Title	Perception of linguistic pitch in Cantonese-English bilingual speakers
Author(s)	Lee, Mei-ling; 李美玲
Citation	Lee, M. [李美玲]. (2012). Perception of linguistic pitch in Cantonese-English bilingual speakers. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.
Issued Date	2012
URL	http://hdl.handle.net/10722/237875
Rights	The author retains all proprietary rights, (such as patent rights) and the right to use in future works.; This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Perception of linguistic pitch in Cantonese-English bilingual speakers

LEE Mei Ling, Meg

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2012.

Abstract

This study investigated the perception of linguistic pitch in Cantonese-English bilingual speakers. Two age groups of bilingual speakers (age 10-13 and age 20-23) were recruited and attended the experiment which included three sessions: Cantonese lexical tone perception test, English lexical stress perception test, the mixed mode test of lexical tone and lexical stress. The stimuli were fifty-five CVCV non-words with fifteen different tone contrasts in lexical tone test, and with initial stressed or final stressed in lexical stress test. The results showed that adults performed significantly better than the children in both lexical tone and lexical stress perception. Children had similar performance in tone and stress perception but adults performed better in stress perception than tone perception. The tone contrasts of mid level-low level, and high rising-low rising were the most difficult to discriminate for both children and adults. Findings suggest that bilingual speakers used the same perception strategy in perceiving linguistic pitch of lexical stress and lexical tone and the language experience still affected and improved the suprasegmental acquisition after age of 10-13.

Keywords: Speech prosody, lexical tones, lexical stress, bilingualism

Introduction

Various cross-linguistic studies have shown that the perception of the second language (L2) at either the segmental or suprasegmental level is affected by the language systems of the first language (L1) to a various degree (Best, McRoberts, & Goodell, 2001; Best & Tyler, 2007; Francis, Cicca, Ma, & Fenn, 2008; Gandour, Wong, & Hutchins, 1998; Krishnan, Xu, Gandour, & Gariani, 2005; Lee, Vakoch, & Wurm, 1996; Zhang, Nissen, & Francis, 2008), depending on the similarities and differences in their sound systems. At the segmental level, the phonological (i.e. organization of speech sounds to form meaningful words) and phonetic (i.e. forms of speech sounds) differences of contrasting consonants and vowels are taken into account in the analysis of perception and discrimination of the L2 sounds (Best & Tyler, 2007). At the suprasegmental level, there is compelling evidence that prosodic categories (i.e. stress, rhythm, fundamental frequency, intensity, duration and pauses, etc.) and intonation of native language experience affects the perception and acquisition of non-native suprasegmental knowledge (Francis et al., 2008). Noted that these studies have been largely focused on perception of speech sounds in monolingual speaker, surprisingly, there is little evidence of the perception of linguistic pitch, in particular, lexical tones and lexical stresses in bilingual speakers. Our study addresses this issue by examining the perception of lexical tones and lexical stresses in Cantonese-English bilingual speakers who are learning Cantonese and English simultaneously.

The Perceptual Assimilation Model proposed by Best (1994) described possible cross-language sound category assimilation patterns depending on phonetic deviation of sound systems in L1 and L2, and hence predicting the difficulty of learning L2 sounds faced by native speakers. The model suggests that the relative difficulty in discriminating the L2 phonemes in particular sound category by the native speakers is constrained by their

language experience, and it highly depends on the degree of the deviation (phonological and phonetic) between the target sound and those of the listener's native sound system. Also, the model suggests that native listeners find it most difficult to discriminate the L2 sounds which assimilate into one native phoneme, because the L2 sounds have equal phonetic similarities from the assimilated native phoneme. For example, Japanese tend to assimilate English /l/ and /r/ into one single Japanese phonemes /r/ (Best & Strange; Takagi & Mann; Yamada & Tohkura as cited in Best et al., 2001), and the English /l/ and /r/ appear to have the same degree of phonetic deviant from the Japanese phoneme /r/ (this is called Single Category Assimilation as cited in Best, 1994), hence the discrimination of the L2 phonemes (/l/ and /r/) was very difficult for the native Japanese. For Cantonese speakers of English, their perception and acquisition of vowels of English was also limited by their native Cantonese phonetic inventory. Previous research found that they had a simpler vowel system than native speakers of English as Cantonese speakers tend to neutralize certain vowel contrasts that only exist in English but not in Cantonese (Bolton & Kwok as cited in Hung, 2000). The vowel contrasts which have very similar phonetic features would be assimilated to one vowel which exists in the native language, therefore the Cantonese speakers of English find it difficult to discriminate different vowels in L2 which have the same degree of phonetic deviant from the L1 vowel.

Studies on the perception of the L2 have been extended from the segmental level to suprasegmental features in recent years, such as tonal language versus non-tonal (word stress) language (Cutler & Chen, 1997; Francis et al., 2008; Gandour et al., 1998; Krishnan et al., 2005; Lee et al., 1996; Qin & Mok, 2011; Zhang et al., 2008). Moreover, recent research has made significant progress by showing that perception of linguistic pitch in L2 was also affected by the prosodic cues experienced in native language (e.g., Francis et al.,

2008; Krishnan et al., 2005). For example, Francis et al. (2008) found a significant difference between native-speaking tonal language group (Mandarin Chinese) and native speaking non-tonal language group (English) on Cantonese lexical tone identification after training to recognize the Cantonese lexical tones. It suggested listener's native language experience in suprasegmental level affected the perception and acquisition of non-native suprasegmental categories. Similarly, a study using Positron emission tomography (PET) by Gandour, Wong and Hutchins (1998) showed that tonal language speaking group (Thai) had brain activation in the left frontal operculum (a region near Broca's area) when discrimination thee pitch patterns in Thai words, but there was no similar activation in the brain was observed in non-tonal language speaking group (English). The authors proposed that the pitch processing mechanisms in the brain is language-specific and affected by native language experience.

What does it happen in Cantonese lexical tones? Cantonese is a tonal language, in which one lexical carries one distinctive tone which conveys different semantic from other lexical with different tones. There are six contrastive lexical tones in Cantonese (Bauer & Benedict, 1997), namely: high level (Tone 1, 55[1]), high rising (Tone 2, 25), mid level (Tone 3, 33), low falling (Tone 4, 21), low rising (Tone 5, 23) and low level (Tone 6, 22). For example, a monosyllable /fu/ can be in six contrastive tones and each lexical represents different meanings, /fu1/膚(skin), /fu2/虎(tiger), /fu3/褲(trouser), /fu4/符(symbol), /fu5/婦(woman), /fu6/父(father). Different tones are different in fundamental frequency, pitch contour and pitch height (as shown in Figure 1). The pitch height can be grouped into high, middle, low, and contour patterns include level, rising and falling (Fok Chan as cited in Khouw & Ciocca, 2007). Although Fok Chan also argued that fundamental frequency is the

[1] The numbers are the pitch values given to the tones to describe their relative onset and offset frequency , as well as their pitch contour (Bauer & Benedict, 1997, p113.).

primary acoustic correlate of determining the lexical tone (as cited in Chan, 2008), previous research found that confusion of perceiving similar tone contrasts still exists in Cantonese speakers. For example, high rising (Tone 2, 25) and low rising (Tone 5, 23) have very similar pitch contour as they have the same pitch onset point but have slightly different pitch offset point (25 vs 23), resulting in higher chance of perceiving tone 2 to tone 5 or vice versa in Cantonese speakers (Khouw & Ciocca, 2007).

