed phonetic realizations.

| Language | Lexical tone | Lexical stress | Intonation system |
|-----------------|---------------------|-----------------------|--------------------------|
| Mandarin | Yes | yes | yes, |
| English | No | yes | different phonetic |
| French | No | no | realizations |

Table 1: Pitch use in the prosodic systems of the three languages.

Although pitch is used universally in intonation contours, pitch is used differently in lexical level across the three languages. As can be seen in Table 1, while Mandarin uses pitch primarily in lexical tones and to some extent in lexical stress, English uses pitch only in lexical stress. In contrast, French does not use pitch contrastively in any lexical prosody. The target languages in this study have different prosodic types in that pitch is introduced into the phonological representation at different levels (Hirst & Di Cristo, 1998; Jun, 2005). In summary, the three languages representing different prosodic types are included in the present study.

² “nucleus” was originally used in Cruttenden (1997). Since “pitch accent” is widely used later to refer to similar thing in the Autosegmental-Metrical Theory, “pitch accent” is used in the present study for the purpose of consistency.

3. Research Questions and Predictions

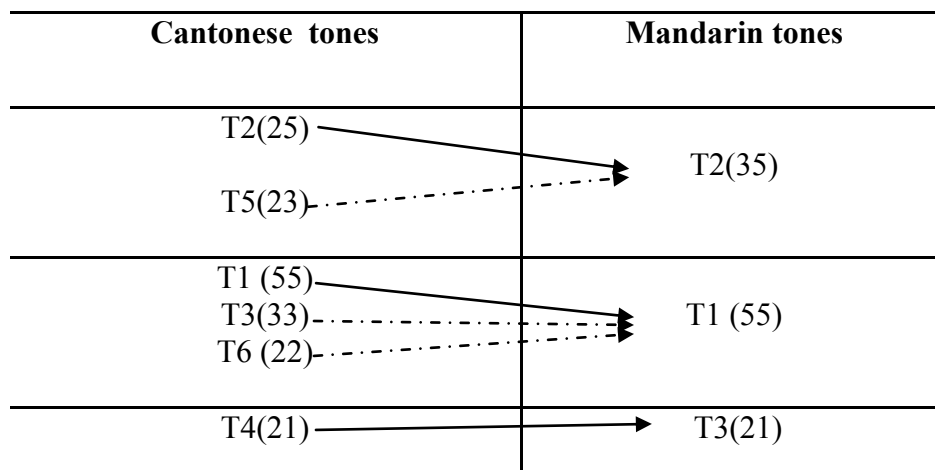
3.1 Research Questions

Based on the pitch use in different prosodic systems, the questions and issues which this study intends to address are formulated as follows. First, is perception of non-native tones affected by pitch use in L1, psychoacoustic similarities of the tonal stimuli, or both? Although the two factors were investigated in studies on non-native tone perception, it is still unclear how L1 pitch use and psychoacoustic similarities of the tonal stimuli would affect the perception of non-native tones. Second, given the effect of L1 experience, does the phonemic status of pitch in L1 separates Mandarin speakers from English and French speakers? Lastly, does different pitch use leads to perceptual differences of English and French speakers? No previous study has investigated the perceptual differences between two groups of non-tone language speakers, so it is unclear if perceptual differences due to different pitch use in L1 between English and French speakers are present.

3.2 Predictions

Concerning the effect of psychoacoustic factors and L1 experience, all the L1 groups are predicted to find it hard to discriminate the Cantonese tones which are acoustically similar (e.g. pairs of T2-T5 and T3-T6) and find it easy to distinguish some tones (e.g. T1) from other tones because of acoustic dissimilarity. On the other hand, Mandarin, English and French speakers are predicted to perceive Cantonese tones in different ways due to their language-specific pitch use in L1.

Regarding performances of different L1 groups, the phonemic status of pitch in L1 may differentiate speakers of tone languages from those of non-tone languages. Mandarin speakers are predicted to outperform the other two L1 groups in overall perceptual performance. It is an empirical question whether English and French speakers will perceive Cantonese tones in the same way.



Note: Bold Line: assimilated to native sound or as a good exemplar;

Dash Line: assimilated to one native sound as a bad exemplar.

Table 2: The assimilation between Cantonese and Mandarin tones as predicted by the PAM.

The performance of different L1 speakers can be predicted in the framework of the PAM. According to the PAM, since Mandarin has lexical tones, Mandarin speaker should perceptually assimilate Cantonese tones to native tones rather than native stress or intonation patterns based on their phonetic similarities. As illustrated in Table 2, the Cantonese rising tone, T2, can be assimilated to the Mandarin rising tone, T2; another Cantonese rising tone, T5, could be assimilated to the Mandarin T2 as a bad exemplar of the Mandarin T2. The case is Goodness-Category (GC) assimilation. The discrimination of the T2-T5 pair ought to be moderate to very good. Second, the three Cantonese level tones (T1, T3, T6) are predicted to be assimilated to the Mandarin high level tone, T1. While Cantonese T1 could be a good exemplar of Mandarin T1, Cantonese T3 and T6 may be both bad exemplars of Mandarin T1. This case is Goodness-Category (GC) assimilation. The discrimination should be moderate to very good as well. Additionally, the pair of two bad exemplars, T3-T6, may be more difficult than the pairs of T1-T3 and T1-T6. Besides Goodness-Category (GC) assimilation, it is likely that Single-Category (SC) assimilation can be applied to the pairs of T2-T5 and T3-T6 as well, since T2 and T5 may be assimilated to the Mandarin rising tone, and T3 and T6 may be assimilated to the Mandarin level tone. The two possibilities of the assimilation type for the T2-T5 and T3-T6 pairs will be discussed with reference with our results. Third, the Cantonese lowing falling tone (T4) could be assimilated as an exemplar of the allotone of Mandarin T3 (21). Since the other five Cantonese tones are unlikely to be assimilated to the T3 allotone, the pairs including T4 are regarded as Two-Category (TC) assimilation, and the discrimination of these pairs should be very good.

Although English and French have intonation contours which are similar to Cantonese tones in pitch patterns, Cantonese tones are predicted to be non-assimilated to English and French intonation contours. Lexical tones convey linguistic meanings in word level whereas intonation conveys pragmatic meaning extending larger units such as phrases or sentences. As lexical tone and intonation use pitch at different levels and domains, the assimilation between non-native tones and native intonation contours appears to be unlikely (see Bent, 2005). Cantonese tones may not be heard as any native prosodic category by English and French speakers.

