Tonal Processing in Cantonese

JIA, Shiwei

A Thesis Submitted in Partial Fulfillment
of The Requirements for the Degree of
Doctor of Philosophy
in
Psychology

The Chinese University of Hong Kong
July 2011

Thesis/ Assessment Committee

Professor Him Cheung (Chair)

Professor Hsuan-Chih Chen (Thesis Supervisor)

Professor Liqiang Huang (Committee Member)

Professor Ellick Kin-Fai Wong (External Examiner)

Acknowledgements

This work was supported in part by grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CUHK4142/04H and CUHK441008) to Prof. Hsuan-Chih Chen and a dissertation grant from the Department of Psychology, The Chinese University of Hong Kong.

I would first like to thank my thesis supervisor, Prof. Hsuan-Chih Chen, for his enduring support and patience through my years of studies, and for his invaluable comments on this thesis. I also like to express my sincere gratitude to my thesis/ assessment committee chair, Prof. Him Cheung, and my thesis committee member, Prof. Liqiang Huang, and the external examiner, Prof. Ellick Kin-Fai Wong for their time and constructive suggestions.

TABLE OF CONTENTS

List of Tables

List of Figures

Abstract

In order to investigate the issues of: 1) how tonal dimensions affect tonal processing, 2) how stimulus type affects tonal processing, and 3) how task requirement affects tonal processing, four experiments were conducted on Cantonese tone perception. Experiments 1 and 2 were two dichotic-listening (DL) experiments. They manipulated the same independent variables (IVs): 1) f0 direction (level and contour) and 2) stimulus type (syllable, pseudo-syllable, and hum), but used different tasks (a discrimination task and an identification task in Experiments 1 and 2, respectively). Experiments 3 and 4 were two event-related potential (ERP) experiments. They manipulated the same IVs: 1) f0 direction (level and contour), 2) f0 height-deviation (large and small), and 3) stimulus type (syllable and hum), but the tasks were different (passive oddball task in Experiment 3 and active oddball task in Experiment 4). The effects of tonal dimension (f0 direction and f0 height) were used to examine how acoustic features (low level) modulate tonal processing. Moreover, the effects of stimulus type and task requirement were used to test the effect of lexical involvement (high level) on tonal processing.

In Experiments 1 and 2, a left-ear advantage (LEA) was found for Cantonese tone processing in both discrimination and identification tasks. However, for the Mandarin tone, an LEA and a right-ear advantage (REA) were found in discrimination and identification tasks, respectively. In the Cantonese condition, the processing speeds of tone for syllables and pseudo-syllables were faster than that for hums, and the performance on the level tone was better than that on the contour tone.

Experiment 3 showed that the MMN and P3a elicited by syllables were larger than those elicited by hums. The level tones elicited earlier and larger MMN and P3a than the contour tones did, and large f0 height-deviations elicited earlier and larger MMN than small f0 height-deviations did. In Experiment 4, the N1, P2, N2, and P3b were elicited in the syllable condition, and the N2 and P3b were elicited in the hum condition. In the syllable condition, the level tone elicited earlier and larger P2 and P3b and earlier N2 than the contour tone did. Large f0 height-deviations elicited earlier and larger N1 and P3b and earlier N2 than small f0 height-deviations did. In the hum condition, the level tone elicited earlier and larger MMN and P3b than the contour tone did, and large f0 height-deviations elicited earlier and larger MMN and P3b than small f0 height-deviations did.

Therefore, both tonal dimension and stimulus type affect lexical tone processing. Tonal processing is modulated by both the low-level acoustic features and the high-level linguistic memory. Neither the *acoustic model*, which only emphasizes the acoustic features, nor the *functional model*, which only emphasizes the linguistic functions, can explain all of the results. The *Auditory Cognitive Science (ACS) framework*, which considers the roles of both acoustic features and linguistic functions, seems more appropriate to describe these results.

摘要

本文研究了粵語聲調加工中：1）聲調維度對聲調加工的影響，2）刺激類型對聲調加工的影響，3） 任務難度對聲調加工的影響。實驗一、二爲雙耳分聽實驗，它們操縱同樣的自變量：1）基頻方向（平調、非平調），2）刺激類型 （音節、假音節、和合成嗡聲），但采用不同任務（實驗一、二分別采用辨別和識別任務）。實驗三、四爲腦電實驗，它們操縱相同的自變量：1）基頻方向（平調、非平調），2）基頻音高偏離幅度（大、小），3）刺激類型（音節、合成嗡聲），但采用不同任務（實驗三、四分別采用被動 oddball 和主動 oddball 任務）。聲調維度（基頻方向和基頻音高）的效應用來檢測聲學特徵對聲調加工的作用。刺激材料與實驗任務的效應用來檢測語義記憶對聲調加工的作用。

實驗一、二發現在粵語聲調加工中，不管是辨別還是識別任務，左耳都具有優勢。但在普通話聲調加工中，辨別和識別任務中分別呈現左耳和右耳優勢。在粵語中，對音節和假音節的反應快于對合成嗡聲的反應，對平調的反應快于對非平調的反應。實驗三發現，與合成嗡聲相比，音節誘發更大的 MMN 和 P3a。與非平調相比，平調誘發更早、更大的 MMN 和 P3a。與小基頻音高偏離相比，大基頻音高偏離誘發更早、更大的 MMN。 實驗四發現，音節條件誘發了 N1，P2，N2，和 P3b，合成嗡聲條件誘發了 N2 和 P3b，這反映了刺激類型效應。在音節條件，與非平調相比，平調誘發更早、更大的 P2 和 P3b 以及更早的 N2。大基頻音高偏離比小基頻音高偏離誘發更早、更大的 N1 和 P3b，以及更早的 N2。在合成嗡聲條件下，與非平調相比，平調誘發更早、更大的 MMN 和 P3b，大基頻音高偏離比小基頻音高偏離誘發更早、更大的 MMN 和 P3b。

結果表明，聲調維度和刺激類型都可以影響声调加工。即，聲調加工受到

聲學特徵和語言記憶的雙重調節。僅强調聲學特徵的聲學特徵模型和僅强調語言記憶的功能模型都不能解釋全部結果。聲學認知模型可以更好的解釋本實驗的結果，因爲此模型同時强調聲學特徵和腦內語言記憶的作用。

Chapter 1

Introduction

Lexical tone is prosodic information[1] that is used to cue linguistic contrasts in tonal languages. It is the most striking feature that distinguishes tonal languages from non-tonal languages. However, most of the existing studies on speech perception have been conducted on non-tonal languages (e.g., English, Dutch), and relatively few studies has been done on tonal languages, especially on lexical tone processing. Consequently, questions about how lexical tone is decoded and how tone and segments are combined during speech perception are sill unanswered. In order to clarify these two questions and reveal the specific features of tonal language processing, this study investigated tonal information processing in Cantonese.

