The Role of Amplitude Envelope in Lexical Tone Perception: Evidence from Cantonese Lexical Tone Discrimination in Adults with Normal Hearing

by

Yining Victor Zhou

A dissertation submitted to the Graduate Program in Speech, Language and Hearing
Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy,
The City University of New York

2012

This manuscript has been read and accepted
for the Graduate Program in Speech, Language and Hearing Sciences
in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

Dr. Brett Martin______________________________________

Date: June 4, 2012______ ____

Chair of Examining Committee

Dr. Klara Marton_____________________________________

Date: June 4, 2012_____ ____

Executive Officer

Dr. Brett Martin____________________________________

Dr. Valerie Shafer__________________________________

Dr. Puisan Wong___________________________________

Supervisory Committee

The City University of New York

Abstract

The Role of Amplitude Envelope in Lexical Tone Perception:
Evidence from Cantonese Lexical Tone Discrimination
in Adults with Normal Hearing

by

Yining Victor Zhou

Advisor: Dr. Brett Martin

Previously published studies on the role of amplitude envelope in lexical tone perception focused on Mandarin only. Amplitude envelope was found to co-vary with fundamental frequency in Mandarin lexical tones, and amplitude envelope alone could cue tone perception in Mandarin which uses primarily tone contour for phonemic tonal contrasts. The purpose of this dissertation is to investigate whether amplitude envelope also co-varies with fundamental frequency in Cantonese, and whether the amplitude envelope cue alone can also aid the perception of lexical tones in Cantonese which uses both tone contour and relative tone height for phonemic tonal contrasts. Signal-correlated noise stimuli were synthesized based on the six Cantonese lexical tones produced naturally in isolation with the carrier syllables /ji/ and /wai/, and contained only the amplitude envelope cue of the six Cantonese lexical tones. The original intensity level of the amplitude envelopes was preserved in one condition of the experiment, but was equalized across tone types in the other condition. Thirty native listeners of Cantonese and thirty native listeners of English were presented pairs of the stimuli, and were instructed to report whether each pair consisted of identical or different Cantonese lexical tones. The results indicated that amplitude envelope co-varied with fundamental frequency in Cantonese. Furthermore, in both conditions of the current Cantonese lexical tone discrimination experiment, the native listeners of Cantonese performed significantly above chance and with greater accuracy and shorter reaction time than the native listeners of English. This suggested that amplitude envelope could cue tone contour and relative tone height for lexical tone perception. Since tone languages in the world are described as using only tone contour, relative tone height, or a combination of both for phonemic tonal contrasts, a theoretical implication of the current study is that amplitude envelope is an additional cue and could cue tone perception in all tone languages. This finding could potentially help improve the encoding of lexical tone contrasts for lexical tone perception in cochlear implant users.

Acknowledgements

This is a great opportunity to express my gratitude to all faculty and staff members in the graduate program of Speech, Language and Hearing Sciences at the Graduate Center of the City University of New York. My special thanks go to the members of my dissertation committee, Drs. Brett Martin, Valerie Shafer and Puisan Wong, for their devotion and guidance. This dissertation would not have been possible without the constant encouragement and the generous support of my advisor, Dr. Brett Martin. The comments and suggestions of Drs. Valerie Shafer, Puisan Wong, Glenis Long, Winifred Strange, Richard Schwartz, Klara Marton and Douglas Whalen have also been crucial to the completion of this dissertation. Last but not the least, the mass brain power and the daily support of my research fellows in the Audiology and Auditory Evoked Potential Laboratory are also greatly appreciated.

Table of Contents

List of Tables **Pages**

List of Figures **Pages**

1. Introduction

The general purpose of this dissertation was to address some of the unresolved issues regarding the independent contribution of amplitude envelope to lexical tone perception in individuals with normal hearing. In Section 1.1, a brief linguistic survey of lexical tones (abbreviated as tones hereafter) will be provided. In Sections 1.2 & 1.3, the physiological and acoustical aspects of tone production will be discussed. In particular, the term ***amplitude envelope*** will be defined and compared with other acoustic attributes of tones. In Section 1.4, previous studies of tone perception in individuals with normal hearing will be reviewed. The rationale of a Cantonese tone discrimination experiment incorporated in this dissertation will be provided in Section 1.5. Section 2 will be dedicated to a detailed description of the methods used in the tone discrimination experiment. The results of the experiment will be described in Section 3 and discussed in Section 4. The theoretical and clinical implications of the results will be discussed in Section 5.

1. 1 A Linguistic Survey of Tones

A tone is a pitch variation pattern that is associated with a word and contributes to the core meaning of the word. In Mandarin, for example, the syllable /ji/ with a high falling pitch contour represents the word ***idea***, whereas the same syllable with a high level pitch contour signifies ***medicine*** (Kleeman, 2010). Pitch is a subjective, auditory attribute of a sound, and can be ordered from low to high on a scale similar to that of a musical melody (American Standards Association, 1960). In order for a pitch pattern to become a tone in a specific language, it must contrast with another pitch pattern phonemically in that language. For instance, the above-mentioned high falling pitch contour and high level pitch contour constitute a phonemic contrast in Mandarin, and therefore are considered two separate tones in Mandarin (Chao, 1947; Yip, 2002). Thus, a tone is a phonological entity and is language-specific (Yip, 2002), whereas pitch is an auditory percept of a sound and is not language-specific (American Standards Association, 1960). The frequency composition of a sound is believed to be the main underlying physical correlate of both tone and pitch. A detailed discussion of the physiological and physical determinants of tones and pitch will be provided in Sections 1.2 & 1.3 below.

Tones exist in approximately 70% of the languages in the world, including most languages in Sub-Saharan Africa, North and South America, East and South-East Asia, and Equatorial Pacific Islands. Some European languages such as Serbo-Croatian, Lithuanian, Swedish and Norwegian also have tones (Yip, 2002). In most tone languages, including Cantonese and Mandarin, the smallest phonological unit that can bear a tone is a syllable, as illustrated in the example of /ji/ in the previous paragraph (Fromkin, 1978; Xu & Wang, 2001; Yip, 2002).

Tones can be categorized as level tones or contour tones. A level tone maintains a relatively steady pitch level throughout the duration of the tone. Tone 1 (High Level Tone) in Mandarin is an example of a level tone. A contour tone is characterized by a significant shift in pitch within the duration of the tone, such as Tone 2 (Rising Tone), Tone 3 (Dipping Tone) and Tone 4 (Falling Tone) in Mandarin.

The relative pitch height and pitch contour of a tone in a specific language can be quantified using a five-point scale, with 1 being the low end and 5 being the high end of a talker's tonal pitch range (Chao, 1947; Yip, 2002). In Mandarin, for example, Tone 4 (Falling Tone) is transcribed as a "51" tone, because it usually starts at the highest point of a talker's tonal pitch range and ends at the lowest pitch point. Thus, the syllable /ji/ with Tone 4 in Mandarin is transcribed as /ji^{51}/, as displayed in Table 1 (Chao, 1947; Yip, 2002).

In Mandarin, every tone has a distinct contour as illustrated in Table 1. Thus, only tone contours are used for phonemic tonal contrasts. In Swahili, on the other hand, there are only level tones (i.e., high, mid and low level tones). Thus, only the relative tone height is used for phonemic contrasts (Hyman, 2007; Yip, 2002).

Tones	**Tone Contours**	**Numerical Representations**
Tone 1	High Level	55
Tone 2	Rising	35
Tone 3	Dipping (Falling-Rising)	214
Tone 4	Falling	51

Table 1. Classification of Mandarin Tones (adapted from Xu, 1997; Yip, 2002)

In a small subset of tone languages such as Cantonese, Taiwanese and Thai, both tone contour and tone height are used for phonemic contrasts (Yip, 2002). In standard Cantonese, there are three basic tone contours that are phonemically contrastive: falling, rising, and level. Relative tone height is also used for phonemic tonal contrasts: high falling vs. low falling, high rising vs. low rising, and upper-middle level vs. lower-middle level. Thus, there are six basic tones in standard Cantonese, as illustrated in Table 2 below (Chao, 1947; Yip, 2002).

Tones	**Tone Height**	**Tone Contours**	**Numerical Representations**
Tone 1	High	High falling/High level	53/55
Tone 2		High rising	25
Tone 3		Upper-Middle level	33
Tone 4	Low	Low falling tone	21
Tone 5		Low rising tone	24
Tone 6		Lower-Middle level	22

Table 2. Classification of Standard Cantonese Tones (adapted from Chao, 1947, and Yip, 2002)

Among older generations of Cantonese speakers, Tone 1 is usually produced as a high falling tone, and sometimes as a high level tone (Bauer & Benedict, 1997; Chao, 1947; Fok, 1974; Yip, 2002). Among young speakers of Cantonese, however, there are dialectal differences in the distribution of the high falling tone and high level tone. In Guangzhou (formerly known as Canton), Tone 1 is still produced mostly as a high falling tone, and sometimes as a high level tone (Rao, Ouyang & Zhou, 1996). In Hong Kong, Tone 1 is produced typically as a high level tone and sometimes as a high falling tone (Bauer & Benedict, 1997; Matthews & Yip, 1994). Regardless of the dialectal differences and age differences, the high falling tone and the high level tone are interchangeable in Cantonese, and therefore are considered allotones in Cantonese (Bauer & Benedict, 1997; Rao et al., 1996; Yip, 2002). For example, the Cantonese word ***clothes*** can be pronounced as /ji/ with either the high falling tone or the high level tone.

As illustrated in Table 3, the syllable /ji/ with each of the six Cantonese tones represents different Cantonese words. These words are not considered homophones because they differ in tones (Matthews & Yip, 1994). In Cantonese, as in other tone languages, homophones are defined as words pronounced with identical phonemes and tones. For instance, the words 衣 *(*clothes), 依 (rely on), 伊 (he or she) and 醫 (medicine) are homophones in Cantonese because they are pronounced as /ji/ with Tone 1 (High Falling or High Level) (Bauer & Benedict, 1997; Matthews & Yip, 1994; Rao et al., 1996).

Phonemic Structures	Cantonese Words (Characters)	Meaning in Cantonese
/ji/ with Tone 1	衣	clothes, membrane
/ji/ with Tone 2	椅	chair
/ji/ with Tone 3	意	idea, intent
/ji/ with Tone 4	儿	child
/ji/ with Tone 5	耳	ear
/ji/ with Tone 6	二	two
/wai/ with Tone 1	威	pride
/wai/ with Tone 2	位	seat
/wai/ with Tone 3	餵	feed
/wai/ with Tone 4	圍	fence, encircle
/wai/ with Tone 5	偉	great
/wai/ with Tone 6	為	for

Table 3. Examples of Cantonese words represented by the syllables /ji/ and /wai/ with each of the six tones in Cantonese (adapted from Rao et al., 1996).

1.2 Physiology of Tone Production

The physiology of tone production has been investigated in several studies (e.g., Erickson, 1976; Garding, Fujimura & Hirose, 1970; Halle, 1994; Sagart, Halle, Boysson-Bardies & Arabia-Guidet, 1986; Simada & Hirose, 1971). The synergistic activation of the cricothyroid and vocalis muscles seems to be the primary physiological mechanism for raising the relative tone height. The activation tightens and stretches the vocal folds, which accelerates the vocal fold vibration and increases the transglottal pressure (Baken & Orlikoff, 2000; Erickson, 1976; Garding et al., 1970; Hirose, 2010; Simada & Hirose, 1971; Titze, 1995). The perceptual consequence of the accelerating vocal fold vibration and the increasing transglottal pressure is an increase in the tone height (Plack, Oxenham, Fay & Popper, 2005).

Several mechanisms for lowering a tone have also been identified (Baken & Orlikoff, 2000; Hirose, 2010; Titze, 1995). First, decreasing subglottal pressure slows down vocal fold vibration and decreases the transglottal pressure (Baken & Orlikoff, 2000). The perceptual consequence of decelerating vocal fold vibration and decreasing transglottal pressure is a descending tone (Plack, Oxenham, Fay & Popper, 2005). Second, the simultaneous de-activation of the cricothyroid muscle and activation of the vocalis muscle also lower a tone. That is, when the vocalis muscle is activated and the cricothyroid muscle is de-activated, the increase in vocalis muscle mass has a greater effect on the vocal fold vibration than the increase in the vocalis muscle stiffness, and hence, the vocal fold vibration slows down (Erickson, 1993; Titze, 1995). The perceptual consequence of this decelerating vocal fold vibration is a descending tone (Plack, Oxenham, Fay & Popper, 2005). Finally, the simultaneous de-activation of the cricothyroid muscle and activation of strap muscles (e.g., sternohyoid, sternothyroid and thyrohyoid muscles) shorten the vocal folds, increases the vocal fold mass, and hence slows down vocal fold vibration (Erickson, 1976; Hirose, 2010). As mentioned above, the perceptual consequence of the decelerating vocal fold vibration is a descending tone.

1.3 Acoustics of Tones

The acoustics of tones have been investigated in numerous studies (see Kuo, Rosen & Faulkner, 2008, for a review). Several acoustic properties of tones have been identified, including fundamental frequency (F0), harmonics, formant structure, amplitude and duration. These properties will be defined and discussed below.

1.3.1 F0 of Tones

The F0 of a tone is defined as the lowest frequency in the spectrum of a tone (Fromkin, 1978; Yip, 2002). It reflects the rate at which the vocal folds vibrate (i.e., close and open alternately) during the production of a tone (Yip, 2002). The F0 contours of the six Cantonese tones produced in isolation are shown in Figure 1. The F0 contour of Tone 1 (High Falling) starts in the upper F0 range of a person’s speaking voice and drops gradually to the low F0 range of the person’s speaking voice. The F0 contour of Tone 2 (High Rising) starts in the mid F0 range of a person’s speaking voice, and gradually rises to the upper F0 range. Tone 3 (Upper-Middle Level) has a flat F0 contour in the upper-middle F0 range of a person’s speaking voice. The F0 contour of Tone 4 (Low Falling) starts in the lower-middle F0 range of a person’s speaking voice and drops gradually towards the lowest F0 range. The F0 contour of Tone 5 (Low Rising) starts in the mid F0 range of a person’s speaking voice, and gradually rises to the upper-middle F0 range. Tone 6 (Lower-Middle Level) has a flat F0 contour in the lower-middle F0 range of a person’s speaking voice.

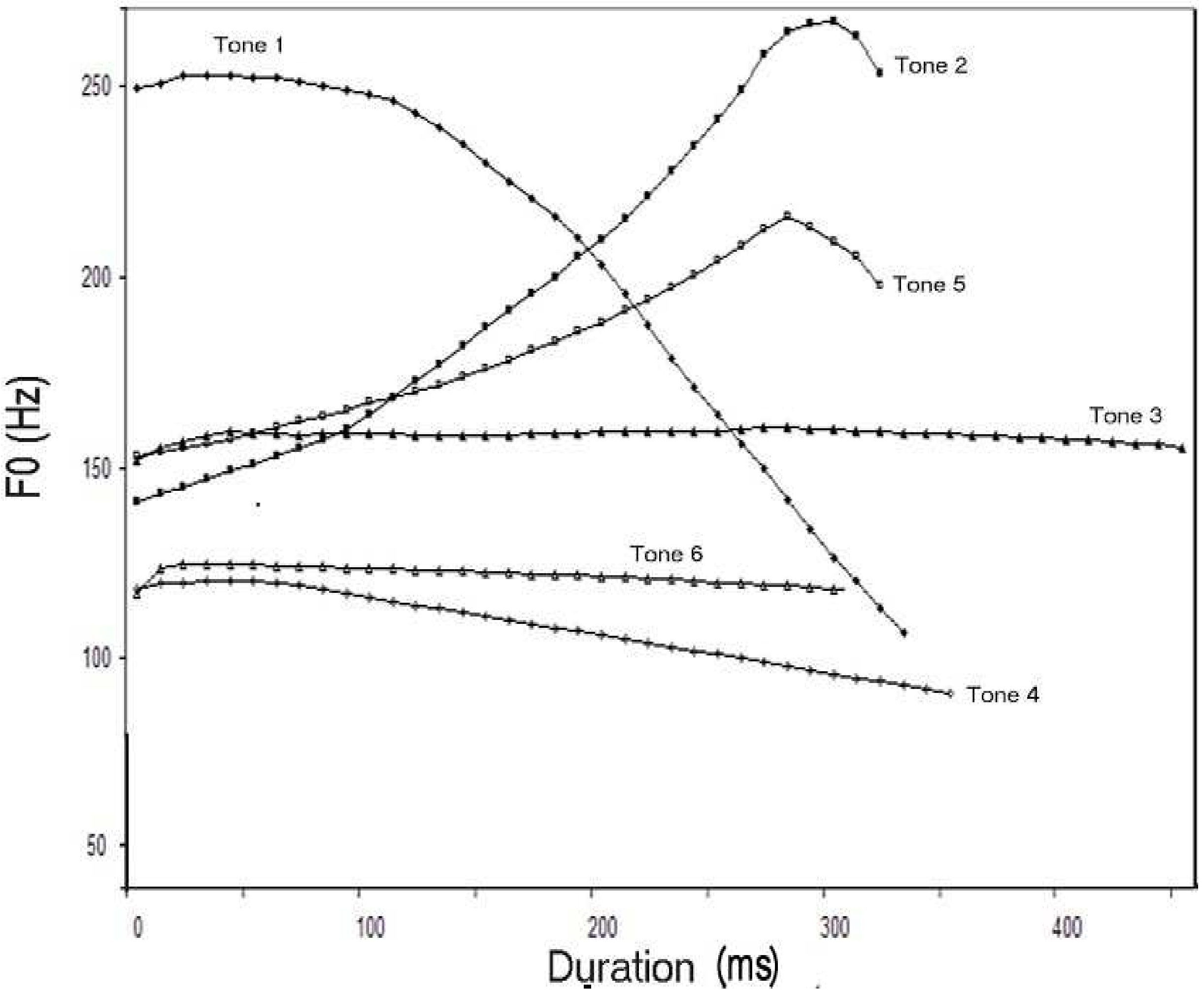

Figure 1: Mean F0 contours (averaged over one male talker and one female talker) of the six Cantonese tones produced with the carrier syllable /fu/ in isolation (adapted from Fok, 1974).

The F0 contours illustrated in Figure 1 are often called the canonical F0 contours (i.e., F0 contours of tones produced in isolation). Tones produced in connected speech may have F0 contours that differ drastically from the canonical forms. Sources of tonal variability in connected speech include coarticulation of adjacent tones (e.g., Li, Le & Qian, 2002; Xu, 1992, 1997 & 2006), intonation (e.g., Shih, 1988; Wong, 1999), and stress (e.g., Gu & Lee, 2007). As an example of coarticulation, the high F0 offset of a high level tone in Cantonese or Mandarin raises the normally low F0 onset of the following rising tone, due to forward assimilation or carry-over (Li et al., 2002; Xu, 1997). As for backward or anticipatory assimilation, the normally high F0 offset of Tone 2 (High Rising) in Cantonese is lowered in anticipation of the low F0 onset of the following Tone 4 (Low Falling) (Li et al., 2002).

Emphatic stress also affects the surface realization of Chinese tones. In Cantonese, for example, the prosodic mechanism for emphasizing key words in an utterance is to lengthen the duration, raise the F0 height and expand the F0 range of the emphasized syllables (Gu, Hirose & Fujisaki, 2005; Gu & Lee, 2007). A similar prosodic mechanism has been found in Beijing Mandarin (Shih, 1988; Xu, 2006).

Intonation also interacts with tones. For example, declination (i.e., downward F0 drift in a declarative sentence) affects the F0 height of tones. Cantonese tones under the influence of declination can be altered to the point where the sentence-initial low level tone has a significantly higher F0 than the sentence-final high level tone. Interrogative sentences have the opposite effect on the F0 contours of tones (e.g., Wong, 1999). In Mandarin, the rising intonation in an interrogative sentence affects Tone 1 (High Level) and Tone 2 (High Rising) the most: it turns the F0 offsets of these two tones into slightly rising tails (e.g., Shih, 1988). In Cantonese, it affects mainly the F0 contours of the low tones: Tone 4 (Low Falling), Tone 5 (Low Rising), and Tone 6 (Lower-Middle Level). It alters the F0 contours of these tones to the point where they resemble that of a high rising tone acoustically and perceptually (Wong, 1999).

1.3.2 Harmonics of Tones

The harmonics of a sound are integer multiples of the F0 of the sound (e.g., Plack et al., 2005). Likewise, the harmonics of a tone are integer multiples of the F0 of the tone (Liang, 1963). In the absence of F0 (e.g., when F0 is masked or filtered out), the harmonics can convey the periodicity of a tone (e.g., Fu & Zeng, 2000; Kong & Zeng, 2006). Thus, harmonics can provide an indirect acoustic cue to the rate of vocal fold vibration during the production of a tone by conveying the periodicity of the vocal fold vibration. This is an example of the “missing fundamental” (Fu & Zeng, 2000; Schouten, 1962).

1.3.3 Formant Structure of Tones

As mentioned in Section 1.1 above, a tone is a pattern of pitch variation superimposed on a syllable. Thus, a naturally produced tone always co-exists with a carrier syllable (e.g., Kong & Zeng, 2006). The formant structure of a given carrier syllable varies depending on the tone, and each tone has its distinct formant structures (e.g., Erickson et al., 2004; Kong & Zeng, 2006).

1.3.4 Duration of Tones

At least in some tone languages, the relative duration of a tone type is consistently shorter or longer than other tone types. In Mandarin, for example, each of the four tones has an intrinsic duration, with Tone 3 having the longest duration, and Tone 4 having the shortest duration (Fu & Zeng, 2000; Kuo et al., 2008; Xu, 1997). The differences in relative duration among the Cantonese tones are controversial (Fok, 1974; Khouw & Ciocca, 2007; Rose, 2000). According to Fok (1074), only Tone 3 (Upper Middle Level) has a relative duration that is significantly longer than that of the other five tones, and there were no significant differences in duration among the other five tones, as illustrated in Figure 1. According to Rose (2000), however, the duration of Tone 4 (Low Falling) is significantly shorter than that of the other five tones, and there are no significant differences in duration among the other five tones.

1.3.5 Amplitude of Tones

Similar to the amplitude of a sound, the amplitude of a tone is the magnitude of the air pressure directly related to the production of the tone, usually measured in pascals at a talker's mouth opening (Coster & Kratochvil, 1984; Sagart et al., 1986). It reflects the air pressure that is originated from the lungs and subsequently shaped by the glottis and the vocal tract (Baken & Orlikoff, 2000). It usually fluctuates during the production of a tone, and each tone type in Mandarin has a distinct amplitude contour (Kong & Zeng, 2006; Kuo et al., 2008; Fu & Zeng, 2000; Whalen & Xu, 1992). For instance, Tone 4 (Falling) has a falling amplitude contour, whereas Tone 2 (Rising) has a rising amplitude contour. Furthermore, Tone 3 (Dipping) has a dipping amplitude contour, and Tone 1 (High Level) has a flat contour of relatively high amplitude (e.g., Whalen & Xu, 1992). Similar patterns were found in tone production studies in Japanese (e.g., Simada & Hirose, 1971), Swedish (Garding, Fujimura & Hirose, 1970), and Thai (e.g., Erickson, 1976 & 1993).

The amplitude cue of a tone can be further broken into at least two components: amplitude envelope and amplitude periodicity (e.g., Rosen, 1992; Fu & Zeng, 2000). Amplitude periodicity is defined as the amplitude fluctuation in the waveform of a sound between the rates of 50 and 500 Hz (Rosen, 1992). It reflects the periodically recurring pattern of amplitude fluctuation in the waveform due to the periodic changes in transglottal pressure as vocal folds close and open during the production of a sound (Baken & Orlikoff, 2000). Thus, amplitude periodicity is related to F0: whereas F0 directly indexes the rate of vocal fold vibration, amplitude periodicity is an indirect acoustic cue of the rate of vocal fold vibration (Kong & Zeng, 2006).

The amplitude envelope of a tone is defined as the amplitude fluctuation in the waveform of a tone at the rate of 2 to 50 Hz (Rosen, 1992). The amplitude envelope of a tone can be estimated from the waveform of the tone in two steps. First, the waveform is divided into a sequence of equal time frames of at least 50 milliseconds long, but no longer than 500 milliseconds. Then, the root-mean-square (RMS) amplitude within each frame is computed and plotted in a two-dimensional line graph, and the line in the graph is the estimated amplitude envelope of the tone (Fu & Zeng, 2000). The amplitude envelope of each of the four Mandarin tones is displayed in Figure 2.

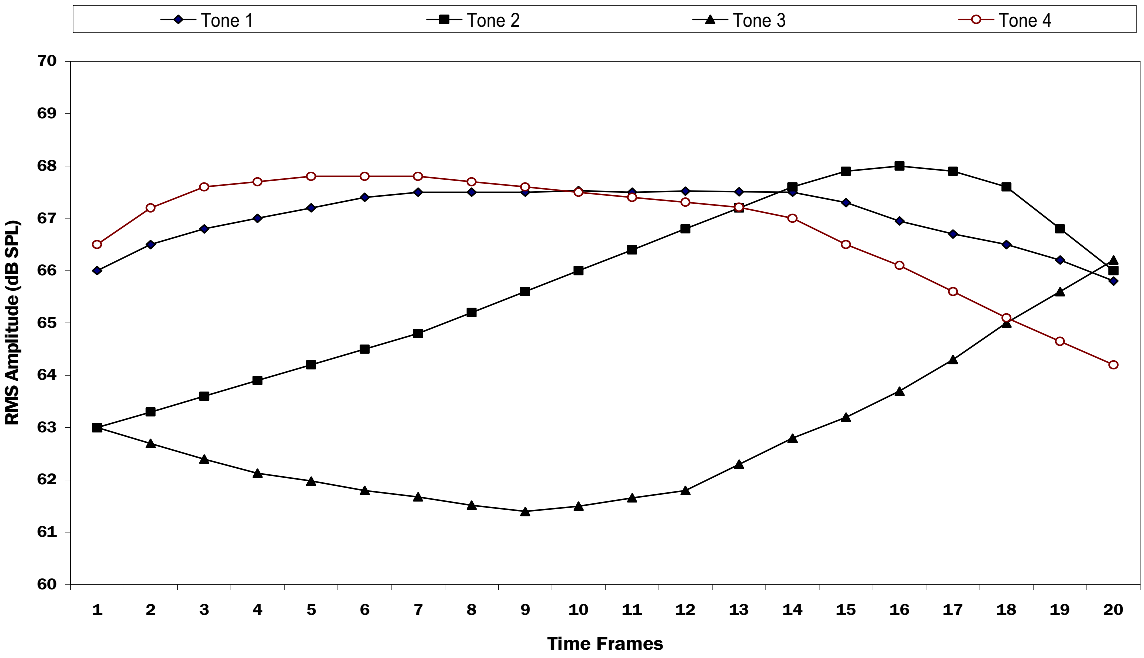

Figure 2. Amplitude envelopes of the four Mandarin tones: each tone was divided into twenty equal time frames, and the root-mean-square (RMS) amplitude within each time frame was averaged across sixty tone tokens produced in isolation by five male talkers and five female talkers with the carrier syllables /a/, /i/, /u/, /o/, /y/ and /ɣ/ (adapted from Fu & Zeng, 2000).

As illustrated in Figure 2, the amplitude envelope of a tone reflects the overall rising, falling or steady trend of amplitude change throughout the production of the tone (Baken &

Orlikoff, 2000; Titze, 1995). In Mandarin, for example, Tone 1 (High Level) has a flat amplitude envelope, whereas Tone 2 (Rising) exhibits a rising amplitude envelope. Tone 3 (Dipping) shows a dipping amplitude envelope, and finally, Tone 4 (Falling) demonstrates a falling amplitude envelope. Amplitude envelope co-varies with F0 contour in Mandarin tones produced in isolation. This covariability was measured using Pearson’s linear correlation method (i.e., Pearson’s Product Moment), and the Pearson’s correlation coefficients for the four Mandarin tones ranged from 0.50 to 0.83 (e.g., Kuo et al., 2008). The correlation coefficients were higher for Tone 3 (Dipping) and Tone 4 (Falling Tone), and lower for Tone 2 (Rising) and Tone 1 (High Level). However, the statistical significance of these differences was not reported.

1.3.6 Summary of the Acoustic Features of Tones

So far, several acoustic features of tones have been identified. They include F0, harmonic structures, formant structures, amplitude envelope, amplitude periodicity and duration. Among these acoustic features, F0 has been investigated the most extensively, and therefore tones are typically defined in terms of their F0 contour and relative F0 height (see Yip, 2002, for a review). In recent years, with the advent of cochlear implant technology, the other acoustic features such as amplitude envelope and amplitude periodicity have gained considerable attention from researchers (e.g., Fu & Zeng, 2000; Kuo et al., 2008) due to the important contribution of amplitude envelope and amplitude periodicity to tone perception in cochlear implant users (e.g., Kong & Zeng, 2006; Luo, Fu, Wei & Cao, 2008).

1.4 Perception of Chinese Tones

Perception of Chinese tones has been studied using word or phrase discrimination or identification tasks (for a comprehensive review, see Krenmayr, Qi, Liu, Liu, Chen, Han, Schatzer & Zierhofer, 2011, and Khouw & Ciocca, 2007). Usually, upon hearing the presentation of two words or phrases differing only in tones, listeners in a discrimination study were to report whether the stimuli presented were identical or not. In an identification study, on the other hand, listeners were presented a spoken word or phrase, and then instructed to identify the word or phrase from a list of written words/phrases, or to point to one of the pictured objects. The stimuli was presented either in isolation, or embedded in a carrier sentence such as "Where is ___?" or "Point to ___?" These word or phrase identification tasks were often called tone identification tasks even though the listeners were not asked to identify the tones specifically (for a comprehensive review, see Krenmayr et al., 2011).