To address the research questions and test the predictions of the PAM model, a paradigm of AX forced choice discrimination was conducted for the subjects with different L1 backgrounds.

4. Methods

4.1 Subjects

As Table 3 shows, 12 native Mandarin speakers (2 male, 10 female) with an average age of 22, 10 native English speakers (7 male, 3 female) with an average age of 21, and 10 native French speakers³ (3 male, 7 female) with an average age of 20 were recruited for the experiment. The Mandarin subjects were students from the Mainland China studying at the Chinese University of Hong Kong. All of them spoke standard Mandarin as L1 and could not speak other Chinese dialects. The English and French subjects were exchange students at the university. As Table 3 shows, the English subjects came from English speaking countries, and most of the French

³ Among the French subjects, F9 is a bilingual speaker of French and English, and F10 is a bilingual speaker of French and Arabic. French is their dominant language and is used in their family and education from childhood.

subjects were from France. The exchange students were paid HKD100 for their participation whereas the Mandarin subjects participated voluntarily.

All of the subjects had been in Hong Kong for less than one and a half months at the time of recording. And they were naïve listeners with no knowledge of Cantonese. Although two of the English subjects and eight of the French subjects had limited experience with Mandarin, they were not proficient in Mandarin in any sense when they participated in the experiment. All the subjects had no or only limited musical training. Finally, the subjects reported no speech or hearing impairments.

| Mandarin Group | Sex | Country | English Group | Sex | Country | French Group | Sex | Country |
|----------------|-----|-----------------|---------------|-----|-----------|--------------|-----|-----------------|
| M1 | F | Main-land China | E1 | M | England | F1 | F | Switzerland |
| M2 | F | | E2 | M | U.S.A | F2 | M | France |
| M3 | F | | E3 | M | Canada | F3 | F | France |
| M4 | F | | E4 | M | England | F4 | M | France |
| M5 | F | | E5 | M | Australia | F5 | M | France |
| M6 | F | | E6 | F | U.S.A. | F6 | F | France |
| M7 | F | | E7 | F | Canada | F7 | F | France |
| M8 | F | | E8 | M | U.S.A. | F8 | F | Canada (Quebec) |
| M9 | M | | E9 | M | U.S.A. | F9 | F | France |
| M10 | M | | E10 | F | U.S.A. | F10 | F | Morocco |
| M11 | F | | | | | | | |
| M12 | F | | | | | | | |

Table 3: Subjects from different countries in three L1 groups.

4.2 Materials

Two monosyllables, [jəu] and [se] (/jau/ and /se/ in Jyutping) were used as test stimuli for two reasons. First, real Cantonese words are attested with all of the six tones with the two syllables (see Table 3). Second, these stimuli were used in previous studies (e.g. Francis et al., 2008).

One female native speaker of Hong Kong Cantonese produced all the stimuli. The recording was carried out in a quiet room. The speaker was required to read a randomized wordlist (see Table 4) in a carrier phrase *ngo5duk6 zi6* “I read the word ___” three times in normal speech rate. The carrier sentence was selected to avoid the final lengthening effect on the tones of our target stimuli, and there is no tone sandhi which applies in the sentence. The recording was done directly on disk using a sampling rate of 44 kHz. The six tones carried by each syllable were excised. 12 tone stimuli (2 syllables × 6 tones) in total were chosen based on the recording quality (e.g. F0 curve clarity, creaky voice).

| Syllable | Tone | Chinese Character | Syllable | Tone | Chinese Character |
|----------|------|-------------------|----------|------|-------------------|
| /jau/ | T1 | 休 rest | /se/ | T1 | 些 some |
| | T2 | 柚 grapefruit | | T2 | 寫 write |
| | T3 | 幼 young | | T3 | 瀉 diarrhea |
| | T4 | 由 from | | T4 | 蛇 snake |
| | T5 | 有 have | | T5 | 社 society |
| | T6 | 又 again | | T6 | 射 shoot |

Table 4: Cantonese words with the six tones carried by the three syllables.

The six tones carried by the two syllables were plotted in Figure 3, which shows that they have similar tonal distribution across the two syllables. T1 stands out from the other tones in terms of pitch height and does not show a significant falling movement during its course. The mid-level tone, T3, is further apart from T1 than from the low level tone, T6. The tonal space in the lower pitch range is very crowded, as T2, T4, T5, and T6 share a similar starting point. The two rising tones, T2 and T5, only differ in the magnitude of rising pitch movement. T4 falls slightly, while T5 rises slightly towards the end. Briefly, the pitch distribution of the tone stimuli here is consistent with the tone letter labels suggested by Bauer & Benedict (1997).

Most of the tones have similar duration, but it is observed that T1 for the syllable /jau/ is shorter than other tones. Although we focus on pitch, it is unclear whether the subtle durational difference distinguishes T1 from other tones better for /jau/ than that for /se/. We try to come back to this possibility later in our results of individual tone pairs. T4 tokens used in this study have no creaky quality. Since these tokens were excised from connected speech, some phonetic variations of tone contours are observed. Tonal coarticulation may account for some modulations of pitch contour. More importantly, we have checked with some native speakers, and they all found these tokens of stimuli acceptable.