Eight chapters are presented in the following text. Chapter 1 presents the definition of lexical tone and a brief review of tonal studies. Based on the results of these studies, three models on tonal processing are introduced in Chapter 2. In Chapter 3, a brief introduction of the present study, including the independent variables (IVs), the two paradigms, and an overview of the four experiments are presented. The four experiments are then reported one by one from Chapter 4 to Chapter 7. At last, in the general discussion in Chapter 8, the main results and theoretical implications of this study are discussed.

---

[1] Prosodic information refers to the changes of f0, duration, and intensity of a syllable, which can be used to distinguish utterances. Stress and tone are the most commonly used pieces of prosodic information for distinguishing utterances.

1. 1 Lexical tone

A language is tonal when fundamental frequency (f0) patterns of syllables are used to cue lexical contrasts (Cutler & Chen, 1997; Yip, 2003; Ladefoged, 2001). "There are numerous tonal languages across East and South Asia, most of Sub-Saharan Africa and among most of the indigenous languages of North and South America" (Pike, 1948, p3). Chinese is one kind of tonal language that is used by a large population. Not only Mandarin Chinese, which is used in the northern part of China, but also dialects of Chinese that are used in other regions of China are tonal languages (e.g., Cantonese, which is used in Hong Kong, Macau, and Guangdong province). In tonal languages, f0 patterns are phonologically contrastive, such that syllable meanings can be distinguished by lexical tone alone. For instance, Mandarin Chinese has four tones: tone 1 is high-level, tone 2 is high-rising, tone 3 is low-falling-rising, and tone 4 is high-falling (Xu, 1997). The syllable /ma/ means "mother", "hemp", "horse", and "to scold" at tones 1, 2, 3, and 4, respectively.

World languages can be broadly divided into tonal and non-tonal languages based on whether they use tonal information to distinguish the meanings of words. Given that most of the current studies on speech processing are conducted on non-tonal languages (e.g., English and Dutch), in which the prosodic information (stress) is not crucial for lexical access (Cutler, 1986), prosodic information is usually ignored in theorizing about speech perception (McClelland & Elman, 1986; Marslen-Wilson, Moss, & van Halen, 1996). In order to know the role of lexical tone and the relation of tonal information and segmental information in speech perception, studies must be done on tonal languages. The results of such tonal studies have some important implications. Firstly, they help people to understand language-specific features in tonal language processing. Secondly, the related results are important for researchers to construct general models of speech perception that can be generalized to both tonal and non-tonal languages (Chen, 1999 and 2006). Thirdly, knowledge of tonal processing is helpful for understanding pitch processing in music and general audition. In the following section, a brief review of the studies on tonal processing is presented.

1.2 Brief review of tonal studies

Recently, more and more studies have been conducted to explore lexical tone processing (Cutler & Chen, 1997; Ye & Connine, 1999; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Tong, Francis, & Gandour, 2008; Malins & Joanisse, 2010; for reviews, see Wong, 2002; Zatorre & Gandour, 2007). Many of the research questions of studies on segmental information—for instance, those centering on phonological role, processing speed, and the neural mechanisms (especially the hemispheric advantage)—are also of relevance to tonal information studies. On one hand, many studies have been conducted on segmental information perception, yielding some useful paradigms and techniques, which can be adopted by tonal studies directly. On the other hand, fruitful results and theoretical hypotheses have been obtained from segmental information studies, which can be applied to tonal processing, allowing researchers to understand the similarities between segment and tone perception and those features specific to tonal processing.

Firstly, the phonological role and the processing speed of tonal processing are studied with both behavioral, event-related potential (ERP), and eye-tracking methods. In particular, the relative phonological weights and processing speeds of tone and segments are compared. Some studies found that the phonological role of tone was weaker than that of segments and that processing speed of tone was slower than that of segments (Cutler & Chen, 1997; Repp & Lin, 1990; Taft & Chen, 1992). Using the auditory lexical decision task, Cutler and Chen (1997) reported that participants were more likely to accept nonwords as words when the nonwords differed from the real words in tone only. Furthermore, the response speed was slower in a same-different judgment task when the only difference between the two syllables was tone. These authors concluded that tone was slowly processed relative to segments and that tone was more likely to be misprocessed in fast response tasks than were segments. In a target dimension classification task, researchers found that tone classification was slower and less accurate than was segment classification (Repp & Lin, 1990). In a homophone judgment task, Taft and Chen (1992) found that participants were more likely to accept two characters as homophones when they differed in tone only.

In addition, there are studies that have found that the phonological importance of tone and segments is comparable, and tone can be processed as quickly as segments (Schirmer, Tang, Penney, Gunter, & Chen, 2005; Malins & Joanisse, 2010). Schirmer et al. (2005) studied Cantonese tone processing with an ERP method using a meaningful judgment task in which semantically correct and incorrect (tone violation, segment violation, and both tone and segment violation) sentences were presented to participants. Results showed similar ERP component patterns in tonal violation and segmental violation conditions, which suggests comparable phonological roles and processing speeds of tone and segments in semantically related tasks. Additionally, in an eye-tracking study using the visual world paradigm, Malins and Joanisse (2010) reported similar competition size and stage between a tonal competitor and a segmental competitor, indicating that the processing speeds and phonological roles of tone and segment are comparable.

In fact, some studies found that the relative phonological importance and processing speeds of tone and segment are modulated by lexical context (Ye & Connine, 1999; Liu & Samuel, 2007). In a monosyllable context, tone plays a weaker role than segment does. But in a sentence or idiom context, tone plays a comparable or an even more important role than segment does.

Secondly, the neural basis (especially the lateralization pattern) of tonal processing is studied using dichotic listening (DL), ERP, and functional Magnetic Resonance Imaging (fMRI) methods (Van Lancker & Fromkin, 1973 and 1978; Wang, Jongman, & Sereno, 2001; Wong, Parsons, Martinez, & Diehl, 2004; Luo et al., 2006; Chandrasekaran, Krishnan, & Gandour, 2007; Gandour et al., 2002; Wong, Parsons, Martinez, & Diehl, 2004). Using the DL method, Van Lancker and Fromkin (1973) found that both tone and segment in Thai were mainly processed by the left hemisphere, which suggests that the phonological function is important in determining hemispheric patterns. Wang et al. (Wang, Jongman et al., 2001; Wong, Parsons et al., 2004) also reported the left hemispheric advantage in Mandarin tone processing. However, using similar DL methods, researchers have failed to find a hemispheric advantage in Cantonese tone processing (Benson, Smith, & Arreaga, 1973; Smith & Shand, 1974) and in Mandarin tone processing (Baudoin-Chial, 1986).

Using the ERP technique, some researchers have recorded the brain responses elicited by tonal processing and have compared the amplitudes of the components on the left and right scalp, revealing that the mismatch negativity (MMN) on the right hemisphere is larger than that on the left hemisphere. The subsequent dipole source analysis also showed greater activity in the right hemisphere (Luo et al., 2006; Ren, Yang, & Li, 2009). These results suggest that the hemispheric pattern is determined by the acoustic features of the tone in the MMN stage. However, some other studies have found that the MMN elicited by cross-category deviation is larger than that elicited by within-category deviation on the left scalp (Xi, Zhang, Shu, Zhang, & Li, 2010). Furthermore, the MMN in native speakers was larger than that found in non-native speakers (Chandrasekaran, Krishnan, & Gandour, 2007). These results reflect the effect of phonological memory in tonal processing in the MMN stage.