Strictly defined tone identification tasks were rarely used in previous studies of Chinese tone perception probably because not every native speaker of a Chinese language was able to label the tone of a spoken word in his/her native language (Bauer & Benedict, 1997; Khouw & Ciocca, 2007; Yip, 2002). Typically, native speakers of Cantonese without extensive professional training in phonetic transcription of Cantonese are unable to name the tone of a spoken Cantonese word at all (Bauer & Benedict, 1997; Matthews & Yip, 1994). Thus, strictly defined tone identification tasks would not be feasible with naïve native speakers of Cantonese. In the following discussion, the term ***tone identification*** will continue to be used loosely to refer to the type of word or phrase identification used in previous studies on Chinese tone perception.

1.4.1 Perception of Chinese Tones in Isolation

Perception of Chinese tones in isolation has been investigated in several studies. When presented individually, Cantonese tones produced by different talkers with a carrier CV syllable in isolation could be identified by native listeners of Cantonese with approximately 75% accuracy in a 6-alternative-forced-choice (6AFC) task (e.g., Fok, 1974; Khouw & Ciocca, 2007). Mandarin tones presented in similar fashion could be identified by native listeners of Mandarin with perfection or near-perfection (e.g., Fu & Zeng, 2000; Kuo et al., 2008; Liang, 1969).

The fact that tone height is used for phonemic tonal contrasts in Cantonese but not in Mandarin may explain the above differences between Mandarin and Cantonese listeners in identifying the tones in their native languages. In the above tone perception studies, the Cantonese listeners identified the contours of Cantonese tones almost as accurately as the Mandarin listeners identified the contours of Mandarin tones. However, the Cantonese listeners sometimes confused tones with similar contours but different height, especially Tone 2 (High Rising) vs. Tone 5 (Low Rising), and Tone 3 (Upper-Middle Level) vs. Tone 6 (Lower-Middle). The confusion between two tones with similar contours but different height could be accounted for given the finding that tone height was a less salient perceptual cue than tone contour in Cantonese (Barry et al., 2004b; Gandour 1981 & 1983). In Gandour’s (1981) multidimensional scaling analysis of Cantonese tone identification in native listeners, for instance, tone height only accounted for 20% of the variance, whereas tone contour could explain 34% of the total variance.

A careful examination of the Cantonese tones illustrated in Figure 1 could explain why tone height is a less salient perceptual cue in Cantonese. Tone 2 (High Rising) and Tone 5 (Low Rising) both have a rising tone contour, with almost identical F0s for the first three quarters of the tone duration. They differ in tone height only in the last quarter of the tone duration (Barry et al., 2002a; Khouw & Ciocca, 2007; Rose, 2000). As for Tone 3 (Upper-Middle Level) and Tone 6 (Lower-Middle Level), they both have a flat tone contour, and only differ in the relative tone height slightly. Thus, the tone height differences between Tone 2 (High Rising) and Tone 5 (Low Rising) and between Tone 3 (Upper-Middle Level) and Tone 6 (Lower-Middle) are rather small-scale acoustic differences. Therefore, it is not surprising that Cantonese listeners sometimes mis-identified these tones with similar contours but different height, especially when they were presented individually with a CV syllable and without contextual cues.

The inter-talker variation in relative tone height could further complicate the identification of Cantonese tones with similar contours but different height. For instance, when three Cantonese level tones produced with a CV carrier syllable in isolation by different talkers were mixed randomly and presented one by one, native listeners identified the three tones with only 48.6% success in a 3AFC task (chance = 33.4%). Their identification accuracy improved to 80.3% when the stimuli produced by each talker were presented in a separate block (Wong & Diehl, 2003). The interpretation of these findings is that native Cantonese listeners needed to hear at least several Cantonese tone tokens produced by a talker in order to determine the talker’s normal Cantonese tone range and correctly perceive the talker’s Cantonese tones (Francis, Ciocca, Wong, Leung, & Chu, 2006; Wong & Diehl, 2003).

While the Chinese tone perception literature indicates a general agreement on the difficulties in perceiving individually presented Cantonese tones with similar contours but different height, there is some controversy over the categorical nature of tone perception. The majority of relevant studies indicated categorical perception of Cantonese and Mandarin tones in native speakers (see Zheng, Minnett & Wang, 2010, for a review). For instance, synthetic level tones that differed only in F0 and fell along an F0 continuum with equal intervals among them were unambiguously identified by native speakers of Cantonese as belonging to distinct Cantonese tone categories. In pairwise discrimination of these synthetic tones, a pair of tones belonging to the same tone category was mostly perceived as identical, whereas a pair of tones belonging to different tone categories was mainly perceived as different. Similar findings were reported for contour tones (see Zheng et al., 2010). However, two studies, one on Thai tones (Abramson, 1979) and one on Cantonese tones (Francis, Ciocca & Ng, 2003), suggested that the discrimination of level tones was not categorical, even though the identification of level tones was. In these studies, synthetic level tones differing only in F0 and lying along an F0 continuum with equal intervals were identified unambiguously as belonging to distinct tone categories, but in pairwise discrimination of these synthetic tones, two tones belonging to the same tone category were not consistently perceived as identical. Rather, the closer the acoustic distance between two level tones (in terms of F0 difference), the more often they were perceived as identical. Likewise, the greater the acoustic distance between two level tones, the more often they were perceived as different (Abramson, 1979; Francis & Ng, 2003). Thus, discrimination of Thai and Cantonese tones was found to be continuous in these two studies. It is difficult to reconcile the conflicting findings from these two studies and the other studies on categorical perception of tones, because different experimental methods were used (e.g., differences in ISI and F0 intervals among synthetic tones). Given that the focus of the current study is not on categorical perception of tones, it suffices to re-iterate that the majority of behavioral and electrophysiological studies indicated categorical perception of Cantonese and Mandarin tones presented in isolation (see Zheng et al., 2010, for a review).

1.4.2 Perception of Chinese Tones in Connected Speech

As discussed in Section 1.3 above, Chinese tones in connected speech may have surface realizations that differ drastically from their canonical forms displayed in Figures 1 & 2 above (Gu, Hirose & Fujisaki, 2005; Li et al., 2002; Ma et al., 2006; Shen, 1990; Shih, 1988; Xu, 1992, 1997 & 2006). Nonetheless, native speakers of Mandarin or Cantonese could identify tones in their native language with perfection or near perfection when the tones were produced and presented in connected speech, even in the absence of syntactic and semantic cues in the carrier sentences (e.g., Li et al., 2002; Wong & Diehl, 2003; Xu, 2004). In Xu's (2004) study, for example, native listeners of Mandarin identified Mandarin tones produced and presented in connected speech with 97 to 99.7% accuracy. In Wong & Diehl's (2003) study on the identification of the three Cantonese level tones discussed above, native listeners of Cantonese demonstrated near-perfect performance (mean = 96%; SD = 4%; chance = 33.4%) when the target tones were embedded in the carrier sentence "Ha^{21} yat^{55} ko^{33} tsi^{21} hai^{21} _____" (The next character is _____). Similar results were found in studies on tone perception in connected speech with different prosodic patterns of intonation and stress (e.g., Gu & Lee, 2007; Shih, 1988; Wong, 1999). These results indicate that perception of Cantonese tones is significantly better in connected speech than in isolation, even when the connected speech provides merely a tonal context with no syntactic or semantic cues. In other words, the tonal context is crucial for tone perception, at least in Cantonese.

1.4.3 Acoustic Cues for Chinese Tone Perception in Native Listeners

The importance of the explicit F0 cue for tone perception has been investigated in studies on filtered speech and auditory chimeras. In studies on auditory chimeras, certain acoustic cues were digitally removed or altered from naturally produced tones, and participants were asked to label the digitally modified tones. In Xu & Pfingst's (2003) auditory chimera study on Mandarin tone perception, for example, when the F0 of a naturally produced Mandarin word with Tone 4 (Falling) was digitally replaced with a rising F0 contour without changing anything else, the word was perceived by native listeners of Mandarin as Tone 2 (Rising). This seemed to suggest that the F0 cue overrode any other acoustic cues for Mandarin tone perception. In Wong & Diehl's (2003) study on Cantonese tone perception in connected speech, when the F0 of a carrier sentence was shifted up by 13% without changing the F0 of the target word embedded in the sentence, the target word with Tone 1 (High Level) was mis-perceived as having Tone 6 (Lower-Middle Level). On the other hand, when the F0 of a carrier sentence was shifted down by 19% without changing the F0 of the target word embedded in the sentence, the target word with Tone 6 (Lower-Middle Level) was mis-perceived as having Tone 1 (High Level). Thus, the F0 cue seemed to dominate other acoustic cues in tone perception in native listeners with normal hearing.

However, tone perception is possible without the explicit F0 cue. In Liang's (1963) study, for example, native listeners of Mandarin identified each of the four Mandarin tones produced and presented in isolation with 94.6% accuracy in a 4AFC task (chance = 25%) even when the tones had been high-pass-filtered with a cut-off frequency of 300 Hz to remove the explicit F0 cue. This was consistent with Schouten's (1938) finding that pitch perception was possible even in the absence of explicit F0. This is often called residue pitch or the phenomenon of missing fundamental.

Liang's (1963) finding has been used as evidence to support that, in the absence of explicit F0, harmonic structures could cue Mandarin tone perception (Kong & Zeng, 2006; Liang, 1963). However, the high-pass filter with a cut-off frequency of 300 Hz only removed explicit

F0, and hence the output of the high-pass filter retained more than just harmonic structures. Among other things, it contained formant structures, amplitude and duration cues which could aid tone perception in Mandarin (see discussion below). Thus, the above-mentioned score of 94.6% might have included the contribution of formant structures, amplitude and duration, and therefore the independent contribution of harmonics to tone perception is still not completely clear. Nevertheless, Liang's (1963) study indicated that Mandarin tone perception was possible without the explicit F0 cue, but was more accurate with the explicit F0 cue (99% correct) than with the harmonics, amplitude, duration and formant structures combined (94% correct).

Indeed, several studies have indicated that formant structures, which are maintained in whispered speech, can allow tone perception, but with poorer accuracy than with the explicit F0 cue (Kong & Zeng, 2006; Liang, 1963; Liu & Samuel, 2004). For example, in Liang's (1963) study on whispered speech which did not contain F0 or its harmonics, native listeners of Mandarin were able to identify the four Mandarin tones in naturally whispered speech in a 4AFC task with 64% accuracy (Liang, 1963). Similar results were reported by Liu & Samuel (2004) and Kong & Zeng (2006). In addition, Liu & Samuel (2004) and Kong & Zeng (2006) compared Mandarin tone perception using naturally whispered speech and synthesized whisper speech. While the naturally whispered speech contained amplitude fluctuation which could cue tone perception (see discussion below), the amplitude in the synthesized whisper speech was kept constant. The naturally whispered speech yielded superior Mandarin tone identification performance, with a mean difference of approximately 5 percentage points between the natural and the synthesized stimuli in both studies. Thus, formant structure could aid Mandarin tone perception, but the formant structure cue and the amplitude cue combined produced better results.

Temporal cues such as duration or amplitude alone can also cue tone perception, but with less accurate results than the spectral cues (Fu & Zeng, 2000; Kon g & Zeng, 2006; Kuo et al., 2008; Whalen & Xu, 1992). In Fu & Zeng's (2000) study, for example, wide-band noise carriers with the intrinsic duration of each Mandarin tone were presented to native speakers of Mandarin for a 4AFC tone identification task. The native speakers identified each tone with 35.6% accuracy (chance level = 25%). Similar results were found in Kuo and colleagues' (2008) study (mean =38%; chance level = 25%). Thus, the duration cue alone could aid tone perception in Mandarin with above-chance accuracy. The role of tone duration in Cantonese tone perception has not been investigated in published studies, and therefore is not clear.

Amplitude can also aid tone perception (Fu & Zeng, 2000; Kong & Zeng, 2006; Whalen & Xu, 1992). In Whalen & Xu's (1992) study on Mandarin tone perception, for example, each of the four Mandarin tones produced in isolation with the carrier syllable /ba/ or /ji/ was used to synthesize a signal-correlated-noise (SCN). The synthesis consisted of several steps. First, the waveform of each tone was digitized with a sampling rate of 10 kHz. Second, each waveform was half-wave-rectified. Finally, half of the sample points of the rectified waveform were randomly selected, and the sign of these sample points was inverted. This produced SCNs which only retained the amplitude cue of the original tone tokens. Native listeners of Mandarin were able to label the Mandarin tone contained in each SCN with approximately 75% accuracy in a 4AFC task. Similar results were found in other studies on the independent contribution of the amplitude cue to tone perception in Mandarin (Fu & Zeng, 2000; Kong & Zeng, 2006).

As mentioned in Section 1.3.4 above, the amplitude cue can be further broken down into at least two components: amplitude envelope and amplitude periodicity. Studies have shown that amplitude envelope or amplitude periodicity alone can aid tone perception (Fu & Zeng, 2000;

Kong & Zeng, 2006; Kuo et al., 2008). In Fu & Zeng's (2000) study on Mandarin tone perception, for example, amplitude envelope and amplitude periodicity were each extracted from Mandarin tones produced naturally in isolation with a single-vowel carrier such as /a/, /i/, /u/, /o/, /y/ or /ɣ/. The extraction was done using Elliptical Infinite Impulse Response (EIIR) filters with a slope of 96 dB/octave. A 50Hz-lowpass filter was used for the extraction of amplitude envelope, and a 50Hz–500Hz-bandpass filter was utilized for the extraction of amplitude periodicity. The amplitude envelope or amplitude periodicity extracted from each tone was used to modulate an SCN using the same approach as in Whalen & Xu (1992). Each SCN was then equalized in duration to remove the duration cue. The SCNs, which only retained either the amplitude envelope or the amplitude periodicity of the original tone tokens, were used as stimuli for a 4AFC tone identification task. Using either the amplitude envelope cue or the amplitude periodicity cue alone, the native listeners of Mandarin were able to label the Mandarin tone contained in each SCN with approximately 60% accuracy. Similar results were found in other studies (e.g., Kong & Zeng, 2006; Kuo et al., 2008). Thus, either the amplitude periodicity cue or the amplitude envelope cue alone could provide sufficient acoustic information for the identification of Mandarin tones in native speakers.

Furthermore, a significant effect of tone type was reported in the above studies on Mandarin tone perception using amplitude envelope alone. In Fu & Zeng’s (2000) study, for example, Tone 4 (Falling), was identified with the highest accuracy (78%), followed by Tone 3 (Dipping) (76%), and Tone 1 (High Level) (44%). Tone 2 (Rising) was identified with the lowest accuracy (35%). These differences were statistically significant ($p < 0.001$). Fu & Zeng (2000) suggested that this effect of tone type could be explained on the basis of the correlation coefficient between the F0 contour and the amplitude envelope of each tone type in Mandarin.

As discussed in Section 1.3.5, F0 contour and amplitude envelope co-vary in Mandarin tones. Fu & Zeng (2000) argued that a Mandarin tone with a higher correlation coefficient between its F0 contour and amplitude envelope was identified with a higher degree of accuracy, as illustrated in Figure 3. However, Fu & Zeng (2000) did not report whether the correlation coefficients between the F0 contour and the amplitude envelope of the four Mandarin tones were significantly different.

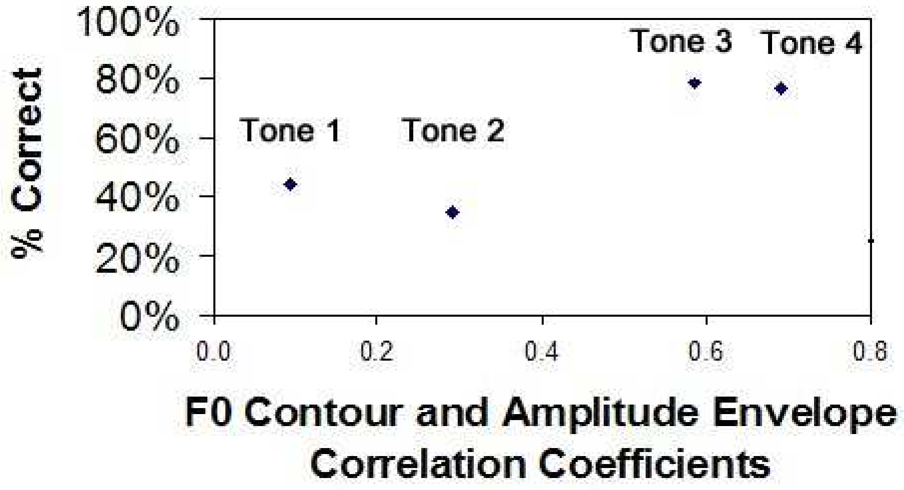

Figure 3: Scatterplot of Mandarin listeners' tone identification performance as a function of the F0 contour and amplitude envelope correlation coefficients of the Mandarin tones. The stimuli were signal-correlated noises containing only the amplitude envelope cue (adapted from Fu & Zeng, 2000).

Fu & Zeng's (2000) argument was further weakened by conflicting findings from Kuo and colleagues' (2008) study. In this later study, Tone 4 (Falling) was identified with the highest degree of accuracy (55%), followed by Tone 2 (Rising) (45%). Tone 3 (Dipping) was identified with the lowest accuracy (31.5%). These differences were statistically significant ($p < 0.001$). Moreover, in Kuo and colleagues' (2008) study, the F0 contour and amplitude envelope correlation coefficients could not be used to predict the outcome of the Mandarin tone identification using amplitude envelope. For instance, Tone 3(Dipping) demonstrated the second highest correlation between its F0 contour and amplitude envelope (r = 0.79) but was identified with the lowest accuracy (31.5%). On the other hand, Tone 1 (High Level) which exhibited the lowest correlation between its F0 contour and amplitude envelope (r = 0.5) was nonetheless identified with higher accuracy (33%) than Tone 3 (Dipping).

It is difficult to reconcile the conflicting findings from Fu & Zeng’s (2000) study and Kuo and colleagues’ (2008) study. First, neither study reported the statistical significance of the F0 contour and amplitude envelope correlation coefficients for the four Mandarin tones. Second, different types of stimuli were used in these two studies. In Fu & Zeng’s (2000) study, the stimuli were SCNs generated using Schroeder’s (1968) method described above. In Kuo and colleagues’ (2008) study, the SCNs were generated by convolving the extracted amplitude envelope with a noise carrier. Thus, it is difficult to determine whether these two types of SCNs conveyed the amplitude envelope of Mandarin tones with equal effectiveness.

Studies of Cantonese tones reveal that amplitude periodicity also allows for perception above chance levels (e.g., Yuan et al., 2007 & 2009; Yuen, Yuan, Lee, Soli, Tong & van Hasselt, 2007; Yuen, Tong, van Hasselt, Yuan, Lee & Soli, 2009). However, no studies have examined the role of amplitude envelope in Cantonese tone perception. In Yuen and colleagues’ (2007) study, for example, Fu and colleagues’ (1998) method was used to extract amplitude periodicity from naturally produced Cantonese tones with the carrier syllable /ji/ or /wai/ in isolation. The amplitude periodicity extracted from each tone was used to synthesize an SCN containing only the amplitude periodicity of the tone. Native listeners of Cantonese in the study were able to identify each Cantonese tone contained in the SCNs with approximately 56% accuracy in a 6AFC task (chance = 16.7%). In a follow-up study (Yuen et al., 2009), disyllabic carriers /hoi gwok/ and /ngoi gwok/ were used instead of the monosyllabic carriers /ji/ and /wai/, and the stimuli were prepared using the same procedure as in the previous study (Yuen et al., 2007). Native listeners of Cantonese identified each Cantonese tone contained in the noise carriers with 64% accuracy in a 6AFC task (chance = 25%). Thus, amplitude periodicity appears to cue tone perception in Cantonese as well as in Mandarin.

It is important to point out that the amplitude and duration cues could only aid tone perception in the absence of explicit F0. When explicit F0 was available, it seemed to override the amplitude and duration cues. In Lin's (1988) study, tones were synthesized using the F0 contour, amplitude contour and duration of naturally produced Mandarin tones. The synthesized tones could have conflicting cues (e.g., having the F0 contour of Tone 1 (High Level), the amplitude contour of Tone 2 (Rising) and the duration of Tone 4 (Falling)). They could also have converging cues (e.g., having the F0 contour, the amplitude contour and the duration of Tone 1 (High Level)). In a 4AFC tone identification task, native listeners of Mandarin recognized the tonal identities of the tones with converging cues with 98.8% accuracy, and labeled the tones with conflicting cues according to the F0 contour 94% of the time. Thus, the explicit F0 cue seemed to override the amplitude cue and the duration cue.

In summary, F0 appeared to be the most dominant acoustic cue for the perception of Mandarin and Cantonese tones in native listeners with normal hearing. In the absence of an explicit F0 cue, tone perception is possible with other acoustic cues such as harmonic structures, formant structures, amplitude and duration. In particular, the amplitude envelope cue or the amplitude periodicity cue alone has been reported to aid tone perception in Mandarin with significantly above-chance accuracy in native listeners with normal hearing.

1.4.4 Acoustic vs. Linguistic Processing of Tones

Tone perception is modulated by both acoustic and linguistic factors (e.g., Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda & Keller, 1996; Lee, Vakoch & Wurn, 1996). In Lee and colleagues' (1996) study on tone discrimination in native listeners and non-native listeners, for example, Cantonese tone pairs with real word and non-word carrier syllables were presented to native listeners of Cantonese and monolingual English listeners. The native listeners of Cantonese performed significantly better in the real word condition (98% correct) than in the non-word condition (89% correct). The native listeners' superior performance in the real word condition could be attributed to the availability of acoustic, phonemic and lexical cues in the real word condition, but only acoustic and phonemic cues in the non-word condition. Thus, the role of lexical processing is important in the perception of Cantonese tones in real words.

The relative contribution of acoustic processing and phonemic processing to Cantonese tone perception could be estimated by comparing the performance of the native Cantonese listeners and that of the monolingual English listeners in Lee and colleagues' (1996) study. In the discrimination of Cantonese tones in the non-word condition, both acoustic and phonemic cues were available to native Cantonese listeners, but only acoustic cues were present for the monolingual English listeners because tonal contrasts are phonemic in Cantonese, but not in English. The availability of phonemic cues in addition to acoustic cues for the native Cantonese listeners is likely to explain why they performed better (mean = 89% correct) than the monolingual English listeners (mean = 65% correct) in the non-word condition.

The reliance on phonemic processing can be manipulated by lengthening the interstimulus interval (ISI) (e.g., Burnham, Francis, Webster, Luksaneeyanawin, Attapaiboon, Lacerda & Keller, 1996; Werker & Logan (1985). In Werker & Logan's (1985) behavioral study on phoneme discrimination in native speakers of English, for example, an ISI of 1500 ms seemed to favor phonemic processing, whereas an ISI of 500 or 250 ms allowed for acoustic processing. The stimuli used in this experiment were two variants of /t/ (i.e., a dental [t] and a retroflex [ṭ]) produced with the vowel /a/ in a CV syllable. In each block of stimuli, pairs of [ta] vs. [ta], [ta] vs. [ṭa], and [ṭa] vs. [ṭa] were included, and the silent gap between the two stimuli within each pair varied from one block to another (i.e., 250, 500 or 1500 ms). The pair [ta] vs. [ṭa], which did not constitute a phonemic contrast in English, was more often perceived as identical with a silent gap of 1500 ms than with a silent gap of 250 or 500 ms. Thus, the pair [ta] vs. [ṭa] was processed more categorically, according to the phonemic system of English, with the silent gap of 1500 ms. But with a shorter silent gap of 250 or 500 ms, English listeners could perceive the within-category distinction. This finding supports the claim that acoustic information is more transient, and the use of longer ISIs will force listeners to rely on phonemic information

Similar findings have been reported in behavioral studies of tone perception. In Burnham and colleagues' (1996) study on Thai tone and musical tone discrimination in native speakers of Thai, for example, pairs of naturally produced Thai tones were discriminated significantly faster with a 1500-ms silent gap between the two tone tokens within a pair than with a 500-ms silent gap. In contrast, pairs of violin pitches were discriminated significantly faster with a 500-ms silent gap between two tones within a pair than with a 1500-ms silent gap. Furthermore, with the 1500-ms silent gap, Thai tones were discriminated significantly faster than the violin pitches. Thus, for the native speakers of Thai, an ISI of 1500 ms seemed to facilitate phonemic processing, whereas an ISI of 500 ms appeared to favor acoustic processing.

In summary, tone perception can be influenced by acoustic and phonemic information. The relative reliance on acoustic versus phonemic information in tone perception may be teased apart by manipulating the ISI in the stimulus presentation, or by comparing the performance of native listeners versus non-native listeners of a tone language. For non-native listeners of a tone language, the processing is presumably acoustically based because tones are not phonemically contrastive.

1.4.5 Measurement Used in Previous Studies on Tone Discrimination

In the tone discrimination studies discussed above, percent correct and reaction time were used to measure participants' perception abilities. Reaction time is easy to measure and is not affected by ceiling effects (e.g., Salthouse & Hedden, 2002). As a measure of response latencies, reaction time is sensitive to between-group and between-condition differences, and therefore can effectively differentiate groups and experimental conditions (e.g., Salthouse & Hedden, 2002). Unfortunately, reaction time cannot measure the response accuracy, and thus has limited use.

In contrast, percent correct measures or number of correct responses can help determine response accuracy in a tone perception experiment. Unfortunately, they are negatively affected by floor or ceiling effects, and do not take participants' perceptual biases into account, which could be problematic in some situations. Consider, for instance, the following hypothetical experiment which presented pairs of identical or different stimuli and required participants to press the "Same" or "Different" button upon hearing each pair of stimuli. 80% of the pairs presented consisted of different stimuli, and the remaining pairs included identical stimuli. At the end of the experiment, two participants both scored 80% correct, but Participant A had no false alarm whereas Participant B's false alarm rate is 100%. Clearly, Participant B had a complete perceptual bias in favor of the ***Different*** pairs at the expenses of the ***Same*** pairs. That is, she always pressed the ***Different*** button regardless of the stimuli presented, and thus her percent correct score of 80% was not a useful estimate of her perceptual ability. On the other hand, Participant A did not have a significant bias for ***Different*** or ***Same*** pairs, and her percent correct score of 80% was likely to be a fair measure of her perceptual ability. This example illustrates how percent correct measures could be misleading due to the existence of perceptual biases.

With the advent of signal detection theory, several measures have been proposed to control for participants' perception bias while measuring their perception ability (e.g., Grier, 1971; Hodos, 1970; Donaldson, 1993). In particular, two sets of these measures have withstood rigorous empirical tests in academic and clinical studies (MacMillan & Creeelman, 1991; Swets, 1996; Wickens, 2002). The first set of measures consists of d' (a measure of perceptual ability) and β (a measure of perceptual bias) (Swets, 1986). The second set includes A' (a measure of perceptual ability) (Grier, 1971) and B' (a measure of perceptual bias) (Hodos, 1970). d' and β measures have two major limitations. First, they should not be computed if there is a ceiling or floor effect in the ***Different*** responses or ***Same*** responses. Second, they assume normal distribution and equal variances in the ***Different*** responses and ***Same*** responses. If these assumptions are violated, d' and β measures will no longer be accurate estimates of perceptual ability and bias. In this case, A' and B' measures would be more appropriate estimates of perceptual ability and bias because they do not assume normal distribution and equal variances. For this reason, A' and B' measures are often considered "nonparametric."

Although recent studies have questioned the "nonparametric" nature of A' and B' measures (e.g., Macmillan & Creelman, 1996; Pastore, Crawley, Berens, & Skelly, 2003), scholars (including the critics of these measures) agree that A' and B' do not require the normal distribution and equal variance assumptions (see Pastore, Crawley, Berens, & Skelly, 2003). More importantly, these measures have a rather simple interpretation: A’ measures range from 0.5 (chance level performance) to 1 (perfect perceptual performance) (e.g., Grier, 1971). B’ measures go from -1 (extreme bias in favor of the ***Same*** responses) to 1 (extreme bias in favor of the ***Different*** responses), with the value of 0 meaning no response bias (e.g., Hodos, 1970).

In summary, as measures of overall perceptual accuracy and bias in a discrimination experiment, d’ and β measures or A’ and B’ measures present several advantages over traditional measures such as number of correct responses and percent correct measures. When the assumption of normal distribution and equal variances does not hold, d’ and β measures lose their advantages, and A’ and B’ measures would be more appropriate. Combined with reaction time measures, A’ and B’ measures can capture response accuracies, biases and latencies even if the assumption of normal distribution and equal variances is violated.

1.5 Rationale for the Current Dissertation

Currently available commercial cochlear implant systems have been reported to be ineffective in encoding the explicit F0 cue and the harmonic structures of a sound due to the poor spectral resolution in the cochlear implant systems (e.g., Loizou, 2006; Wilson & Dorman, 2008). Experimental speech processing strategies (e.g., Luo & Fu, 2004) have demonstrated promising clinical results in improving tone perception in cochlear implant users by enhancing the encoding of the amplitude envelope cue. Thus, studies on the role of amplitude envelope in tone perception may help improve speech processing strategies for cochlear implant users who speak a tone language.

Given the importance of research on the role of amplitude envelope in tone perception, this dissertation was designed to address some of the unresolved issues regarding the independent contribution of amplitude envelope to tone perception in individuals with normal hearing. As mentioned above, previously published studies on the role of amplitude envelope in tone perception focused on Mandarin. Each of the four Mandarin tones has a distinct contour, and tone contour alone can be used to perceive tone categories. Therefore, previously published studies did not investigate whether amplitude envelope could be used to cue tone perception in tone languages such as Cantonese, Thai & Taiwanese which must use both tone contour and tone height for accurate perception of phonemic tonal contrasts.