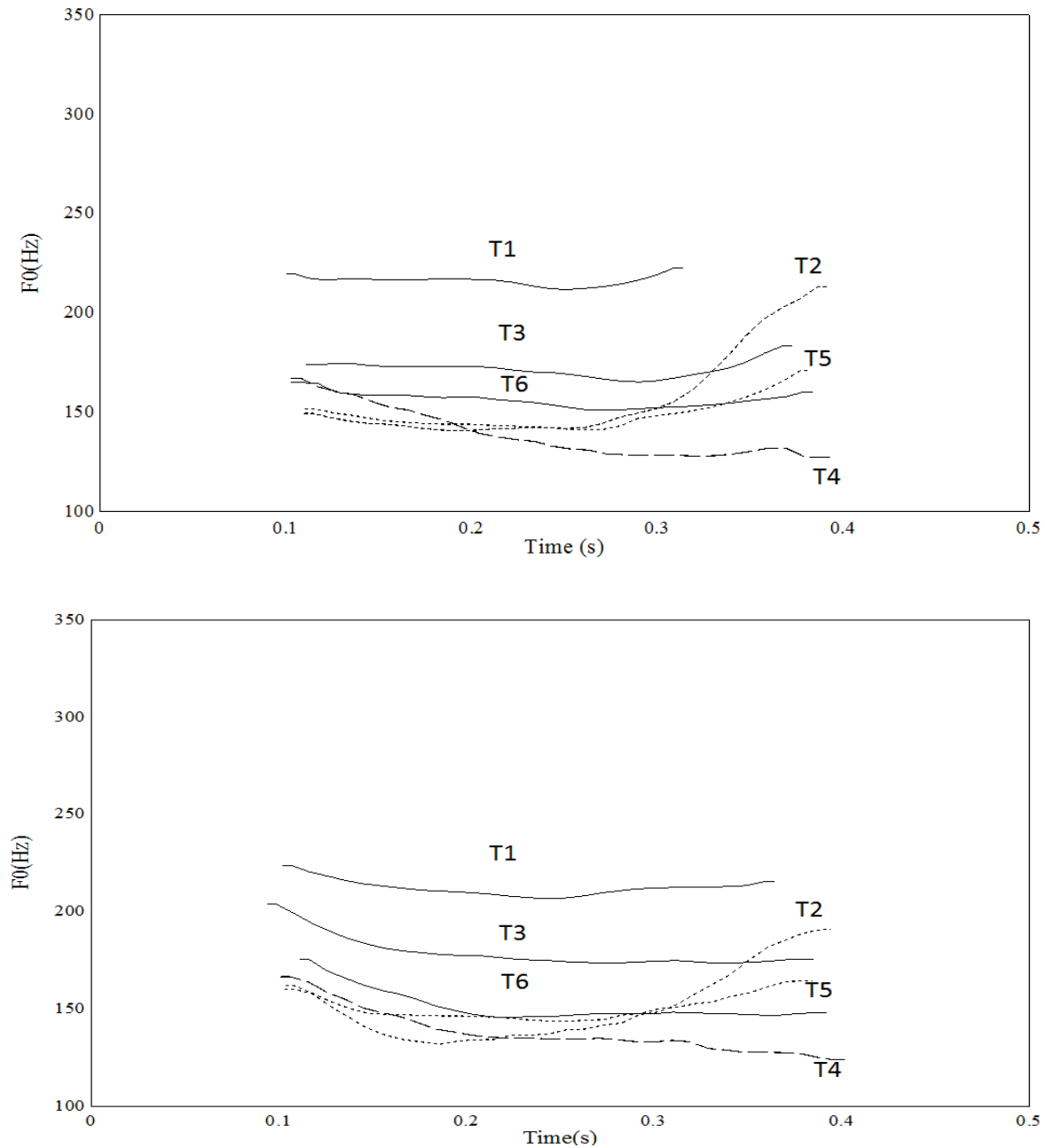


Figure 3: Pitch track of the Cantonese tones carried by /jau/ (top panel) and /se/ (bottom panel).

4.3 Procedures

An AX forced-choice discrimination task was conducted for the subjects. There are two types of tone pairs, AA pairs (pairs with the same tone) and AB pairs (pairs with different tones). All the possible pairings of the six tones for each syllable, 6 AA pairs and 15 AB pairs, were presented randomly to the subjects. Each AB pair was presented two times with presentation order counter-balanced. 60 tokens of AB pairs (15 AB pairs \times 2 syllables \times 2 orders) and 12 tokens of AA pairs (6 AA pairs \times 2 syllables) were used⁴. The 72 tokens in total were grouped into 7 blocks with 10 tokens in each block and 12 pairs in the last block.

The subjects were required to sit in front of a laptop computer in a quiet room at the Chinese University of Hong Kong. The stimuli were presented to them via a stereo headphone with the volume adjusted to a comfortable level. The subjects were told that they would hear pairs of sounds from a certain language, and their task was to judge whether two speech tones in each pair were the same or different. The subjects were asked to use their left and right index fingers to press the button “F” labeled “same” and the button “J” labeled “different” on the keyboard of the laptop as soon as possible after hearing two speech tones. The task was controlled by the software DMDX. The test session lasted for approximately 20 minutes. The inter-stimulus interval (ISI) was 500 milliseconds (ms)⁵. The time-out time was 3000 ms and missing responses were excluded from analysis. There was a short break between every two blocks. The same instruction was given orally and was also displayed visually on the laptop screen before the test. No feedback was given. Before the actual experiment, the subjects familiarized themselves with practice trials randomly selected from the test stimuli. Error rates⁶ were collected.

5. Results

In the discrimination task, all the participants made very few errors for the AA pairs (error rates under 1.7 % for each L1 group) in the task. The three L1 groups did not differ in the discrimination of AA pairs. Therefore, only results for the AB pairs are reported below.

5.1 Overall performance in the speech task versus non-speech task

The average error rates of all AB pairs for the two syllables are shown in Figure 4. Among the three L1 groups, the Mandarin group performed the best with the lowest error rates for both syllables. While the English group had a slightly lower rate than the French group in /jau/, it had a slightly higher rate than the French group in /se/. In terms of syllables, all the L1 groups appeared to have higher error rates for /jau/ than those for /se/.

⁴ The unbalance of the AB and AA pairs may have induced bias to the “different” responses resulting in more errors for the AA pairs. However, only very few errors of the AA pairs were found for each L1 group. Additionally, this study focuses on the results of AB pairs. Therefore, the unbalanced design did not appear to have affected the results adversely.

⁵ The ISI of 500 ms used in our study may be too short. The task may have mainly tapped the auditory and phonetic knowledge of listeners and restricted the retrieval of phonological knowledge at higher levels (Werker & Logan, 1985). Future studies should include both ISIs of 500 and 1500 ms to investigate the effect of L1 experience.

⁶ Reaction time was also collected. But only the data of reaction time were not reported here due to page limit.

One GLM Repeated-Measures test was conducted for error rates with L1 group (Mandarin, English, and French) as a between-subject factor and Syllable (/jau/ vs. /se/) as a within-subject factor. The main effects of L1 group [$F(2, 29) = 5.2, p = 0.012$] and Syllable [$F(1, 29) = 11.6, p = 0.002$] are significant. However, the interaction effect [$F(2, 29) = 1.2, p = 0.305$] was not significant. Since a significant interaction effect of L1 group and syllable was not found for error rates, we assumed that the pattern of tonal discrimination holds for the three different L1 groups by collapsing the two syllables.