In one fMRI study using the tone identification task, researchers presented Mandarin syllables with different tones to native Mandarin speakers and English speakers who did not know Chinese. Results showed greater activity in the left temporal lobes in Mandarin speakers only (Wong, Parsons, Martinez, & Diehl, 2004). Gandour et al. (2002) presented vowel syllables (varying the vowel) and tone syllables (varying the tone) in Thai to Thai speakers and Mandarin speakers, revealing that the left hemispheric advantage of tonal processing was found in native speakers only. The above fMRI studies suggest that tone is mainly processed by the linguistic areas of the left hemisphere.

In summary, previous studies have found that the phonological roles and processing speeds of tone and segment are modulated by context and task requirement. In pre-lexical tasks (e.g., syllable-monitoring, same-different judgment, homophone judgment, and target-dimension classification) and in a monosyllable condition, tone is usually processed slower than is segment, and tone is usually phonologically less important than is segment. In tasks that are more lexically involved (e.g., meaningful judgment) and in a sentence or idiom condition, the processing speeds and phonological roles of tone and segment are usually comparable.

In terms of the neural basis of tonal processing, some studies revealed a left hemispheric advantage, whereas other found a right hemispheric advantage. Some studies failed to find a hemispheric advantage on tonal processing at all. The discrepancies might come from different task demands in different studies and different temporal resolutions of techniques in different studies (e.g., the temporal resolutions of the ERP technique, the fMRI, and the DL are different, and the ERP technique can explore very early processing stage).

Chapter 2

Models

Tonal information has mainly been studied in two fields: speech perception (lexical tone) and general audition (pure tone). Correspondingly, two main theories have been developed from past studies: the *functional model* (Liberman, 1996; Liberman & Whalen, 2000) and the *acoustic model* (Zatorre, Belin, & Penhune, 2002; Poeppel, 2003). Although the rationales of the two models conflict with one another to some extent, both of them are supported by selective evidence from lexical tone studies. Recently, researchers have begun to view this dichotomy as unnecessary and have tried to raise new hypotheses to integrate the functional and acoustic model. Among these new hypotheses, the *auditory cognitive science (ACS) framework* is quite influential (Holt & Lotto, 2008; Diehl, Lotto, & Holt, 2004). In the following section, the functional and acoustic models and the ACS framework are introduced.

2.1.1 The functional model

The functional model was developed from studies on Categorical Perception[2] and Motor Theory (Liberman & Mattingly, 1985). The mapping between speech signals and linguistic units such as phoneme is complex. Although the acoustic signals of the same phoneme vary much along with its context and speakers' gender, age, and emotional stage, listeners can identify the phoneme without effort. In the 1950s, researchers began to study the complex mapping between acoustic signals and

---

[2] According to Categorical Perception, linguistic signals are categorically rather than continuously represented in the brain. Moreover, there are category boundaries between any two categories. Acoustic variation across boundaries is easily detected, but acoustic variation of equal size within a given category is difficult to detect.

linguistic units and revealed the phenomenon of Categorical Perception (Liberman, 1957; Liberman & Mattingly, 1985). Categorical Perception indicates that the categorical memory in the brain rather than the acoustic signal inputted through ears is important to speech perception. In order to explain this phenomenon, Liberman et al. proposed that the processing objects in speech perception are not acoustic signals but motor commands that construct articulatory gestures. Moreover, the motor commands are processed by a language-specific system in the brain. This theory was coined Motor Theory (Liberman, 1996; Liberman & Whalen, 2000).

The functional model adopts the motor theory, claiming that it is the high-level function rather than the low-level acoustic features of the inputted sound that determines the manner of processing and the brain mechanisms engaged. The linguistically related signals are mainly processed by language-specific mechanisms in the left hemisphere, and music and emotion-related signals are mainly processed by the right hemisphere (Van Lancker, 1980; Liberman, 1996; Liberman & Whalen, 2000).

These predictions are supported not only by Categorical Perception, but also by studies on neural mechanisms of vowel processing (Näätänen, 1997), consonant + vowel syllables (Hugdahl, 1995), and musical notes (Kimura, 1967; Goodglass & Calderon, 1977; Bryden, 1986; Piazza, 1980). These studies reported that linguistic and music-related sounds were mainly received by the left and right hemispheres, respectively, which are in line with the functional model.

In terms of lexical tone, given that it is linguistically contrastive, it should be processed by language-related mechanisms according to the functional model, and indeed this has been supported by some studies with DL, ERP, and fMRI methods. Van Lancker and Fromkin (1973; 1978) conducted the first DL experiment to test the lateralization of lexical tone processing. In this study, tone words (varying only in lexical tone), consonant words (varying only in consonant), and hums (pitch-changing hums) in Thai were identified by native Thai speakers, English speakers who did not know Thai, and musically trained English speakers. Native Thai speakers showed the left hemisphere advantage for tone and consonant processing, English speakers showed the left hemisphere advantage for consonant processing, and no musical training effect was found. These results indicate that the neural basis of tonal processing is affected by its linguistic function rather than the familiarity that comes from training. The left hemisphere advantage of tonal processing was also found in Mandarin (Wang, Jongman et al., 2001; Wang, Behne et al., 2004) and Norwegian (Moen, 1993) via the DL paradigm.

In addition, one ERP study revealed that linguistic memory was more important than acoustic similarity for Mandarin tone processing in the pre-attentive stage. In that study, which used the passive oddball paradigm[3], two types of Mandarin tone pairs (i.e., dissimilar pairs (tone 1/3) and similar pairs (tone 2/3)) were presented to Mandarin speakers and English speakers who were Chinese-naive. Results showed larger MMN amplitude in Mandarin speakers relative to English speakers, and the

---

[3] In the passive oddball paradigm, a stimulus sequence containing "standard" and "deviant" stimuli is presented to participants while they are actively involved in another distracter task (e.g., typically reading a book or watching a silent movie). By subtracting the brain's response to "standard" stimuli from the brain's response to "deviant" stimuli, the MMN can be obtained. The MMN reflects automatic detection for change and is an index of pre-attentive processing (Näätänen, 2001).

similarity effect only existed in the Mandarin group (Chandrasekaran, Krishnan et al., 2007). Thus, linguistic experience can override acoustic similarity in tonal processing in the MMN stage. Apart from that, another study examined the categorical perception of Mandarin tone by using the ERP method (Xi et al., 2010). Results showed that f0 variations across categorical boundaries elicited larger MMN than equal size of f0 variations within categories on the left scalp, which further suggests that it is not the acoustic features, but rather the linguistic functions that are crucial in tonal processing.