Cantonese is a good candidate for the investigation of whether amplitude envelope can systematically cue tone contour and tone height for tone perception. As discussed in Section 1.1 above, tone languages use tone height, tone contour or a combination of both for phonemic tonal contrasts, and Cantonese is one of the few languages using both tone height and tone contour for phonemic tonal contrasts. Thus, if amplitude envelope can cue tone perception in Cantonese, it should theoretically be possible to cue tone perception in any tone language,

The independent contribution of amplitude envelope to Cantonese tone perception was examined in a preliminary study in preparation for this dissertation (Zhou, 2010). Using the same approach as in Fu & Zeng (2000), amplitude envelope was extracted from six Cantonese tones produced naturally in isolation with the carrier syllable /ji/. As illustrated in Figure 4, the falling tones (Tone 1 and Tone 4) had amplitude envelopes with a falling slope; the rising tones (Tone 2 and Tone 5) exhibited amplitude envelopes with a rising slope, whereas the amplitude envelopes of the level tones (Tone 3 and Tone 6) were relatively flat. Moreover, the high tones demonstrated amplitude envelopes in the relatively higher amplitude range, whereas the low tones display amplitude envelopes in the relatively lower amplitude range. Thus, each Cantonese tone type has an amplitude envelope with a specific contour and a relative height resembling its F0 contour and relative F0 height, as illustrated in Figure 4. More specifically, the relative amplitude height of a Cantonese tone is associated with its F0 height, whereas the contour of the amplitude envelope of a Cantonese tone is related to its F0 contour.

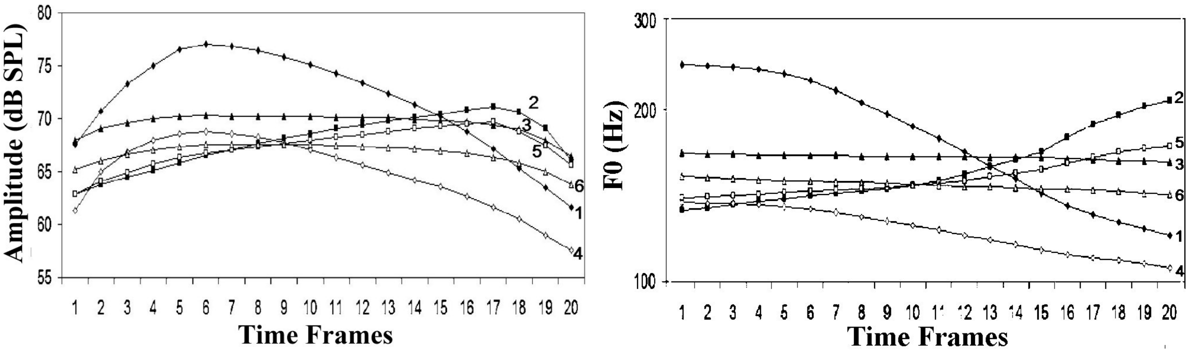

Figure 4. Amplitude envelope (left panel) and F0 contour (right panel) of each Cantonese tone produced with the syllable /ji/ in isolation. Each tone was divided into 20 equal time frames. The root-mean-square (RMS) amplitude and mean F0 within each time frame was averaged across 20 tone tokens produced by one male talker and one female talker (adapted from Zhou, 2010).

A linear correlation analysis revealed a fair correlation between the F0 contour and the amplitude envelope (abbreviated as ***F0-AE covariability*** hereafter) for each of the six Cantonese tones, with Pearson's correlation coefficients ranging from 0.45 to 0.71, as illustrated in Figure 6. There were significant differences across tone types ($F_{(1,4)} = 29.62$; $p < 0.001$). In particular, the falling tones exhibited a higher correlation coefficient ($r = 0.71$) than the rising tones ($r = 0.53$; $t_{(3)} = 8.53$; $p < 0.006$) which in turn demonstrated a higher coefficient than the level tones ($r = 0.45$; $t_{(3)} = 9.45$; $p < 0.005$). These correlation patterns were similar to those observed in Fu & Zeng (2000) and Kuo et al. (2008). That is, the falling tones exhibited a higher degree of F0-AE covariability than the rising tones which in turn demonstrated a higher degree of F0-AE covariability than the level tones.

In order to determine whether the amplitude envelope cue alone could aid Cantonese tone perception in a preliminary study of the currently proposed dissertation, the amplitude envelope of each Cantonese tone was used to modulate an SCN according to Fu & Zeng’s (2000) approach. As discussed in Section 1.3.5 above, each SCN contained only the amplitude envelope of a Cantonese tone. Yet, native listeners of Cantonese (n=20) were able to identify the Cantonese tone contained in each SCN with approximately 40% accuracy in a 6AFC tone identification task in the preliminary study (chance level = 16.7%). There was a significant effect of tone type ($F_{(5,95)} = 29.59$; $p < 0.001$). The falling tones and the rising tones were identified with higher accuracy than the level tones ($p < 0.001$), as illustrated in Figure 5. This could be accounted for on the basis of the F0-AE covariability. As illustrated in Figure 6, the falling tones (Tone 1 and Tone 4) and the rising tones (Tone 2 and Tone 5), which were identified with higher accuracy than the level tones (Tone 3 and Tone 6), had higher F0 and amplitude correlation coefficients than the level tones (Tone 3 and Tone 6). However, the relationship between the F0-AE covariability and Cantonese tone identification using amplitude envelope was not linear. For instance, as discussed above, the falling tones (Tone 1 and Tone 4) had significantly higher F0-amplitude correlation coefficients than the rising tones (Tone 2 and Tone 5), but there was no statistically significant difference in the identification of falling tones versus rising tones using amplitude envelope alone ($p = 0.28$). Thus, the degree of F0-AE covariability could not be used to fully predict the identification of Cantonese tones using amplitude envelope alone.

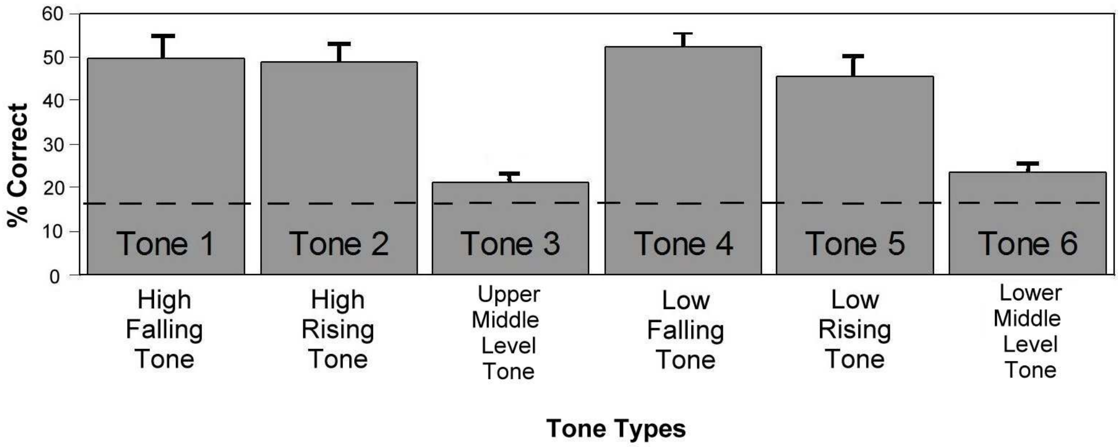

Figure 5: Cantonese tone identification using amplitude envelope. The chance level, represented by the dotted line, was 16.7%. The error bars represent one standard error from the group average (adapted from Zhou, 2010).

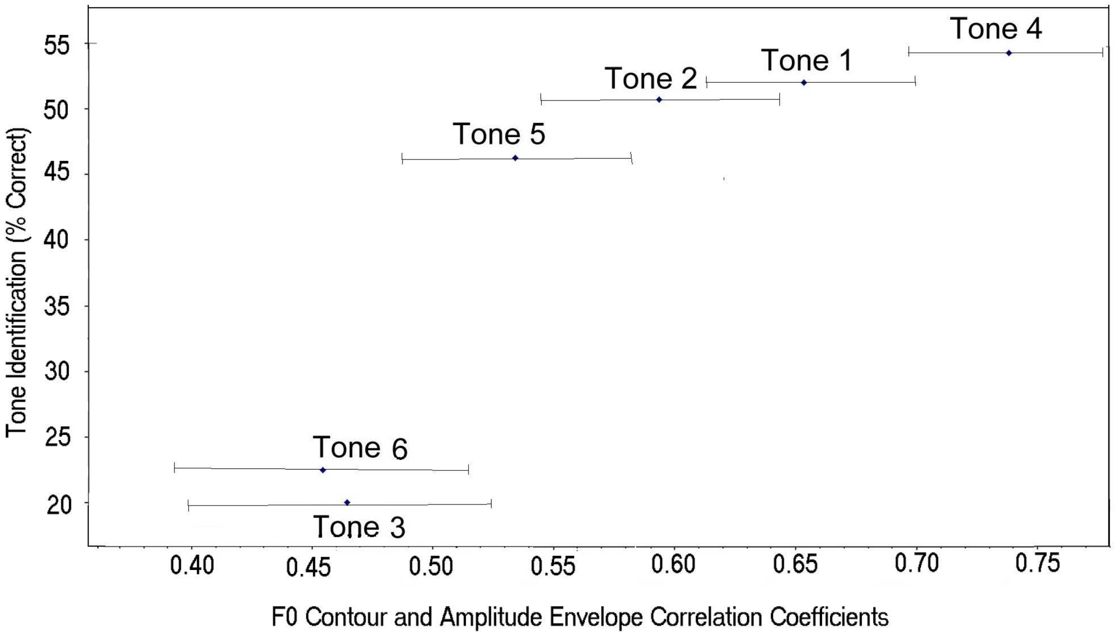

Figure 6: Scatterplot of twenty native Cantonese listeners’ tone identification performance as a function of the F0 contour and amplitude envelope correlation coefficients of Cantonese tones. The horizontal whiskers represent the 95% confidence intervals of the correlation coefficients. The stimuli used were signal-correlated noises containing only the amplitude envelope cue (adapted from Zhou, 2010).

Analysis of the perceptual patterns indicated that a falling tone (i.e., Tone 1 or Tone 4) was identified as a falling tone 74% of the time, a rising tone (i.e., Tone 2 or Tone 5) was identified as a rising tone approximately 83% of the time, and a level tone was identified as a level tone only 36% of the time (chance level = 33.4%). Thus, on average, the tone contour was identified with 54.2% accuracy (SD = 0.07; chance = 33.4%), which was significantly above chance ($t_{(19)} = 13.8$; $p = 0.001$). However, it was puzzling that the level tones were mis-identified as rising tones 45% of the time.

Analysis of the perceptual patterns also revealed that the high tones (i.e., Tone 1, Tone 2 & Tone 3) were differentiated from the low tones (Tone 4, Tone 5 & Tone 6) with 59.1% success using the AE cue (chance = 50%; SD = 0.08; $t_{(19)} = 5.2$; $p = 0.002$). Thus, on average, the tone height was recognized by the listeners significantly above chance level. However, it was puzzling that Tone 1 (High Falling) was mis-identified as Tone 4 (Low Falling) 39% of the time, and Tone 5 (Low Rising) was mis-identified as Tone 2 (High Rising) 37% of the time (chance = 16.7%).

It was also puzzling that, in previously published studies on tone perception using amplitude envelope alone, some tone types were identified with much lower accuracy than others, and no consistent patterns were observed across these studies. In Fu & Zeng’s (2000) study on Mandarin, for example, the rising tone was identified with significantly the lowest accuracy (mean = 35.5%; chance = 25%). In Kuo and colleagues’ (2008) study on Mandarin, the dipping tone produced significantly lower identification performance (mean = 31.5%; chance = 25%) than all other tones. The puzzling findings in these previously published studies could not be fully explained yet, as discussed in Section 1.4.3 above.

In these previously published studies as well as in the preliminary study of this dissertation, only word identification tasks were used, and the stimuli were SCNs re-synthesized using either Fu & Zeng’s (2000) method or Kuo and colleagues’ (2008) approach. Upon hearing each stimulus, participants were to identify, from a list of written Chinese characters differing only in tones, the character representing the stimulus. As mentioned above, the SCNs were wide-band noises re-synthesized from single-syllable words that had been low-pass-filtered with the cut-off frequency of 50 Hz. These SCNs contained nothing but the amplitude envelope of the original words. Yet, participants were asked to identify, from the list of written Chinese words differing in only tones, the original word on the basis of which each SCN had been re-synthesized. This was a very difficult task because the participants had to use the auditorily presented SCN to retrieve a lexical item from their mental lexicon, and then match the retrieved lexical item with one of the written Chinese words presented on the computer screen or on paper. In fact, a few participants in the preliminary experiment of the current study complained about the difficulty of the task.

One way to alleviate this problem is to ask the listeners to name the tone contained in an auditorily presented SCN. Unfortunately, as mentioned in Section 1.4, naïve native listeners of Cantonese are typically unable to name the tone of a spoken Cantonese word. Thus, a better alternative would be to use a pair-wise tone discrimination paradigm. This paradigm would only require listeners to report whether two auditorily presented SCNs contain the same tone or two different tones.

In the current study, a tone discrimination paradigm was used to investigate the independent contribution of amplitude envelope to Cantonese tone perception. The stimuli were SCNs re-synthesized as in the preliminary study (see Section 2 for more detail). Each SCN was a wide-band noise containing only the amplitude envelope of a Cantonese tone produced in isolation with a carrier syllable. There were two experimental conditions: in Condition One, each SCN was presented at the original intensity level of the Cantonese tone based on which the SCN was re-synthesized (see Section 2 for more detail). This was to preserve the original height and contour of the amplitude envelope of each Cantonese tone in order to determine the combined contribution of the contour and the relative height of the amplitude envelope to Cantonese tone perception.

In Condition Two, the amplitude of the SCNs was equalized to the grand average peak amplitude of the Cantonese tones based on which the SCNs were re-synthesized. This equalization would eliminate the peak amplitude differences among the six Cantonese tones and hence reveal the effect of removing the relative peak amplitude differences on Cantonese tone perception using amplitude envelope. This would in turn isolate the independent contribution of the contour of the amplitude envelopes to Cantonese tone perception using amplitude envelope. For instance, if Condition Two produces significantly inferior results than Condition One, it would suggest that the relative peak amplitude differences among the Cantonese tones is important for Cantonese tone perception using amplitude envelope alone. Thus, the comparison of the results from Condition One and Condition Two would help tease apart the relative contribution of the contour and the relative height of amplitude envelope to Cantonese tone perception in native listeners of Cantonese.

An alternative approach to tease apart the relative contribution of the height and the contour of the amplitude envelope would be to rove the RMS or peak amplitude of the SCNs randomly in Condition Two (i.e. to shift the RMS or peak amplitude of some SCNs up or down randomly). However, this alternative approach produced undesirable side effects in a pilot study involving five native listeners of Cantonese. For instance, when the peak amplitude of some SCNs was randomly shifted up or down, a pair of tone tokens such as Tone 1 (High Falling) vs. Tone 1 (High Falling) might or might not be perceived as identical by the native listeners, depending on whether the peak amplitude of the two tone tokens were shifted in the same direction (i.e., up or down) and by the same magnitude. Due to this kind of undesirable side effect, the amplitude roving approach was rejected, and the amplitude equalization was used instead in the current study. The amplitude equalization approach had the advantage of avoiding the above-mentioned undesirable side effects while achieving the objective of teasing apart the relative contribution of the height and the contour of amplitude envelope to Cantonese tone perception in the current study.

The current study not only involved native listeners of Cantonese (abbreviated as ***the Cantonese group*** hereafter), but also English-speakers with no formal or informal learning of a tone language (abbreviated as ***the English group*** hereafter). The reason for including the English group in the current study was the following. As discussed in Section 1.4.4 above, native speakers of a tone language discriminate native tone contrasts in non-words more accurately than non-native speakers, probably because the native speakers have access to phonemic cues in addition to acoustic information, whereas the non-native speakers only have access to acoustic cues (e.g., Lee et al., 1996). The SCNs used in the current study were comparable to non-words since they only contained the amplitude envelopes of the original words based on which they were re-synthesized. Thus, if the Cantonese group in the current study could discriminate the Cantonese tone contrasts contained in the SCNs more accurately than the English group, it would suggest that the Cantonese group had access to phonemic cues in addition to acoustic information, whereas the English group only had access to acoustic cues. It would also suggest that the SCNs were effective in conveying the phonemic contrasts of the Cantonese tones.

Four specific questions were to be answered in the current study:

(1) **Can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts?**

a) More specifically, can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts involving a level tone and a contour tone?

b) In addition, can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts involving tones with similar contours but different height as listed in Appendix Four?

(2) **Using the amplitude envelope cue alone, do native listeners of Cantonese discriminate**

a) Cantonese tone contrasts in one carrier syllable better than in the other?

b) some pairs of Cantonese tone contrasts better than others?

If so, can the results be accounted for on the basis of the degree of F0-AE covariability or other factors?

(3) **Using the amplitude envelope cue alone, does the Cantonese group discriminate Cantonese tone contrasts more accurately than the English group?**

(4) **What are the relative contributions of the contour and the height of amplitude envelope to Cantonese tone perception using amplitude envelope alone?**

2. Methods

2.1 Participants

Seventy adults between twenty-five and forty years of age participated in the current study. Two native speakers of Cantonese (one male and one female) produced Cantonese tones based on which the stimuli for the study were re-synthesized (see Sections 2.2 and 2.4 for details). Another seven native speakers of Cantonese served as naïve judges to rate the perceptibility of the Cantonese tones produced by the two native speakers of Cantonese. Thirty additional native speakers of Cantonese participated in the perceptual experiment of the current study, and were randomly coded from C1 to C30. Thirty English-speaking adults with no formal or informal learning experience of a tone language also participated in the perceptual experiment of this study. They were randomly coded from E1 to E30. All participants had a pure-tone air conduction thresholds of 25 dB HL or better bilaterally at octave frequencies between 250 Hz and 8000 Hz (ASHA, 1997). They also had Type A tympanograms (ASHA, 1997) and present ipsilateral acoustic reflexes at 90dB HL at 1000 Hz.

2.2 Recording and Selection of the Stimuli

The stimuli used in the current study were SCNs re-synthesized from Cantonese tones produced in isolation with the carrier syllable /ji/ or /wai/. The use of tones produced in isolation as stimuli for this study was to control for several variables affecting Cantonese tone perception, such as co-articulation (e.g., Li, Lee & Qian, 2002), intonation (e.g.,Wong, 1999) and emphatic stress (e.g., Gu & Lee, 2007). The rationale for using the carrier syllables /ji/ and /wai/ was many-fold. First, the carrier syllables /ji/ and /wai/ were used in previous studies on the role of amplitude in Cantonese tone perception (i.e., Yuan et al., 2007; Yuen et al., 2007). Thus, using these two carrier syllables in the current study would facilitate the comparison of the results of this study with those of previous studies on Cantonese tone perception using amplitude cues only. Second, each of the two carrier syllables could be combined with each of the six Cantonese tones to constitute a meaningful and common word in Cantonese (Yuan et al., 2007; Yuen et al., 2007), as illustrated in Table 3 above. This would ensure that the carrier syllables used in the current study represent familiar words occurring frequently in Cantonese daily conversation. Third, the two carrier syllables have very different formant structures (Yuan et al., 2007; Yuen et al., 2007). As discussed in Section 1.3.3 above, formant structures affect tone perception. Therefore, using carrier syllables with different formant structures in the current study would shed light on the possible effect of formant structures on tone perception using amplitude envelope alone. Lastly, in the previous studies using /ji/ and /wai/ as the carrier syllables, a significant effect of carrier syllable was found (Yuan et al., 2007; Yuen et al., 2007). Although the authors fell short of providing further detail on the significant effect of carrier syllable, the mere finding of a significant effect of carrier syllable suggested that the amplitude of a tone might be affected by the type of carrier syllable. Therefore, it would be advantageous to include several types of carrier syllables in the current study. Unfortunately, using too many carrier syllables would lengthen the time required for the perceptual experiment exponentially and could cause excessive fatigue to the participants. Using only /ji/ and /wai/ as the carrier syllables would limit the experiment to within three hours, thereby reducing the potential effect of fatigue.

Stimulus recording was done with each of the Cantonese talkers individually, with a Technica Audio AT892CT4 head-mounted microphone positioned directly in front of the talker in a sound-attenuated booth. The microphone was connected to a digital recording device, a Dell Dimension E521 computer with a Sigma C-Major Audio sound card, located in an adjacent sound-attenuated booth. The recording software Sound Forge was used, with a sampling rate of 44010Hz, and a resolution of 16 bits. Twelve written Chinese characters representing the twelve target syllables (see Table 3) were individually presented on a computer screen to each talker, with a two-second pause between any two characters. Each talker was instructed to read each character aloud and as naturally as possible, as if he/she were reading the letters on an eye exam chart during a vision examination. Each character was presented ten times, in a random order with the constraint that no character be presented twice or more in a row. Thus, the two talkers each produced ten tokens of each target syllable, yielding a total of 240 tone tokens recorded digitally.

The 240 tone tokens were further analyzed using the software Praat and according to Fu & Zeng's (2000) procedure. Only two tokens produced by each talker for each target syllable were selected as stimuli, using the following procedure. First, each of the ten tone tokens produced by each talker with each target syllable was divided into twenty equal time frames, and the RMS amplitude and mean F0 within each time frame were computed. The F0 was estimated using a short-term autocorrelation method (Rabiner, 1977). Then, the mean F0 within each time frame was averaged across the ten tokens. Similarly, the RMS amplitude within each time frame was averaged across the ten tokens. Lastly, the two tokens with the least deviation from the ten-token average were selected as the exemplars or prototypes of that target syllable of that talker. With twelve target syllables and two talkers in the current experiment, a total of forty-eight target syllable prototypes were selected.

The forty-eight target syllable prototypes were then equalized in duration to remove the duration cue for the tone identification task in the experiment because the purpose of this experiment was to determine whether native listeners of Cantonese can use the amplitude envelope cue alone to identify Cantonese tones. The duration-equalizing procedure consisted of two steps. First, the grand average duration of all of the tone tokens produced by both talkers were computed. Then, the duration of each of the forty-eight target syllables were equalized to the grand average duration using a linear interpolation method (Fu & Zeng, 2000). For instance, suppose $f(n)$ is the output function of a target syllable at any sample point n. If the duration of the target syllable was t ms and the grand mean duration of all target syllables was T ms, then the duration of the target syllable was increased to T ms using the following linear interpolation formula (Fu & Zeng, 2000):

$$g(n) = f(m) + a[f(m+1)-f(m)],$$

where $g(n)$ is the output of the linear interpolation at a given sample point n,

m is the integral part of the product $n*t/T$, and

a is the remainder of the product $n*t/T$.

The impact of this duration equalization on the perceptibility of the target syllables was assessed using an identification task performed by seven naïve native listeners of Cantonese. The stimuli were presented to the naïve listeners using a Dell Optiplex computer with the Excel/Visual Basic software. While seated in a sound-attenuated booth at the Graduate Center, the judges listened to the stimuli binaurally via TDH-50 headphones, with the peak intensity level of stimuli calibrated to 70 dB SPL. This intensity level was used in the preliminary study (Zhou, 2010), and was considered comfortably audible by all participants. The stimuli were separated by talker because previous studies suggested that native listeners of Cantonese needed to hear several tokens produced by a talker in order to familiarize themselves with the talker's pitch range and to recognize the talker's tones (Wong & Diehl, 2003). With one male talker and one female talker in the current study, there were two sets of tone prototypes. Within each set, the tone prototypes were further divided into two blocks according to the carrier syllable (i.e., /ji/ and /wai/). Thus, there were a total of four blocks of tone prototypes, and these blocks were presented to the judges in a random order using Excel/Visual Basic. Each block contained ten identical tokens of each tone prototype produced by a talker. Each tone token was embedded in the carrier sentence “/m^{41} gɔi^{51} hyn^{55} tsy^{22} gɔ^{33} _____ tsi^{22}/ (“Please circle the character _____”). The presentation of each block consisted of the following parts:

1. A pre-recorded introduction in Cantonese was given by the talker of the stimuli. The English translation of the introduction was as follows:

   > “There is an answer sheet (see Appendix One) in front of you. Please look at the six large-print words on the answer sheet as I read them aloud one by one.”

   Each large-print word represented the written form of each of the stimuli to be presented in the block. The talker read the six words twice and in the same order as they were printed on the answer sheet. The purpose of this introduction was to familiarize the listeners with the talker's voice and tones.
2. Pre-recorded instructions in Cantonese produced by the talker were then presented. The English translation of the instructions was:

   > “Now you will hear each of the above Cantonese words many times. Upon hearing each word, you must circle one and only one character on the answer sheet, even if you have to guess.”
3. Each stimulus was presented with the carrier sentence in a random order with an ISI of three seconds.

This was a 6AFC identification task, and the a priori passing criteria was that at least six of the seven judges must each identify each tone prototype produced by each talker with 90% accuracy (chance level = 16.67%) . The actual results were better than the a priori passing criteria: each of the seven judges identified each tone type produced by each talker with each carrier syllables with 90% or higher accuracy, which was comparable to those reported in previous studies (e.g., Khowe & Ciocca, 2007). Thus, the forty-eight duration equalized tone prototypes presented in this rating session were later used to generate the SCN stimuli for the tone discrimination experiment in the current study.

2.3 Synthesis of the Stimuli

For ease of discussion, the forty-eight duration-equalized target tone prototypes will be labeled ***natural tones*** below. Having passed the above rating procedure, the natural tones were further processed using Praat software according to the procedure used in Fu & Zeng (2000) to extract the amplitude envelope. Each natural tone was first half-wave rectified, and low-pass filtered using an EIIR filter with a cut-off frequency of 50 Hz and a slope of 96 dB/octave (Fu & Zeng, 2000). The waveform of the extracted amplitude envelope was transformed into an SCN by randomly changing the sign of half of the sample points in the waveform (Fu & Zeng, 2000; Whalen & Xu, 1992). This type of SCN is a wide-band noise retaining only the amplitude envelope cue of a natural tone. In the current study, each SCN was labeled a ***processed tone***. Given forty-eight natural tones selected for the current study, forty-eight processed tones were generated, and were used as stimuli in the tone discrimination experiment.

For Condition One of the tone discrimination experiment, the original amplitude (or intensity level) of each of the processed tone was preserved. Thus, both the peak amplitude of each processed tone was maintained at the original intensity level of the natural tone token based on which the processed tone was generated. For Condition Two, the peak amplitude (or intensity level) of each processed tone was equalized to the grand mean peak amplitude of all the forty-eight processed tones in order to eliminate the peak amplitude differences among the six Cantonese tones. As discussed in Section 1.5 above, this would help determine the independent contribution of the contours of the amplitude envelopes to Cantonese tone perception. At the same time, the amplitude equalization could reveal the effect of the relative height of amplitude envelope to Cantonese tone perception.

2.4 Experimental Procedure

The twenty-four processed tones based on the female talker's natural tones were separated from those of the male talker's, which yielded two blocks of stimuli for each condition of the tone discrimination experiment, and a total of four blocks for the whole experiment. Thus, each block contained twenty-four processed tone tokens produced by the same talker, twelve of them with the carrier syllable /ji/, and the remaining twelve with the carrier syllable /wai/. The twelve processed tone tokens with each carrier syllable fell into six Cantonese tone categories, with two exemplars for each category. The six tone categories with each carrier syllable were arranged into thirty-six pairs of tone contrasts as shown in Appendix Four. Each tone pair occurred randomly ten times in a block. Thus, there were 2 (carrier syllables) x 36 (tone pairs) x 10 (tokens), totaling 720 pairs of stimuli in each talker block in each condition.

The thirty-six tone pairs consisted of thirty ***Different*** pairs (e.g., Tone 1 vs. Tone 2), and six ***Same*** pairs (i.e., Tone 1 vs. Tone 1, Tone 2 vs. Tone 2, …, Tone 6 vs. Tone 6) produced by the same talker with the same carrier syllable. For each ***Same*** pair, two different exemplars of the same tone type were used. The amplitude envelopes of the two exemplars were similar but not identical as shown in Figure 7, which would favor phonemic processing instead of acoustic processing in the Cantonese group in the current tone discrimination experiment (Strange, 2010). For each ***Different*** pair, on the other hand, only the exemplar with the least deviation from the ten-token average (see Section 2.2) was used.

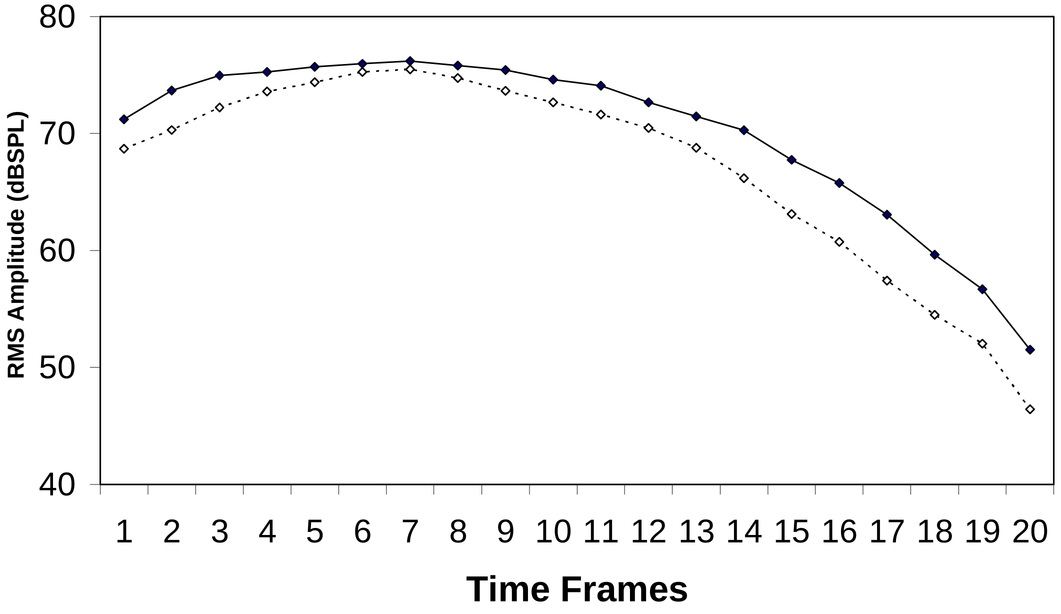

Figure 7. Amplitude envelopes of the two tokens of Tone 1 (High Falling) produced by the male talker with the carrier syllable /ji/.