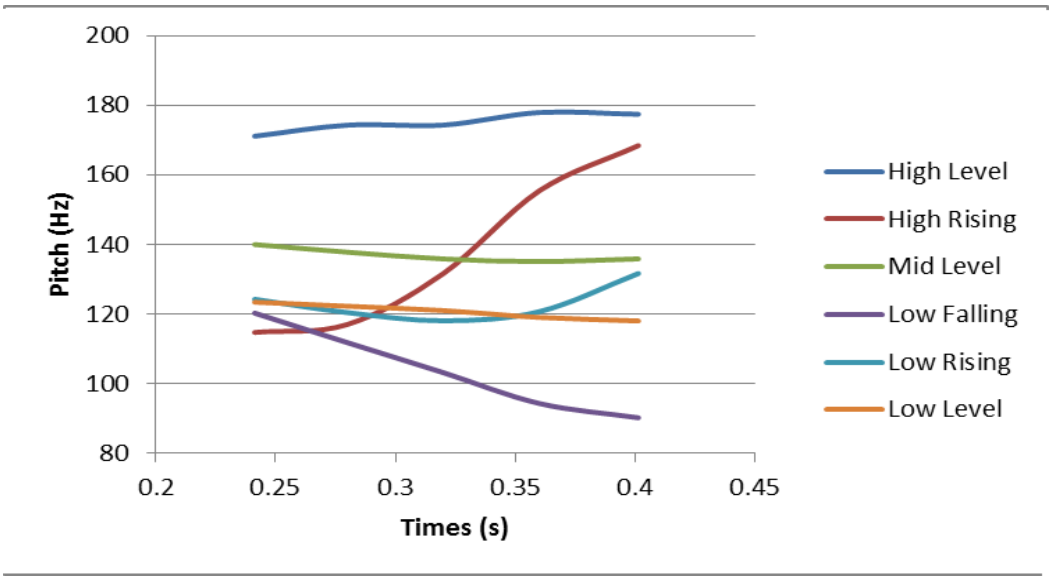

Figure 1. Pitch contour lines of 6 contrastive tones on monosyllabic word /fu/ produced by a native male Cantonese speaker.

Previous studies showed that native Cantonese children have already acquired the lexical tone as early as before age 2 (So & Dodd, 1995) but they can only identify Cantonese lexical tone as well as adults when they are around 10-11 years (Ciocca & Lui, 2003). Ciocca and Lui (2003) conducted a perception test on Cantonese lexical tones on four age groups of 4, 6, 10 and adults. The subjects would listen to a Cantonese carrier phrase which carried a target tone and they were asked to choose one out of two pictures which was represented the

target meaning. Results showed that children at age of 10 and adults performed significantly better than children aged 4 and 6, and there was no significant difference between the performance of children aged 10-11 and adults. Findings suggested that children could reach adults' performance in discriminating lexical tone when they are at the age of 10-11.

In non-tonal language, such as English, stress is a linguistic prosody on different position of a word in which determines the meaning of the word (e.g. reCORD vs REcord). Acquisition of suprasegmental knowledge was shown to be earlier than complete segmental acquisition. Studies have shown that English infants as young as 2 month old are already able to discriminate different stress patterns in pseudo-words (Jusczyk & Thompson as cited in Skoruppa et al., 2009) . This finding suggested that the young infants were already able to discriminate stressed from unstressed pattern due to language exposure to suprasegmental experience since born. The stressed and unstressed syllables are perceived differently in terms of acoustic properties of fundamental frequency, intensity and duration (Cutler & Chen, 1997). However, previous research found that some acoustic parameters are more important than the others in perceiving the stressed syllables from unstressed syllables. Sluijter and van Heuven (1996) carried out a production study in which ten speakers produced Dutch lexicals and stress pairs. Duration, overall intensity, formant frequencies and spectral levels in four frequency bands were measured. Spectral balance is the distribution of intensity over different frequency bands (Sluijter & van Heuven, 1996). The results showed that stressed syllables are produced with stronger vocal effort resulting in more intensity increased on higher frequencies than lower frequencies (spectral balance), and the findings suggested that spectral balance and duration are strong and reliable acoustic correlates to discriminate stressed syllable from unstressed syllable in Dutch but overall intensity and vowel quality were the poor cues, and the conclusion was extended to English. There was

another research showing the same findings. Chan (2008) conducted a lexical stress perception experiment on native Cantonese speakers and native English speakers by manipulating the acoustic parameters of duration, fundamental frequency and spectral balance, the results showed that native Cantonese speakers used fundamental frequency as primary prosodic cue in stress perception (as the way they use for tone perception) while native English speakers utilized spectral balance as primary cue instead. The findings suggested that spectral balance is an important acoustic cue for stress perception and the native language experience will affect the perception strategy used to perceive non-native linguistic pitch.

It has been an interesting issue for researchers that whether bilinguals are able to use a systematic way to differential two languages in lexical and semantic aspects. According to Cenoz and Genesee (2001, p.95), there are three stages that bilingual children involved to encode lexicons in both language. Cenoz and Genesee (2001) stated that in the first stage, bilingual child begins with one lexical system to encode lexicons in both languages. In the second stage, the child starts to apply the same syntactic rules to both language but distinguishes two different lexicons. Finally, the young bilingual speakers are able to apply two separate linguistic strategies in coding both lexicon and syntax. However, there is little research on the bilingual children's language development on suprasegmental level that whether they also use the same perception strategies or two separate perception systems in later stage as what they have for the lexicon development in both languages. Although there is evidence showing the effect of language experience on perception of suprasegmental level of speech as previously discussed (e.g., Francis et al., 2008; Krishnan et al., 2005), fewer studies have investigated the perception of suprasegmental features in bilingual speakers and the comparison between different bilingual age groups. Furthermore, there is little evidence

of the mutual effect or relationship of lexical tone and lexical stress on each other in bilingual speakers.

The present study will fill in the previous research gap by focusing on the perception of linguistic pitch in bilingual speakers and extend the previous research of Ciocca and Lui (2003) that if bilingual children can also achieve the adult performance in tone and stress perception at age of 11-13. Two primary research questions are addressed: 1) whether there is a difference between the perception of lexical tone and lexical stress in bilingual speakers; 2) whether bilingual sparker's perceptual abilities in lexical tone and lexical stress still improve after age of 10-13 due to language maturity.