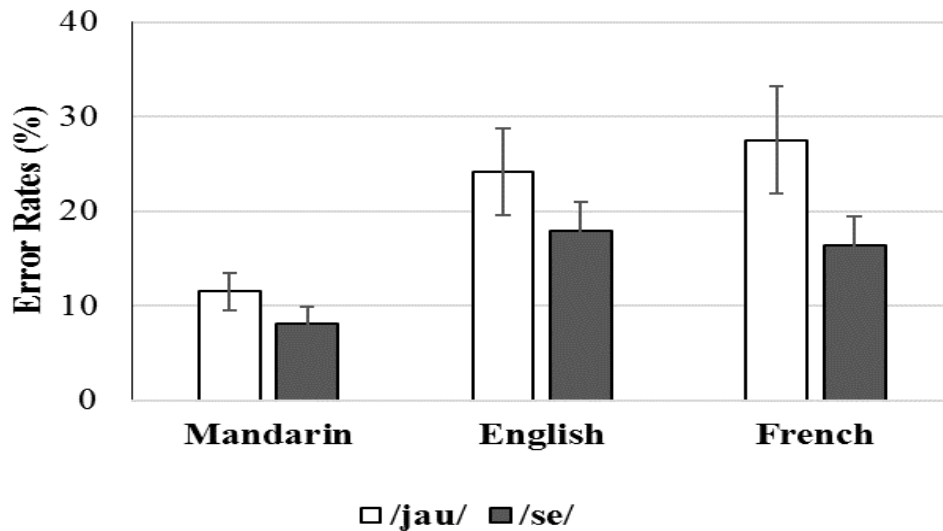


Figure 4: Error rates of tonal discrimination by different L1 groups. Error bars show one standard error.

Post-hoc tests were conducted to explore the significant effects. One-Way ANOVA tests (Bonferroni tests)⁷ with L1 group as between-subject factor was conducted to investigate the effects of L1 group. The Mandarin group had a significantly better performance in terms of error rates than either the English ($p = .04$) or French ($p = .023$) group. In addition, no significant difference ($p = .99$) was found between the English and French groups in terms of error rates.

In summary, the overall effects of L1 group were found in terms of error rates in the tonal discrimination task. While the Mandarin group had a better performance than the other two L1 groups, the English and French groups did not appear to have any consistent differences in their overall performance.

5.2 Performance for individual tone pairs

Since different combinations of tones were involved in the present study, the discrimination pattern of each L1 group was examined in detail with respect to individual tone pairs. Since all

⁷ Equal group variances were found for the data of error rates, so Bonferroni tests were done with them.

the groups made few errors in AA pairs, as can be seen in Table 4, we only focus on AB pairs in our analysis.

First, error patterns of all the tone pairs for the three L1 groups are shown in Table 5⁸. Among the three L1 groups, the Mandarin group performed best as they had the lowest error rates for most of tone pairs. The subjects had different error rates in these individual tone pairs.

| Mandarin | Tone1 | Tone2 | Tone3 | Tone4 | Tone5 | Tone6 |
|-----------------|-------|-------------|-------------|-------|-------|-------|
| Tone1 | 0.0 | — | — | — | — | — |
| Tone2 | 2.1 | 0.0 | — | — | — | — |
| Tone3 | 16.7 | 0.0 | 0.0 | — | — | — |
| Tone4 | 0.0 | 6.3 | 0.0 | 0.0 | — | — |
| Tone5 | 0.0 | 72.9 | 0.0 | 6.3 | 0.0 | — |
| Tone6 | 4.2 | 0.0 | 36.1 | 6.3 | 10.4 | 0.0 |

| English | Tone1 | Tone2 | Tone3 | Tone4 | Tone5 | Tone6 |
|----------------|-------|-------------|-------|-------|-------------|-------|
| Tone1 | 0.0 | — | — | — | — | — |
| Tone2 | 5.0 | 0.0 | — | — | — | — |
| Tone3 | 17.5 | 5.0 | 0.0 | — | — | — |
| Tone4 | 5.0 | 23.3 | 10.0 | 5.3 | — | — |
| Tone5 | 10.8 | 75.0 | 12.5 | 27.5 | 0.0 | — |
| Tone6 | 7.5 | 28.3 | 27.5 | 10.0 | 60.0 | 5.0 |

| French | Tone1 | Tone2 | Tone3 | Tone4 | Tone5 | Tone6 |
|---------------|-------|-------------|-------|-------|-------------|-------|
| Tone1 | 0.0 | — | — | — | — | — |
| Tone2 | 2.5 | 0.0 | — | — | — | — |
| Tone3 | 10.0 | 10.0 | 0.0 | — | — | — |
| Tone4 | 5.0 | 17.5 | 12.5 | 0.0 | — | — |
| Tone5 | 8.3 | 82.5 | 17.5 | 18.3 | 0.0 | — |
| Tone6 | 6.7 | 27.5 | 20.8 | 25.8 | 70.0 | 5.0 |

Table 5: Error rates of all tone pairs in a matrix for the Mandarin (top panel), English (mid panel) and French (bottom panel) subjects. Shading area shows statistical significance.

One Repeated-Measures ANOVA tests were conducted on the error rates with L1 group as a between-subject factor (3 levels) and Tone pairs (15 levels) as a within-subject factor. The results revealed the main effects of L1 group [F (2, 29) =5.1, p=0.012], Tone pair [F (5.7⁹, 167) =41.3, p<0.001], and the interaction between them [F (11.5, 167) =11.5, p<0.001]. Post-hoc

⁸ Since the results of individual tone pairs were not different across the two syllables, the matrix of error rates were reported by collapsing syllable types.

⁹ Greenhouse-Geisser values are used here for the tone pair and interaction effects because the assumption of sphericity was violated. The degree freedom is adjusted accordingly.

analysis, 15 One-Way ANOVA tests (15 AB tone pairs) with L1 group as a between-subject factor (3 levels), were conducted on the error rates of each AB pair.

For the Mandarin subjects, they had very low error rates for most of the pairs. Effect of L1 group was found for T5-T6 ($p < .001$), T2-T6 ($p = .008$), T3-T5 ($p = .05$), T4-T5 ($p < .05$), as illustrated in the top panel in Table 5. The Mandarin speakers had significantly lower error rates on these tone pairs than the English and French speakers. Moreover, few errors were found for the pairs with T4 among the Mandarin speakers. For instance, no error was found for the T1-T4 and T3-T4 pairs, and the error rates of T4-T6, T2-T4, and T4-T5 were the lowest among the three L1 groups.