There was a practice session before the tone discrimination experiment. Each listener was seated in a sound-attenuated booth at the Graduate Center, and listened to the stimuli via a pair of TDH-50 headphones. The four blocks of stimuli were presented in a random order via Excel/Visual Basic, and the presentation of each block consisted of the following parts:

1. A pre-recorded introduction was given by the talker of the stimuli. The introduction was in Cantonese for the Cantonese group and in English for the English group, but the

content of the introduction was the same for both groups. The English version of the introduction was as follows:

- “There are six tones in Cantonese. Here they are.” (The six processed tones contained in the block were then presented in a random order, with a one-second pause between two tones.)
- “Two of these tones have a falling pitch.” (The High Falling Tone and the Low Falling Tone were then presented respectively, with a one-second pause between them.)
- “The first tone has a higher pitch level than the second tone. Please listen again.” (The High Falling Tone and the Low Falling Tone were then presented once again, with a one-second pause between them.)
- “Another two Cantonese tones have a rising pitch.” (The High Rising Tone and the Low Rising Tone were then presented respectively, with a one-second pause between them.)
- “The first tone has a higher pitch level than the second tone. Please listen again.” (The High Rising Tone and the Low Rising Tone were then presented once again, with a one-second pause between them.)
- “The remaining two tones have a level pitch. (The Upper-Middle Level Tone and the Lower-Middle Level Tone were then presented respectively, with a one-second pause between them.)
- “The first tone has a higher pitch level than the second tone. Please listen again.” (The Upper-Middle Level Tone and the Lower-Middle Level Tone were then presented once again, with a one-second pause between them.)
- “Once again, here are the six tones in Cantonese.” (The six tones were then presented once again in the following order with a one-second pause between any two tones: High

Falling Tone, Low Falling Tone, High Rising Tone, Low Rising Tone, Upper-Middle Level Tone, and Lower-Middle-Level Tone.)

2. Pre-recorded instructions were given by the talker of the stimuli. The instructions were in Cantonese for the Cantonese group and in English for the English group, but the content of the instructions was the same for both groups. The English version was as follows:

   > "You will hear pairs of Cantonese tones now. Upon hearing each pair, you must decide whether the pair consists of the same Cantonese tone or two different Cantonese tones, and then you must click the left or the right button on the mouse accordingly."

   For the odd-numbered listeners in each language group, the left button on the mouse was labeled "Same", whereas the right button was labeled "Different." For even-numbered listeners, the left button on the mouse was labeled "Different", whereas the right button was labeled "Same." All participants were right-handed, and decided on their own to use their right hands to press the mouse buttons. They were instructed to position their right index finger between the left button and the right button before and after pressing a mouse button.

3. Each of the thirty-six pairs of tone contrasts produced by the talker with each of the carrier syllables were presented once in a random order. There was a 1500-ms silence between the two tones of each pair, and a 4-second interval between the onsets of two adjacent pairs. The responses were recorded and processed in Excel/Visual Basic, and no feedback was provided.

The practice session lasted approximately twenty minutes per listener. The actual experiment was conducted in the same way as the practice session with two exceptions. First, the pre-recorded introduction was not presented in the actual experiment. Second, ten identical tokens of each pair of tone contrasts produced by each talker with each carrier syllable were presented in each condition in the actual experiment. The actual experiment lasted approximately three hours per listener, including a brief break between blocks.

At the end of the experiment, there was a brief survey in which each listener verbally explained their decision-making criteria used in the discrimination experiment. After the survey, each listener was given a brief tone identification test in order to assess their ability to process natural tones in Cantonese. The procedure and the stimuli were the same as those used in the rating of the natural tones (see Section 2.3 above), with only two exceptions. First, each of the stimuli used in the brief tone identification test was not embedded in the carrier sentence, but presented in isolation. Second, the English listeners in the brief tone identification test were instructed to report on the answer sheet (see Appendix Three) whether each stimulus sounded like a question or statement. The results of this tone identification test and the survey were used to help interpret the results from the tone discrimination experiment.

2.5 Data Processing and Analysis

2.5.1 Acoustical Analysis of the Cantonese Tones Used in the Current Experiment

2.5.1.1 F0 Contours and Amplitude Envelopes of the Cantonese Tones

Using the procedure described in Section 2.2 above, the average F0 contour and amplitude envelope of each of the six Cantonese tones produced by each talker with each carrier syllable were computed. The correlation between the F0 contour and the amplitude envelope of each tone type produced by each talker with each carrier syllable was estimated using the SPSS-19 quadratic regression analysis. The quadratic regression was used in stead of the linear regression because the amplitude envelopes and F0 contours resembled quadratic curves more than straight lines, as illustrated in Figure 8.

As mentioned in Section 2.4 and described in Appendix Four, the six tone types produced by each talker with each carrier syllable were arranged into thirty-six tone pairs, totaling 144 tone pairs for the discrimination experiment (i.e., 36 tone pairs x 2 talkers x 2 carriers). For each of these 144 tone pairs, an ***F0-AE covariability index*** was estimated by averaging the F0-AE correlation coefficients of the two tones in each pair. For example, the F0-AE correlation coefficient of Tone 1 produced by the male talker with the carrier syllable /ji/ and that of the corresponding Tone 2 were averaged to derive the F0-AE covariability index for this pair of Tone 1 vs. Tone 2.

2.5.1.2 Acoustical Distance Indices

The acoustical distance between the two tones within each of the thirty-six tone pairs produced by each talker with each carrier syllable was estimated using two acoustical distance indices. The first acoustical distance index was the peak amplitude difference (in dBSPL) between the two tones within a tone pair. It measured the acoustical distance between the two tones in terms of the relative height of their amplitude envelopes, and was therefore labeled ***amplitude envelope height difference index***, or ***AE height difference index***. The second acoustical distance index was computed by correlating, frame by frame as illustrated in Figure 8, the amplitude envelopes of the two tones within a tone pair using the SPSS-19 quadratic regression analysis. The quadratic correlation coefficients measured the similarity between the contours of the amplitude envelopes of the two tones within a tone pair, and were therefore labeled ***amplitude envelope contour similarity indices***, or ***AE contour similarity indices*** for short.

2.5.2 Processing and Analysis of Tone Discrimination Data

To determine whether native listeners of Cantonese could use amplitude envelope alone to discriminate pairs of Cantonese tone contrasts, each Cantonese participant's average A' and B' measures in Condition One (i.e., stimuli with original amplitude) were computed and compared with the theoretical chance level. A' and B' were used for the following reasons. First, as discussed in Section 1.4.5, A' and B' present certain advantages over traditional measures such as percent correct and number of correct responses, especially in the presence of response biases. Second, unlike d' and β measures, A' and B' do not assume normal distribution and equal variances in the responses. This is particularly important because the pilot data of the current study did not appear to be normally distributed with equal variances.

A' and B' were calculated using the following formulae:

Let H = Hit Rate (%), and F = False Alarm Rate (%).

Then, $A' = 0.5 + (H-F)(1+H-F) / [4H(1-F)]$, if $H \geq F$, or

$A' = 0.5 + (F-H)(1+F-H) / [4F(1-H)]$, if $H \leq F$ (Grier, 1971);

and $B' = H(1-H)/[F(1-F)] - 1$, if $H \geq 1 - F$, or

$B' = 1 - F(1-F) / [H(1-H)]$, if $H \leq 1 - F$ (Hodos, 1970).

In order to determine whether native listeners of Cantonese could use amplitude envelope alone to differentiate level tones from contour tones, each Cantonese participant's average A' measures for the pairs involving a level tone and a contour tone in Condition One were compared with the theoretical chance level. Lastly, in order to determine whether native listeners of Cantonese could use amplitude envelope cue alone to differentiate low tone from high tones, each Cantonese participant’s average A’ measures for the pairs involving tones of similar contours but different height in Condition One were compared with the theoretical chance level.

For a detailed item analysis of the current tone discrimination experiment, the thirty-six tone pairs used in the experiment were divided into four mutually exclusive groups as follows:

1. the six ***Same*** pairs;
2. the six pairs consisting of tones with similar contours but different height (SCDH Pairs);
3. the twelve pairs consisting of tones with different contours but same height (DCSH Pairs);
4. the twelve pairs consisting of tones of different contours and different height (DCDH Pairs).

This grouping was based on the contour and the relative height of the amplitude envelopes of the two tones within each tone pair. A complete list of these four categories of tone pair types is included in Appendix Four. Each participant’s percent correct measures, A’ measures, B’ measures and reaction time measures for each of the four tone pair types produced by each of the two talkers with each of the two carrier syllables in each of the two conditions were computed. The reaction time measures of the two groups of participants were analyzed using a mixed design ANOVA with repeated measures on 2 conditions x 2 talkers x 2 carrier syllables x 4 tone pair types for main effects of group, experimental condition, talker, carrier syllable and tone pair type, and interactions among these factors.

If the B’ measures indicate a perceptual bias in either group of participants, the A’ measures of the two groups of participants would be analyzed using a mixed design ANOVA with repeated measures on 2 conditions x 2 talkers x 2 carrier syllables x 4 tone pair types. If the B’ measures indicate no perceptual bias in either group of participants, the percent correct measures of the two groups would be used instead of the A’ measures in the mixed design ANOVA with repeated measures on 2 conditions x 2 talkers x 2 carrier syllables x 4 tone pair types. As discussed in Section 1.4.5, in the absence of a perceptual bias, A’ measures and percent correct measures would be equally appropriate measures of tone discrimination accuracy. Due to their computation simplicity and ease of interpretation, percent correct measures would replace A’ measures for the detailed item analysis if the B’ measures indicate no perceptual bias in either group of participants.

Finally, the relative contribution of the contour and the height of amplitude envelope to Cantonese tone discrimination using amplitude envelope was determined using two different methods. The first method was based on the detailed item analysis of the above-mentioned four tone pair types. As mentioned in Section 2.4 above, there were two conditions in the current tone discrimination experiment. In Condition One, the original amplitude or intensity level of each of the processed tone was preserved. Thus, the peak amplitude of each processed tone was kept the same as that of the original natural tone token based on which the processed tone was generated. In Condition Two, the peak amplitude of each processed tone was equalized to the grand mean peak amplitude of all the forty-eight processed tones in order to eliminate the peak amplitude differences among the six Cantonese tones. Thus, in Condition Two, the contribution of the relative height of amplitude envelope to Cantonese tone discrimination was eliminated, and the results from Condition Two would reveal the independent contribution of the contours of the amplitude envelopes to Cantonese tone discrimination. In addition, the comparison of the results from both conditions would help determine the independent contribution of the relative height of amplitude envelope to Cantonese tone perception.

The second method was based on the SPSS-19 multiple regression analysis. It would quantitatively correlate the thirty Cantonese participants' discrimination results (i.e., the percent correct measures or A' measures) in Condition One with the three types of acoustical indices (i.e., the F0-AE covariability indices, the AE height difference indices and AE contour similarity indices). The results of this multiple regression analysis would reveal, in quantitative terms, the relative contributions of the F0-AE covariability, the contour and the height of amplitude envelope to Cantonese tone discrimination using amplitude envelope alone.

2.5.3 Analysis of the Survey and the Tone Identification Test

The participants' responses to the brief survey at the end of the current tone discrimination experiment were grouped into meaningful categories of similar responses. The participants' performance in the tone identification test at the end of the current tone discrimination experiment was analyzed in terms of individual average percent correct, group average percent correct, and standard deviations. The individual percent correct measures from this tone identification test were compared with those of the tone discrimination experiment using the SPSS-19 linear regression analysis.

3. Results

3.1 Acoustical Analysis

3.1.1 F0 Contours and Amplitude Envelopes of the Cantonese Tones

The female talker's average F0 contour and amplitude envelope for each tone produced with each carrier syllable were highly correlated with those of the male talker according to the SPSS-19 quadratic regression analysis (r= 92%, p <0.0001). Therefore, the two talkers' F0 contour and amplitude envelope for each tone type produced with each carrier syllable were combined for further analysis, as displayed in Figure 8. A visual inspection of Figure 8 suggested that the amplitude envelopes produced with the /wai/ syllable exhibited a significantly wider amplitude range than those produced with the /ji/syllable. That is, the amplitude range spanned from 45.5 to 78 dBSPL for the tones produced with the /wai/ syllable, but only from 47 to 73.5 dBSPL for those produced with the /ji/ syllable. In contrast, the F0 contours produced with the /wai/ syllable did not show a wider F0 range than those produced with the /ji/ syllable. The F0 range was between 107 and 220 Hz for the tones produced with the /wai/ syllable, and between 106 and 230 Hz for those produced with the /ji/ syllable.

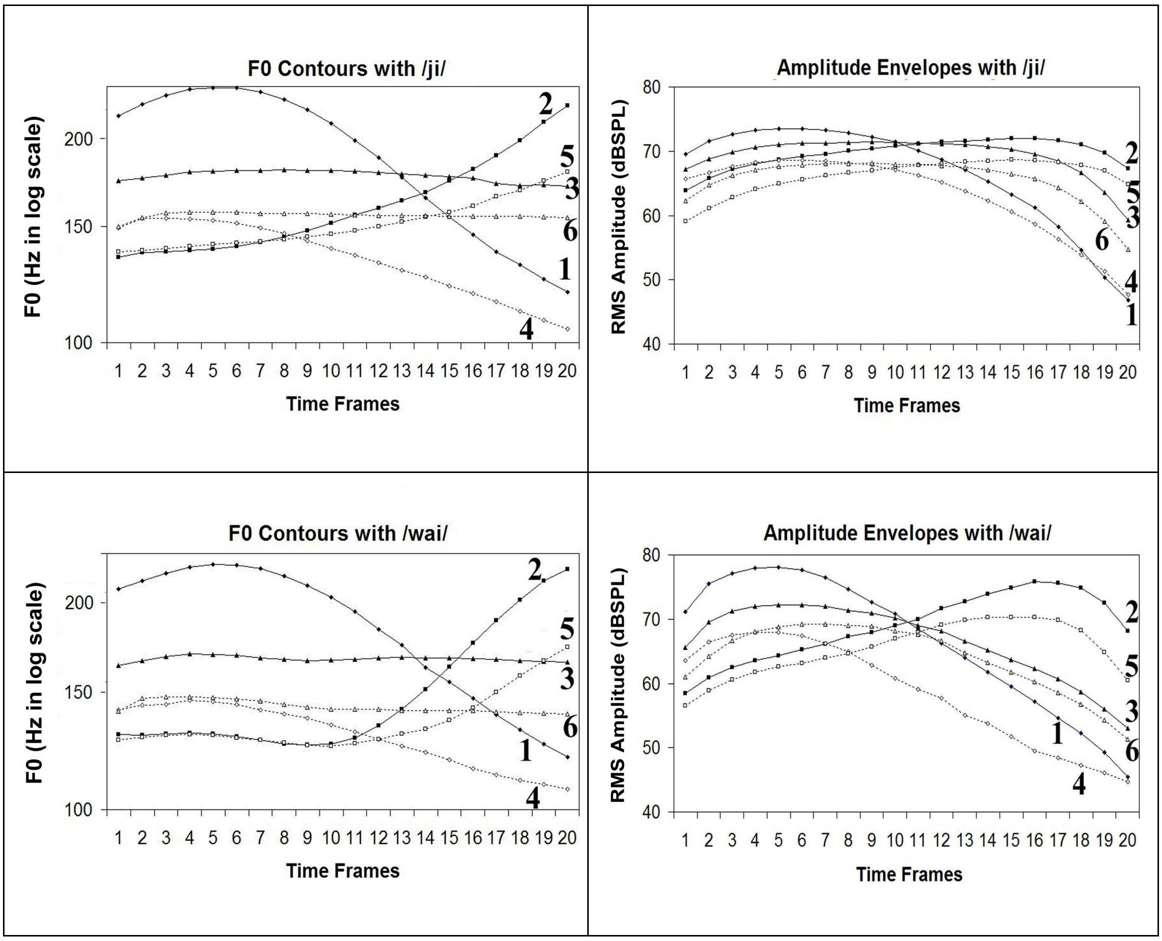

Figure 8. Amplitude envelopes and F0 contours of Cantonese tones produced with the syllable /ji/ or /wai/ in isolation. Each tone was divided into 20 equal time frames. The root-mean-square (RMS) amplitude and mean F0 within each time frame was averaged across 20 tone tokens produced by one male talker and one female talker.

The correlation between the F0 contour and the amplitude envelope of each tone produced by each talker with each carrier syllable was measured using the SPSS-19 quadratic regression analysis discussed in Section 2.5.1. The F0-AE correlation coefficient r for each tone type produced by each talker with each carrier syllable is displayed in Table 4. These F0-AE correlation coefficients measures were higher than those reported in previous studies (e.g., Fu et al., 2000; Kuo et al., 2007; Zhou, 2010). Previous studies used a linear regression to measure the

F0-AE correlation and significantly under-estimated the F0-AE correlation because the amplitude envelopes and F0 contours resembled quadratic curves more than straight lines.

Tone	Talker	Carrier Syllable	r	r-squared
Tone 1 (High Falling)	Female	/ji/	0.96	0.92
		/wai/	0.97	0.94
	Male	/ji/	0.91	0.83
		/wai/	0.94	0.88
Tone 2 (High Rising)	Female	/ji/	0.91	0.83
		/wai/	0.90	0.81
	Male	/ji/	0.86	0.74
		/wai/	0.89	0.79
Tone 3 (Upper Middle Level)	Female	/ji/	0.89	0.79
		/wai/	0.86	0.74
	Male	/ji/	0.85	0.72
		/wai/	0.84	0.71
Tone 4 (Low Falling)	Female	/ji/	0.96	0.92
		/wai/	0.96	0.92
	Male	/ji/	0.93	0.86
		/wai/	0.94	0.88
Tone 5 (Low Rising)	Female	/ji/	0.92	0.85
		/wai/	0.92	0.85
	Male	/ji/	0.89	0.79
		/wai/	0.91	0.83
Tone 6 (Lower Middle Level)	Female	/ji/	0.83	0.69
		/wai/	0.89	0.79
	Male	/ji/	0.87	0.76
		/wai/	0.89	0.79

Table 4. Quadratic correlation coefficients between the F0 contour and the amplitude envelope of each Cantonese tone produced in isolation with the carrier syllable /ji/ or /wai/ by a male talker and a female talker.

As mentioned in Section 2.4 above, the six tone types produced by each talker with each carrier syllable were arranged into thirty-six tone pairs, totaling 144 tone pairs for the discrimination experiment (i.e., 36 tone pairs x 2 talkers x 2 carriers). For each of these 144 tone pairs, an ***F0-AE covariability index*** was estimated by averaging the F0-AE correlation coefficients of the two tones in each pair. For example, the F0-AE correlation coefficient of Tone 1 produced by the male talker with the carrier syllable /ji/ and that of the corresponding Tone 2 were averaged to derive the F0-AE covariability index for the pair of Tone 1 vs. Tone 2. The F0-

AE covariability indices of these 144 thirty-six pairs ranged from 0.83 to 0.97, with a grand mean of 0.91 and a standard deviation of 0.3, as illustrated in Figure 9.

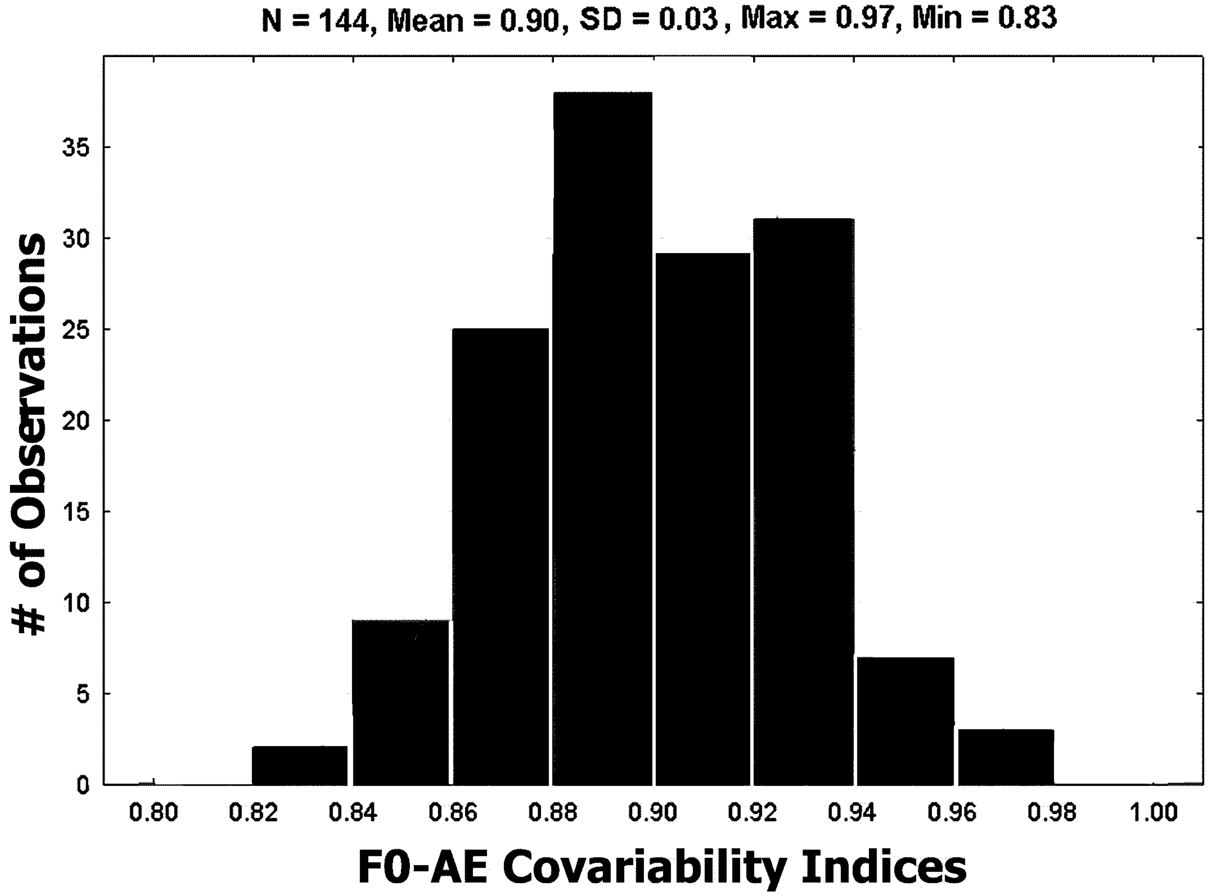

Figure 9. Histogram of the 144 F0-AE covariability indices

The 144 F0-AE covariability indices were subsequently normalized using arcsine-transformation for further statistical analysis. In particular, the 144 arcsine-transformed F0-AE covariability indices were analyzed using a 2 (talkers) x 2 (carriers) ANOVA. The main effects of talker and carrier syllable were statistically significant as displayed in Table 5. Overall, the tones produced by the male talker yielded significantly higher F0-AE covariability indices (mean = 0.93; SD = 0.03) than those produced by the female talker (mean = 0.89; SD = 0.02), and the tones produced with the /wai/ syllable yielded significantly higher F0-AE covariability indices (mean = 0.91; SD = 0.02) than those produced by the /ji/ syllable (mean = 0.90; SD = 0.02). The interaction between talker and carrier syllable did not reach statistical significance, as illustrated in Table 5.

Factors	SS	DF	MS	F	p-value	Power
Talker	0.02	1	0.02	75.88	< 0.001	0.99
Carrier	0.01	1	0.01	25.63	< 0.001	0.99
Talker x Carrier	0.01	1	0.01	3.39	0.074	0.99

Table 5. Effects of talker and carrier syllable on the F0-AE covariability indices.

3.1.2 Acoustical Distance Indices

As defined in Section 2.5.1.2, the acoustical distance indices consisted of AE contour similarity indices and AE height difference indices. The AE contour similarity index of a tone pair measured the acoustical similarity between the AE contours of the two tones within the tone pair. It was calculated by correlating, time frame by time frame as illustrated in Figure 8 and Appendix 5, the amplitude envelopes of the two tones within a pair using the SPSS-19 quadratic correlation method. The AE contour similarity indices for the thirty-six tone pairs produced by the two talkers ranged from -0.54 to 1, with a grand mean of 0.48 and standard deviation of 0.56, as illustrated in Figure 10. The AE contour similarity indices were the highest for the ***Same*** pairs (i.e., close or equal to 1.0), slightly lower for the pairs consisting of two tones with similar contours but different height (i.e., ~ 0.95), and the lowest for pairs consisting of a falling tone and a rising tone. These data will be further described in the next paragraph.

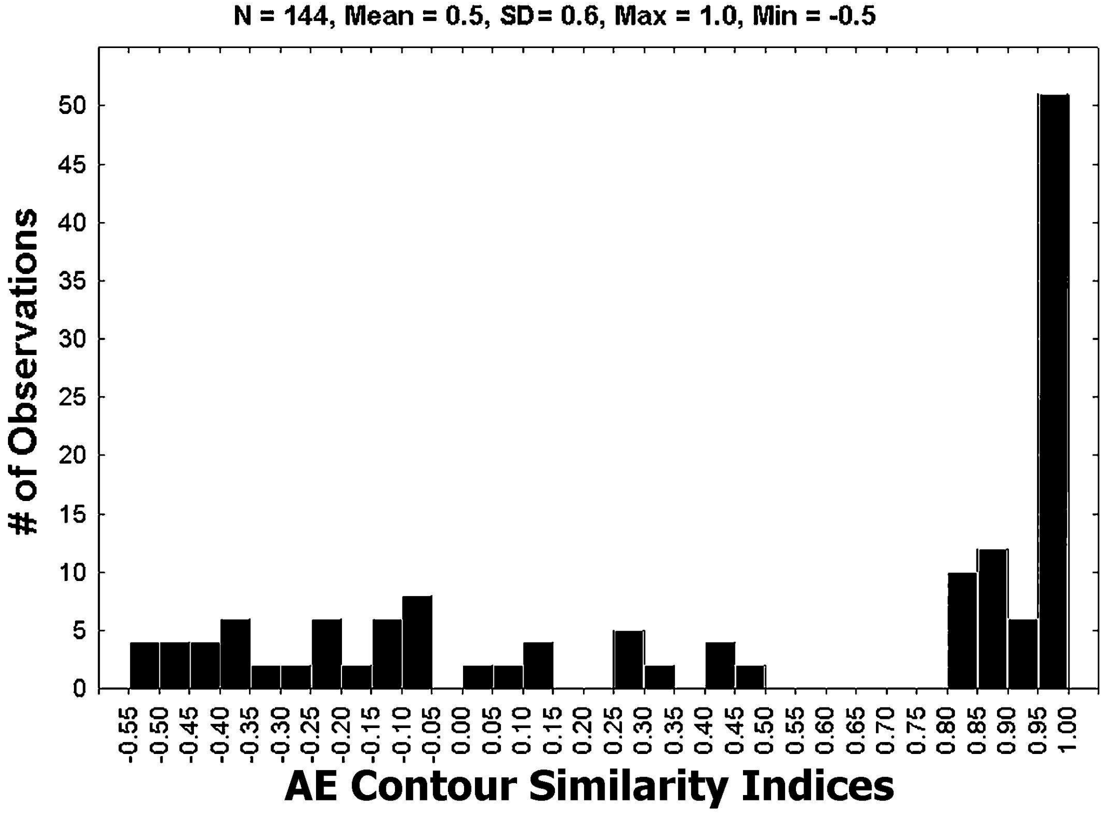

Figure 10. Histogram of the 144 AE Contour Similarity Indices

The AE contour similarity index for each tone pair produced by each talker with each carrier syllable was normalized using arcsine-transformation before further statistical analyses. The AE contour similarity indices of the tone pairs produced by the male talker were then compared with those produced by the female talker using a 2 (talkers) x 2 (carriers) ANOVA. The main effects of talker and carrier syllable were statistically significant as displayed in Table 6. Overall, the tones produced by the male talker yielded significantly higher AE contour similarity indices (mean = 0.50; SD = 0.52) than those produced by the female talker (mean = 0.42; SD = 0.59), and the tones produced with the /wai/ syllable yielded significantly higher AE contour similarity indices (mean = 0.53; SD = 0.52) than those produced with the /ji/ syllable (mean = 0.40; SD = 0.59). The interaction between talker and carrier syllable was also significant. That is, the above-mentioned carrier syllable effect was greater for the male talker (means = 0.59 for /wai/ and 0.45 for /ji/) than for the female talker (means = 0.50 for /wai/ and 0.39 for /ji/), as indicated by the effect size measured by Cohen's d (0.27 for the male talker, and 0.17 for the female talker).

Factors	**SS**	**DF**	**MS**	**F**	**p-value**	**Power**
Talker	0.84	1	0.84	34	< 0.001	0.99
Carrier	7.86	1	7.86	10	0.003	0.87
Talker x Carrier	0.86	1	0.86	23	< 0.001	0.99

Table 6. Effects of talker and carrier syllable on the AE contour similarity indices.

On the other hand, the AE height difference index of a tone pair was computed as the peak amplitude difference (in dBSPL) between the two tones within the pair, and thus measured the acoustical distance between the two tones in terms of the relative height of their amplitude envelopes. The AE height difference indices for the thirty-six tone pairs produced by the two talkers ranged from 0.06 to 6.46 dBSPL, with a grand mean of 2.5 dBSPL and standard deviation of 1.9, as illustrated in Figure 11 and Appendix 5. The AE height difference indices were the lowest for the ***Same*** pairs (i.e., close or equal to 0), slightly higher for the pairs consisting of tones with different contours but similar height, and the highest for the remaining two categories of tone pairs (i.e., pairs involving tones with similar contours but different height, and pairs consisting of tones with different contours and different height).