Moreover, in one fMRI study (Wong et al., 2004), Chinese words that varied along Mandarin tones were identified by native Mandarin speakers and English speakers who did not know Chinese. Results showed greater brain activity in the left temporal lobes only in the native speakers. In another fMRI study (Gandour et al., 2002), researchers presented tone words (varying in tone) and vowel words (varying by vowel) in Thai to Thai native speakers and Chinese speakers who did not know Thai, and found that both the tone and the vowel words evoked greater brain activity in the left hemisphere in Thai speakers, whereas the left hemisphere advantage in Chinese speakers only existed for vowel words. These results indicate that the phonological significance of the stimuli determines the activated brain areas.

Thus, convergent evidence from both behavioral and neuroimaging studies suggest that it is not the acoustic feature, but rather the linguistic function that is crucial in determining the manner of tonal processing.

2.1.2 The acoustic model

This model originates from studies on general auditory information processing. In one pure tone study using positron emission tomography (PET) (Zatorre & Belin, 2001), researchers presented standard sound that has two frequencies and two versions of deviant sounds to participants: one version was a spectrally-varying version in which the frequency number increased to 2, 4, 8, and 16 times of the standard sound, though the duration was kept constant; the other version was a temporally-varying version, in which the frequency was not changed, but the duration was shortened to 1/2, 1/4, 1/8, 1/16 times of the original standard one. The results showed that the left auditory cortex and the left inferior frontal cortex were more activated by the temporally-changing version, and the right auditory cortex and the right inferior frontal cortex were preferentially responsible to the spectrally-changing version.

The authors of the above study developed the acoustic model to explain their findings. According to this model, the processing manner and the engaged neural systems of auditory information processing are determined by the acoustic features of sound rather than the function of sound. Fine temporally-varying signals are predominately processed by the left hemisphere, whereas fine spectrally-varying signals are predominately processed by the right hemisphere (Zatorre, Belin, & Penhune, 2002). Similar to the acoustic model, the "asymmetric sampling in time" (AST) theory (Poeppel, 2003) also uses acoustic features to explain the engaged brain mechanisms. This model posits that short temporal (~20-40 ms) and long temporal

(~150-250 ms) integrations are mainly extracted by the left and right hemispheres, respectively.

With the exception of the pure tone study (Zatorre & Belin, 2001), the predictions of the acoustic model has been supported by research on perception of phonetic and music-related information (Zatorre, Belin, & Penhune, 2002). From the acoustic perspective, consonants and vowels in speech are mainly characterized by the varying of the voice onset time (VOT) and formant transition. The variation ranges of both are in tens of milliseconds, indicating that they are fine temporal changing signals, such that they should be preferentially processed by the left hemisphere according to the acoustic hypothesis. Melody information should be primarily processed by the right hemisphere, as it entails frequency-varying signals from the acoustic perspective. Indeed, consistent with the acoustic hypothesis, a number of studies have found a left hemisphere advantage in consonant and vowel perception and a right hemisphere advantage in music information processing (Hugdahl, 1995; Kimura, 1967; Goodglass & Calderon, 1977; Bryden, 1986; Piazza, 1980).

Regarding lexical tone, which is characterized by f0 level and the pattern of f0 variance across a vowel or syllable, it is a spectrally-varying signal and should be predominately processed by the right hemisphere according to the acoustic model. In one ERP study (Luo et al., 2006), Mandarin syllables with varied tones were presented to native listeners by using the passive oddball paradigm. Results showed that MMN on the right scalp was larger than that on the left scalp. Another ERP experiment reported similar results (Ren, Yang, & Li, 2009). The researchers concluded that the acoustic feature played a dominant role in lexical tone processing in the MMN stage.

In summary, the acoustic model that was developed from studies on general audition can also be used to explain studies on phoneme (e.g., consonants and vowels) and musical information processing. Moreover, the results of some studies on tonal processing are also in line with the acoustic model.

2.1.3 The ACS framework

Although both the functional and the acoustic models are supported by selective evidence, their underlying rationales and resulting predictions conflict to some extent. Take tonal processing for example: according to the acoustic model, tonal information should be primarily received by the pitch-related areas in the right hemisphere because lexical tone is an f0-varying signal, whereas the functional model predicts that lexical tone is mainly processed by the language-related areas in the left hemisphere, given that lexical tone is phonologically contrastive. It seems that studies on tonal processing could provide an opportunity to disentangle these conflicting predictions, but unfortunately both models have been empirically supported by a number of studies, thus, neither one can explain the results across all of those studies.

In addition, regarding the categorical perception of lexical tone, mixed results have been reported. Using the traditional discrimination and identification steps, Abramson (1979) found that a level tone continuum in Thai was not perceived categorically. However, Wang (1976) reported that a Mandarin tone continuum from contour to level tones was perceived categorically. Recently, a study (Francis, Ciocca, & Ng, 2003) conducted on Cantonese tones revealed that contour tones rather than level tones were perceived categorically. Thus, which factor is more important to tonal processing, either the acoustic feature or the categorical memory in the brain, is still a point of debate.

Thus, neither the acoustic nor the functional model can describe all of the empirical data collected from tonal studies. In fact, the two models hold opposite predictions. With certain results, researchers can support either one model or the other. In order to resolve these conflicting conclusions, researchers have attempted to develop new models that integrate the previous studies' results. Among them, the ACS framework is an influential one (Holt & Lotto, 2008; Diehl et al., 2004).

In the ACS framework, three important notions are proposed: 1) speech is not special, and speech and general audition are not two separate topics; speech processing must begin with auditory signal processing, 2) the ACS framework considers that both the memory trace in the brain and the acoustic feature inputted through the ears are important to tonal processing, and 3) regarding the relationship between acoustic feature and memory trace, there are two possibilities. One is that the acoustic feature and the linguistic memory of tone play crucial roles independently in different processing stages or in different brain mechanisms. For instance, the acoustic feature plays a dominant role in the early stage of processing and in the primary auditory areas of the brain. Linguistic memory plays a dominant role in the late stage of processing and in the semantically related areas of the brain (Zatorre & Gandour, 2007; Gandour, 2004; Luo et al., 2006). Another possibility is that the acoustic feature and linguistic memory work interactively rather than independently to modulate tonal processing from beginning to end (Zatorre & Gandour, 2007; Gandour et al., 2004). The third notion is similar to the notions of the TRACE model (McClelland & Elman, 1986), according to which both low-level feature and high-level memory are crucial in speech perception, and different levels work interactively in speech perception.

As mentioned in Cahpter 1, processing speed, phonological importance, and related neural mechanisms of tonal processing are modulated by many factors, such as task requirement, the language experiences of the participants, and the materials in which tones are realized. The current study was designed to examine tonal processing in Cantonese, particularly to test the effects of tonal dimension (which is introduced in Chapter 3), stimulus type, and task requirement on tonal processing. Through these manipulations, this study aimed to clarify in which condition acoustic feature is dominant to tonal processing and in which condition memory trace is crucial to tonal processing, and to reveal the overall picture of tonal processing.