Besides the difference, the three L1 groups had some error patterns in common, as shown in Table 5. On the one hand, the T2-T5 pair was the most difficult pair with the highest error rates for the three groups. A similar pattern was also found for the pairs of T3-T6. The T3-T6 pair had a higher error rates than the other level tone pairs, T1-T3 and T1-T6, for all the L1 groups. The difficulty of T2-T5 and T3-T6 in the two tasks can be attributed to the acoustic similarity of these tones. On the other hand, the tone pairs with T1 (e.g. T1-T2, T1-T3 etc.) had fewer errors than the other pairs for the three L1 groups. It can be explained by the clear acoustic distance of T1 from the other tones. The discrimination pattern of T1 is similar for the two different syllables (i.e. /jau/ and /se/), so it is not case that the subtle durational difference distinguishes T1 from other tones better for /jau/ than that for /se/. Thus, pitch pattern rather than duration pattern is responsible for the discrimination easiness of T1.

The shared and different patterns of individual tone pairs are found among the three L1 groups. While the three L1 groups had differences in specific pairs, they shared some error patterns as well.

5.3 Level versus contour tone pairs

The analysis of individual tone pairs suggested that the three L1 groups may differ in their sensitivity to different types of tone pairs. It was observed that, besides T2-T5, the Mandarin subjects had low error rates for most of the pairs. However, they did not find the T3-T6 pair easy to discriminate, resulting in the second highest error rates among all the pairs. On the contrary, although T3-T6 was not easy for the English and French groups to distinguish due to the acoustic similarity of the two tones, the pair was not relatively difficult compared with other tone pairs (e.g. T5-T6). Based on our observation, we assumed that the Mandarin group may find pairs of level tones more difficult than those of contour tones. On the contrary, the English and French groups may find pairs of contours tones more confusable than those of level tones.

To confirm our observation, the AB tone pairs were divided into two categories: Level tone pairs (T1-T3, T1-T6, and T3-T6) versus Contour tone pairs (the pairs including contour tones), to investigate the possible effects of tone pair types. Since the T2-T5 pair was highly confusable for all the L1 groups, it was excluded from this analysis.

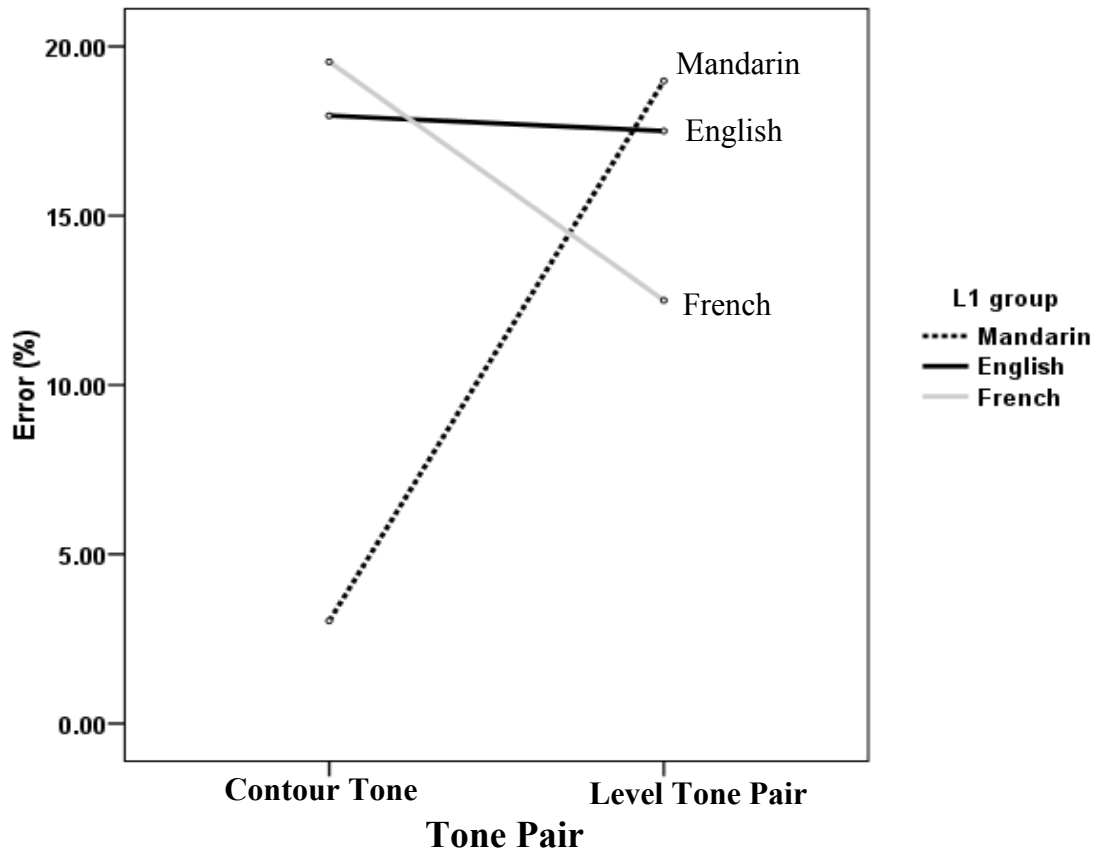


Figure 5: Average error rates of contour and level tone pairs by the three L1 groups.

Three paired t-tests (3 groups) were conducted for error rates of the different tone pair types. Figure 5 shows the average error rates for level and contour tone pairs across the three L1 groups. The Mandarin group had significantly more errors for the level tone pairs than the contour pairs ($p=0.002$); the English group appeared to have equally difficulty distinguishing the level and contour pairs ($p=0.948$); although no significant difference was found between the two types of tone pairs ($p=0.09$) for the French group, they distinguished the level tone pairs better than the contour pairs, supported by a large effect size ($d=0.52$).

Briefly, the Mandarin speakers differed from the English and French groups in their discrimination of the level and contour pairs. While the English group had equal difficulty distinguishing the two types of tone pairs, the French group appeared to distinguish the level tone pairs better than the contour tone pairs.

5.4 Individual differences

Despite the fact that the subjects shared L1 in each L1 group, not all of them had the same language background as discussed in 4.1. Some uncontrolled factors such as language varieties may exert influences on the subjects' performances and result in large individual differences of each L1 group. Thus, it is necessary to analyze the individual differences of each L1 group and the individual data of each subject.