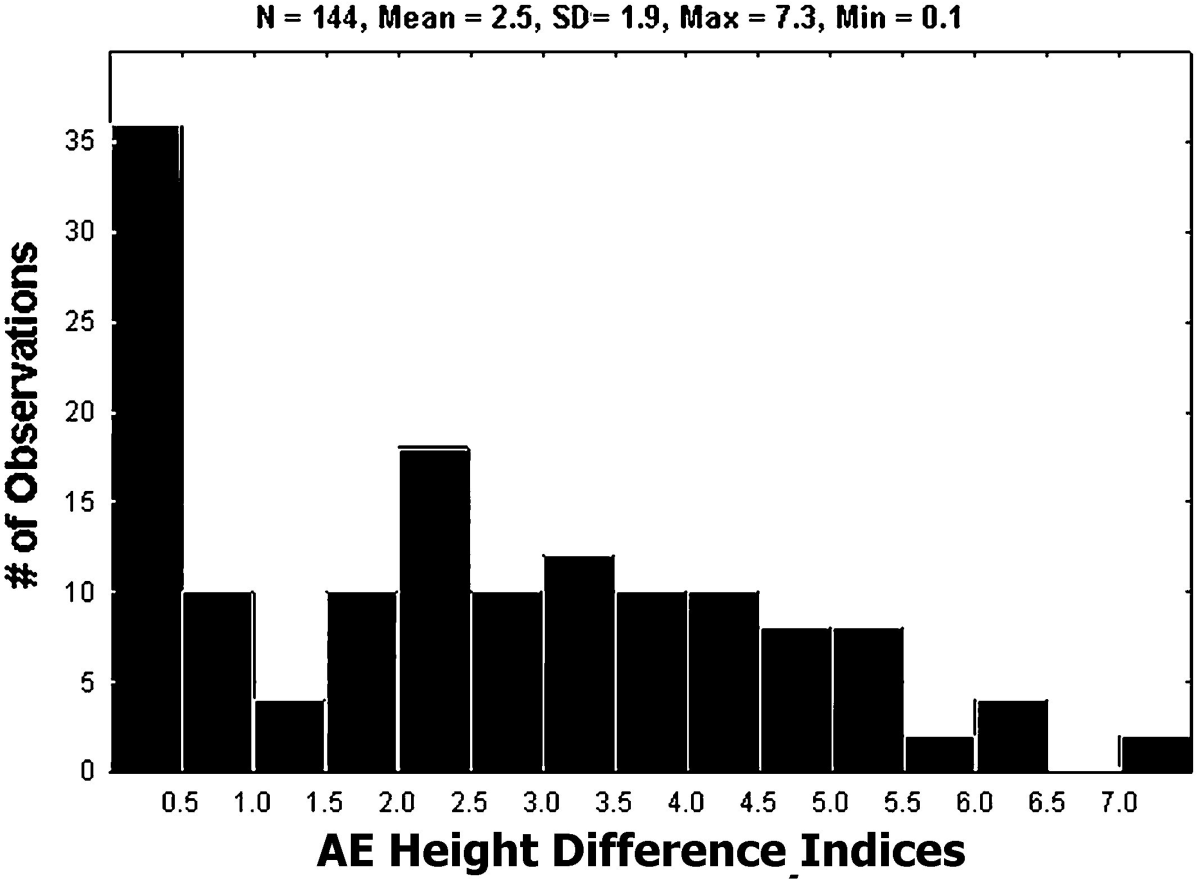

Figure 11. Histogram of the 144 AE Height Difference Indices

The AE height difference indices of the tones produced with each carrier syllable by the male talker were arcsine-transformed and then compared with those produced by the female talker using a 2 (talkers) x 2 (carriers) ANOVA. The main effects of talker and carrier syllable were statistically significant as displayed in Table 7. Overall, the tones produced by the male talker yielded significantly higher AE height difference indices (mean = 2.61; SD = 2) than those produced by the female talker (mean = 2.44; SD = 1.86), and the tones produced with the /wai/ syllable yielded significantly higher AE height difference indices (mean = 2.79; SD = 1.95) than those produced with the /ji/ syllable (mean = 2.3; SD = 1.8). The interaction between talker and carrier syllable did not reach statistical significance. That is, the above-mentioned carrier syllable effect was greater for the male talker (means = 2.87 for /wai/ and 2.25 for /ji/) than for the female talker (means = 2.56 for /wai/ and 2.25 for /ji/), as indicated by the effect size measured by Cohen's d (0.31 for the male talker, and 0.16 for the female talker).

	SS	**DF**	**MS**	**F**	**p-value**	**Power**
Talker	0.18	1	0.18	23	< 0.001	0.99
Carrier	0.56	1	0.56	15	< 0.001	0.96
Talker x Carrier	0.02	1	0.02	27	< 0.001	0.99

Table 7. Effects of talker and carrier syllable on the AE height difference indices.

3.2 Results from the Discrimination Experiment

3.2.1 Overview of B' Measures

The thirty Cantonese participants' individual B' measures for Condition One (i.e., stimuli with original amplitude) ranged from -0.09 to 0.11, and those of the English group varied from -0.05 to 0.32. The group average and standard deviation were 0.02 and 0.06 respectively for the Cantonese, and 0.01 and 0.06 for the English, as illustrated in Figure 12. Neither group's average B' measure was statistically different from 0 ($t_{(29)}$ = -1.62 and p = 0.11 for the Cantonese group; $t_{(29)}$ = 0.14 and p = 0.27 for the English group). As mentioned in Section 1.4.6, the theoretically possible range of B' measures spans from -1 to 1, with 0 indicating no perceptual bias, -1 suggesting extreme bias in favor of the ***Same*** pairs, and + 1 meaning maximum bias for the ***Different*** pairs. Thus, neither the Cantonese group nor the English group demonstrated statistically significant bias in Condition One of the current tone discrimination using amplitude envelope alone.

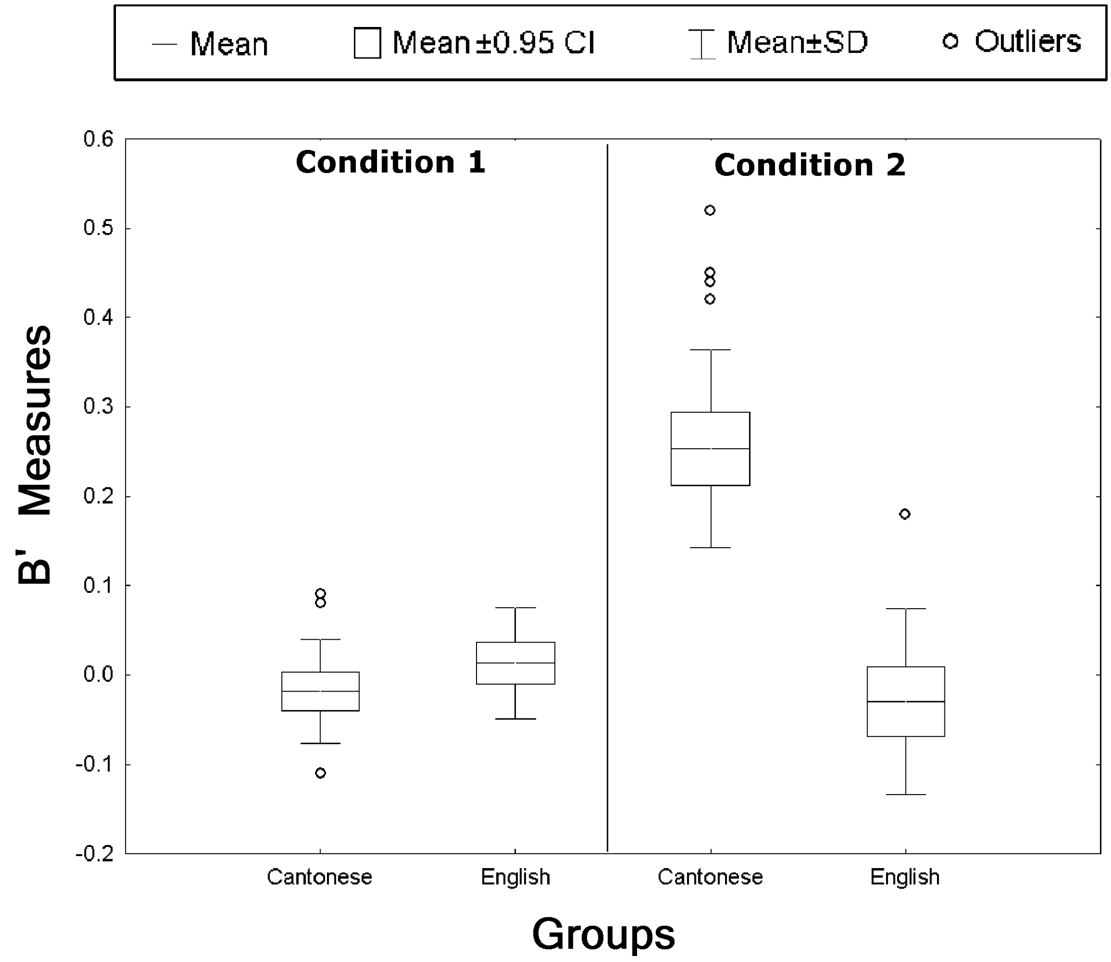

Figure 12. B' measures for the tone discrimination using amplitude envelope (0 = no perceptual bias; -1 = maximum bias in favor of ***Same*** pairs; + 1 = maximum bias for ***Different*** pairs)

The thirty Cantonese participants' individual B' measures for Condition Two (i.e., stimuli with equalized peak amplitude) ranged from 0.11 to 0.42, and those of the English group varied from -0.18 to 0.33, as illustrated in Figure 12. The group average and standard deviation were 0.25 and 0.11 respectively for the Cantonese, and -0.02 and 0.10 for the English. A 2 (conditions) x 2 (groups) repeated measures ANOVA indicated that the B' measures in Condition Two were significantly worse than those in Condition One for either group as shown in Table 8. However, these results did not necessarily suggest that the tone discrimination in Condition Two was more biased in Condition One. As mentioned in Section 2.4, the stimuli in Condition Two had been equalized with respect to their peak amplitude level, and hence the peak amplitude differences among the six Cantonese tones had been removed in Condition Two. Thus, some ***Different*** pairs in Condition One (e.g., the pairs consisting of two tones with similar contours but different height) might sound more like ***Same*** pairs in Condition Two, as illustrated in Figure 13. This factor was not taken into account in the computation of the B’ measures in Condition Two in order to maintain consistency in the computation of B’ measures across conditions. Thus, the worse B’ measures in Condition Two should probably be interpreted as a negative effect of the removal of the peak amplitude differences among the six Cantonese tones on the current tone discrimination using amplitude envelope.

Factors	**SS**	**DF**	**MS**	**F**	**p-value**	**Power**
Group	0.44	1	0.44	107	< 0.001	0.97
Condition	0.42	1	0.42	37	< 0.001	0.95
Group x Condition	0.81	1	0.81	165	< 0.001	0.94

Table 8. Effects of language group and experimental condition on the B’ measures.

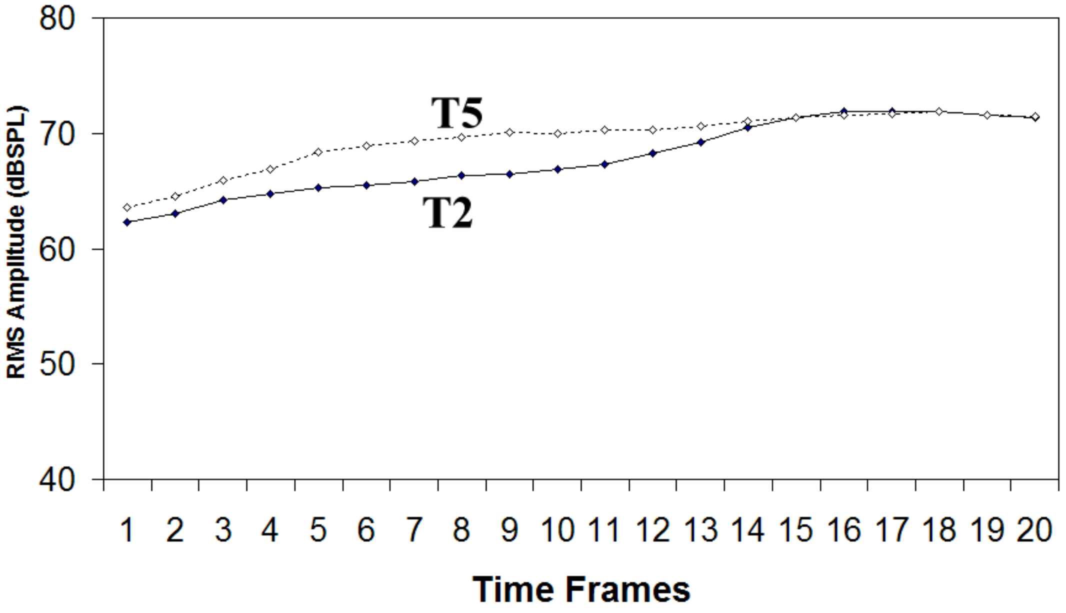

Figure 13. Amplitude envelopes of Tone 2 (High Rising) and Tone 5 (Low Rising) produced by the female talker with the carrier syllable /wai/. The peak amplitudes of both tones have been equalized.

3.2.2 Overview of A' Measures

Each participant's A' measure for Condition One (i.e., stimuli with original amplitude) is displayed in Figure 14. These A' measures ranged from 0.86 to 0.95 among the Cantonese participants, and from 0.59 to 0.91 among the English participants, with the possible range going from 0.5 (chance level) to 1 (perfect discrimination). The group average and standard deviation were 0.9 and 0.03 respectively for the Cantonese group, and 0.7 and 0.09 for the English group. A one-way ANOVA suggested that the Cantonese group outperformed the English group ($F_{(1, 58)} = 143.72$; $p < 0.001$). However, each group's A' measure in Condition One was significantly above the chance level of 0.5 (Cantonese group: $t_{(29)} = 78.6$ and $p < 0.001$; English group: $t_{(29)} = 12.9$ and $p < 0.001$).

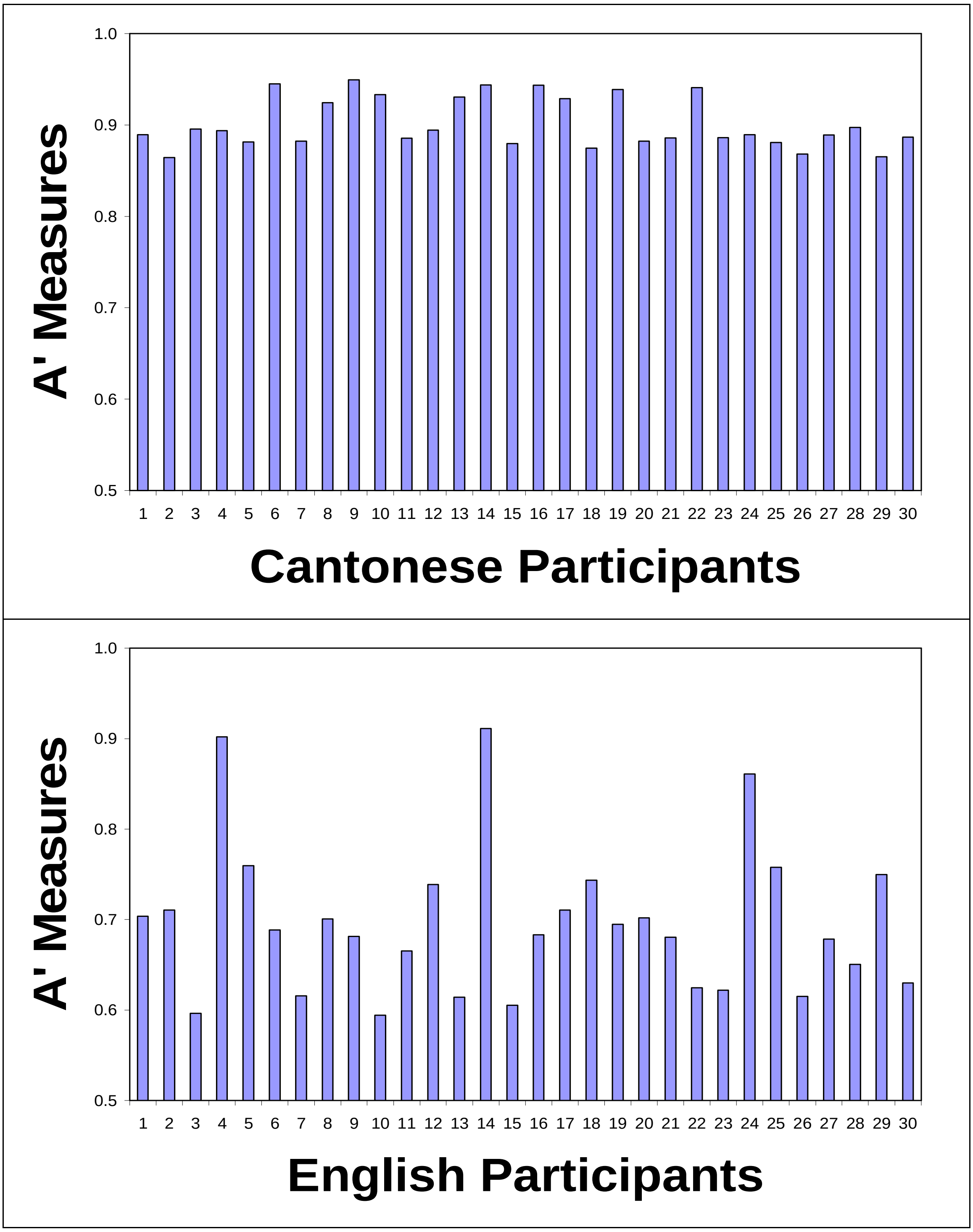

Figure 14. The sixty participants’ A’ measures in Condition One of tone discrimination using amplitude envelope (chance = 0.5; perfect discrimination = 1).

The thirty Cantonese participants' individual A' measures for Condition Two (i.e., stimuli with equalized peak amplitude) ranged from 0.81 to 0.88, and those of the thirty English participants from 0.55 to 0.84, as illustrated in Figure 15. The group mean and standard deviation were 0.84 and 0.03 respectively for the Cantonese group, and 0.65 and 0.08 for the English group. A 2 (conditions) x 2 (groups) repeated measures ANOVA suggested that both groups performed significantly better in Condition One than in Condition Two, and the Cantonese group outperformed the English group in both conditions, as shown in Table 9.

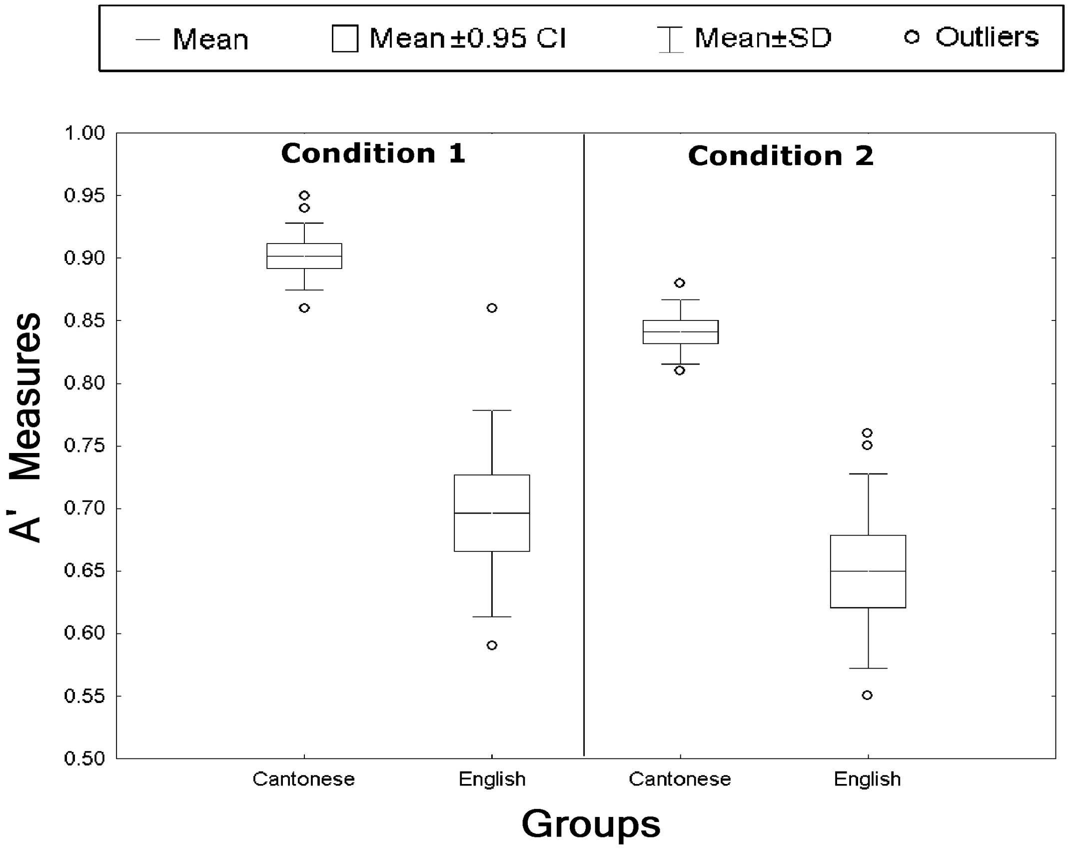

Figure 15. A' measures for tone discrimination using amplitude envelope (0.5 = chance level; 1 = perfection)

Factors	SS	DF	MS	F	p-value	Power
Group	0.44	1	0.44	107	< 0.001	0.97
Condition	0.42	1	0.42	37	< 0.001	0.95
Group x Condition	0.81	1	0.81	165	< 0.001	0.98

Table 9. Effects of language group and experimental condition on the A’ measures for tone discrimination using amplitude envelope.

3.2.3 Overview of Reaction Time Measures

The thirty Cantonese participants’ individual average reaction time measures in Condition One (i.e., stimuli with original amplitude) ranged from 276 to 369 ms, and those of the thirty English participants from 261 to 475 ms among the English participants, as illustrated in Figure 16. The group average and standard deviation were 328 ms and 68 ms respectively for the Cantonese group, and 375 and 80 ms for the English group. A one-way ANOVA suggested that the Cantonese group’s grand mean reaction time measure was significantly shorter than that of the English group ($F_{(1, 58)} = 7.93$; $p < 0.007$).

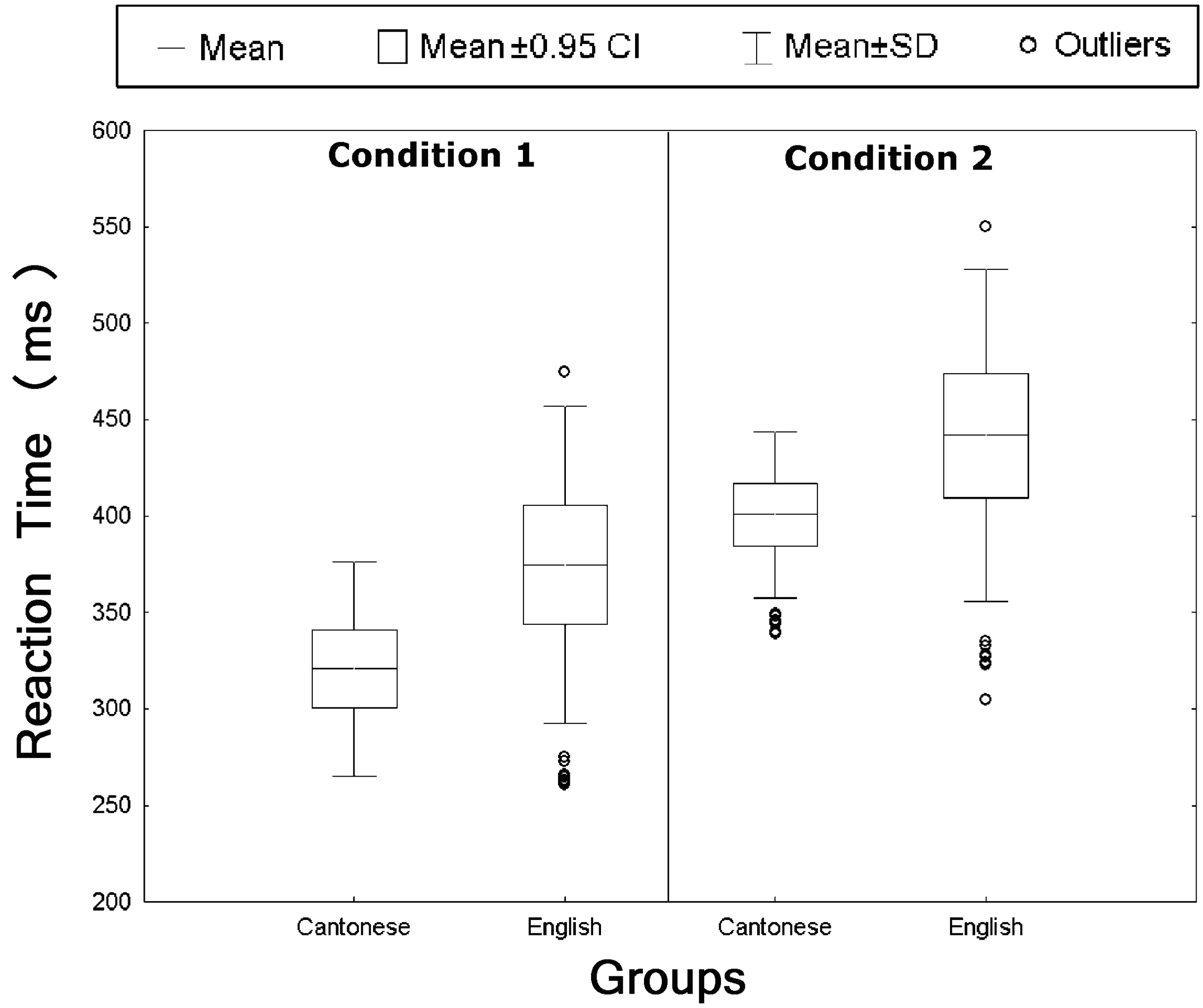

Figure 16. Reaction time measures for the tone discrimination using amplitude envelope

The thirty Cantonese participants' individual average reaction time measures in Condition Two (i.e., stimuli with equalized peak amplitude) ranged from 339 to 444 ms, and those of the English participants varied from 305 to 550 ms, as illustrated in Figure 16. The grand mean and standard deviation were 401 ms and 71 ms respectively for the Cantonese group, and 442 and 83 for the English group. A 2 (conditions) x 2 (groups) repeated measures ANOVA of the sixty participants' individual average reaction time measures for Condition One and Condition Two suggested that both groups responded faster in Condition One than in Condition Two, and the Cantonese group was consistently faster than the English group in both conditions.

Effect	SS	DF	MS	F	p-value	Power
Group	0.02	1	0.02	76	< 0.001	0.97
Condition	0.01	1	0.01	26	< 0.001	0.61
Talker x Carrier	0.01	1	0.01	19	< 0.001	0.41

Table 10. Effects of language group and experimental condition on reaction time measures for tone discrimination using amplitude envelope.

Thus, the repeated measures ANOVA of the sixty participants' individual average reaction time measures and A' measures for Condition One and Condition Two exhibited similar patterns. That is, the sixty participants performed more accurately and faster in Condition One than in Condition Two, and the Cantonese group outperformed the English group in both conditions in terms of accuracy and speed. In order to quantify the correlation between the reaction time measures and the A' measures, the sixty participants' individual average A' measures in Condition One and Condition Two were compared with their corresponding reaction time measures using the SPSS-19 linear regression analysis. The correlation coefficient was -0.68 and was statistically significant ($t_{(119)} = 101.93$; $p < 0.001$). Thus, the A' measures were fairly well correlated with the reaction time measures in the current tone discrimination experiment.

3.2.3 Pairs Involving a Level Tone and a Contour Tone in Condition One

Each participant's A' measure for the pairs involving a level tone and a contour tone in Condition One (i.e., stimuli with original amplitude) was also computed. These A' measures ranged from 0.79 to 0.93 among the Cantonese participants, and from 0.56 to 0.91 among the English participants, as illustrated in Figure 17. The group average and standard deviation were 0.87 and 0.05 respectively for the Cantonese, and 0.68 and 0.09 for the English. A one-way ANOVA suggested that the Cantonese group's average A' measure was significantly higher than that of the English group ($F_{(1, 58)}$ = 107.96; p < 0.001). However, each group's average A' measure was significantly above the chance level of 0.5 (Cantonese group: $t_{(29)}$ = 43.4; p < 0.001; English group: $t_{(29)}$ = 10.97; p < 0.001).