Method

Participants

Two groups of participants were recruited. For the adult group (n=30), they were all undergraduate students of The University of Hong Kong who haven't studied linguistic before and didn't have any hearing impairment (Mean age =21.7, Age range = 20-23, male =19, female = 11). For the children group (n=30), they were all studying in international schools or English as teaching medium schools who haven't had any knowledge about linguistic either and didn't have any hearing impairment (Mean age =11.7, Age range =10-13, male =10, female = 20). The adult participants and the children's parents had rated the participants' understanding and speaking of Cantonese and English in a 5-point scale (1 means poor and 5 means excellent), as well as on the Cantonese-English bilingualism level (1 means speaking predominantly on one language and 5 means fluently speaking two languages) in a questionnaire (see appendix). The following table has summarized the average rating of participants' abilities of understanding and speaking of Cantonese, English and the level of Cantonese-English bilingualism.

Table 1.

The Average Rating and Standard Deviation on a 5-point Scale of Participants' Understanding Ability, Speaking Ability and Bilingualism Level of Cantonese and English.

	Understanding ability		Speaking ability		Overall self-rating
	Cantonese	English	Cantonese	English	Bilingualism level
Children	4.03(0.93)	4.33(0.80)	3.80(1.32)	4.36(0.72)	3.80(0.87)
Adults	4.73(0.52)	3.63(0.93)	4.67(0.48)	3.43(0.97)	3.9(0.55)

Materials

CVCV sounding sequence was constructed because CV structure is common in Cantonese and is the simplest phonological structure in both Cantonese and English. The CVCV structure has been successfully used to assess the perception of suprasegmental speech in previous research (e.g., Skoruppa et al., 2009). Also, to avoid the confounding effect of intrinsic knowledge of pitch, loudness and duration of vowels in the real lexical words on the perception linguistic pitch (Sluijter & van Heuven as cited in Chan, 2008), these CVCV sound sequences are created as pseudo-words and none of them were real words in either Cantonese or English.

There are 55 CVCV pseudo-words which are created by combining 11 consonants /b, p, d, t, g, k, m, l, h, s, f/ and five vowels /i, u, ɛ, ɑ , ɔ/. There were several reasons for choosing these consonants and vowels.

First, they are not clusters as clusters are less common phonemes in both Cantonese and English. Second, they are pronounced the same (or similar phonetic realization) in both Cantonese and English, as well as most commonly used initial consonants in both languages so that the unfamiliarity in the segmental features exert the least influence on the

participants' perception. Third, there is no variation in their pronunciation across different syllabic contexts. Fourth, According to the norms on the Hong Kong Cantonese Articulation Test (Tsou et al., 2006), all our subjects (age group 10-13 year old and 20-23 year old) have acquired the above consonants before school age. Fifth, So and Dodd have found that " the order of acquisition of (Cantonese)consonants was similar to that reported for English, the rate of acquisition was more rapid" (1995), therefore, the consonants chose both exist and are acquired in similar order by the Cantonese-speaking and English-speaking children.

Consonants /l/ and /n/ were not included. According to Bauer and Benedict (1997, p.329), in the recent decades Hong Kong people have merged the initial consonants /n-/ and /l-/ into /l-/ (especially for people younger than age 45). Therefore, for some younger generations, they may not be familiar with the initial consonant /n/ or it doesn't even exist for them. In that case, /n/ is not included in this experiment.

According to Cheung (2000), there are 7 long vowels in CV structure in the analysis of Cantonese vowels, they are /i, u, ɛ, ɑ, ɔ/ which are primary vowels and /y, œ/ which are secondary vowels. The 5 primary vowels (but not the secondary vowels) in Cantonese also exist in English phonology, therefore the 5 primary vowels were used in this experiment. The /i, u, ɛ, ɑ, ɔ/ in Yue Pin representations are /i, u, e, a, o/.

The 55 pseudo-words are the stimuli for both Cantonese tone perception test and English stress test. We described each test in turn.

Cantonese tone perception test

There were two types of Cantonese stimuli. They had the same combination of consonants and vowels, but only differed in the position of constant tone. In the first type, the first syllable (CV) for all stimuli was constant and the second syllable was with different tones (Tone 1 to Tone 6). In the second type, the second syllable was constant while the first

syllable was with other tone. Tone 1 (high level) was used as constant tone because it had the highest frequent occurrence in Cantonese dialect (Ng & Kwok, 2004), the recorder and the participants would be most familiar with this tone and it would exert least influence for the recorder's production in the recording and the participants' perception in the experiment.

A total of 15 tone contrasts were constructed with 6 tones (for detail, see Table 2). These 15 tone contrasts were carried by 55 CVCV sequence. A balanced *latin-square design* (see table 3) was used to control the variation of tone contrasts and CVCV structure. The 15 tone contrasts and 55 CVCV sequences were both randomly grouped into 5 blocks, each block comprised of 11CVCV carrying 3 tone contrasts. With this design, we generated 5 versions of stimuli, each having 165 sets of stimuli (3 tone contrasts ×11 CVCV sequences ×5 blocks). This design enabled each version included all 55 stimuli and 15 tone contrasts. In addition, to counterbalance the order of the position of the constant tone, two types of Cantonese stimuli (1st syllable with the constant tone; 2nd syllable with the constant tone) was used. Thus, a total of 10 versions of Cantonese stimuli were used.

Table 2.

The Matrix of 15 Tone Contrasts

	High level	High rising	Middle level	Low falling	Low rising	Low level
High level	-	-	-	-	-	-
High rising	1	-	-	-	-	-
Middle level	2	6	-	-	-	-
Low falling	3	7	10	-	-	-
Low rising	4	8	11	13	-	-
Low level	5	9	12	14	15	-

Table 3.

Latin-square design of Tone Contrast Blocks and CVCV Blocks

Version	Blocks				
	Ⅰ*	Ⅱ	Ⅲ	Ⅳ	Ⅴ
1	A#	B	C	D	E
2	B	C	D	E	A
3	C	D	E	A	B
4	D	E	A	B	C
5	E	A	B	C	D

Note. ***** Ⅰ, Ⅱ, Ⅲ, Ⅳ, Ⅴ, are the blocks of tone contrasts with each block have 3 tone contrasts. #A, B, C, D, E are the blocks of CVCV sequences with each block have 11 CVCV sequences.

English stress perception test

The combination of consonants and vowels was the same as those used in the Cantonese tone test. The CVCV non-words were initial-stressed or final stressed. A total of 110 sets of stimuli were generated (55 CVCV X 2 stressed conditions).