First of all, we need to compare the individual differences of the three L1 groups by

referring to the data of standard errors in Figure 5 in section 4.1. It was found that the subjects in the English and French groups, who had more diverse linguistic experience, had larger individual variations than those in the Mandarin group, who had similar linguistic experience. Thus, data of standard errors demonstrated that the individual variations of the English and French groups were larger than those of the Mandarin group.

Since the English and French subjects came from different countries, they spoke different English or French varieties. The large individual variations may come from their dialects.

For the English speakers, we grouped the seven subjects from England and Australia on the one hand, and the three subjects from Canada and U.S., whose L1 is Northern American English, on the other hand. The two subgroups did not have a prominent difference with a difference of only 6% in terms of error rates.

Among the French speakers, one subject speaks Canadian French, which may be different from Standard French in suprasegmental domain (Kaminskaia, 2005). Moreover, one subject from Morocco is a bilingual speaker of French and Arabic. In order to explore the individual differences within the French speakers, we grouped the two subjects who differed from the others in their background (i.e. the Quebec French speaker and the French-Arabic speaker) at the one hand, and grouped the other French speakers at the other hand. The error rates of the first French subgroup is higher by 17% than those of the later French subgroup. It is difficult to draw a firm conclusion whether language varieties did influence the French speakers' perceptual performance here, as we had only two French speakers who speak different French varieties. Also, it is unclear if the difference of error rates is large enough to indicate they behave differently.

Although the English and French subjects came from different countries, different language varieties, at least for the English speakers, did not appear to exert a direct impact on the subjects' performance.

6. Discussion

6.1 Overall performance

Regarding the subjects' overall performance, the effect of L1 experience was found in discriminating non-native tones. The Mandarin group performed much better than the English and French groups in terms of error rates. It was attributed to the effect of L1 experience, that is, Mandarin speakers' experience of native tones.

6.2 Performance on individual tone pairs

Concerning performance on individual tone pairs, both L1 experience and psychoacoustic factors were found to affect the subjects' perception of Cantonese tones.

6.2.1 Psychoacoustic similarity of tone pairs

In the discrimination task, some tone pairs were easier or harder to discriminate than others. These asymmetrical patterns were shared by the three L1 groups. Among the 15 tone pairs, while the tone pairs of T2-T5 and T3-T6 were more difficult to discriminate than the other pairs, the tone pairs with T1 were easier to distinguish than the other pairs. The pattern can be explained by the psychoacoustic similarities of the stimuli.

First, the T2-T5 pair was the most confusable for all L1 groups, because T2 and T5 are acoustically similar and only differ in the magnitude of the final rising movement. Even native adult speakers found the T2-T5 pair confusable (Mok & Wong, 2010) and children also had difficulty in acquiring this pair (Ciocca & Lui, 2003). Second, among the level tones, the T3-T6 pair was more difficult to discriminate than the T1-3 and T1-6 pairs for all the L1 groups. Because the mid-level tone (T3) is further apart from T1 than from the low level tone (T6) in the acoustic space, there is a shorter acoustic distance between T3 and T6, which contributes to the relative difficulty of this pair among the level tone pairs. In addition, it is not surprising that all the subjects found the pairs with T1 easy to distinguish in both tasks, because T1 is well separated from the other tones in the acoustic space (see Figure 3).

Thus, the psychoacoustic similarity or dissimilarity of the two tones in each pair is one of the determining factors of perceptual difficulty in the discrimination task.

6.2.2 Effect of L1 experience

Although the three L1 groups shared some difficult and easy pairs due to the acoustic similarity of the tonal stimuli, they had different performance in discriminating individual tone pairs under the influence of their L1 experience.

First, the phonemic status of pitch differentiated the Mandarin group from the other two L1 groups. Contrary to the English and French groups who made more errors in most tone pairs, the Mandarin group found most of the tone pairs easy to distinguish in the task. The better performance of the Mandarin group in the T5-T6, T2-T6, T3-T5, and T4-T6 pairs can be explained by their L1 experience with Mandarin tones. For instance, T5 and T6, which are acoustically similar, were not difficult for the Mandarin group to discriminate. A rising pitch contour at the offset, which differentiates T5 from T6, is used in the Mandarin tonal inventory, so the Mandarin speakers were able to use this cue to distinguish the T5-T6 pair as well as other pairs, T2-T6 and T3-T5. Similarly, a falling pitch contour used in the Mandarin tonal inventory helped the Mandarin speakers discriminate T4 and T6, which differ in slight falling pitch contour at the offset.

However, how pitch is used in the native tonal inventory hindered the Mandarin speakers from distinguishing some Cantonese level tones. Although the acoustic similarity accounted for in part the difficulty of the T3-T6 pair, T3-T6, with the second highest error rates, was also regarded to be a difficult pair for the Mandarin group in part due to L1 experience. The finding is consistent with Francis et al. (2008) who found that Mandarin speakers often confused the three level tones. In their study, the Mandarin speakers misidentified T3 as T1 or T6 and T1 as T3 or T6 more frequently than English speakers. Since height differences among level tones are within-category differences, the Mandarin speakers have to give up some sensitivity to pitch height to facilitate categorical perception of tones (Leather, 1987; Stagray & Downs, 1993). Therefore, we found that the Mandarin speakers had a greater difficulty discriminating the level tone pairs than the contour tone pairs, because they were not sensitive enough to pitch height.

Different from the Mandarin speakers, the English and French speakers did not have lexical tones in their L1. On the one hand, their L1 experience hindered their perception of speech tones, especially pairs of contour tones such as T5-T6, T2-T6, T3-T5, and T4-T6. That is why much more pairs showed high error rates for the English and French speakers than the Mandarin speakers. On the other hand, the prosodic systems without the constraint of tonal categorization allowed the English and French speakers to detect the minor difference between the level tones.

Compared with contour tone pairs, the pairs of level tones such as T3-T6 were not highly difficult for them. We argue that the English and French speakers may perceive tones mainly relying on psychoacoustic aspects of the stimuli. For instance, they could hardly discriminate some similar tones such as T5-T6. T5 and T6 are acoustically similar as they share the same pitch onset.

There is no significant difference between English and French groups in tonal discrimination. The results disconfirm the prediction in Hallé et al. (2004) that French speakers may outperform English speakers because French prosody has “no constraint by lexical accentuation and stress patterns as English does”. Without the use of lexical tones in their L1, the English and French groups perceived tones in a similar way.