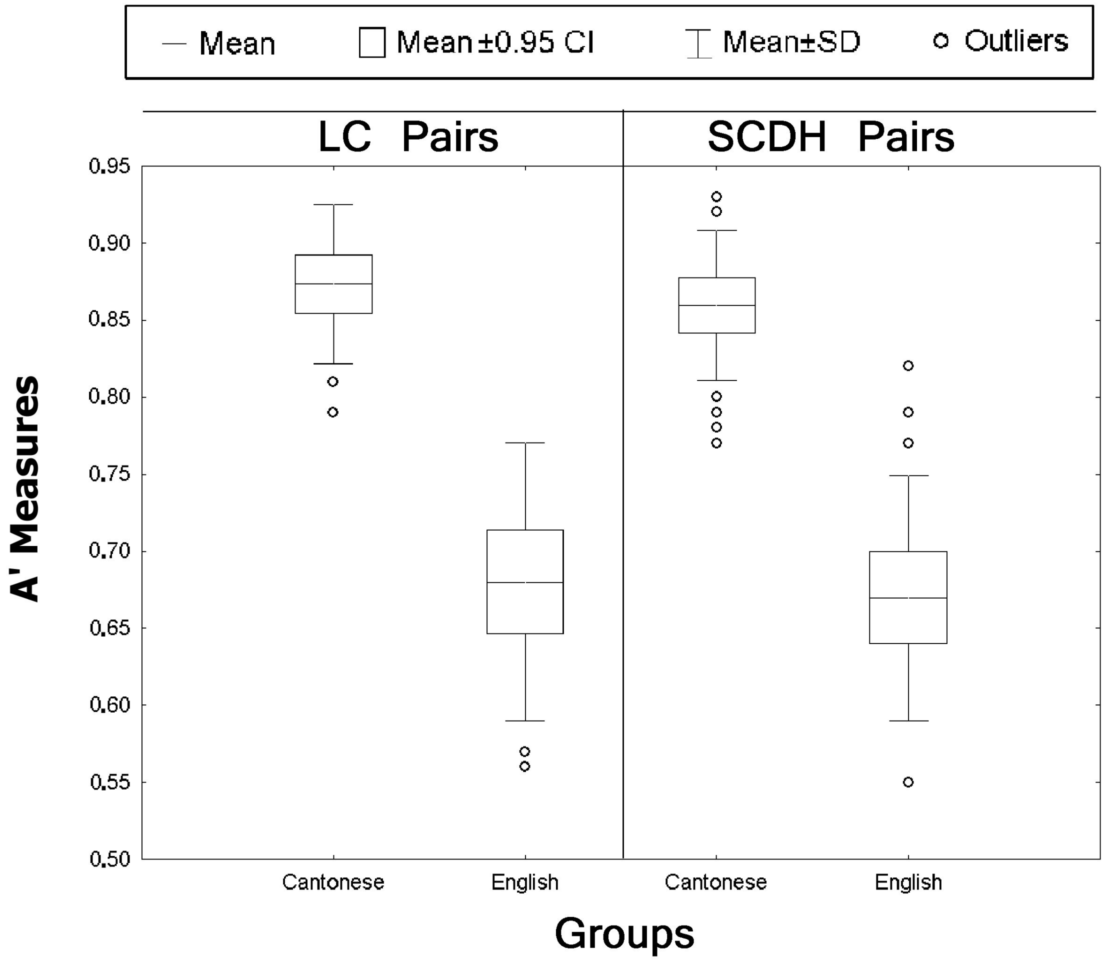

Figure 17. Discrimination of the pairs involving a level tone and a contour tone (LC Pairs), and the pairs consisting of two tones with similar contours but different height (SCDH pairs) in Condition One (A' of 0.5 = chance; 1 = perfect discrimination)

3.2.4 Tones with Similar Contours but Different Height in Condition One

Each participant's A' measure for the pairs involving two tones with similar contours but different height in Condition One (i.e., stimuli with original amplitude) was also computed. These A' measures ranged from 0.76 to 0.93 among the Cantonese participants, and from 0.55 to

0.89 among the English participants. The group average and standard deviation were 0.86 and 0.05 respectively for the Cantonese, and 0.67 and 0.08 for the English, as illustrated in Figure 17. A one-way ANOVA suggested that the Cantonese group's grand mean A' measure was significantly higher than that of the English group ($F_{(1, 58)} = 115.83$; $p < 0.001$). However, each group's A' measure was significantly above the chance level of 0.5 (Cantonese group: $t_{(29)} = 41.42$; $p < 0.001$; English group: $t_{(29)} = 10.04$; $p < 0.001$).

3.2.5 Detailed Item Analysis

3.2.5.1 Overview

As described in Section 3.2.1, the B' measures of the current tone discrimination experiment indicated no perceptual bias in favor of either the ***Different*** pairs or ***Same*** pairs among the sixty participants. In the absence of perceptual bias, A' measures and percent correct measures would be equally appropriate measures of tone discrimination accuracy in the current experiment, as discussed in Section 1.4.5 above. Due to their computation simplicity and ease of interpretation, percent correct measures were used in this detailed item analysis.

Each participant's percent correct measures and reaction time measure for each of the four tone pair types produced by each of the two talkers with each of the two carrier syllables in each of the two conditions were computed, yielding 32 percent correct measures and 32 reaction time measures per participant, and a total of 1920 percent correct measures and 1920 reaction time measures. The 960 percent correct measures in Condition One (i.e., stimuli with original amplitude) ranged from 0.75 to 0.93 among the Cantonese participants, and from 0.29 to 0.97 among the English participants. The grand mean and standard deviation in Condition One were respectively 0.84 and 0.18 for the Cantonese group, and 0.67 and 0.17 for the English group. In

Condition Two (i.e., stimuli with equalized peak amplitude), the percent correct measures ranged from 0.01 to 0.93 among the Cantonese participants, and from 0.22 to 0.9 among the English participants. The mean and standard deviation in Condition Two were respectively 0.57 and 0.16 for the Cantonese group, and 0.41 and 0.15 for the English group, as illustrated in Figure 18.

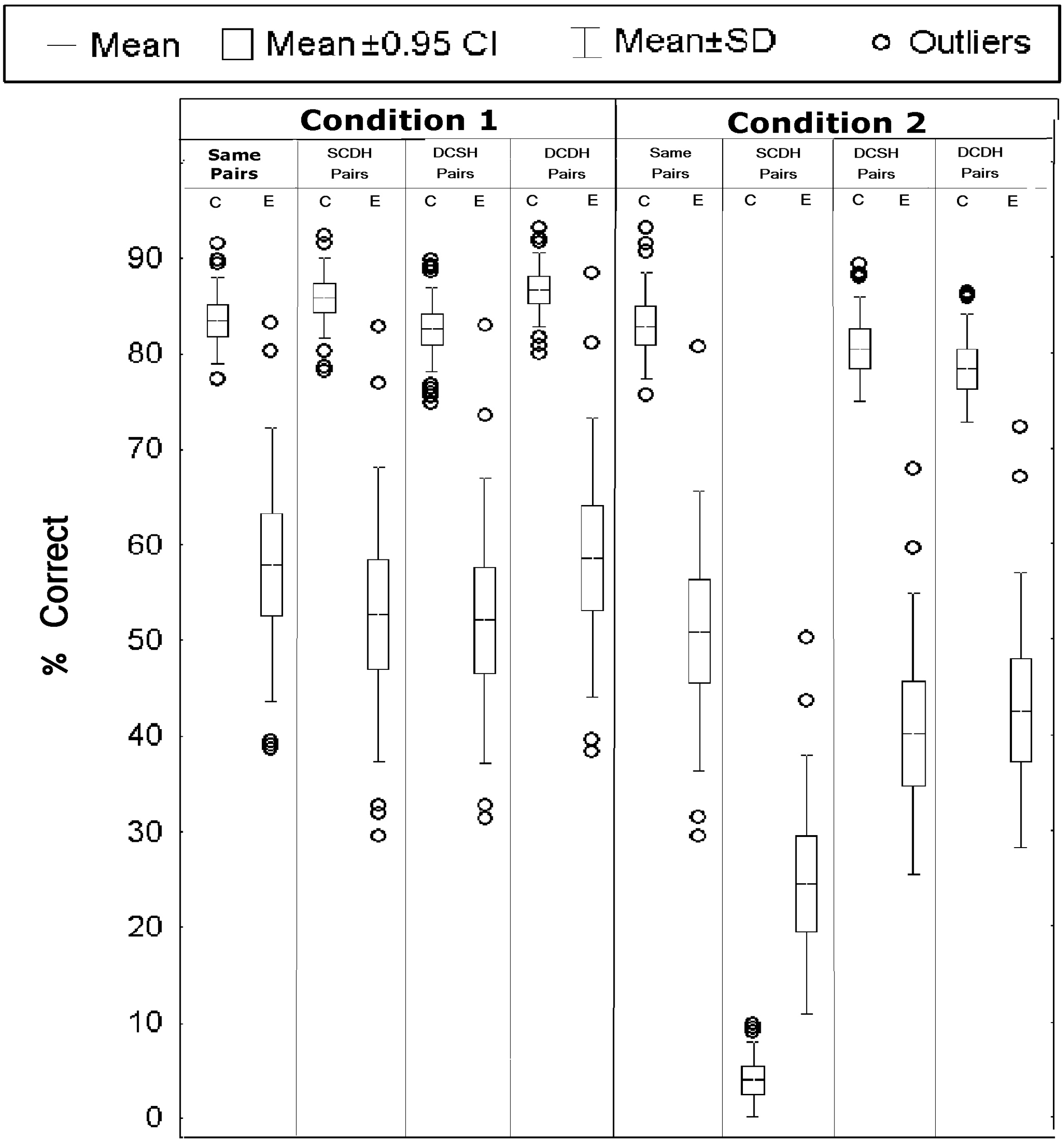

Figure 18. Cantonese (C) and English (E) Participants’ Discrimination of four different types of tone pairs in Condition One and Condition Two

The 1920 percent correct measures were arcsine-transformed and then analyzed using a 2 (groups) x 2 (conditions) x 2 (talkers) x 2 (carrier syllables) x 4 (tone pair types) repeated measures ANOVA, and the results are displayed in Table 11. The talker effect was the only main effect that did not reach statistical significance, and the interaction between talker and other factors were also statistically insignificant. The remaining main effects were all statistically significant, and the majority of the interactions among these remaining factors were also statistically significant. The details of the significant main effects and important interactions will be discussed in the next section.

Factors	SS	DF	MS	F	p-value	Power
group	17.99	1	17.99	2432	< 0.001	1.00
condition	0.01	1	0.01	9	0.005	0.83
talker	0.00	1	0.00	3	0.080	0.42
carrier	4.65	1	4.65	4109	< 0.001	1.00
tonetype	40.63	3	13.54	106	< 0.001	1.00
group*condition	0.06	1	0.06	37	< 0.001	1.00
group*talker	0.00	1	0.00	0	0.943	0.05
condition*talker	0.00	1	0.00	0	0.760	0.06
group*carrier	4.62	1	4.62	2906	< 0.001	1.00
condition*carrier	0.03	1	0.03	26	< 0.001	0.99
talker*carrier	0.00	1	0.00	2	0.176	0.27
group*tonetype	12.89	3	4.30	1724	< 0.001	1.00
condition*tonetype	0.02	3	0.01	5	0.003	0.91
talker*tonetype	0.02	3	0.01	5	0.073	0.40
carrier*tonetype	14.60	3	4.87	2608	< 0.001	1.00
group*condition*talker	0.00	1	0.00	3	0.099	0.38
group*condition*carrier	0.04	1	0.04	27	< 0.001	0.99
group*talker*carrier	0.00	1	0.00	1	0.361	0.15
condition*talker*carrier	0.00	1	0.00	0	0.832	0.06
group*condition*tonetype	0.01	3	0.00	2	0.094	0.54
group*talker*tonetype	0.00	3	0.00	1	0.649	0.16
condition*talker*tonetype	0.01	3	0.00	2	0.068	0.59
group*carrier*tonetype	11.80	3	3.93	2519	< 0.001	1.00
condition*carrier*tonetype	0.00	3	0.00	1	0.260	0.35
talker*carrier*tonetype	0.01	3	0.00	2	0.074	0.58
group*conditionion*talker*carrier	0.00	1	0.00	4	0.056	0.49
group*conditionion*talker*tonetype	0.01	3	0.00	1	0.300	0.32
group*condition*carrier*tonetype	0.03	3	0.01	11	< 0.001	0.99
group*talker*carrier*tonetype	0.00	3	0.00	1	0.617	0.17
condition*talker*carrier*tonetype	0.01	3	0.00	3	0.072	0.44
group*condition*talker*carrier*tonetype	0.00	3	0.00	1	0.377	0.27

Table 11. Effects of language group, experimental condition, talker, carrier syllable and tone pair type on the percent correct measures of tone discrimination using amplitude envelope.

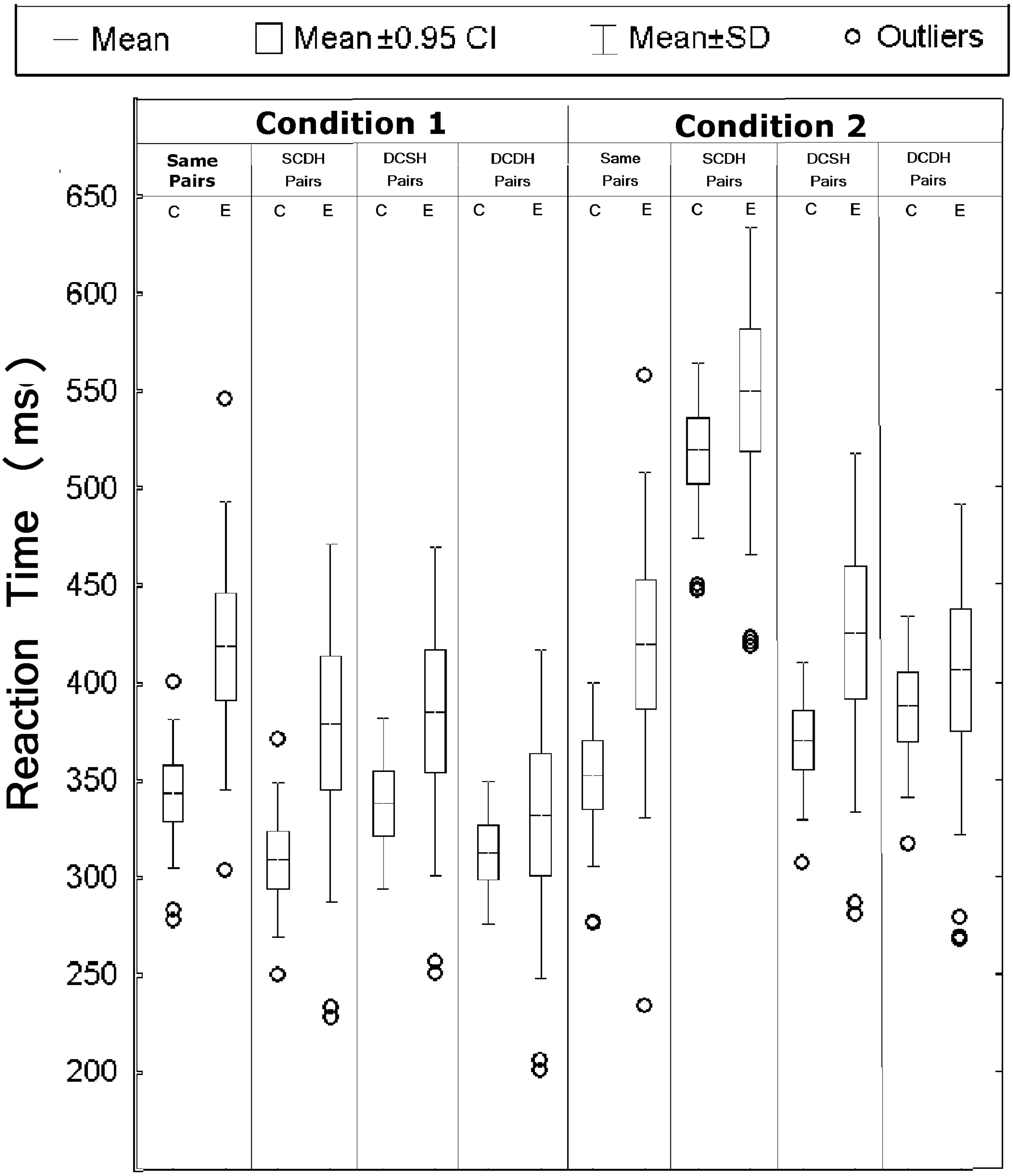

Figure 19. Cantonese (C) and English (E) Participants' Discrimination of four different types of tone pairs in Condition One and Condition Two

As illustrated in Figure 19, the 1920 reaction time measures ranged from 171 to 579 ms among the Cantonese participants, and from 151 to 726 ms among the English participants. The mean and standard deviation were 369 and 79 for the Cantonese group, and 417 and 109 for the English group. The 1920 reaction time measures were also analyzed using a 2 (groups) x 2 (conditions) x 2 (talkers) x 2 (carrier syllables) x 4 (tone pair types) repeated measures ANOVA, and the results are displayed in Table 12. The talker effect was the only main effect that did not reach statistical significance, and the interaction between talker and other factors were also statistically insignificant. The remaining main effects were statistically significant, and the majority of the interactions among these remaining factors were also statistically significant. The details of the significant main effects and important interactions will be discussed in the next section.

Factors	SS	DF	MS	F	p-value	power
group	2830000	1	2830000	1743	< 0.001	1.00
condition	4976	1	4976	7	0.015	0.66
talker	2557	1	2557	4	0.071	0.50
carrier	833000	1	833000	1043	< 0.001	1.00
tonetype	1350000	3	451000	8	< 0.001	1.00
group*condition	31400	1	31400	23	< 0.001	0.99
group*talker	69	1	69	0	0.813	0.05
condition*talker	98	1	98	0	0.779	0.12
group*carrier	215000	1	215000	174	< 0.001	1.00
condition*carrier	24500	1	24500	16	< 0.001	0.92
talker*carrier	1939	1	1939	2	0.190	0.19
group*tonetype	1530000	3	509000	377	< 0.001	1.00
condition*tonetype	8481	3	2827	3	0.034	0.56
talker*tonetype	15700	3	5239	5	0.072	0.40
carrier*tonetype	774000	3	258000	208	< 0.001	1.00
group*condition*talker	2402	1	2402	3	0.123	0.26
group*condition*carrier	22200	1	22200	18	< 0.001	0.99
group*talker*carrier	544	1	544	1	0.467	0.35
condition*talker*carrier	801	1	801	1	0.258	0.15
group*condition*tonetype	4099	3	1366	1	0.308	0.24
group*talker*tonetype	1635	3	545	0	0.761	0.11
condition*talker*tonetype	11000	3	3682	4	0.071	0.41
group*carrier*tonetype	679000	3	226000	186	< 0.001	1.00
condition*carrier*tonetype	1669	3	556	1	0.646	0.22
talker*carrier*tonetype	6729	3	2243	2	0.128	0.45
group*conditionion*talker*carrier	2633	1	2633	3	0.123	0.34
group*conditionion*talker*tonetype	5692	3	1897	1	0.251	0.31
group*condition*carrier*tonetype	23700	3	7914	8	< 0.001	0.88
group*talker*carrier*tonetype	1930	3	643	1	0.658	0.08
condition*talker*carrier*tonetype	8700	3	2900	3	0.082	0.34
group*condition*talker*carrier*tonetype	4030	3	1343	1	0.393	0.08

Table 12. Effects of language group, experimental condition, talker, carrier syllable and tone pair type on the reaction time measures of the tone discrimination using amplitude envelope.

Thus, the repeated measures ANOVA of the 1920 reaction time measures and the 1920 percent correct measures concurred to suggest that the talker effect did not reach statistical significance but the main effects of group, condition, carrier syllable and tone pair type were all statistically significant. The 1920 percent correct measures and the 1920 reaction time measures were then compared using the SPSS-19 linear regression analysis in order to quantify the correlation between the two types of measures. The correlation coefficient was 0.76, and was statistically significant ($F_{(1, 3838)} = 11946$; $p < 0.001$). Due to this high correlation between the reaction time measures and the percent correct measures, the following discussion will focus more on percent correct measures

3.2.5.2 Main Effects of Group and Condition and their Interaction

The between-group differences in the percent correct measures and the reaction time measures of the current tone discrimination experiment were both statistically significant as illustrated in Tables 13 and 14 ($F_{(1, 58)} = 2432$ and $p < 0.001$ for percent correct measures; $F_{(1, 58)} = 1743$ and $p < 0.001$ for reaction time). That is, the Cantonese group was more accurate and faster than the English group when the results from both conditions were combined, as shown in Tables 13 and 14. The main effect of condition was also statistically significant ($F_{(1, 58)} = 9$ and $p = 0.005$ for percent correct measures; $F_{(1, 58)} = 7$ and $p < 0.015$ for reaction time). Thus, the sixty participants as a group were more accurate and faster in Condition One (i.e., stimuli with original amplitude) than in Condition Two (i.e., stimuli with equalized peak amplitude), as indicated in Tables 13 and 14. The interaction between group and condition was also statistically significant for both the A’ and reaction time measures ($F_{(1, 58)} = 37$ and $p < 0.001$ for percent correct measures; $F_{(1, 58)} = 23$ and $p < 0.001$ for reaction time). That is, although the Cantonese participants were more accurate and faster than the English participants in both conditions, the between-group difference in accuracy and speed was reduced in Condition Two, as displayed in Tables 13 and 14.

	Condition 1	**Condition 2**	**2 Conditions Combined**
Cantonese Group	0.84 (0.18)	0.67 (0.17)	0.75 (0.04)
English Group	0.57 (0.16)	0.41 (0.15)	0.49 (0.15)
2 Groups Combined	0.69 (0.18)	0.54 (0.17)	0.62 (0.18)

Table 13. Group average percent correct measures and standard deviation (in parentheses) in Conditions One and Two of tone discrimination using amplitude envelope.

	Condition 1	**Condition 2**	**2 Conditions Combined**
Cantonese Group	328 (68)	401 (71)	364 (40)
English Group	375 (80)	442 (84)	408 (81)
2 Groups Combined	352 (68)	421 (71)	386 (69)

Table 14. Group average reaction time and standard deviation (in parentheses) in Conditions One and Two of tone discrimination using amplitude envelope.

3.2.5.3 Effect of Carrier Syllable and its Interaction with Group and Condition

The main effect of carrier syllable was statistically significant ($F_{(1, 58)} = 4109$ and $p < 0.001$ for percent correct measures; $F_{(1, 58)} = 1043$ and $p < 0.001$ for reaction time). The carrier syllable /wai/ produced significantly faster and more accurate responses than the carrier syllable /ji/, as shown in Figures 20 and 21. The interaction between carrier syllable and group was also statistically significant ($F_{(1, 58)} = 2906$ and $p < 0.001$ for percent correct measures; $F_{(1, 58)} = 174$ and $p < 0.001$ for reaction time). That is, although the carrier syllable /wai/ produced significantly more accurate and faster responses than the carrier syllable /ji/ in both groups of participants, the carrier syllable effect was more pronounced in the Cantonese group than in the English group as illustrated in Figures 20 and 21.

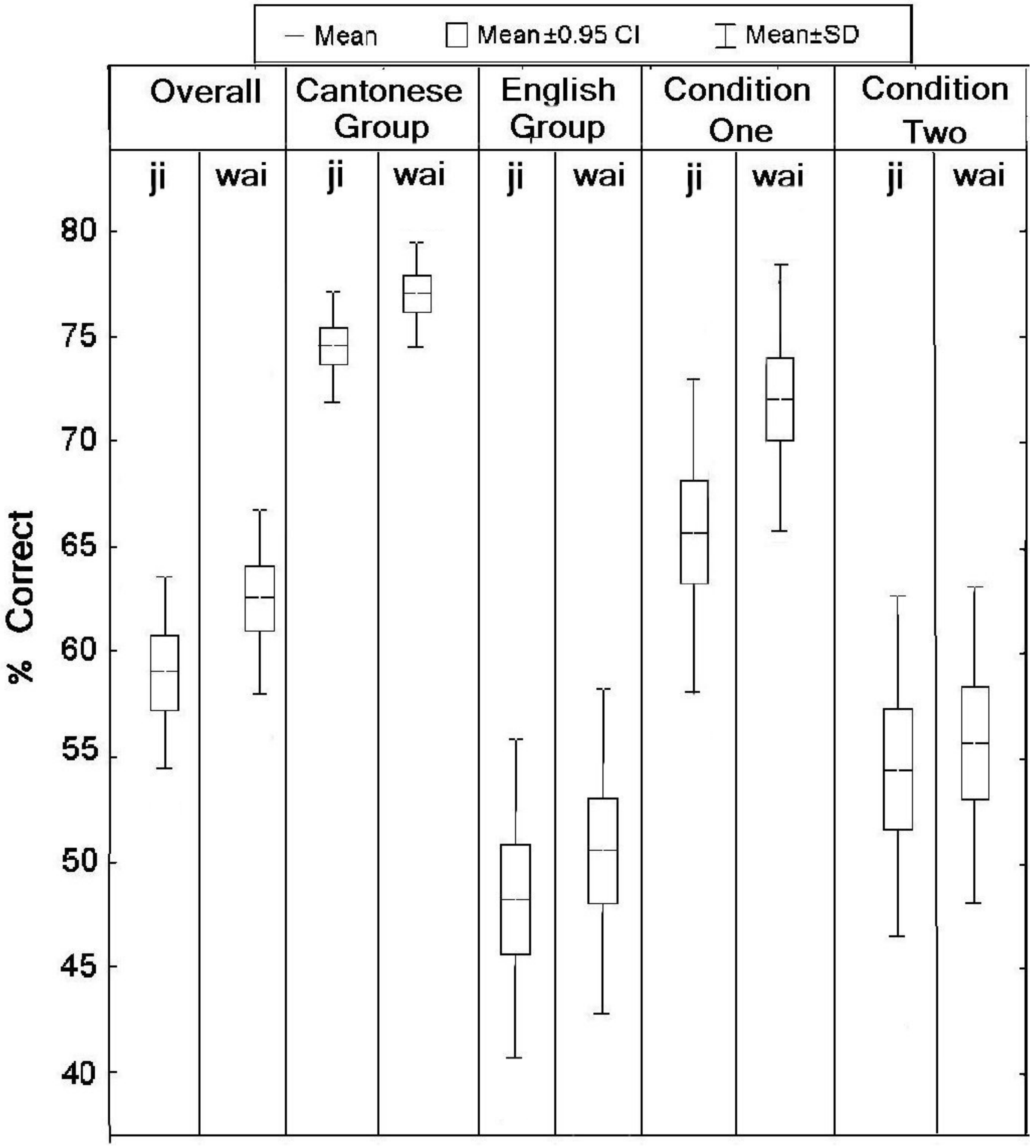

Figure 20. Effects of carrier syllable (in percent correct) and its interactions with group and condition in Cantonese tone discrimination using amplitude envelope

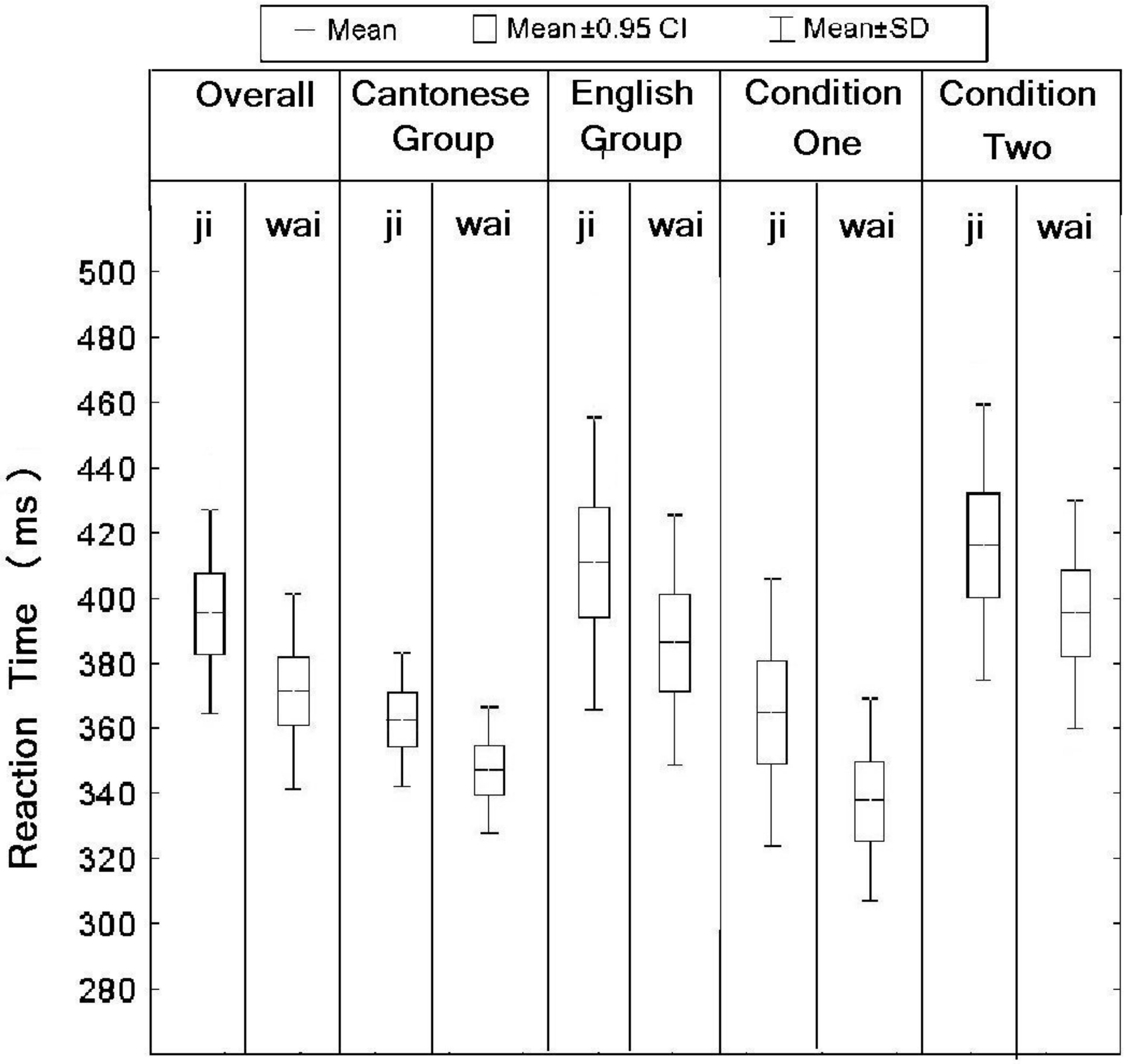

Figure 21. Effects of carrier syllable (in reaction time) and its interactions with group and condition in Cantonese tone discrimination using amplitude envelope

The interaction between carrier syllable and condition was also statistically significant ($F_{(1, 58)} = 26$ and $p < 0.001$ for percent correct measures; $F_{(1, 58)} = 16$ and $p < 0.001$ for reaction time). In other words, although the carrier syllable /wai/ produced better results than the carrier syllable /ji/ in both conditions, this carrier syllable effect was greater in Condition One (i.e., stimuli with original amplitude) than in Condition Two (i.e., stimuli with equalized peak amplitude), as indicated in Figures 20 and 21. Lastly, the three-way interaction among carrier syllable, group and condition was also significant statistically, but they did not seem to be relevant to the research questions of the current study. Therefore they will not be discussed further.

3.2.5.4 Effect of Tone Pair Types and its Interaction with Group, Condition, and Carrier

The main effect of tone pair type was statistically significant ($F_{(3, 177)} = 106$ and $p < 0.001$ for percent correct measures; $F_{(3, 177)} = 8$ and $p < 0.001$ for reaction time). Tukey's HDS post hoc tests were used for pairwise comparison of the percent correct measures for the four categories of tone pairs. As illustrated in Figure 22, the pairs consisting of tones with similar contours but different height were discriminated with significantly the lowest accuracy, whereas the other three types of tone pairs did not produce statistically significant differences in discrimination accuracy among themselves. Tukey's HDS post hoc tests were also used for pairwise comparison of the reaction time measures for the four categories of tone pairs, and the results were parallel to those of the percent correct measures as illustrated in Figures 22 and 23. That is, the pairs consisting of tones with similar contours but different height were discriminated with significantly the longest reaction time, whereas the other three types of tone pairs did not yield statistically significant differences in discrimination accuracy among themselves.