Recording and Segmenting

A fluent female Cantonese–English bilingual speaker was recruited. She was born in Hong Kong but had lived in an English-speaking country for 10 years. She had basic phonetic training and linguistic background. She was informed about the background of the experiment and aware that the items recorded would be used as stimuli in the experiment.

All stimuli were recorded in a sound proof booth at the University of Hong Kong using phonetic analysis software Praat (Boersma & Weenink, 2012) with 41, 000 HZ at mono tone setting. The tone and stress stimuli were presented to the speaker separately in

slides with each slide showing one stimulus. The speaker was asked to count '1, 2' after production of each stimuli and the counting was to let the speaker get ready for the next stimuli. The pacing of slide shows was controlled by the experimenter. The whole set of tone and stress stimuli were produced 3 times and in each time the order of stimuli in tone or stress recording was randomly presented to the speaker. The recording of tone and stress was done alternately for each time. The practice trials were provided before the recording. The speaker was asked to produce the stimuli at her comfortable tone level and produce the constant tone (Tone 1) for all stimuli to be perceived constantly. A total of 2310 stimuli were recorded (55 CVCV X 6 tones X 2 types X 3 repetition + 55 CVCV X 2 stressed conditions X 3 repetition = 1980 + 330 = 2310). The recordings were done on two separate days.

All sound editing was done using Praat (Boersma & Weenink, 2012) by inspecting the spectrogram and waveform of the stimuli at the same time. The stimuli with disfluency, unclear, unstable loudness and too long or too short duration (longer than 1000 ms or shorter than 700 ms) were excluded. The acoustic parameters of duration and intensity of the syllables were measured using Praat. A total of 770 stimuli were selected (55 CVCV X 2 types X 6 tones + 55 CVCV X 2 stressed conditions). The following table summarized the average duration and average intensity of the selected stimuli.

Table 4.

Average Duration (s) and Intensity (dB) of Stimuli with Stress.

	Duration	Intensity
Initial stressed	0.78	63.13
Final stressed	0.80	62.82

Table 5.

Average Duration (s) and Intensity (dB) of Stimuli with Cantonese Tones.

	1st Syllable Constant		2nd Syllable Constant	
	Duration	Intensity	Duration	Intensity
Tone 1	0.80	63.60	0.80	63.60
Tone 2	0.92	64.50	0.93	63.28
Tone 3	0.83	63.46	0.85	64.73
Tone 4	0.90	63.88	0.82	64.45
Tone 5	0.95	63.00	0.87	63.85
Tone 6	0.88	66.55	0.81	63.58

Procedures

We used the AXB paradigm which has been widely used in previous research on speech perception across both adults and children (Best et al., 2001; Fullana & Mora, 2007; Levy & Strange, 2008; Sawusch & Gagnon, 1995; van Hessen & Schouten, 1999; Yeon, Wayland, Harnsberger, & Silver, 2004) . There were 4 orders of AXB (AAB, ABB, BAA, BBA) in term of target position and choice position, in order to avoid any bias of choice due to positions, 4 orders of choice and target arrangement were included in the test and were randomly assigned to the stimuli. The experiment was carried out in a sound proof room at the University of Hong Kong for adults and in a silent classroom or place inside the school for children. In the tasks, AXB stimuli were played by a laptop computer and automatically run by the Java program with a randomized order. The participants heard three audio non-word stimuli in a set (A, X and B), which differed only in tone or stress, via headphones connected to the laptop computer. The participants were then asked to judge whether the second stimulus sounded more like (or the same as) the first stimulus or the last stimulus by

respectively pressing the ‘1’ or ‘3’ on the laptop keyboard to indicate the answer. Written instructions and practice trials were provided before the experiment. There were 3 sections in each version of testing stimuli. Each participant would do on one version, in which they responded to one version of Cantonese stimuli (165 sets), followed by 110 sets of English stress stimuli, and then the mixed mode of Cantonese tone and English stress stimuli (165+110=275 sets). The participants were given 1 min for break between each section. The accuracy and reaction time would be automatically recorded by the program.

The interstimulus interval was 300 ms and the response time was 3000 ms (used by van Hessen & Schouten, 1999) . The inter-trial interval was 400 ms. The rationale for choosing 300ms as interstimuli interval and 400ms as inter-trial interval lied in the consideration of the auditory span and auditory processing time. Gerrits and Schouten (2004) stated in their study that the auditory span was 200ms to 300 ms and the processing time of 250ms was sufficient for recognizing a speech signal, therefore the inter-trial interval of 400 ms would be enough to let the auditory features of previous trial stiumuli fade out and 300ms for interstimulus would be adequate for processing the auditory features of the heard stimulus.

Result

Analyses were conducted on both response accuracy and reaction time on correctly answered items. The mean response accuracy and mean reaction time were calculated. Table 4 and Table 5 presented mean percentage accuracy and reaction time of Cantonese lexical tones and English lexical stress for children and adults under different modes respectively.

Table 6.

Mean Percentage Correct and Standard Deviations of Tone Perception and Stress Perception in Single and Mixed Modes by Participants.

	Single Mode		Mixed Mode	
	Tone	Stress	Tone	Stress
Children	92(5)	91(6)	88(7)	89(7)
Adults	95(5)	98(2)	96(3)	98(4)

Table 7.

Mean Reaction Time (ms) and Standard Deviations of Tone Perception and Stress Perception in Single and Mixed Modes by Participants

	Single Mode		Mixed Mode	
	Tone	Stress	Tone	Stress
Children	540 (180)	550(280)	560(280)	550(320)
Adults	420(130)	340(130)	340(140)	360(160)

Perception in lexical tone and lexical stress

To examine the perceptual difference of lexical tone and lexical stress by children and adults under different modes, we conducted a 2(linguistic pitch perception task: lexical tone and lexical stress) X 2 (Mode: Single vs. Mixed) X 2(group: Children vs. adults) mixed factorial ANOVA on both response accuracy and reaction time. Here, linguistic pitch and mode were within-subject variables and group was a between-subject variable. The dependent variables were the mean percentage accuracy and reaction time on correctly answered items respectively.

In terms of accuracy, there was significant main effects of linguistic pitch (lexical tones and lexical stress), $F(1, 58)=6.86$, $p<0.05$, $\eta_p^2=0.11$, with a higher score on lexical stress perception. Also, the main effect of mode was significant, $F(1, 58)=7.02$, $p<0.05$, $\eta_p^2=0.11$, as the performance in single mode was better than the mixed mode. The main effect of the group was significant, $F(1, 58)=42.87$, $p<0.05$, $\eta_p^2=0.43$, in which the adults generally performed better than the children.

There was a significant two-way interaction of mode and group, $F(1, 58)=12.21$, $p<0.05$, $\eta_p^2=0.17$ and linguistic pitch x group was marginal significant, $F(1,58)=3.49$, $p=0.07$, $\eta_p^2=0.06$. A three-way interaction of linguistic pitch, mode and group was also significant, $F(1,58)=6.58$, $p<0.05$, $\eta_p^2=0.10$.