Chapter 3

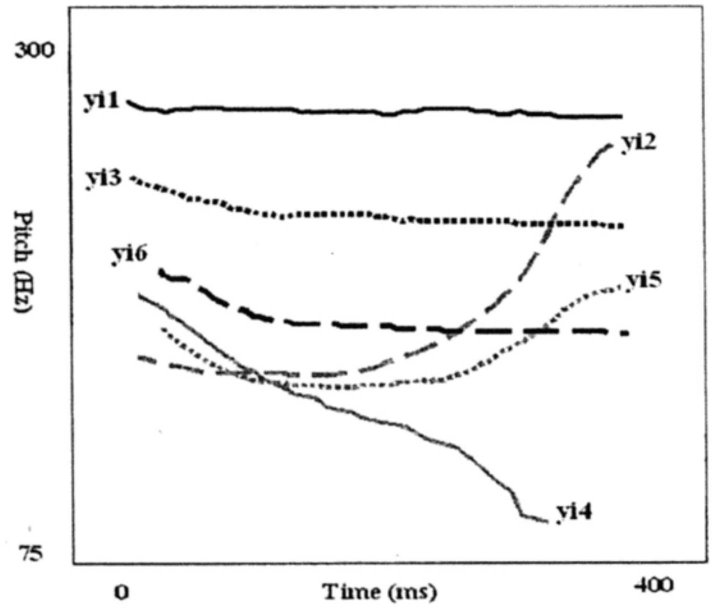
The Present Study

In Chapter 2, three models of tonal processing were introduced along with their different rationales and predictions. In order to examine these models, four experiments were conducted in this study. In Chapter 3, the two indenendent variables are presented firstly. And then the DL paradigm and the oddball paradigm in ERP technique are introduced. In the last section, an overview of the four experiments is given.

3.1 Two IVs

3.1.1 Tonal dimension

Figure 1. F0 lines of the six Cantonese tones on the syllable /yi/, which were drawn based on recorded syllables.



Note: The three level tone in Cantonese (T1, T3, and T6) also have additional abbreviated versions realized in short syllables with a voiceless stop coda (/-p, -k, -t). As was the case with the majority of previous studies, only these six complete tones were considered in this study.

Phonetic studies have revealed that the f0 patterns of syllables are the primary acoustic correlate of lexical tone perception (Yip, 2003; Fok, 1974; Vance, 1977). Further studies have examined sub-dimensions of f0 changing. In a lexical tone identification experiment, naturally produced Cantonese tones were presented to 511 listeners (Fok, 1974). Results showed that listeners divided the six Cantonese tones into two groups based on overall f0 change: flat f0 (tones 1, 3, and 6) and fluctuating f0 (tones 2, 4, and 5). Furthermore, the flat group was divided into high (tone 1), intermediate, (tone 3) and low (tone 6) tones based on average f0 level. The changing f0 group was divided into rising f0 (tones 2 and 5) and falling f0 (tone 4) based on the direction of change, and the rising f0 subgroup was divided into tones 2 and 5 based on the magnitude of change. Ultimately, Fok concluded that, in Cantonese tone perception, listeners made use of average f0 level (high, low) and direction of f0 change (rising, falling) to identify tones.

By reanalyzing Fok's (1974) data with a multidimensional scaling (MDS) analysis, Gandour (1981) found three perceptual *dimensions*—namely height, direction, and contour—that are related to Cantonese tone perception. Gandour interpreted the height dimension as whether the average f0 level was high or low; the direction as whether the direction of f0 change was rising, falling, or flat; and the contour[4] dimension as whether the magnitude of f0 change was large or small. In another study (Gnadour, 1983), 19 types of synthetic tone pairs of Mandarin, Cantonese, Taiwanese, and Thai were presented to native and non-native listeners. Participants completed a dissimilarity rating task for tone pairs. Results revealed only

---

[4] The "contour" has different meaning in different places. When contour denotes one kind of tonal dimension, it means the magnitude of change, or the slope, of f0. But when it denotes a tonal type, it means the type of tone that has glide movements, which is distinguished from a level tone.

two dimensions, direction and height. Gandour attributed the discrepancies to the differences in stimuli (natural vs. synthetic) and task (identification vs. dissimilarity rating). Whatever the case, the height and direction dimensions[5] were reliably found in these studies.

The above phonetic research suggests that Cantonese tone perception relies on two main dimensions of f0. Each dimension denotes one kind of acoustic feature in the changing of f0. The acoustic difference between the two tonal dimensions may lead to different processing speed or/and engaged brain areas. However, the two dimensions were usually mixed together in previous studies. Thus, the effects of the two dimensions were ignored. Therefore, this study manipulated tonal dimensions as IVs, and investigated the effects of tonal dimensions on tonal processing.

3.1.3 Stimulus type

Past studies have shown that processing speed and the phonological weight of tones were modulated by lexical status (Ye & Connine, 1999; Liu & Samuel, 2007). The lateralization pattern of tonal processed was affected by lexical status, too (Van Lancker & Fromkin, 1973).

In order to determine how lexical status affects tonal processing in Cantonese, three kinds of stimuli were used: real syllables, pseudo-syllables, and hums. A real syllable is the existing syllable in Cantonese, which is pronounceable (has segmental information) and has meaning (*Chinese Character Database:* with Word-formations,

---

[5] Because there is inconsistency in dimension labels in the literature, this paper will use f0 height to denote the average pitch level of f0 and f0 direction to denote the direction of f0 changes. In addition, this study focused on the height and direction dimensions because they are reliably found in some studies.

2003). A pseudo-syllable is pronounceable but does not have meaning. The hum is synthesized from the real syllable, for which the segmental information was removed and only the f0 pattern remains. Thus, the hum is unpronounceable and does not covey meaning. The lexical involvement decreases gradually from syllable to pseudo-syllable to hum.

3.2 Paradigms

Hemispheric advantage is an important issue in this study because its result can be used to infer acoustic processing or linguistic processing, thus disentangling the theoretical debates centering on the acoustic versus the functional model. Therefore, the DL paradigm, the most commonly used paradigm in perceptual lateralization studies, was employed in Experiments 1 and 2. In addition, the ERP method was employed in Experiments 3 and 4. On one hand, the brain responses on the left and right scalps could be compared to indicate a hemispheric advantage, and on the other hand, the combination time points of segment and tone could be explored with the ERP method.

3.2.1 The DL paradigm

In the traditional DL task, two different stimuli are presented independently to both ears simultaneously, and participants are required to report the first one they hear as quickly and accurately as possible. Anatomical and neuroimaging studies (Brancucci et al., 2004; Kimura, 1967) have found that the sounds inputted to the ears were predominantly conveyed to the contralateral cortices with suppression of the ipsilateral pathway. In other words, the performance of the right ear can be used to denote the function of the left hemisphere cortices, and vice versa. The language-related sounds that are inputted through the right ear can be transferred to the left hemisphere directly, whereas the linguistic sounds inputted to the left ear are transferred to the right hemisphere first and are then retransferred to the left hemisphere across the corpus callosum. This retransfer attenuates the available information and increases the response time (RT) for the left ear signals. Hence, the right ear advantage (REA) is usually found for verbal stimuli processing in right-handed people.