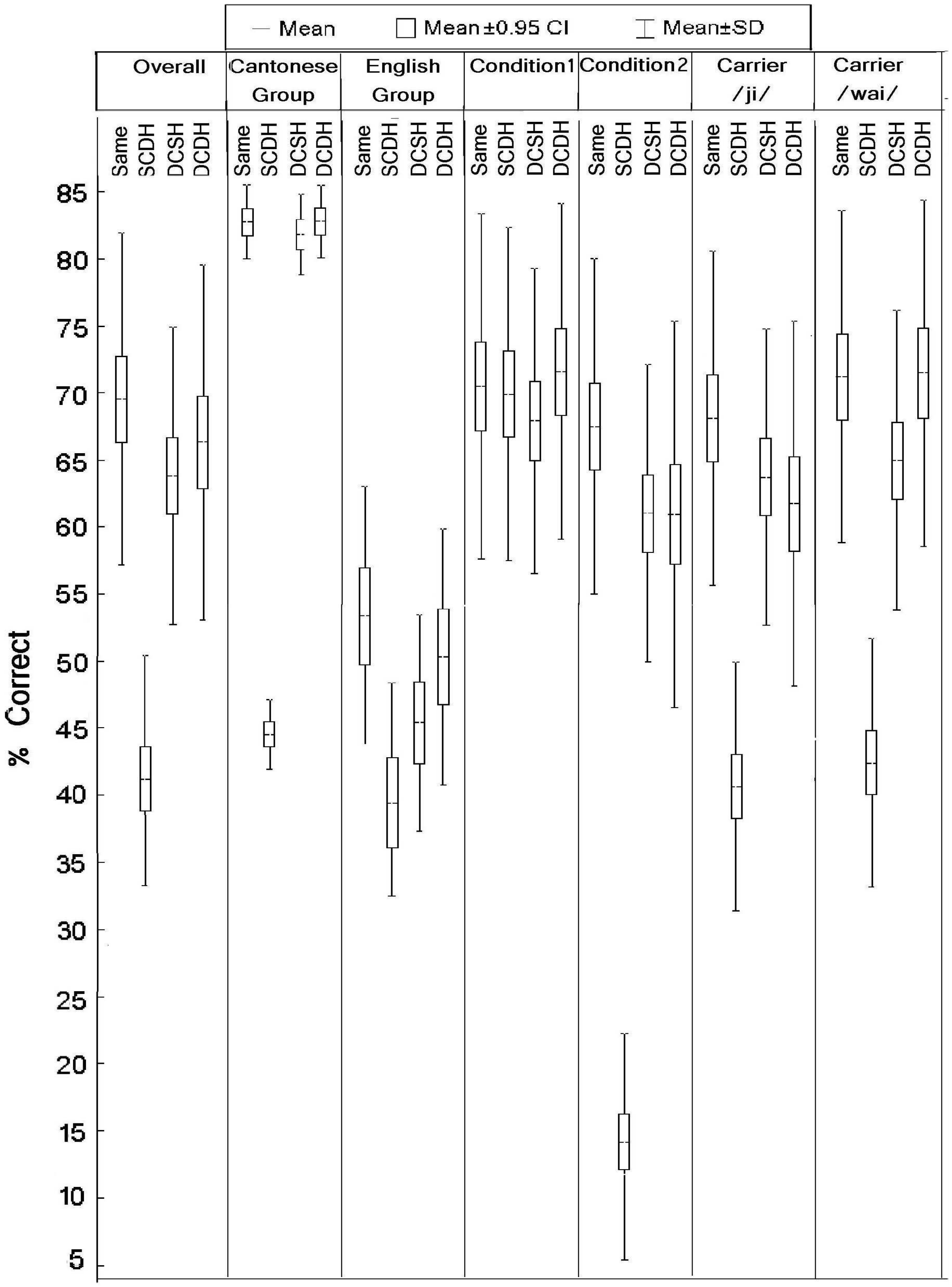

Figure 22. Effects of tone pair types (in percent correct) and interactions with language group, experimental condition and carrier syllable in Cantonese tone discrimination using amplitude envelope

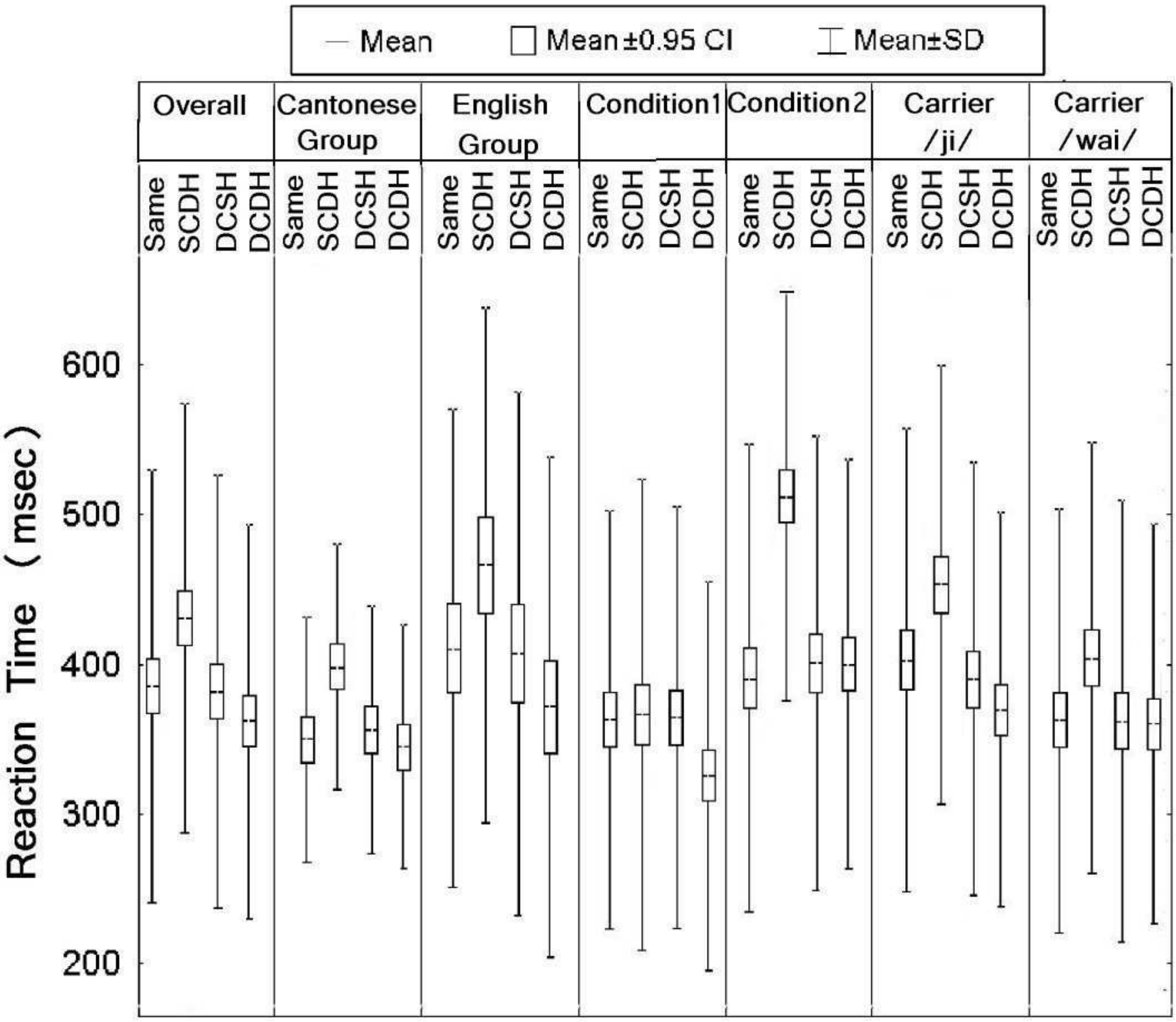

Figure 23. Effects of tone pair types (in reaction time) and interactions with language group, experimental condition and carrier syllable in Cantonese tone discrimination

The interaction between tone pair type and group was statistically significant ($F_{(3,177)}$ = 1724 and $p < 0.001$ for percent correct measures; $F_{(3,177)}$ = 377 and $p < 0.001$ for reaction time). Tukey's HDS post hoc tests were used to compare the performance of the two groups for each of the four categories of tone pairs. As indicated in Figures 22 and 23, although the Cantonese group outperformed the English group for each category of tone pairs, the performance gap between the two groups was smaller for one category (i.e., tones with similar contours but different height) than for the remaining three categories. The theoretical implication of this result will be discussed in Section 4.2.3.

The interaction between tone pair type and condition was also statistically significant ($F_{(3,177)}$ = 5 and p = 0.003 for percent correct measures; $F_{(3,177)}$ = 3 and p = 0.034 for reaction time). Tukey's HDS tests were used to compare the results of the four categories of tone pairs in each of the two conditions. In Condition One (i.e., stimuli with original amplitude), there was no significant difference in percent correct measures among the four categories of tone pairs at the alpha level of 0.05, as illustrated in Figures 21 and 22. In Condition Two (i.e., stimuli with equalized peak amplitude), however, the pairs consisting of tones with similar contours but different height were discriminated with significantly lower accuracy than the other three categories of tone pairs, and the tones with different contours but similar height were discriminated with lower accuracy than the **Same** pairs, as illustrated in Figure 22. In terms of reaction time (see Figure 23), the pairs consisting of tones with different contours and different height were discriminated with significantly shorter reaction time than the other categories of tone pairs in Condition One. In Condition Two, however, the pairs consisting of tones with similar contours but different height were discriminated with longer reaction time than the other categories of tone pairs. The theoretical implication of these findings will be discussed in Section 4.

The interaction between carrier syllable and tone pair type was also statistically significant ($F_{(3,177)}$ = 2608 and p < 0.001 for percent correct measures; $F_{(3,177)}$ = 208 and p < 0.001 for reaction time). Tukey's HDS tests were used to compare the results of the two carrier syllables for each of the four categories of tone pairs. With the carrier syllable /ji/, only the tones with similar contours but different height were discriminated with significantly lower accuracy than the other categories of tone pairs as shown in Figure 22. With the carrier syllable /wai/, however, the pairs involving tones with similar contours but different height and the pairs consisting of tones of different contours but similar height were discriminated with significantly lower accuracy than the other categories of tone pairs. In terms of reaction time (see Figure 23), the tones with similar contours but different height were discriminated with longer reaction time than the other categories of tone pairs with either carrier syllable as displayed in Figure 23.

Lastly, among all the three-way interactions involving tone pair type, only the interaction among tone types, carrier and group was significant statistically ($F_{(3, 177)} = 2519$ and $p < 0.001$ for percent correct measures; $F_{(3,177)} = 186$ and $p < 0.001$ for reaction time). Among all the four-way interactions involving tone types, only the interaction among tone pair type, carrier, condition and group was significant statistically ($F_{(3,177)} = 11$ and $p < 0.001$ for percent correct measures; $F_{(3,177)} = 8$ and $p < 0.001$ for reaction time). However, these interactions did not seem to be relevant to the research questions of the current study, and will not be discussed further.

3.3 Correlation between Tone Discrimination Results and the Three Acoustical Indices

Each Cantonese participant’s percent correct measure for each of the thirty-six tone pairs produced by each talker with each carrier syllable in Condition One (i.e., stimuli with original amplitude) was computed. This yielded 144 percent correct measures per participant, and a total of 4320 percent correct measures for the thirty Cantonese participants. These 4320 percent correct measures were then normalized using the arcsine transformation.

An SPSS-19 multiple regression analysis was used to correlate the Cantonese participants’ 4320 percent correct measures in Condition One with the three types of acoustical indices described in Section 3.1 (i.e., the F0-AE covariability indices, the AE height difference indices, and the AE contour similarity indices) using the. The coefficient of the correlation between the F0-AE covariability and the Cantonese participants' percent correct measures in Condition One was 0.06 when the other two acoustical indices were partialled out. This coefficient was statistically significant ($t_{(4319)} = 27$; $p < 0.001$), and thus the F0-AE covariability indices accounted for 0.4% of the variance in the Cantonese participants' tone discrimination performance in Condition One.

The coefficient of the correlation between the AE height difference indices and the Cantonese participants' percent correct measures in Condition One was 0.28 when the other two acoustical indices were partialled out. This coefficient was also statistically significant ($t_{(4319)} = 7$; $p < 0.001$), and thus the AE height difference indices alone accounted for 8% of the variance in the Cantonese participants' tone discrimination performance in Condition One. The coefficient of the correlation between the AE contour similarity indices and the Cantonese participants' percent correct measures in Condition One was 0.57 when the other two acoustical indices were partialled out. This coefficient was also statistically significant ($t_{(4319)} = 7$; $p < 0.001$), and thus the AE contour similarity indices alone accounted for 32% of the variance in the Cantonese participants' tone discrimination performance in Condition One.

Of all three acoustical factors examined, the AE contour similarity indices accounted for the most variance in the Cantonese participants' percent correct measures in Condition One, and the AE height difference indices accounted for the second most variance. Combined together, the AE height difference indices and the AE contour similarity indices accounted for 37% of variance in the Cantonese participants' tone discrimination performance in Condition One when the F0-AE covariability indices were partialled out. All three acoustical indices together accounted for approximately 37.5% of the variance in the Cantonese participants' percent correct measures in Condition One.

3.4 Results of the Survey and the Tone Identification Test

3.4.1 Survey

All participants stated that Condition Two of the current tone discrimination experiment (i.e., stimuli with equalized peak amplitude) was more difficult than Condition One (i.e., stimuli with original amplitude). The decision as to whether a tone pair contained two identical or different tones was reportedly easy and fairly clear-cut in Condition One, but often difficult in Condition Two. Even the participants who performed around chance level in Condition One said the decision was not difficult in Condition One, but much more difficult in Condition Two.

Before the experiment, none of the sixty participants knew about the six-tone classification of Cantonese tones. After the experiment, forty of the participants (twenty-nine native Cantonese speakers and eleven native English speakers) somewhat remembered the six-tone classification, and stated that there were two falling tones, two rising tone, and two tones that were neither falling nor rising. These eleven English participants' individual overall A' measures for the current discrimination experiment were compared with those of the remaining nineteen English participants using a one-way ANOVA. The between-group difference in A' measures (0.75 – 0.62 = 0.13) was statistically significant ($F_{(1, 28)} = 157$; $p < 0.001$).

Twenty-two participants (seventeen native Cantonese speakers and five native English speakers) even remembered the terms ***flat tone*** or ***level tone,*** and used them appropriately. As for the low versus high tone dichotomy, fifteen participants (twelve native Cantonese speakers and three native English speakers) remembered the terms ***high tone*** and ***low tone***, and described the high tones as being higher in louder, or more intense than the low tones. Thus, some participants from both language groups learned from the brief introduction to the six Cantonese tones during the practice session.

For the forty participants who remembered the six-tone classification, their decision as to whether a tone pair presented in the current discrimination experiment contained two identical or different tones was based on two criteria: the similarity or difference in the slope of the tones (i.e., falling, rising or neither), and the similarity or difference in the loudness or intensity level of the tones. As for the twenty participants who did not remember all six Cantonese tones, five native speakers of English said they just listened to the two tones within a pair carefully, and then decided whether the two tones sounded different or not. The remaining fifteen participants (one native speaker of Cantonese and fourteen native speakers of English) reported that, to a certain extent, they were able to pay attention to the intensity and the rising or falling slope of a tone, and then decide whether a tone pair contained two tones of identical intensity and slope.

3.4.2 Identification of Natural Tones in Cantonese

As described in Section 2.4 above, each of the sixty participants in the current Cantonese tone discrimination experiment was given a brief identification test of naturally produced Cantonese tones at the conclusion of the tone discrimination experiment. The Cantonese participants were instructed to identify each of the naturally produced Cantonese tones on the answer sheet (see Appendix Two). On the other hand, the English listeners were instructed to report on the answer sheet (see Appendix Three) whether each of the naturally produced Cantonese tones sounded like a question or statement. The purpose of the identification test was to assess each participant’s ability to process natural tones in Cantonese (see Section 2.4 for details).

Each Cantonese participant’s average percent correct measure for the tone identification test was provided in Figure 24. The measures ranged from 0.79 to 0.94, with a group mean of 0.85 and a standard deviation of 0.04. The group mean was significantly above the chance level of 16.7% ($t_{(29)} = 121$; $p < 0.001$). These thirty Cantonese participants’ tone identification scores were compared with their tone discrimination scores in Condition One using the SPSS-19 linear regression analysis. The correlation coefficient (Pearson’s r) was 0.69 and was statistically significant ($t_{(29)} = 27$; $p < 0.001$). Thus, the thirty Cantonese participants’ tone identification scores were fairly well correlated with their tone discrimination scores.

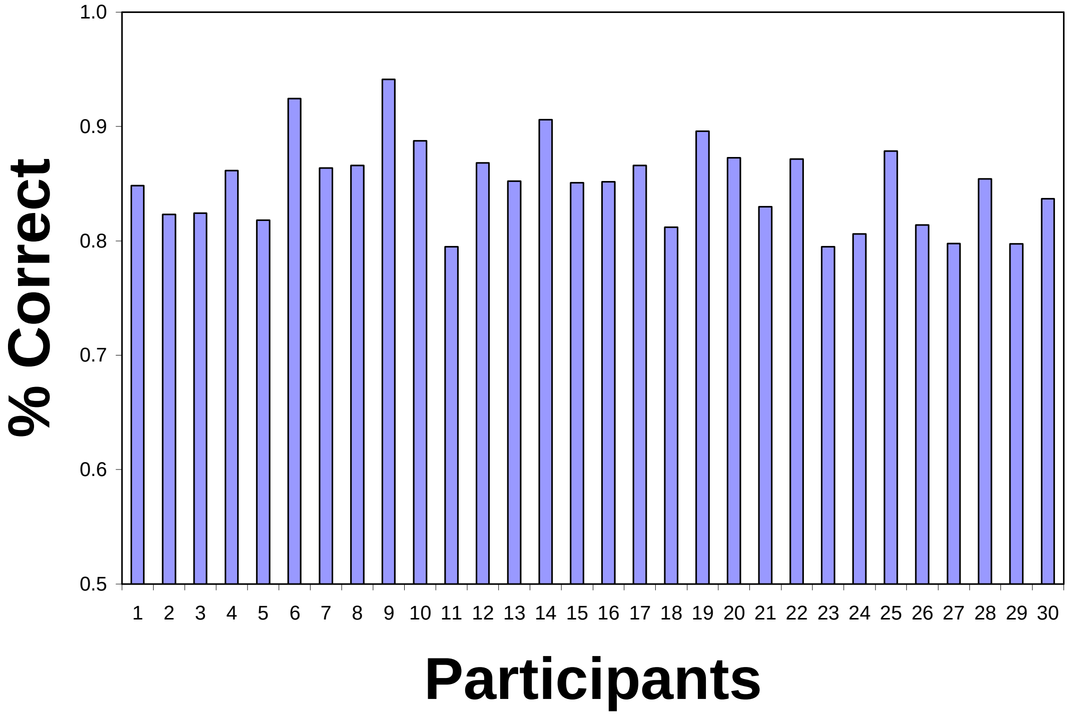

Figure 24. Thirty Cantonese participants’ performance in tone identification (chance = 16.7%).

For the English participants, 83% of the syllables with a falling tone or a level tone were identified as statements, whereas 89% of the syllables with a rising tone were recognized as questions. In order to derive a percent correct score for the English participants' tone identification performance, these identification patterns were operationally defined as correct responses because they were consistent with English prosody rules. That is, statements are typically produced with a falling or level F0 contour, and questions are generally produced with a rising F0 contour. The operationally defined percent correct measures for the thirty English participants are illustrated in Figure 25. The measures ranged from 0.69 to 0.95, with a group mean of 0.86 and a standard deviation of 0.07. The group mean was significantly above the chance level of 50% ($t_{(29)} = 66$; $p < 0.001$). The thirty English participants' tone identification scores were compared with their tone discrimination scores in Condition One using the SPSS-19 linear regression analysis. The correlation coefficient (Pearson's r) was 0.38, and was statistically significant ($F_{(1, 28)} = 5$; $p= 0.039$). Thus, the thirty English participants' tone identification scores were somewhat correlated with their tone discrimination scores.

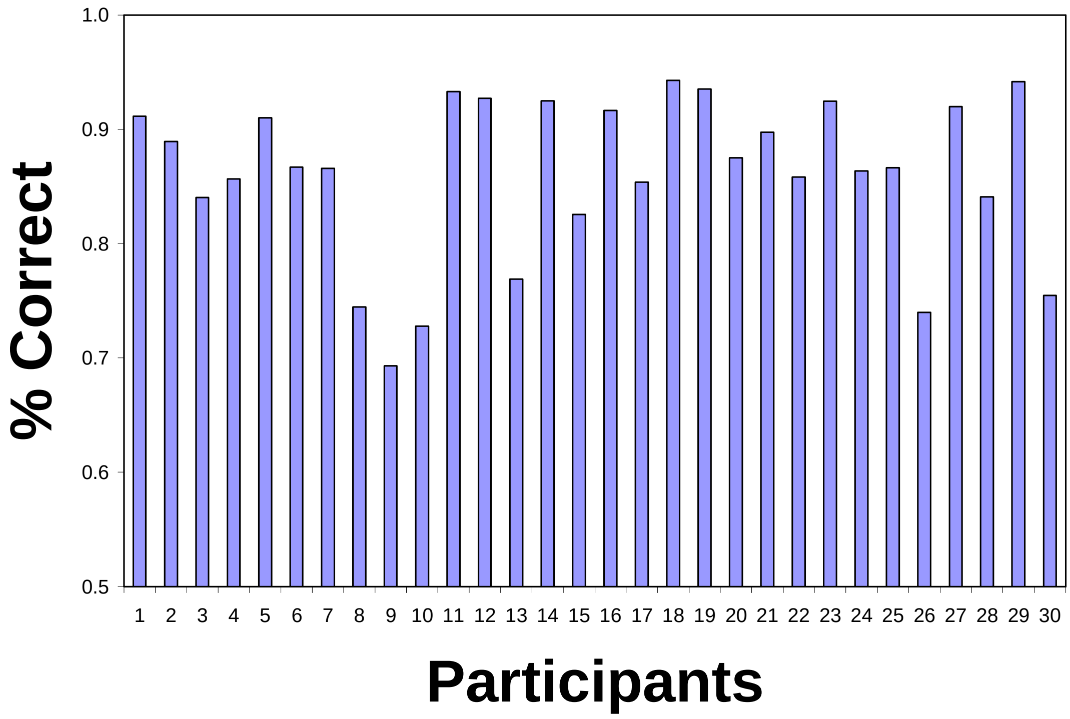

Figure 25. Thirty English participants' performance in tone identification (chance = 50%).

4. Discussion

As mentioned in Section 1.5 above, this dissertation was designed to answer four specific questions:

(1) **Can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts?**

a) More specifically, can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts involving a level tone and a contour tone?

b) In addition, can native listeners of Cantonese use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts involving tones with similar contours but different height as listed in Appendix Four?

(2) **Using the amplitude envelope cue alone, do native listeners of Cantonese discriminate**

a) Cantonese tone contrasts in one carrier syllable better than in the other?

b) some pairs of Cantonese tone contrasts better than others?

If so, can the results be accounted for on the basis of the degree of F0-AE covariability or other factors?

(3) **Using the amplitude envelope cue alone, does the Cantonese group discriminate Cantonese tone contrasts more accurately than the English group?**

(4) **What are the relative contributions of the contour and the height of amplitude envelope to Cantonese tone perception using amplitude envelope alone?**

The answers to these research questions will be discussed below.

4.1 Cantonese Tone Discrimination using Amplitude Envelope

As mentioned in the introduction, tone discrimination tasks have not been used in previously published studies on tone perception using amplitude envelope. Thus, the current experiment may be the first one to investigate tone discrimination using amplitude envelope alone, and any meaningful outcome from the current study would shed new light on tone perception using amplitude envelope alone. As described in Section 3.2.1 above, all Cantonese participants and 40% of the English participants in the current tone discrimination experiment were able to use the amplitude envelope cue alone to discriminate pairs of Cantonese tone contrasts with significantly above-chance accuracy. That is, they were able to differentiate tones with different contours, and distinguish low tones from high tones (including tones with similar contours but different height). This suggested that amplitude envelope alone provided a sufficient cue for differentiating tones with different contours, and tones with similar contours but different height. This was a unique contribution of the current study to the research on the role of amplitude envelope in tone perception because previous studies (e.g., Fu & Zeng, 2000; Kuo et al., 2008) focused on Mandarin which only used tone contour to differentiate tones.

The current study also indicated that native speakers of Cantonese were able to use amplitude envelope to discriminate Cantonese tones with significantly greater accuracy and shorter reaction time than native speakers of English. As reviewed in Section 2.4.4 above, previous studies (e.g., Lee et al., 1996) reported that native speakers of a tone language discriminate tones in their native language better than non-native speakers of that language due to the use of both linguistic and acoustic processing in the native speakers of the tone language, but the use of only acoustic processing in the non-native speakers of the tone language. Thus, the superior performance of the Cantonese group compared to the English group was likely to be related to the phonemic status of tones in Cantonese and not in English. That is, the Cantonese listeners may have used the phonemic tone categories to help discriminate the tones.

Alternatively, one may argue that experience with tones in their native language may have resulted in better loudness perception in the Cantonese listeners, and therefore gave the Cantonese listeners some advantage over the native English listeners. However, this alternative explanation would have difficulties explaining other findings from the current experiment. As described in Section 2.4 above, there were two experimental conditions in the current experiment. In Condition One, the amplitude of each stimulus remained the same as that of the original natural tone based on which the stimulus had been re-synthesized. In Condition Two, the peak amplitude of all stimuli was equalized, which eliminated the peak amplitude differences among the six Cantonese tones. Thus, the tones which had similar contours but different height in Condition One became tones with similar contours and ***equal*** height in Condition Two due to the removal of the peak amplitude differences among the six different tones. An example of these tones with similar contours and ***equal*** height in Condition Two was illustrated in Figure 13 in Section 3.2.1 above. Clearly, these tones had similar but not identical acoustics. If experience with tones in their native language were to result in better loudness/amplitude envelope perception in the Cantonese listeners, they should have been able to detect the difference between tones such as Tone 2 (High Rising) vs. Tone 5 (Low Rising) in Figure 13 with greater accuracy than the English listeners, but they actually performed with significantly lower accuracy than the native English listeners as illustrated in Figure 18 in Section 3.2.5.1 above. A reasonable explanation for the Cantonese listeners' poorer performance came from the finding from the post-experiment survey in the current study (see Section 3.4.1). That is, most Cantonese listeners reported that they had used the phonemic tone categories in Cantonese to help discriminate the tones in the current experience. The acoustic difference between tones such as Tone 2 (High Rising) vs. Tone 5 (Low Rising) in Figure 13 had been reduced drastically as a result of the peak amplitude equalization in Condition Two, and therefore the two tones were perceived as two tone tokens belonging to the same tone category in Condition Two. As for the English listeners, they relied mainly on acoustic criteria, and therefore were able to detect the minute difference between tones such as Tone 2 (High Rising) vs. Tone 5 (Low Rising) in Figure 13 with greater accuracy than the Cantonese listeners, as displayed in Figure 18.

In summary, the current study suggests that native Cantonese listeners used both linguistic and acoustic cues for the current tone discrimination using amplitude envelope, whereas native English speakers relied mainly on acoustic cues. This is consistent with previous studies on discrimination of naturally produced tones (e.g., Lee et al., 1996). However, none of the previously published studies on tone discrimination used stimuli containing only the amplitude envelope cue. Thus, the current study makes a unique contribution to the research on tone discrimination by revealing the role of amplitude envelope in tone discrimination, and by teasing apart the relative contribution of acoustic and linguistic processing to tone perception using amplitude envelope alone.

4.2 Acoustic Factors in Cantonese Tone Discrimination using Amplitude Envelope

4.2.1 F0-AE Covariability

Previous studies on tone perception using amplitude envelope have reported that the degree of F0-AE covariability is the most important factor affecting tone perception using amplitude envelope (e.g., Fu & Zeng, 2000). That is, in tone perception using amplitude envelope alone, a tone with an amplitude envelope that closely resembles its F0 contour would be recognized with greater accuracy than a tone whose amplitude envelope does not closely resemble its F0 contour (Fu & Zeng, 2000). However, as discussed in Section 1.4.3, a careful examination of the previous studies on tone perception using amplitude envelope revealed that the perceptual patterns in these previous studies could not be fully accounted for on the basis of F0-AE covariability. Furthermore, as described in Section 3.1 above, the degree of F0-AE covariability only accounted for 0.4% of the variance in the current tone discrimination experiment. Thus, the results in the current Cantonese tone discrimination using amplitude envelope could only be minimally accounted for by the degree of F0-AE covariability in Cantonese tones. This indicates that there must be other variables that affected Cantonese tone discrimination using amplitude envelope in the current study.

4.2.2 Acoustical Distance Indices

As discussed in Section 2.5.1.2, the acoustical distance indices consisted of the AE height difference index and the AE contour similarity index. These acoustical distance indices accounted for 37% of the variances in the current tone discrimination experiment. This is consistent with findings from previously published studies which suggested that tone contour and tone height were important acoustic cues for Cantonese tone perception (e.g., Barry, 2004b; Gandour, 1981 & 1983) as discussed in Section 2.1. However, these previous studies did not use stimuli containing amplitude envelope alone. Therefore, the current study makes new contribution to the research on tone discrimination by revealing the role of amplitude envelope contour and height in tone discrimination, and by creating the terms and concepts of ***acoustical distance indices, AE height difference indices*** and ***AE contour similarity indices***. These terms and concepts have never been discussed or examined in previously published studies on tone perception using amplitude envelope (e.g., Fu & Zeng, 2000; Kuo et al., 2008). Not only were these terms created out of nomenclature necessity during the current experiment, but their relative contribution to the current Cantonese tone discrimination using amplitude envelope was also investigated.