The simple main effect analysis was conducted to examine the perceptual difference of linguistic pitch in Cantonese and English, separately for children and adults. For children, the main effect of mode was significant, with higher scores in single mode relative to the mixed mode, $F(1,29)=12.49$, $p<0.05$, $\eta_p^2=0.30$. But the main effect of linguistic pitch was not significant, $F(1, 29)=0.27$, $p=0.61$, showing similar performance in lexical tone and lexical stress. There was no significant interaction between linguistic pitch (lexical tone vs lexical stress) and mode (single vs. mixed), $F(1,29)=2.24$, $p=0.15$, $\eta_p^2=0.07$.

For the adults, a significant difference was found between tone and stress perception, $F(1, 29)=10.52$, $p<0.05$, $\eta_p^2=0.27$. However, the single and mixed mode condition didn't show any significant effect as the adults had similar performance in two different modes, $F(1, 29)=0.73$, $p=0.40$, $\eta_p^2=0.03$. The two way interaction of tasks and modes was significant, $F(1, 29)=5.00$, $p<0.05$, $\eta_p^2=0.15$. The paired sample T-test was used to analyze the effect of tasks under two modes. In single mode, adults scored better in stress perception

than tone perception, $T(29)=-3.49$, $p<0.05$. In mixed mode, adults had similar performance in tone and stress perception, $T(29)=-1.15$, $p=0.26$.

In order to find the effect of groups on the accuracy in linguistic pitch under different modes, simple main effect analysis was conducted separately in single and mixed mode. In single mode, the main effect of group was significant, the adult participants significantly did better than children participants both in lexical tone perception, $F(1, 58)=7.59$, $p<0.05$, and lexical stress perception, $F(1, 58)=37.46$, $p<0.05$. In mixed mode, the adults also scored better than children in both lexical tone, $F(1,58)=42.13$, $p<0.05$, and lexical stress $F(1,58)=33.07$, $p<0.05$.

In terms of reaction time on correctly answered items, there was no significant difference in the perceptual performance on linguistic pitch, $F(1, 58)=1.49$, $p=0.23$, $\eta_p^2=0.03$, as well as on mode, $F(1, 58)=0.50$, $p=0.48$, $\eta_p^2=0.01$, respectively. However, the main effect of group was significant, $F(1, 58)=13.28$, $p<0.05$, $\eta_p^2=0.19$, in which adults reacted faster than children in responding the trials. The two-way interaction of task and mode was significant, $F(1, 58)=4.66$, $p<0.05$, $\eta_p^2=0.074$. The three-way interaction of tasks, mode and group was also significant, $F(1, 58)=7.96$, $p<0.07$, $\eta_p^2=0.12$.

The simple main effect analysis was conducted to examine the reaction time in responding the linguistic pitch in Cantonese and English, separately for children and adults again. For children, the main effect of linguistic pitch was not significant, as children had similar reaction time in responding lexical tone and lexical stress, $F(1, 29)=0.02$, $p=0.89$, $\eta_p^2=0.001$. Also, the main effect of mode was not significant as the children responded in a similar speed in single mode and mixed mode, $F(1, 29)=0.06$, $p=0.81$, $\eta_p^2=0.002$.

For adults, the main effect of linguistic pitch was significant as they reacted faster in lexical stress perception than lexical tone perception, $F(1, 29)=5.88$, $p<0.05$, $\eta_p^2=0.17$. The main effect of mode was not significant, as the adults responded in a similar speed in single mode and mixed mode, $F(1, 29)=2.31$, $p=0,14$, $\eta_p^2=0.07$. The two way interaction of tasks and modes was significant, $F(1, 29)=22.73$, $p<0.05$, $\eta_p^2=0.44$. The paired sample T-test was used to analyze the effect of tasks under two modes. In single mode, adults reacted faster in stress perception than tone perception, $T(29)=3.79$, $p<0.05$. In mixed mode, adults reacted faster in tone perception than stress perception, $T(29)=-2.57$, $p<0.05$.

Again, in order to find the effect of groups on reaction time in linguistic pitch in different modes, simple main effect analysis was conducted in two modes. In single mode, it showed that the adult participants reacted faster than children participants in both lexical tone, $F(1, 58)=8.68$, $p<0.05$, and lexical stress, $F(1, 58)=14.84$, $p<0.05$. In mixed mode, the adults also performed faster than children in lexical tone, $F(1,58)=13.97$, $p<0.05$, and lexical stress, $F(1,58)=8.42$, $p<0.05$.

Lexical Tone Contrasts Analyses

In each version of Cantonese tone perception test, there were 15 tone contrasted included. In order to find out the main effect of tone contrast on the participants' performance in accuracy and reaction time, a 15(Tone contrasts: 15 pairs) x 2(Group: children and adults) two-way ANOVA was carried out on response accuracy and reaction time on correctly answered items. The tone contrast pairs was within-subject variable and group was between-subject variable. The percentage accuracy and reaction time were dependent variables.

In terms of accuracy, it was found that the main effect of tone contrast was significant, $F(14, 45)=6.16$, $p<0.05$, $\eta_p^2=0.66$, and so was the groups, $F(1, 58)=7.59$,

$p<0.05$, $\eta_p^2=0.12$. The interaction between tone contrasts and groups was not significant $F(14, 45)=1.17$, $p=0.33$, $\eta_p^2=0.27$. It was found the children and adults could score more than 90% mean accuracy in all tone contrasts, except for tone contrast T3 vs T6 (mid level - low level) and T2 vs T5 (high rising-low rising). The adults performed better than children in all tone contrasts.

In terms of reaction time, the main effect of tone contrast was significant, $F(14, 45)=7.38$, $p<0.05$, $\eta_p^2=0.69$, and so was the groups, $F(1,58)=8.82$, $p<0.05$, $\eta_p^2=0.13$, but the interaction of tone contrast and group was not significant $F(14,45)=1.74$, $p=0.08$, $\eta_p^2=0.35$. The following charts have summarized the means accuracy percentage and reaction time of the participants in the 15 tone contrasts.