English stress system may explain why the English speakers did not outperform the French speakers. Besides suprasegmental cues to stress (e.g. pitch), vowel reduction is strongly correlated with stress placement. Thus, English speakers may not rely very much on pitch to recognize words (Cooper, Cutler, & Wales, 2002). Cooper, Cutler, and Wales (2002) found that the suprasegmental information of stress did not play a role in English as important as in Spanish and Dutch lexical access, although lexical stress was indeed found to be used by English speakers in an early stage to reduce the sets of words activated by the same study and shown to be important in speech perception by other studies (Tremblay, 2008). That may account for why English speakers did not do better than French speakers in discriminating tones. Moreover, since most of the French subjects in the present study learn English as L2, which could potentially explain why we did not find differences between the English and French speakers.

L1 experience indeed had an impact on the discrimination of non-native tones. The phonemic status of pitch contributed to the better overall performance of the Mandarin group than the other two L1 groups. Language-specific use of pitch in L1 can explain the difference among the three L1 groups in discriminating individual tone pairs and types of tones.

6.2.3 The two main factors

Although L1 experience and psychoacoustic factors were found to have an impact on the subjects' perceptual performance it may be hard to tease apart the two main factors in some cases. For instance, the Mandarin speakers found T3-T6 difficult to discriminate due to both of acoustic similarity and their native tonal system. Therefore, L1 experience and psychoacoustic factors may not always be independent from each other.

6.3 Revisiting PAM

The PAM predicts that non-native tones may be assimilated to native prosodic categories. Regarding the Mandarin speakers, as demonstrated in Table 6 below, they assimilated the Cantonese tones to their similar native tones, and the assimilation predicted by the PAM partially accounts for the Mandarin speakers' performance.

Cantonese T1 is phonetically similar to Mandarin T1 whereas the other two Cantonese levels, T3 and T6, are less similar to Mandarin T1. Thus, while T3-T6 is a pair of equally bad exemplars of Mandarin T1, T1-T3 and T1-T6 are pairs of one good exemplar and a bad exemplar of Mandarin T1. As shown in Table 6, the results show that the Mandarin speakers had difficulty discriminating T3-T6, but they found it easier to discriminate the pairs of T1-T3 and T1-T6 than the T3-T6 pair. The results are consistent with the Goodness-Category assimilation predicted by

the PAM. The Mandarin speakers found the T2-T5 pair (assumed to be Goodness-Category assimilation) difficult to discriminate. The results of the T2-T5 pair do not match well with our predictions. In addition, the Mandarin speakers found the tone pairs including T4 (assumed to be Two-Category assimilation) easy to discriminate. The results of T4 match with the PAM predictions.

| Tones pairs | PAM Predictions | Findings |
|---|---|---|
| T3-T6, T1-T3, T1-T6 | Goodness-Category (GC) assimilation; Discrimination from moderate to very good | T3-T6 poor discrimination; T1-T3 & T1-T6 good discrimination |
| T2-T5 | Goodness-Category (GC) assimilation; Discrimination from moderate to very good | Poor discrimination |
| Tone pairs including T4 (T1-T4, T2-T4, T3-T4, T4-T5, T4-T6) | Two-Category (TC) assimilation, Excellent discrimination | Excellent discrimination |

Table 6: Evaluation of the PAM predictions concerning the assimilation between Cantonese and Mandarin tones.

Although the PAM is good in predicting the assimilation between the Mandarin tones and Cantonese tones, a circularity problem is found for the predictions of the T3-T6 and T2-T5 pairs. It is also likely that the two pairs are Single-Category assimilation rather than Goodness-Category assimilation. Because T3 and T6 as well as T2 and T5 are acoustically similar, the Mandarin speakers may have assimilated both of the level tones to the native level tone and assimilated both of the two rising tone to the native rising tone. Assuming the pairs of T3-T6 and T2-T5 to be Single-Category assimilation matches with the PAM predictions too, as the two pairs were confusable for the Mandarin speakers. In this case, both Goodness-Category assimilation and Single-Category assimilation are possible for the two pairs. The PAM does not give an objective method to measure the similarity of native and non-native tones based on the articulatory gestures in order to choose the assimilation type for a given tonal contrast. We have to choose one assimilation type of tone pairs over another on the basis of the listeners' discrimination performance. Then, the process is more concerned with predicting the assimilation types on the basis of the listeners' discrimination performance rather than predicting the discrimination performance based on the assimilation types. In this sense, the PAM becomes less powerful as it has a problem of circularity in predicting assimilation types and subjects' discrimination performance.

Regarding the English and French speakers, there was no evidence supporting the assimilation between Cantonese tones and English/French prosodic categories (i.e. intonation and stress

contours). The results agree with our prediction that the two L1 groups do not assimilate non-native tones to L1 prosodic categories and perceived tones mainly relying on the psychoacoustic aspects of the stimuli. Therefore, it is unlikely for speakers of non-tone languages to assimilate non-native tones to any L1 phonemic unit, because lexical tones are absent in their L1.

In the PAM terms, the tonal contrasts may be Non-Assailable (NA) to any speech sounds or perceived as Uncategorized-Uncategorized (UU) contrasts by the English and French speakers (Hallé et al., 2004). The English and French speakers did perceive tones as speech sounds, as pitch is used in their linguistic system. The possibility of Non-Assailable case is rejected. Therefore, Cantonese tones are just not categorized by the English or French speakers to any native prosodic category and the Uncategorized-Uncategorized assimilation is proposed to account for the perception of tones by speakers of non-tone languages.

7. Conclusion

To sum up, both psychoacoustic similarity of the tonal stimuli and L1 experience affected the perception of Cantonese tones by non-native speakers.

Due to psychoacoustic distance of the tonal stimuli, the subjects found the pairs with T1 easier to distinguish than the other tones and the pairs of T2-T5, T3-T6, or T5-T6 confusable.

L1 experience was found to differentiate the performance of the different L1 groups. While the phonemic status of tones in L1 helped the Mandarin speakers outperform the other two L1 groups, language-specific use of pitch in L1 caused the difference of the three L1 groups in discriminating specific tone pairs. The level tone pairs (e.g. T3-T6) were more difficult than contour tone pairs for the Mandarin subjects, but some contour tone pairs (e.g. T5-T6) were more difficult than level tone pairs for the English and French subjects.