Since the 1950s', the DL has been commonly used in perceptual lateralization studies (Broadbent, 1954; Kimura, 1967; Bryden, 1988; Hugdahl, 1995), especially in speech perception studies. Kimura (1967) studied the ear advantage of speech and non-speech sounds processing, and found the REA for speech processing in right-handed participants. Hugdahl and his colleagues at Bergen University provided strong evidence of the REA for speech processing in right-handed participants. In their study, consonant + vowel (CV) syllable pairs (e.g., the same dichotic pairs, /ba/-/ba/, and different dichotic pairs, /ba/-/ga/) were presented to more than 1,000 participants (including both right- and left-handed participants, aged 6 to 88). Participants were required to report the first syllable they heard. In this study, only the different dichotic pairs were included in the data analysis, and an obvious REA trend was found for the majority of right-handed people (the authors named this task the

Bergen test).

Except for the traditional DL task, the "probe-target" discrimination DL task is widely used in perceptual lateralization studies. In the traditional DL test, each trial has one dichotic pair, and participants need to report the first one they hear. Then, the report proportions for both ears are compared. However, in the "probe-target" discrimination DL task, each trial contains two dichotic pairs with an interval (around 800 ms) between them. The first pair is the target and the mask, and the second pair is the probe and the mask. The target and probe are task-related sounds that are presented to the same ear. Participants are required to compare them and press a key indicating whether or not they are the same. Task-related sounds are presented to the two ears randomly with the equal chance. RTs and accuracies for the two ears are recorded and compared to denote ear advantage.

Using this paradigm, Brancucci and his colleagues found that the duration was mainly processed by the left hemisphere no matter the duration was realized by speech or music (Brancucci, Anselmo, Martello, & Tommasi, 2008), and that intensity information was mainly processed by the right hemisphere no matter it was realized by music or speech (Brancucci, Babiloni, Rossini, & Romani, 2005). Researchers concluded that it was the acoustic parameter rather than the function of the sounds that determined the lateralization patterns in the "target-probe" discrimination task.

In the current study, the Bergen test and the "probe-target" discrimination task were adopted in Experiments 1 and 2, respectively. The two kinds of DL tests differ in processing depth. In the "target-probe" discrimination DL task, participants do not need to identify the items. They just make global comparison of the target and probe. But in the Bergen identification test, participants are required to identify and report the targets. Thus, the discrimination task involves less linguistic information relative to the Bergen test. With the two tests, Experiment 1 and 2 would reveal the task requirement effect on tonal processing.

3.2.2 The oddball paradigm in ERP

The ERP is one kind of noninvasive neuroimaging method. The striking advantage of this method is its high temporal resolution, which is useful for revealing the time course of cognitive processing. In addition, the topography of the scalp and source analysis based on dense ERP signals can provide some spatial information on mental processes. Thus, the ERP method is a useful technique for cognitive neuroscience studies (Luck, 2005).

The oddball paradigm is a classic paradigm used in ERP studies and is usually used in general audition and speech perception studies. In this paradigm, a sound sequence that contains a standard sound (large proportion) and a deviant sound (small proportion) is presented to participants. According to the task at hand, it can be divided into two types: the passive oddball paradigm and the active oddball paradigm. In the passive oddball paradigm, participants are required to pay attention to a silent movie or a book without making any active response to imported sounds. In the active oddball task, participants are required to engage in a classification task by pressing one key for deviant sounds and pressing another key for standard sounds.

In offline analysis, the electroencephalographs (EEGs) elicited by standard and deviant stimuli are averaged separately to get the standard and deviant ERPs. Then, the standard ERP is subtracted from the deviant ERP to get the difference wave. The elicited components in the passive and active oddball paradigms are different. In the passive oddball paradigm, usually the MMN and the P3a are evoked. In the active oddball paradigm, the N2 and P3b (sometimes also the N1 and P2) are usually evoked (Polich, 2007).

The MMN usually begins at approximately 100 ms and peaks at around 150 ms (syllable-elicited MMN occurs relatively later), mainly distributed around the frontal scalp. It is an index of pre-attentive processing because it can be elicited in inattentive states, even in sleeping and unconscious states. The latencies of the P3a and P3b are similar, but the scalp distributions and their neural and cognitive mechanisms are different. The P3a is usually largest at the frontal scalp and located at frontal cortices, reflecting stimuli-driven attention mechanisms during task processing. The P3b mainly distributes around the central scalp and originates from the temporal and partial areas, suggesting attention and subsequent memory processing in the temporal-partial areas. The N2 elicited in the active oddball task also is largest around the central scalp and reflects attention and memory processing related to the stimuli (Näätänen, 2001; Polich, 2007).

In order to examine the effects and processing time points of acoustic information (might on MMN and P3a) and linguistic memory information (might on N2 and P3b) in tonal processing, and to classify how and when tone and segment are combined, the passive and active oddball paradigms were employed in Experiments 3 and 4, respectively.

3.3 Overview of the four experiments

In order to reconcile the debate concerning the models and to clarify the relationship between tone and segment during speech perception, the IVs of tonal dimension and stimulus type were manipulated in two (1 and 2) DL experiments and two (1 and 2) ERP experiments.

In the two DL experiments, the manipulation of tonal dimension and stimulus type was similar, but they differed in task requirements. In the two ERP experiments, again the manipulation of tonal dimension and stimulus type was similar, but they also differed in task requirements. Thus, Experiments 1 and 2 were introduced together, as were Experiments 3 and 4.

3.3.1 Experiments 1 and 2

Experiments 1 and 2 focused on the lateralization pattern of tonal processing. In these two experiments, two IVs were manipulated. One was the tonal dimension, for which the f0 direction[6] (level tones and contour) was manipulated. The other was lexical status, for which the stimulus type (real syllables, pseudo-syllables, and hums) was manipulated.

---

[6] Tonal direction denotes the direction of f0 movement—either flat, rising, or falling—based on which the six Cantonese tones were divided into flat tones (T1, T3, and T6) and non-flat tones (T2, T4, and T5). Likewise, according to the phonetic tonal type, the six Cantonese tones can be divided into the same two groups (level tones and contour tones). Thus, tonal direction and tonal type are equivalent in the following. F0 height was not considered in DL experiments because it was not likely to affect the lateralization of tonal processing.

In Experiment 1, participants completed a "probe-target" discrimination task, and in Experiment 2, participants completed a Bergen test.

For data analysis, three-way ANOVAs (ear, f0 direction, and stimulus type) were conducted separately for each of the two experiments. The effects of f0 direction, stimulus type, and ear advantage (lateralization) were analyzed. The modulation of the tonal type and the stimulus type on the ear advantage were also important. Additionally, the results of Experiments 1 and 2 were compared to test the effect of task requirement on tonal processing.