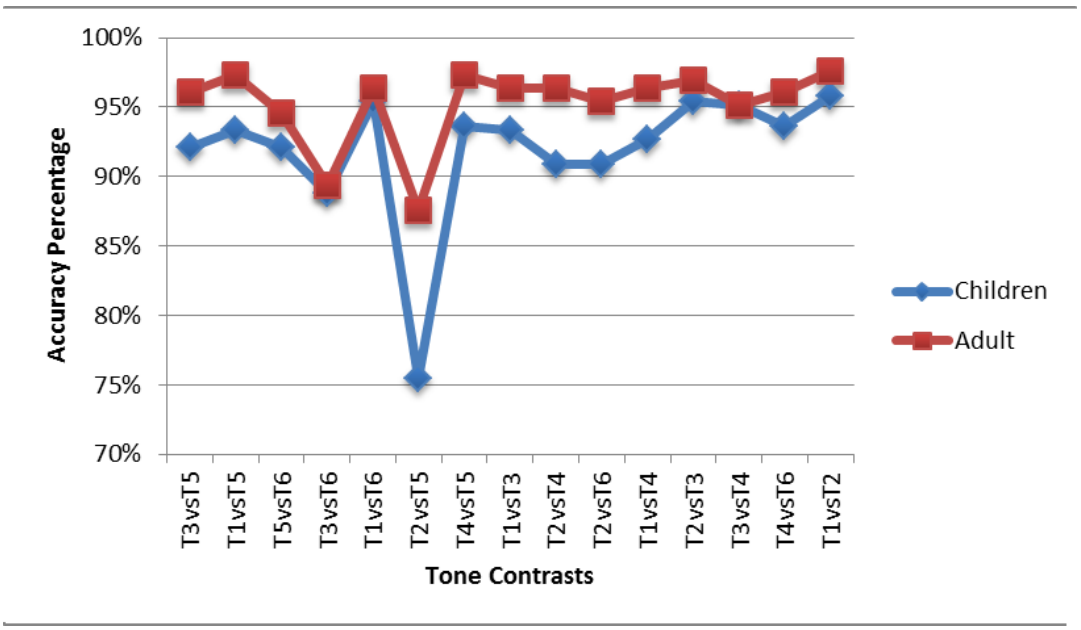

Figure 2. The participants' accuracy performance in 15 tone contrasts.

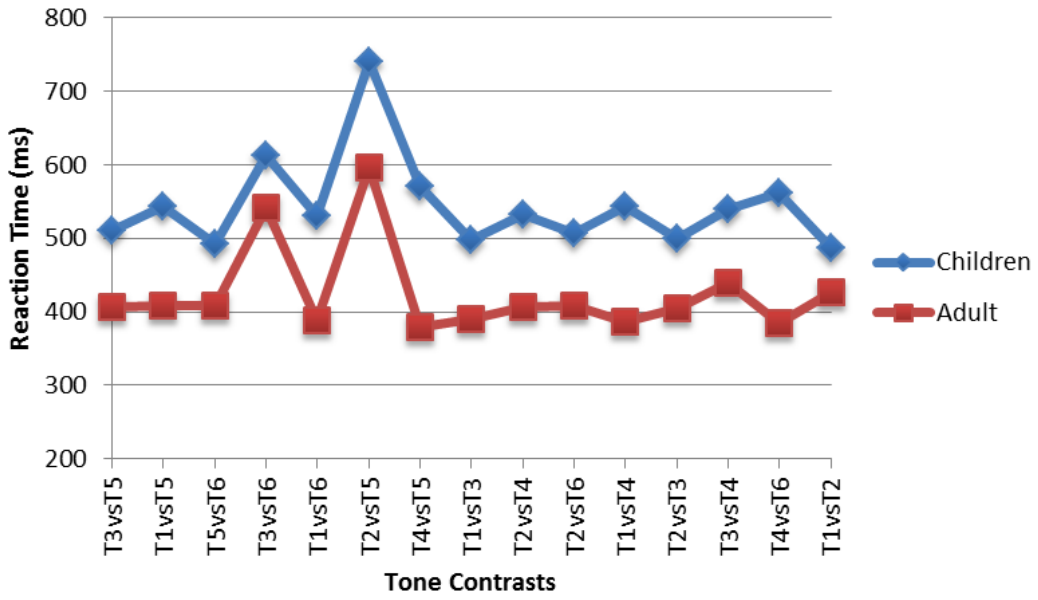

Figure 3. The participants' reaction time in 15 tone contrasts.

Position of constant tone analysis

There were two types of Cantonese stimuli, the first type was with constant tone on the first syllable, the second type was the constant tone on the second syllables. In order to find out if there was any effect of the position of constant tone on the participants' performance in accuracy and reaction time, a 2 (positions of constant tone: 1st syllable and 2nd syllable) x 2 (groups : children and adults) ANOVA was used to calculate the accuracy and reaction time difference in single mode. The position of constant tone was within-subject variable and group was between-subject variable. The response accuracy and reaction time were dependent variables. It was found that there was no main effect of constant tone position in both accuracy, $F(1, 56)=1.34$, $p=0.25$, $\eta_p^2=0.023$, and reaction time, $F(1, 56)=0.16$, $p=0.69$, $\eta_p^2=0.003$ respectively. In other words, the position of constant tone didn't affect the participants' perceptual performance on lexical tone. The group had significant effect as the adults performed better than children in both accuracy response,

$F(1, 56)=7.68$, $p<0.05$, $\eta_p^2=0.12$, and reaction time, $F(1, 56)=8.54$, $p<0.05$, $\eta_p^2=0.13$.

Blocks analysis

The 55 CVCV was grouped into 5 blocks, in order to find out if the blocks had any effect on the participants' performance in accuracy and reaction time, a 5 (blocks of stimuli) x 2 (groups of participants) ANOVA was carried out. The results showed that there was no significant main effect of blocks in response accuracy, $F(4, 55)=0.72$, $p=0.58$, $\eta_p^2=0.05$ and reaction time, $F(4, 55)=1.85$, $p=0.13$, $\eta_p^2=0.12$. In other words, the CVCV combination had no effect on participants' perceptual performance in linguistic pitch. Again, there was significant main effect of groups in accuracy, $F(1, 58)=7.59$, $p<0.05$, $\eta_p^2=0.12$, and reaction time, $F(1, 58)=8.81$, $p<0.05$, $\eta_p^2=0.13$, as the adults performed better and reacted faster than the children in all blocks. The following charts have summarized the means of accuracy and reaction time of the participants in 5 blocks.

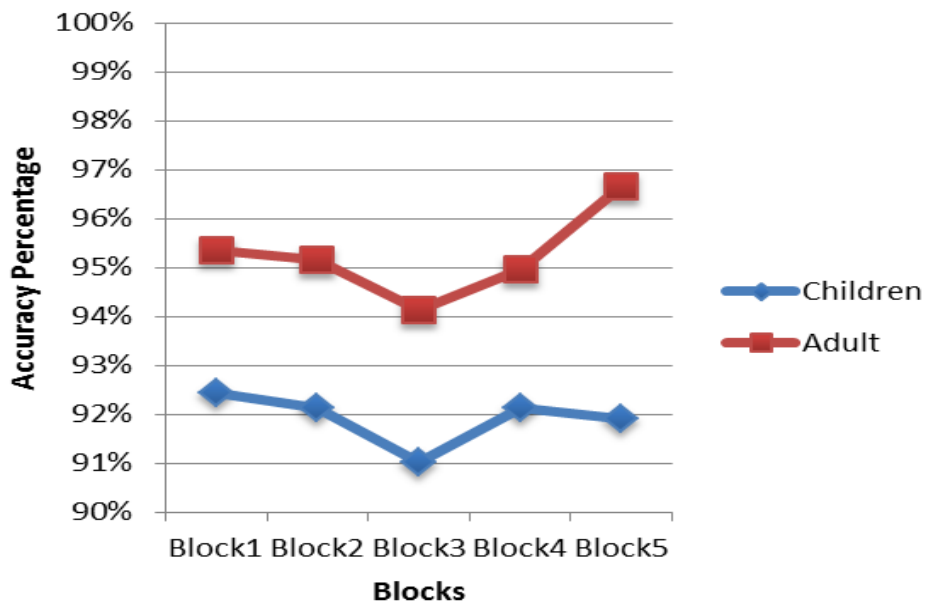

Figure 4. The participants' accuracy performance in 5 blocks.