4.2.3 Relative Contribution of the Acoustical Distance Indices to Cantonese Tone Discrimination using Amplitude Envelope

As discussed in Section 1, Cantonese uses both tone contour and tone height for phonemic contrasts (Fok, 1974; Yip, 2002). If amplitude envelope alone can provide sufficient acoustic cues for the perception of tone contour and tone height in Cantonese, it would be reasonable to hypothesize that the contour of amplitude envelope provides an acoustic cue for the perception of tone contour, and the relative height of amplitude envelope serve as an acoustic for tone height. This hypothesis was tested empirically in the current experiment which included two experimental conditions. In Condition One, each stimulus was presented with the original amplitude (or intensity level) of the natural tone based on which the stimulus had been re-synthesized. In Condition Two, the peak amplitude of all stimuli was equalized, which eliminated the peak amplitude differences among the six Cantonese tones. Thus, the role of amplitude envelope in Cantonese tone discrimination was investigated in Condition One, whereas the independent contribution of the contour of amplitude envelope to Cantonese tone discrimination was examined in Condition Two. In other words, the relative contribution of the ***contour*** and ***height*** of amplitude envelope to Cantonese tone discrimination could be teased apart by comparing the results from the two conditions. This experimental design had not been used or discussed in previous studies on tone perception using amplitude envelope (see Kuo et al., 2008, for a review).

As described in Section 3.2 above, Condition One produced significantly better results than Condition Two. Thus, the removal of the peak amplitude differences among the six Cantonese tones impeded Cantonese tone discrimination using amplitude envelope. This was true for both the Cantonese group and the English group. Item analysis of both the Cantonese group and the English group revealed that the pairs involving tones with similar contours but different height were the most affected by the removal of the peak amplitude differences among the six Cantonese tones as illustrated in Figure 18 and described in Section 3.2.5. Thus, the relative height of amplitude envelope is an indispensable cue for the perception of tone height in Cantonese. This is consistent with findings from previous studies which suggested that the tone height cue was an indispensable cue for Cantonese tone perception (e.g., Barry, 2004b; Gandour, 1981 & 1983). However, these previous studies did not use stimuli containing amplitude envelope alone. Therefore, the current study makes another unique contribution to the research on tone discrimination by revealing the role of amplitude envelope height in tone discrimination.

The discrimination of Cantonese tone contours was not affected as much as by the removal of the peak amplitude differences among the six tones as described in Section 3.3 above. Thus, the contour of amplitude envelope alone provided a fairly robust acoustic cue for the discrimination of tone contour in Cantonese, in the presence or absence of the relative amplitude height cue. This is a new finding given that the relative amplitude envelope height had not been controlled for in previously published research on tone perception using amplitude envelope (e.g., Fu & Zeng, 2000; Kuo et al., 2008).

The relative contributions of the contour and height of amplitude envelope to the current Cantonese tone discrimination experiment were quantified using the SPSS-19 multiple regression analysis. As described in Section 3.3.2, the contour of amplitude envelope contributed more than the relative height of the amplitude envelope to the current Cantonese tone discrimination using amplitude envelope. This is also a unique contribution of the current study to the research on tone perception using amplitude envelope since the relative contributions of amplitude envelope height and contour had not been investigated in previously published studies on tone perception using amplitude envelope (see Kuo et al., 2008, for a review).

4.2.4 Summary of the Acoustical Factors

In the current study, three acoustical variables have been identified to have significantly contributed to Cantonese tone discrimination using amplitude envelope: the F0-AE covariability, the contour of amplitude envelope, and the relative height of amplitude envelope. Only the first two of these variables (i.e., the F0-AE covariability and the contour of amplitude envelope) had been investigated in previously published studies (see Kuo et al., 2008, for a review). In the current study, not only were the three acoustical variables investigated, but their relative contributions to tone perception using amplitude envelope were also examined. That is, amplitude envelope contour contributed the most to Cantonese tone discrimination using amplitude envelope, accounting for 32% of the variance in the current tone discrimination experiment. Amplitude envelope height had the second most contribution, accounting for 8% of the variances when the amplitude envelope contour was controlled for. The degree of F0-AE covariability had the least, but statistically significant contribution, accounting for 0.4% of the variances. Together, the three factors accounted for only 37.5% of the variance in the current Cantonese tone discrimination using amplitude envelope, and hence there should be other factors in Cantonese tone perception using amplitude envelope. In fact, as discussed in Section 4.1 above, there seemed to be linguistic variables affecting Cantonese tone discrimination using amplitude envelope.

4.3 Linguistic Factors in Cantonese Tone Discrimination using Amplitude Envelope

As described in Section 3.2, carrier syllable and tone type seemed to affect Cantonese tone discrimination using amplitude envelope. That is, the carrier syllable /wai/ produced significantly better results than the carrier syllable /ji/ in Cantonese participants, and some tone pairs were discriminated with significantly greater accuracy than others. Previous studies on tone perception using amplitude envelope suggested that the underlying mechanism for these carrier syllable effect and tone type effect was the degree of F0-AE covariability (e.g., Fu & Zeng, 2000). That is, carrier syllables that produced tones with amplitude envelopes closely resembling the F0 contours would be recognized with greater accuracy than carrier syllables that produced tones with amplitude envelopes greatly deviating from the F0 contours. Likewise, tone pairs whose amplitude envelopes closely resembled the F0 contours would be recognized with greater accuracy than tone pairs that greatly deviated from the F0 contours. However, as described in Section 4.2.1 above, the degree of F0-AE covariability only accounted for 0.4% of the variances in the current tone discrimination experiment. Thus, the degree of F0-AE covariability could only minimally account for the carrier syllable effect and tone type effect in the current study. The acoustical distance mechanism (i.e., the acoustical distance indices) accounted for more but not even half of the variance in the current tone discrimination experiment. Thus, there should be other factors involved in Cantonese tone perception using amplitude envelope.

Indeed, there was at least one linguistic factor that seemed to constrain the acoustical distance mechanism in the current Cantonese tone discrimination experiment. This linguistic factor appeared to be related categorical perception. As discussed in Section 1.4.1 above, categorical perception of tones has been a controversial issue, and previous studies provided empirical evidence for and against the existence of categorical perception of tones (see Zheng et al., 2010, for a review). Although the purpose of the current study was not to investigate the issue of categorical perception of tones, the results of the current study did provide some evidence in favor of categorical perception of tones. The first piece of evidence came from the discrimination of tones with similar contours but different height in the current study as discussed in Section 4.1. The second piece of evidence arose from the discrimination of the ***Same*** pairs. As discussed in Section 2.3 above, each ***Same*** pair used in this experiment consisted of two different tokens of a Cantonese tone produced with a specific carrier syllable by a native Cantonese talker. For example, one ***Same*** pair used in the current tone discrimination experiment consisted of two different tokens of Tone 1 (High Falling) produced with the carrier syllable /ji/ by the male talker. As illustrated in Figure 7 in Section 2.3 above, the acoustics of the two tokens were very similar, but not identical. Twenty of the thirty English participants mistakenly considered these two tokens as representing two different Cantonese tones at significantly above chance level ($t_{(19)} = 2.08$; $p = 0.028$), whereas all of the Cantonese participants correctly perceived these two tokens as representing the same Cantonese tone with significantly above chance accuracy, as displayed in Figure 18. This suggests that the Cantonese participants tended to perceive these two tokens phonemically, instead of using pure acoustic criteria, whereas the English participants seemed to rely more on acoustic cues.

The Cantonese participants' phonemic perception of the Cantonese tones may have somewhat constrained the acoustical distance mechanism described in the previous section. That is, two tone tokens differing within certain acoustic boundaries (as those illustrated in Figures 7 and 13) were perceived by Cantonese participants as tokens belonging to the same tone category, which prevented the acoustical distance mechanism from operating within those boundaries. In contrast, two tones differing beyond certain acoustic boundaries were treated as belonging to two different tone categories by the Cantonese participants, in which case the acoustical distance mechanism seemed to be operating. These findings were consistent with results from previous studies in favor of categorical perception of naturally produced tones (see Zheng et al., 2010, for a review). It would be interesting to investigate, in a future study, the boundaries among Cantonese tones when amplitude envelope is the only available cue for Cantonese tone perception in native speakers of Cantonese. For the time being, it suffices to say that the Cantonese participants' phonemic perception of Cantonese tones may have somewhat constrained the operation of the acoustical distance mechanism in the current tone discrimination study, and could explain why the acoustical distance mechanism could only account for 37.5% of the variance in the current tone discrimination experiment.

4.4 Summary of the Findings from the Current Study

In the current tone discrimination study, the Cantonese participants were able to use the amplitude envelope cue to discriminate Cantonese tone pairs above chance level, and with greater accuracy and shorter reaction time than the native speakers. This suggested that amplitude envelope alone was effective in conveying Cantonese tone contrasts to native listeners of Cantonese. In particular, when only the amplitude envelope cue was available in Cantonese tone perception, the relative amplitude height was indispensable for the perception of tone height, whereas the contour of amplitude envelope provided a robust and sufficient cue for the discrimination of tones with different contours. F0-AE covariability in Cantonese tone production played a weak but statistically significant role in Cantonese tone discrimination using amplitude envelope alone. Together, these three acoustical factors accounted for 37.5% of the variances in the current Cantonese tone discrimination. In addition, linguistic factors such as phonemic perception of Cantonese tones seemed to have played an important role in native speakers of Cantonese in the current experiment, and somewhat constrained the operation of the acoustic factors.

5. Theoretical and Clinical Implications

The current study indicated that amplitude envelope alone can aid tone perception in a language such as Cantonese which uses both tone contour and tone height for phonemic tonal contrasts. Since tone languages in the world are described to use only tone height, tone contour or a combination of both for phonemic tonal contrasts (Yip, 2002), it is theoretically possible for amplitude envelope to cue tone perception in any tone language. The clinical implication of this finding could be far-reaching. As discussed above, cochlear implant users who speak a tone language have difficulties perceiving tones and therefore overall speech due to the limitations of currently available commercial speech processing strategies in encoding F0. Experimental speech processing strategies (e.g., Luo & Fu, 2004) have demonstrated promising results in improving tone perception in cochlear implant users by enhancing the encoding of the amplitude envelope cue. These experimental strategies could potentially be adapted to enhance the encoding of the amplitude envelope cue in Cantonese tones and thus improve tone perception in Cantonese-speaking implantees, given the current finding that amplitude envelope could also cue Cantonese tone perception. In broader terms, since the current dissertation suggests that it is theoretically possible for amplitude envelope to cue tone perception in any tone language, the above-mentioned experimental speech processing strategies could possibly be adapted to enhance the encoding of the amplitude envelope cue in any tone language, and thus improve tone perception in cochlear implantees who speak any tone language.

Furthermore, the current finding that amplitude envelope could cue certain tone types better than others for tone perception could help design better speech processing strategies to improve tone perception in cochlear implant users. That is, when designing algorithms for encoding the amplitude envelopes of tones, one should bear in mind that an algorithm that works well for one tone type might not be suitable for other tone types. Therefore, various tone types should be taken into consideration.

Similarly, given the current finding that some carrier syllables produced consistently better tone discrimination results than others, it could imply that amplitude envelope may cue some carrier syllables better than others for tone perception. This outcome could also help improve speech processing strategies for a better tone perception in cochlear implant users. For example, when designing algorithms for encoding the amplitude envelopes of tones, one should keep in mind that an algorithm that works well with one carrier syllable might not be suitable for other carrier syllables. Therefore, various carrier syllables should be taken into consideration. Since carrier syllables consist of vowels and consonants, various combinations of vowels and consonants should be taken into account.

6. Conclusion

This dissertation has shed new insight into the very nature of the role of amplitude envelope in tone perception. In particular, it has made several unique contributions to the research on the role of amplitude envelope in tone perception. These unique contributions could potentially help improve tone perception in cochlear implant users, which could in turn contribute to the improvement of overall speech perception in cochlear implant users. Since speech perception is an indispensable part of daily verbal communication, this dissertation could potentially help improve the verbal communication skills in cochlear implant users. Thus, the clinical implications of this dissertation could be far-reaching.

However, there were several limitations in this dissertation. First, the tone discrimination experiment lasted approximately three hours per participant, and thus the results might have been affected by the effect of fatigue. In other words, the results of the current experiment may have under-estimated the contribution of amplitude envelope to Cantonese tone discrimination.

Second, the stimuli used in the current tone discrimination and identification experiments were based on tones produced with the carrier syllables /ji/ and /wai/ only. As discussed in Section 3.2 above, there was significant effect of carrier syllable in the current tone discrimination experiment. That is, the carrier syllable /wai/ produced significantly faster and more accurate responses than the carrier syllable /ji/. Thus, further studies using all possible Cantonese carrier syllables as stimuli may be necessary to determine whether the amplitude envelopes of Cantonese tones produced with any Cantonese carrier syllables can effectively cue tone perception in Cantonese.

Third, the tones used in the current tone discrimination and identification experiments were produced by one male talker and one female talker only. Although the main effect of talker did not reach statistical significance in the current tone discrimination experiment, the statistical power was relatively low (~0.42) as discussed in Section 3.2. Therefore, it is difficult to interpret this statistically insignificant main effect of talker and its interactions with other factors as discussed in Section 3.2. More talkers may be needed in future studies in order to increase the statistical power, and hence facilitate the interpretation of the main effect of talker and its interactions with other factors.

Lastly, each tone used in the current discrimination and identification experiments was produced with a monosyllabic carrier in isolation. It is difficult to extrapolate the findings of the current study to connected speech in Cantonese. Thus, further studies using tones produced in connected speech are necessary to investigate whether amplitude envelope alone can cue tone perception in connected speech in Cantonese.

Appendix One:

Answer Sheet for Cantonese Listeners’ Tone Identification (with Carrier Syllable /ji/)

請圈上正确的字.謝謝. (Please circle the appropriate character upon hearing each word.)

1.	衣	椅	意	儿	耳	二
2.	衣	椅	意	儿	耳	二
3.	衣	椅	意	儿	耳	二
4.	衣	椅	意	儿	耳	二
5.	...					

Appendix Two

Answer Sheet for Cantonese Listeners’ Tone Identification (with Carrier Syllable /wai/)

請圈上正确的字.謝謝. (Please circle the appropriate character upon hearing each word.)

1. 威 位 餵 圍 偉 為
2. 威 位 餵 圍 偉 為

…

Appendix Three

Answer Sheet for the English Group’s Tone Identification

Upon hearing each word, please decide whether it sounds like a statement or question, and circle the appropriate answer below.

1. Question Statement
2. Question Statement
3. Question Statement
4. Question Statement
5. ...

Appendix Four

A complete list of the four categories of tone pair types used in the current study

1. **Same** Pairs

Tone 1 (High Falling) vs. Tone 1 (High Falling)
Tone 2 (High Rising) vs. Tone 2 (High Rising)
Tone 3 (Upper Middle Level) vs. Tone 3 (Upper Middle Level)
Tone 4 (Low Falling) vs. Tone 4 (Low Falling)
Tone 5 (Low Rising) vs. Tone 5 (Low Rising)
Tone 6 (Lower Middle Level) vs. Tone 6 (Lower Middle Level)

2. Pairs consisting of tones with similar contours but different height (SCDH Pairs)

Tone 1 (High Falling) vs. Tone 4 (Low Falling)
Tone 2 (High Rising) vs. Tone 5 (Low Rising)
Tone 3 (Upper Middle Level) vs. Tone 6 (Lower Middle Level)
Tone 4 (Low Falling) vs. Tone 1 (High Falling)
Tone 5 (Low Rising) vs. Tone 2 (High Rising)
Tone 6 (Lower Middle Level) vs. Tone 3 (Upper Middle Level)

3. Pairs consisting of tones with different contours but similar height (DCSH Pairs)

Tone 1 (High Falling) vs. Tone 2 (High Rising)
Tone 1 (High Falling) vs. Tone 3 (Upper Middle Level)
Tone 2 (High Rising) vs. Tone 1 (High Falling)
Tone 2 (High Rising) vs. Tone 3 (Upper Middle Level)
Tone 3 (Upper Middle Level) vs. Tone 1 (High Falling)
Tone 3 (Upper Middle Level) vs. Tone 2 (High Rising)
Tone 4 (Low Falling) vs. Tone 5 (Low Rising)
Tone 4 (Low Falling) vs. Tone 6 (Lower Middle Level)
Tone 5 (Low Rising) vs. Tone 4 (Low Falling)
Tone 5 (Low Rising) vs. Tone 6 (Lower Middle Level)
Tone 6 (Lower Middle Level) vs. Tone 4 (Low Falling)
Tone 6 (Lower Middle Level) vs. Tone 5 (Low Rising)

4. Pairs consisting of tones with different contours and different height (DCDH Pairs)

Tone 1 (High Falling) vs. Tone 5 (Low Rising)
Tone 1 (High Falling) vs. Tone 6 (Lower Middle Level)
Tone 2 (High Rising) vs. Tone 4 (Low Falling)
Tone 2 (High Rising) vs. Tone 6 (Lower Middle Level)
Tone 3 (Upper Middle Level) vs. Tone 4 (Low Falling)

Tone 3 (Upper Middle Level) vs. Tone 5 (Low Rising)
Tone 4 (Low Falling) vs. Tone 2 (High Rising)
Tone 4 (Low Falling) vs. Tone 3 (Upper Middle Level)
Tone 5 (Low Rising) vs. Tone 1 (High Falling)
Tone 5 (Low Rising) vs. Tone 3 (Upper Middle Level)
Tone 6 (Lower Middle Level) vs. Tone 1 (High Falling)
Tone 6 (Lower Middle Level) vs. Tone 2 (High Rising)

Bibliography

American Standards Association (1960). Acoustical Terminology SI, 1-1960. New York.

American Speech-Language-Hearing Association (1997). *Guidelines for Audiologic Screening.* Rockville, MD.

Baken, R.J. & Orlikoff, R. (2000). *Clinical Measurement of Speech and Voice.* San Diego: College Hill Press.

Barry, J. G., Blamey, P. J., Martin, L. F. A., Lee, K. Y.S., Tang, T., Ming, Y. Y. & van Hasselt, C. A. (2002). Tone discrimination in Cantonese-speaking children using a cochlear implant. *Clinical Linguistics and Phonetics, 16*, 79-99.

Bauer, R.S., & Benedict, P.K. (1997). *Modern Cantonese Phonology.* The Hague: Mouton de Gruyter.

Burnham, D., Francis, E., Webster, D., Luksaneeyanawin, S., Attapaiboon, C., Lacerda, F., & Keller, P. (1996). Perception of lexical tones across languages: Evidence for a linguistic mode of processing. *International Conference on Spoken Language Processing, 1996*, 2514-2517.

Chao, Y. (1947). *Cantonese Primer.* Cambridge, MA: Harvard University Press.

Cohen, J (1992). A power primer. *Psychological Bulletin, 112*, 155–159

Connell, B., Hayward, R.J., & Ashkaba, J.A. (2000). Observations in Kunama tone (Barka dialect). *Studies in African Linguistics, 29,* 1-41.

Donaldson, W. (1993). Accuracy of d’ and A’ as estimates of sensitivity. *Bulletin of the Psychonomic Society, 31*, 271-274.

Erickson, D. (1976). *Physiological Analysis of the Tones of Thai. Unpublished Ph.D. Dissertation.* University of Connecticut.

Erickson, D. (1993). Laryngeal muscle activity in connection with Thai Tones. *Annual Bulletin RILP, 27*, 135-150.

Fok, Y.Y. (1974). *A Perceptual Study of Tones in Cantonese.* Hong Kong: University of Hong Kong Press.

Francis, A.L., Ciocca, V.C., Wong, N.K.U., Leung, W.H.Y., & Chu, P.C.Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America, 119*, 1712-1726.

Fromkin, A. (1978). *Tone: a linguistic survey.* New York: Academic Press.

Fu, Q.J., & Zeng, F.G. (2000). Identification of temporal envelope cues in Chinese tone recognition. *Asia Pacific Journal of Speech, Language and Hearing, 5,* 45-47.

Fu, Q.J., Zeng, F.G., Shannon, R.V., & Soli, S.D. (1998). Importance of tone envelope cues in Chinese speech recognition. *Journal of Acoustical Society of America, 104,* 505-510.

Gandour, J.T. (1978). The perception of tone. In A. Fromkin (Ed.), *Tone: a linguistic survey.* New York: Academic Press.

Gandour, J.T. (1981). Perceptual dimensions of tone: evidence from Cantonese. *Journal of Chinese Linguistics, 9,* 20-36.

Gandour, J.T. (1983). Tone perception in far eastern languages. *Journal of Phonetics, 11*, 149-175.

Garding, E., Fujimura, O., & Hirose, H. (1970). Laryngeal control of Swedish word tones. *Annual Bulletin of Research Institute of Logopedics & Phoniatrics, 4,* 45-54.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75,* 424-429.

Gu, W., Hirose, K., & Fujisaki, H. (2005). Analysis of the effects of word emphasis and echo questions on pitch contours of Cantonese utterances. *Interspeech 2005,* 1825-1828.

Gu, W. & Lee, T. (2007). Effect of tonal context and focus on Cantonese pitch. *Proceedings of the 16th International Congress of Phonetic Science.*

Halle, P. A. (1994). Evidence for tone-specific activity of the sternohyoid muscle in modern standard Chinese. *Language & Speech, 37*, 103-123.

Hirano, M., Ohala, J., & Vennard, W. (1969). Function of laryngeal muscles in regulating fundamental frequency and intensity in phonation. *Journal of Speech and Hearing Research, 12*, 616-628.

Hirose, H., Simada, Z., & Fujimura, O. (1970). An electromyographic study of the activities in laryngeal muscles during speech utterances. *Annual Bulletin RILP, 4,* 9-25.

Hirose, H. (2010). Investigating the physiology of laryngeal structures. In W. Hardcastle, J. Laver & I.E. Gibbon (Ed.). *Handbook of Phonetical Sciences.* Malden, MA: Blackwell Publishing.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74,* 351-354.

Hyman, L.M. (2007). Universals of tone rules: 30 years later, C. Gussenhoven & T. Riad (eds.), *Tones and Tunes*. Berlin: Mouton de Gruyter.

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

Kleeman, J. (2010). *Oxford Chinese Dictionary*. Oxford: Oxford University Press.

Kong, Y.Y., & Zeng, F.G. (2006). Temporal and spectral cues in Mandarin tone recognition. *Journal of the Acoustical Society of America, 120*, 2830-40.

Krenmayr, A., Qi, B., Liu, B., Liu, H., Chen, X., Han, D., Schatzer, R., & Zierhofer, C.M. (2011). Development of a Mandarin tone identification test: Sensitivity index d' as a performance measure for individual tones. *International Journal of Audiology, 50*, 155-163.

Kuo, Y., Rosen, S., & Faulkner, A. (2008). Acoustic cues to tonal contrasts in Mandarin: Implications for cochlear implants. *Journal of the Acoustical Society of America, 122*, 2815-2824.

Lee, P. Y. (2009). *Pocket Cantonese Dictionary.* Berkeley: Periplus Publishing.

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

Li, Y., Lee, T. & Qian, Y. (2002). Acoustical pitch analysis of continuous Cantonese speech. *Proceedings of International Symposium on Chinese Spoken Language Processing, P1B.04.*

Liang, Z.A. (1963). Auditory basis of tone recognition in Standard Chinese. *Acta Physiologica Sinica, 26,* 85-90.

Liu, S., & Samuel, A.G. (2004). Perception of Mandarin lexical tones when pitch information is neutralized. *Language and Speech, 47, 109-138.*

Loizou, P. C. (2006). Speech processing in vocoder-centric cochlear implants. *Advances in Otorhinolaryngology, 64*, 109-143.

Luo, X., & Fu, Q. (2004). Enhancing Chinese tone recognition by manipulating amplitude enveloped: implications for cochlear implants. *Journal of the Acoustical Society of America, 116*, 3659-3667.

Luo, X, F, Q.J., Wei, C.G., & Cao, K.L. (2008). Speech recognition and temporal amplitude modulation processing by Mandarin-speaking cochlear implant users. *Ear & Hearing, 29*, 957-970.

Ma, J.K., Ciocca, V., & Whitehill, T.L. (2006). Effects of intonation on Cantonese lexical tones. *Journal of the Acoustical Society of America, 120*, 3978-3987.

Macmillian, N. A. & Creelman, C.D. (1991). *Detection Theory: A User's Guide*. Cambridge, MA: Cambridge University Press.

Matthew, S., & Yip, V. (1994). *Cantonese: A Comprehensive Grammar.* New York: Routledge.

Pastore, R. E., Crawley, E. J., Berens, M. & Skelly, M. (2003). "Nonparametric" A' and other modern misconceptions about signal detection theory. *Psychonomic Bulletin & Review, 10*, 556-569.

Plack, C. J., Oxenham, A. J., Fay, R. R., & Popper, A. N. (2005). *Pitch: Neural Coding and Perception*. New York: Springer.

Rao, B., Ouyang, J., & Zhou, W. (1996). *Guangzhouhua Fangyan Cidian* (Guangzhou Cantonese Dialect Dictionary). Hong Kong: Commercial Press.

Rabiner, L. R. (1977). On the use of autocorrelation analysis for pitch detection. *IEEE Transacctions in Acoustics and Speech Signal Processing, 26,* 24-33.

Rose, P. (1984). Role of subglottal pressure and vocal cord tension in a Chinese dialect. *New Papers on Chinese Language Use.* Camberra: Australian National University Press.

Rose, P. (2000). Hong Kong Cantonese citation tone acoustics: a linguistic tonetic study. *Proceedings of the 8th Australian International Conferences on Speech Science and Technology,* 198-203.

Rosen, S. (1992). Temporal information in speech: Acoustic, auditory and linguistic aspects. *Philosophical Transactions: Biological Sciences, 336*, 367-373.

Sagart, L. Halle, P., Boysson-Bardies, D. D, & Arabia-Guidet, C. (1986). Tone production in modern standard Chinese: An electromyographic investigation. *Cahiers de Linguistique – Asie Orientale, 15*, 205-221.

Salthouse, T.A. & Hedden, T. (2002). Interpreting reaction time measures in between-group comparisons. *Journal of Clinical and Experimental Neuropsychology, 24,* 858-872.

Schouten, J. F., Ritsma, R. J., & Cardozo, B. L. (1962). Pitch of the residue. *Journal of the Acoustical Society of America, 34,* 1418-1424.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44*, 1735-1736.

Shen, X.S. (1989). Interplay of the four citation tones in Mandarin and intonation. *Journal of Chinese Linguistics, 17,* 61-73.

Shih, C.L. (1988). Tone and intonation in Mandarin. *Working Papers of Cornell Phonetics Laboratory, 3,* 83-109.

Simada, Z. B. & Hirose, H. (1971). Physiological correlates of Japanese accent patterns. *Annual Bulletin of RILP, 5, 41-49.*

Strange, W. (2010). Personal communication.

Swets, J.A. (1996). *Signal Detection Theory and ROC Analysis in Psychology and Diagnosis.* Mahwah, NJ: Erlbaum.

Titze, I.R. (1995) *Principles of Voice Production.* Iowa City: National Center for Speech and Voice.

Tong, M.C., & Lee, K.Y. (2009). Do Chinese speakers need a specialized cochlear implant system? *Journal of Otorhinolaryngology & Related Specialties*. *71*, 184-6.

Vance, T. J. (1976). An experimental investigation of tone and intonation in Chinese. *Phonetica, 33,* 368-392.

Vance, T. J. (1977). Tonal distinction in Cantonese. *Phonetica, 34,* 93-107.

Werker, J.F., & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37,* 35-44.

Whalen, D.H. & Xu, Y. (1992). Information for mandarin tones in amplitude contour and in brief segments. *Phonetica, 49,* 25-47.

Wickens, T.D. (2002). *Elementary Signal Detection Theory*. New York: Oxford University Press.

Wilson, B.S., & Dorman, M. (2008). *Cochlear implants*: A remarkable past and a brilliant future. *Hearing Research, 242,* 3-21.

Wong, P. (1999). Effects of duration on the perception of Cantonese level tones. *Journal of the Acoustical Society of America, 106,* 2150-2153.

Wong, P. & Diehl, R. (2003). Perceptual normalization for inter- and intra-talker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research, 46,* 413-421.

Xu, Y. (1994). Production and perception of coarticulated tones. *Journal of the Acoustical Society of America, 95,* 2240-2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25,* 61-83.

Xu, Y. (2006). Tone in connected discourse. In K. Brown (Ed.). Encyclopedia of Language and Linguistics. Oxford: Elsevier.

Xu, L.& Pfingst, B.E. (2003). Relative importance of temporal envelope and fine structure in lexical-tone perception. *Journal of the Acoustical Society of America, 114,* 3024-3027.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33,* 319-337.

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

Yuan, M., Lee, T., Yuen, K.C.P., Soli, S.D., Tong, M.C.F. & van Hasselt, C.A. (2007). Band-specific temporal periodicity enhancement for Cantonese tone perception with noise-excited vocoder. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007,* 694-697.

Yuen, K.C., Tong, M.C.F., van Hasselt, C., Yuan, M., Lee, T, & Soli, S.D. (2009). Cantonese lexical tone recognition from frequency-specific temporal envelope and periodicity components in the same versus different noise band carriers. *Cochlear Implants International, 10,* 148-158.

Zheng, H. Y., Minett, J.W., & Wang, W. S. Y. (2010). The impact of tone systems on the categorical perception of lexical tones: An event-related potentials study. *Language and Cognitive Processes, 1,* 1-31.

Zhou, Y. V. (2010). *The Role of Amplitude Envelope in Cantonese Lexical Tone Perception in Cantonese-speaking Adults with Normal Hearing.* Unpublished Master of Philosophy Thesis. The City University of New York.