References

- Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. New York: Mouton de Greyter.
- Beckman, M. E. (1986). *Stress and Non-stress Accent*. Dordrecht: Foris.
- Bent, T., Bradlow, A. R., & Wright, B. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and non-speech sounds. *Journal of Experimental Psychology: Human Perception and Performance*. 32(1), 97-103.
- Bent, T. (2005). *The perception and production of non-native prosodic categories*. Unpublished Ph.D. thesis, Department of Linguistics, Northwestern University, Evanston, IL.
- Best, C. T. (1995). A direct realistic view of cross-language speech perception. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). Baltimore: York Press.
- Best, C. T., & Tyler M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins.
- Blum V. (1999). Intonation in English, Finnish and French: ethology versus typology. *International Congress of Phonetic Science*, San Francisco.
- Brown, C. (1998). The role of L1 grammar in the L2 acquisition of segmental structure. *Second*

- Language Research* 14, 136-193.
- Burnham, D. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7, 207-239.
- Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.-S Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp.259-280). Amsterdam: John Benjamins.
- Chao, Y. R. (1947). *Cantonese primer*. Cambridge: Cambridge University Press.
- Chao, Y-R (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.
- Chen, Y., & Xu, Y. (2006). Production of weak elements in speech—Evidence from F0 patterns of neutral tone in Standard Chinese. *Phonetica*, 63, 47–75.
- Ciocca, V., & Lui, J., (2003). The development of lexical toneperception in Cantonese. *Journal of Multilingual Communication Disorders*, 1:141-147.
- Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech*, 45(3), 207-228.
- Cutler, A. (2005). Lexical stress. In D. B. Pisoni, & R. E. Remez (Eds.). *The handbook of speech perception* (pp. 264-289). Oxford: Blackwell.
- Delattre, P. (1951). *Principes de phonétique française*. Middlebury: Middlebury College.
- Di Cristo, A. (1998). Intonation in French. In D. Hirst, & A. Di Cristo (Eds.), *Intonation systems: A survey of twenty languages* (pp. 195–218). Cambridge: Cambridge University Press.
- Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing "deafness" in French? *Journal of Memory and Language*, 36, 406-421.
- Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, 106(2), 682-706.
- Flege, J. E. (1987). The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, 15(1), 47–65.
- Flege, J. E. (1995). Second-language speech learning: theory, findings, and problems. In W. Strange (Eds), *Speech perception and linguistic experience: Issues in cross-language research* (pp.233-273). Baltimore: York Press.
- Flege, J. E., & Hillenbrand (1986). Differential use of temporal cues to the /s/-/z/ contrast by native and non-native speakers of English. *Journal of the Acoustical Society of America*, 79, 508-517.
- Fok Chan, Y. Y. (1974). *A perceptual study of tones in Cantonese*. Hong Kong: Centre of Asian Studies, University of Hong Kong.
- 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, 268-294.
- Francis, A. L., Ciocca, V., & Ng, B.K.C. (2003). On the (non)categorical perception of Cantonese lexical tones. *Perception and Psychophysics*, 65(6), 1029-1044.
- Frost, D. (2011) Stress cues in English and French: a perceptual study. *Journal of the International Phonetic Association*. 41(01): 67-84.
- Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America*, 27(4), 765–768.
- Gandour, J. T. (1983). Tone perception in far Eastern languages. *Journal of Phonetics*, 11, 149–175.
- Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University Press.
- Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin

- Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics*, 32, 395–421.
- Hancin-Bhatt, B., & A. Govindjee. (1999). A computational model of feature competition in L2 transfer. P Broeder, & J Murre (Eds.), *Language and thought in development: Cross-linguistic studies* (pp. 145-161). Gunter Narr: Tuebingen.
- Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.
- Huang, T. (2004). *Language-specificity in auditory perception of Chinese tones*. Unpublished Ph.D. thesis, Ohio State University.
- Iverson, P., Kuhl, P. K., Akahane-yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87, B47-B57.
- Jun, S.A. (Ed.). (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford University Press.
- Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, 35, 104–117.
- Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp.121-154). Baltimore: York Press.
- Lado, R. (1957). *Linguistics across cultures*. Ann Arbor, MI: The University of Michigan Press.
- Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 59-81). Dordrecht: Foris Publications.
- Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, 25, 527–542.
- Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.
- Lin, T., & Wang, W. S.-Y. (1984). The perception of tones. *Journal of Chinese Linguistics*, 2, 59-69.
- Mok, P., Zuo, D. & Wong, P. (2013) Production and perception of a sound change in progress: tone merging in Hong Kong Cantonese. *Language Variation and Change*, 25: 341-370.
- Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics*, 34(1): 134-154.
- Peng, G., Zheng, H-Y., Gong, T., Yang, R-X., Kong, J-P., & Wang, W. S-Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics*, 38, 616-624.
- Polka, L. (1991). Cross-language speech perception in adults: Phonemic, phonetic and acoustic contributions. *Journal of the Acoustical Society of America*, 89, 2961–2977.
- Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, 21, 143-163.
- So, C. K. (2010). Cross-language Perception of Non-native Tonal Contrasts: Effects of Native Phonological and Phonetic Influences. *Language and Speech*, 53(2), 273-293.
- Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics*, 29, 553–584.
- Vaissière, J. (1991). Rhythm, accentuation, and final lengthening in French. In J. Sundberg, & R. Carlson (Eds.), *Music, language, speech, and brain* (pp. 108–120). New York: Macmillan Press.

- Vaissière, J. (2002). Cross-linguistic prosodic transcription: French versus English. *Problemy i metody eksperimental'no-foneticheskikh issledovanij, In honour of the 70th anniversary of Prof.L.V. Bondarko., N. B. Volskaya, N. D. Svetozarova and P. A. Skrelin.* (pp.147-164). St. Petersburg, St.-Petersburg State University.
- Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences*, 28, 61–72.
- Wang, Y., Jongman, A., & Sereno, J. (2001). Dichotic perception of linguistic tones by Mandarin and American listeners. *Brain and Language*, 78, 332–348.
- Wang, Y., Behne, D. M., Jongman, A., & Sereno, J. (2004). The role of linguistic experience in the hemispheric processing of lexical tones. *Applied Psycholinguistics*, 25, 449–466.
- Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning*, 54, 681–712.
- Welby, P. (2006). French intonational structure: Evidence from tonal alignment. *Journal of Phonetics*, 34(3), 343–371.
- Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America*, 75, 1866-1878.
- Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception and Psychophysics*, 37(1), 35-44.
- Xu, Y., Gandour, J., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, 120(2), 1063–1074.

Author contact information:

Zhen Qin: qinzhenquentin2@ku.edu
Peggy P. K. Mok: peggymok@cuhk.edu.hk



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).