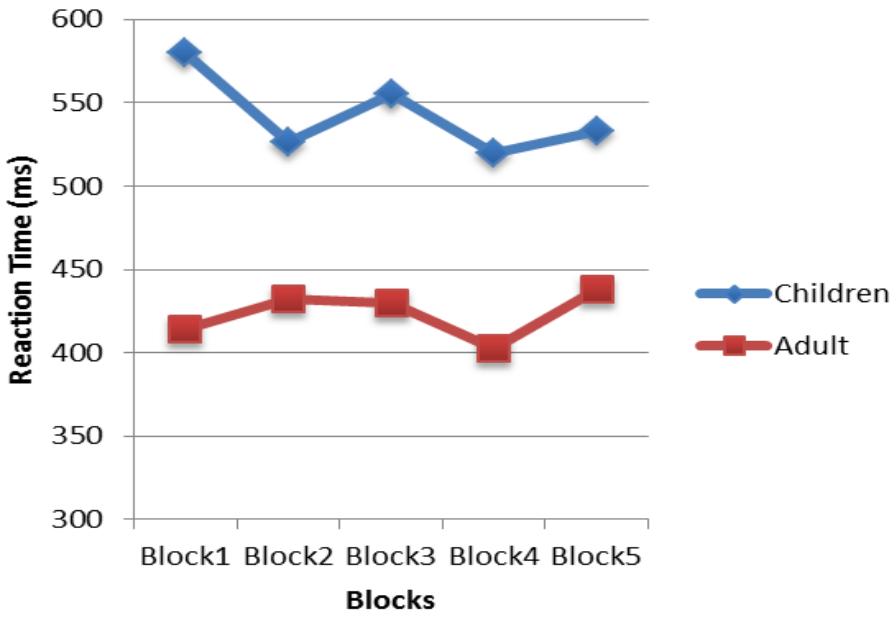

Figure 5. The participants' reaction time in 5 blocks.

Discussion

Adult bilingual speakers generally performed better than children bilinguals in discriminating both lexical tones contrasts and stress contrasts. For children, lexical tone perception and lexical stress perception are approximately equal, but adults performed better in lexical stress than lexical tone. Also, children's perception of lexical tones and lexical stresses depends on the presentation mode, with a better performance in single mode relative to mixed mode. No such difference was found in adults.

One of the most striking findings of our study is that there was no difference between the perception of lexical tone and lexical stress in bilingual children but adults performed better in lexical stress than lexical tone. The similar performance in accuracy and reaction time in lexical tone and lexical stress by bilingual children suggests that they have acquired the perception strategies well in both lexical tone and lexical stress perception. This finding suggests that the bilingual children might have used the same perception strategy in

perceiving tone and stress. It would be further explained in the below paragraphs. For bilingual adults, better performance in lexical stress than lexical tone may be attributed that perception of lexical stress is easier than lexical tone as the prosodic cues needed to perceive stress might be less than that in perceiving lexical tone. In perceiving stress, the stressed syllables tend to have higher frequency and longer duration than the unstressed syllables (Chan, 2008; Sluijter & van Heuven, 1996). Whereas in perceiving lexical tone, there are 6 contrastive tones which already increase the perception load. Furthermore, more prosodic cues including onset and offset frequency, as well as pitch contour are all contributed in perceiving the corresponding tones (Fok Chan as cited in Chan, 2008), therefore, perceiving lexical stress might be easier than lexical tone.

There were two possible explanations of using the same perception strategies by the Cantonese-English bilingual speakers. The first one lies in the similar acoustic cues used for perceiving tone and stress. As discussed before, Cantonese speakers with advanced English (Cantonese-English bilingual speakers with English as second language) would still use fundamental frequency as prosodic cue in perceiving English stress as if in the way they perceive Cantonese tone (Chan, 2008). In English, the stressed syllables tend to have higher fundamental frequency than the unstressed syllables (Chan, 2008; Sluijter & van Heuven, 1996), therefore even Cantonese-English bilingual speakers tend to use fundamental frequency as a primary cue to perceive English stress, they could still achieve well performance in English stress perception test. It showed that the Cantonese-English bilingual speakers utilized the same perception strategies in perceiving tone and stress using the fundamental frequency cue.

Alternatively, it is also plausible that Cantonese-English bilingual speakers have already developed two perception strategies as the native speakers do in perceiving their own

native language. In other words, the Cantonese-English speakers might be able to use fundamental frequency as primary cue to perceive lexical tone contrast in Cantonese tone perception task and meanwhile they were able to switch to another strategy of using spectral balance to perceive lexical stress in English stress perception task, therefore they could achieve equally well in both lexical tone and stress. However, this possibility was less likely to happen. Comparing the results of reaction time in lexical tone and stress test in single mode to mixed mode, it showed that there was no significant difference between the reaction time within the subject groups, it might show that the participants used the same perception mechanism or the same strategies in perceiving the lexical tone and lexical stress, leading to no competition of processing time or the switch of strategies. Therefore, bilingual speakers used the same perception strategies in perceiving tone and stress.

The second possible explanation of using the same perception strategies by the Cantonese-English bilingual speakers was the similar functional use of tone and stress. Lexical tone and lexical stress are both used to denote the meaning of words in suprasegmental level. Bilingual speakers might have perceived the tone and stress as the same type of linguistic pitch, therefore using the same perception mechanism to perceive tone and stress, leading to the similar performance in tone and stress perception.

We also found that bilingual adults performed better in accuracy and reaction time than bilingual children, and it suggests that bilingual speaker's perceptual ability in lexical tone and lexical stress improved even after the age of 10-13 due to longer language exposure to suprasegmental experience. Unlike the findings from Ciocca and Lui's study on Cantonese monolingual speakers(2003), this implies that bilingualism might play a role in perceiving linguistic pitch, or the tasked used by the Ciocca and Lui (2003) was easy which caused ceiling effect. Ciocca and Lui (2003) used existing Cantonese word which might already

provide the intrinsic knowledge of prosodic cues exerting influence on the perception. Second, participants were given self-paced mode to respond to the trials. In the current study, the participants were required to respond to the trials as quickly as possible, the increased difficulty of the tasks in the present study could reveal the performance difference between the children and adult groups.