3.3.2 Experiments 3 and 4

Experiments 3 and 4 focused on the time course of the effects of tonal dimension and stimulus type, and the time course of the segment and tone combination. In these two experiments, three IVs were manipulated. The first was tonal dimension 1 (f0 direction was level or contour), the second was tonal dimension 2 (change in f0 height was large or small), and the third was stimulus type (tonal patterns were realized on a real syllable and a hum (DL experiments showed that result patterns for the pseudo-syllable and syllable were similar. Thus, for simplicity and clarity, the pseudo-syllable was eliminated from the ERP experiments)).

The passive oddball paradigm was used in Experiment 3. The latencies and amplitudes of MMN and P3a were analyzed. The effects of f0 direction, f0 height, and stimulus type on amplitudes, latencies, and the lateralization of MMN and P3a were analyzed. The active oddball paradigm was employed in Experiment 4. The latencies and amplitudes of N1, P2, N2, and P3b were analyzed. The effects of f0 direction, f0 height, and stimulus type on amplitudes, latencies, and the lateralization of N1, P2, N2, and P3a were analyzed. In addition, the results of the two experiments were compared to test the effect of task requirement on tonal processing.

Chapter 4

Tonal Processing in Discrimination DL

Experiment 1A- Cantonese tone

Experiment 1A was designed to study the lateralization pattern of tonal processing in discrimination task, and also wanted to study how tonal dimension and stimulus type affect lateralization of tonal processing.

Two IVs were manipulated in this experiment. The first one was the f0 direction (level and contour). There were six tones in Cantonese (Figure 1), three of them are level tones (T1, T3, and T6) and another three are contour tones (T2, T5, and T4), which provides ideal opportunity to study the role of f0 direction in tonal processing and how f0 direction influences lateralization of tonal processing. The second IV is the stimulus type (syllable, pseudo-syllable, and hum). For the three kinds of materials, the lexical information in them decreases gradually. The syllables have full lexical information (both meaning and segmental information), pseudo-syllables do not have meaning information, and the hums have neither meaning nor segmental information.

Experiment 1 adopted the "probe-target" discrimination DL task. The accuracies and RTs were recorded as DVs.

If the lateralization of tonal processing depends on its acoustic features (acoustic hypothesis), lexical tone would be more likely to be processed by the right hemisphere, and the lateralization pattern would not be influenced by stimulus type. If the lateralization of tonal processing is determined by its linguistic functions (functional hypothesis), tonal processing would show a left hemisphere advantage in syllable condition and might show a right hemisphere advantage in hum condition.

Participants

Twenty seven undergraduates (13 males, aged 18-24) from the Chinese University of Hong Kong (CUHK) who were native speakers of Cantonese participated in this experiment. All of them were right handed as indicated by the Edinburgh inventory of handedness (Oldfield, 1971). None of them had more than three years of musical training. And none of them had hearing impairment. In addition, an audiometric test[7] was conducted to guarantee that the hearing ability of the two ears was not different significantly (5dBs). They participated in the experiment to get course credits.

Materials and design

Four sets of Cantonese syllables (the four syllables /si/, /fu/, /yi/, and /se/ were pronounced at six tones in each set) were used as materials. Hence all the four sets were minimal contrasts, in which only the lexical tone was different while the consonant and vowel were kept constant. Considering the potential effects of frequency/familiarity and number of homophone (NOH) to tonal perception (Ganong, 1980; McQueen, 1991), the frequencies and NOH across experimental conditions

---

[7] In the audiometric test, the complex tones with different intensities (increasing from 0 dB to 30 dB with steps of 3 dBs) were randomly presented to participants. The test order of the left and right ears was also random. Participants should indicate the ear in which the sound appeared when the sound was perceivable. When there was no difference (3 dB) between the hearing thresholds between the two ears, the participant was recruited.

were matched (*F*s ≦ 1.67 and *ps* ≧ .43) according to the *Chinese Character Database*: with Word-formations (2003). Table 1 showed the four syllable sets.

Table 1. Four syllable sets used in Experiment 1.

Tone	Set 1	Set 2	Set 3	Set 4
1	/si1/ e.g., 詩 "poem"	/fu1/ e.g., 夫 "husband"	/se1/ e.g., 些 "some"	/yi1/ e.g., 姨 "aunt"
2	/si2/ e.g., 史 "history"	/fu2/ e.g., 父 "father"	/se2/ e.g., 寫 "write"	/yi2/ e.g., 椅 "chair"
3	/si3/ e.g., 試 "try"	/fu3/ e.g., 褲 "trousers"	/se3/ e.g., 瀉 "spell"	/yi3/ e.g., 意 "meaning"
4	/si4/ e.g., 時 "time"	/fu4/ e.g., 符 "symbol"	/se4/ e.g., 蛇 "snake"	/yi4/ e.g., 兒 "son"
5	/si5/ e.g., 市 "city"	/fu5/ e.g., 婦 "woman"	/se5/ e.g., 社 "society"	/yi5/ e.g., 耳 "ear"
6	/si6/ e.g., 氏 "surname"	/fu6/ e.g., 負 "negative"	/se6/ e.g., 射 "shoot"	/yi6/ e.g., 二 "two"

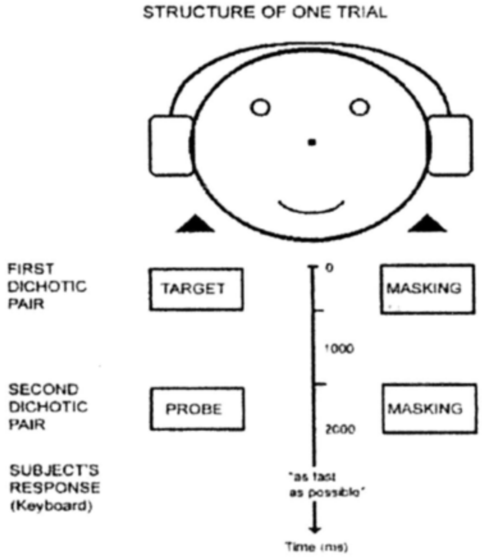
Note: Following each syllable, one Chinese character example and the English equivalent of the character are given.

And four sets of pseudo-syllables (syllable /bi/, /bu/, /di/, and /du/ at six tones) that were pronounceable but did not have meaning according to the *Chinese Character Database:* with Word-formations (2003) were also chosen.

The 24 syllables and 24 pseudo-syllables were read in isolation by phonetically trained female native speaker of Cantonese, and were recorded on a DAT tape, which were then digitalized at a sampling rate of 44.1 kHz, and were further edited with Cool Edit Pro 2.0 for normalization of duration (400 ms, including the 15 ms rise and fall ramps) and peak intensity (60dB). Additionally, each syllable was synthesized (Praat) to a corresponding hum, in which only the f0 was retained and the segmental information was removed.

Procedure

Figure 2. The structure of the "target-probe" discrimination task in one trial. This figure was adopted from Brancucci et al. (2008).