Perceptual saliency of different tone contrasts

Our findings are consistent with previous research that mid level-low level, and high rising-low rising are the most difficult tone contrasts. Ciocca and Lui (2003) pointed out that the mid level and low level have similar flattened contour with only 10-30Hz difference in the frequency level (Baucer and Benedict, 1997, p.250-252; White, Ciocca and Chow as cited in Ciocca & Lui, 2003). As for the high rising-low rising tone contrasts, they have the same onset pitch level with similar contour pattern but differed only slightly in the offset point of the frequency level. The frequency patterns of these two tone contrasts were typical for a more difficult perception among other tone contrasts (Bauer and Benedict, 1997, p.250-252; Ciocca, Whitehill and Ng as cited in Ciocca & Lui, 2003). However, the results in the present study can extend the fact that the most difficult pairs of contrasts were the same for native Cantonese children and Cantonese-English bilingual children.

Beside accuracy, the longer reaction time in mid level- low level and high rising-low rising tone contrasts among other tone contrasts might also reveal there was a close relationship between the ease of tone contrasts and the processing time for the perception, as well as suggesting that this situation also occur in bilingual adulthood.

In addition, we found that the positions of the target tonal syllables in the Cantonese tone tests did not affect the perception accuracy and reaction time, which might show that the perception of tone is independent of other phonetic context. Another possibility was that

bilingual speakers could interpret the difference of the prosodic cues not only within a syllable, but also for a word with two syllables as a whole thing. Also, the combination of consonants and vowels (CVCV sequence) in the segmental level did not have influence on the suprasegmental perception; it showed that the perception of tone is independent of phonetic context which is consistent with the findings previously. It could be suggested that future research could consider to use different CVCV structures with the tone contrasts, e.g. ki2fe5, ba2mo5, se3fu6 (the numbers are the tones), in the AXB task which may have exerted more variation on the perception. These findings collaborate our idea that the perceptual difference found in our study reflects bilingual speakers' perception of linguistic pitch instead of other contextual information (position of lexical tone and the block).

In conclusion, bilingual speakers use the same perception strategies underpinning the perception of lexical tone and lexical stress. The perception ability of lexical tones and lexical stresses can still be improved by the continuous language learning experience after the age of 10-13.

Reference

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese Phonology*. Berline: mouton deGruyter.

Best, C. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. . In J. Goodman & H. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. pp.167-224). Cambridge, MA: MIT Press.

Best, C., McRoberts, G., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimiation to the listener's native phonological system. . *Journal of the Acoustical Society of America, 109*(2), 775-794.

Best, C., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: in honor of James Emil Flege*. Amsterdam: John Benjamins Publishing Company.

Boersma, P., & Weenink, D. (2012). Praat: doing phonetics by computer (Version 5.2.42). Retrieved 24 January, 2012, from http://www.praat.org/

Cenoz, J., & Genesee, F. (2001). Trends in langauge acquisition research. In J. Cenoz & F. Genesee (Eds.), *Trends in Bilingual Acquisition*. Amsterdam: John Benjamins Publishing Company.

Chan, M.-K. K. (2008). *The perception and production of lexical stress by Cantonese speakers of English*. Unpublished master degree thesis. . The University of Hong Kong.

Cheung, P. S. (2000). *The development of vowel production in Cantonese-speaking infants*. Unpublished undergraduate thesis. The University of Hong Kong. Hong Kong.

Ciocca, V., & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders, 1*(2), 141-147.

Cutler, A., & Chen, H. C. (1997). Lexical tone in Cantonese spoken word processing. *Perception & Psychophysics, 59*(2), 165-179.

Francis, A. L., Cicca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268-294.

Fullana, N., & Mora, J. C. (2007). *Production and Perception of Voicing contrasts in English Word-Final Obstruents: Assessing the Effects of Experience and Starting Age.* Paper presented at the the Fifth International Symposium on the Acquisition of Second Language Speech.

Gandour, J., Wong, D., & Hutchins, G. (1998). Pitch processing in the human brain is influenced by language experience. *Cognitive Neuroscience, 9*(9), 2115-2119.

Gerrits, E., & Schouten, M. E. H. (2004). Categorical perception depends on the discrimination task. *Perception & Psychophysics, 66*(3), 363-376.

Hung, T. N. (2000). Towards a phonology of Hong Kong English. *World Englishes, 19*(3), 337-356.

Khouw, E., & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics, 35*, 104-117.

Krishnan, A., Xu, Y. S., Gandour, J., & Gariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*, 161-168.

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

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics, 36*(1), 141-157.

Ng, M. L., & Kwok, C. (2004). Frequency of occurrence of Cantonese word initials, finals and tones. *Asia Pacific Journal of Speech, Language and Hearing, 9*, 189-199.

Qin, Z., & Mok, P.-K. (2011). *Perception of cantonese tones by mandarin, english and french speakers.* . Chinese University of Hong Kong. Hong Kong.

Sawusch, J. R., & Gagnon, D. A. (1995). Auditory Coding, Cues and Coherence in Phonetic Perception. *Journal of Experimental Psychology: Human Perception and Performance, 21*(3), 635-652.

Skoruppa, K., Pons, F., Christophe, A., Bosch, L., Dupoux, E., Sebastian-Galles, N., . . . Peperkamp, S. (2009). Language-specific stress perception by 9-month-old French and Spanish infants. *Developmental Science, 12*(6), 914-919.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of Acoustical Society of America, 100*(4), 2471-2485.

So, L. K. H., & Dodd, B. (1995). The acquisition of phonology by Cantonese-speaking children. *Journal of Child Language, 22*(3), 473-495.

Tsou, B., Lee, T., Tung, P., Cheung, P., Ng A., To, K. S., . . . Chan, Y. (2006). Hong Kong Cantonese Articulation Test (HKCAT) 《香港粵語發音測試》. Hong Kong: City University of Hong Kong / Department of Health.

van Hessen, A. J., & Schouten, M. E. H. (1999). Categorical Perception as a Function of Stimulus Quality. *Phonetica, 56*, 56-72.

Yeon, S. H., Wayland, R., Harnsberger, J., & Silver, J. (2004). Stability in perceptual assimilation: Talker and vowel context effects. *Journal of Acoustical Society of America, 116*(4), 2571.

Zhang, Y. H., Nissen, S. L., & Francis, A. L. (2008). Acoustic characteristics of English lexical stress produced by native Mandarin speakers. *The Journal of the Acoustical Society of America, 123*(6), 4498-4513.