In Experiment 1 the "target-probe discrimination" paradigm (Brancucci et al., 2004; Brancucci et al., 2008; Brancucci, di Nuzzoaet, & Tommasia, 2009) is adopted. As Figure 2 shows, two dichotic pairs were presented with 1 second interval between them. The first pair was target and mask, and the second pair was probe and mask. The target and probe appeared in the same ear, which was the task-related ear. Participants were asked to pay attention to the task-related ear only and discriminate the probe and target were the same or not, and press corresponding keys immediately after probe onset. The maximal response time was 2 seconds after probe offset. After participant's response, there was 1 second rest, and then next trial appeared. The stimuli duration was about 400 ms. Therefore, one trial lasted for about 3- 4 seconds. The procedures in this study were run using the E-prime 1.1 software (Psychology Software Tools, Inc).

The three level tones constituted 6 "target-probe" pairs (1-3, 1-6, 3-6, 1-1, 3-3, and 6-6). The probe and target in each pair were inverted once (1-3 → 3-1). Thus, we had 12 level-tone trials for each set and they were presented randomly in one block. The four sets of syllables made up four blocks of this kind of trials. Similarly, four blocks of contour-tone trial were also made up. Both of them were tested 2 times, one with the left ear as task-related ear and another with the right ear as task-related ear. Before each block, a signal was given to cue the task-related ear. Totally, Experiment 1 had 192 trials in 16 blocks (4 blocks of level tone-left ear, 4 blocks of contour tone-left ear, 4 blocks of level tone-right ear, and 4 blocks of contour tone-right ear). There was 1-minute rest between two blocks. The duration for syllable condition was about 30 minutes. Similar procedures were used in pseudo-syllable and hum conditions. The three kinds of materials were tested separately in different days, and the order of them was random in participants.

Before the formal experiment, the instruction was given to them, and a practice section was done to guarantee he/she understood the rule. During the formal experiment, the accuracies and RTs were recorded as DVs.

Results and discussion

The incorrect trials (13.9 %) were discarded from RT analysis. The trials with accuracy and RT that exceed 3 SDs of means were eliminated for each participant (2.2 %). Correlation between accuracy and RT was checked to guarantee that no speed-accuracy trade off happened.

Two three-way ANOVAs (ear, stimulus type, and tonal type) were conducted on RT and accuracy to test the main effect of ear (lateralization), stimulus type, and tonal type, and how stimulus type and tonal type modulate lateralization. For interaction, I was not interested in the three way interaction, but interested in how stimulus type and tonal type affect lateralization pattern. Therefore, the two two-way interactions were reported separately and illustrated by figures. The Greenhouse-Geisser correction was applied when it is necessary to avoid the Type 1 errors.

Accuracy

ANOVA revealed a significant main effect of lateralization, (F (1, 26) = 5.8, MSE = 0.068, $p < .05$, $\eta^2 = 0.11$[8]), accuracy of the left ear (86 %) was better than that of the right ear (82 %). The main effect of stimulus type was significant (F (2, 52) = 15.34, MSE = 0.058, $p < .001$, $\eta^2 = 0.14$). The mean accuracy of syllable, pseudo-syllable, and hum were 83 %, 85 %, and 82 % respectively. Post hoc analysis showed that performance of pseudo-syllable was better than that of syllable ($p < .05$), in turn, performance of syllable was better than that of hum ($p < .05$). The main effect of tonal type was also significant (F (1, 26) = 113.36, MSE = 0.68, $p < .001$, $\eta^2 = 0.35$). Performance of level tone (89 %) was significantly better than that of contour tone (78 %).

Results showed that the interaction between stimulus type and lateralization was not significant (Figure 3). The LEA appeared on syllable, pseudo-syllable, and hum.

Results showed a significant interaction effect between tonal type and

---

[8] In repeated-measure ANOVA, the η^2 is usually used to test the effect size. And the effect sizes of about 0.01, 0.06, and 0.14 are termed as small, medium, and large, respectively (Cohen, 1988).

lateralization (Figure 4), (F (2, 52) = 5.04, MSE = 0.015, $p < .05$, $\eta^2 = 0.10$). Further analysis showed significant lateralization effect on contour tone (t (26) = 3.07, p < .05), the left ear (80 %) performed better than the right ear did (75 %). The null effect of lateralization on level tones may be induced by the ceiling effect (about 90 %).

Figure 3. Lateralization patterns (ACC) across stimulus type in Experiment 1 A.

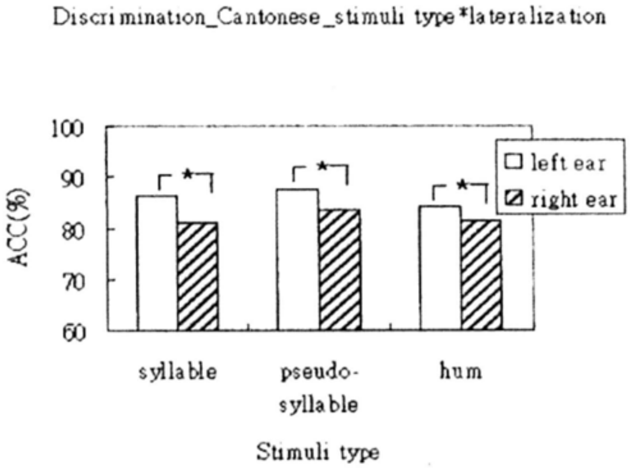
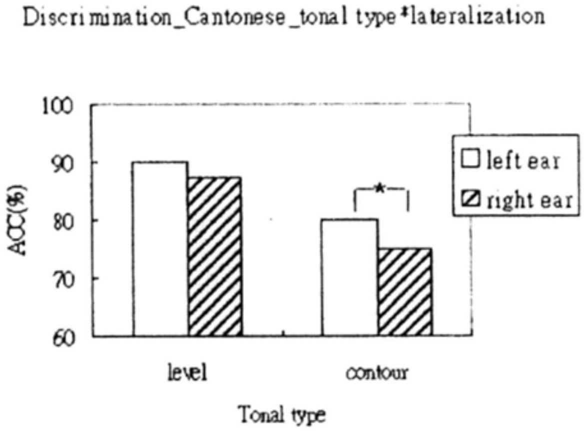


Figure 4. Lateralization patterns (ACC) across tonal type in Experiment 1 A.



RT

ANOVA revealed a significant main effect of lateralization (F (1, 26) = 9.6, MSE = 49250.76, $p < .05$, $\eta^2 = 0.18$) that due to a left ear advantage (728 ms) relative to the right ear (759 ms). And a significant main effect of stimulus type, (F (2, 52) = 14.9, MSE = 45188.51, $p <.001$, $\eta^2 = 0.29$). The mean RT on syllable, pseudo-syllable, and hum were 723ms, 720 ms, and 740 ms respectively. The response on syllable and pseudo-syllable was significantly faster that on hum (*ps* < .05), and the speeds on syllable and pseudo-syllable were comparable. The main effect of tonal type was significant, (F (1, 26) = 108.28, MSE = 58478.58, $p < .001$, $\eta^2 = 0.32$). Response at level tone (692 ms) was faster than that at the contour tone (771 ms).

However, no significant interaction effect was found on RT analysis (Figures 5 and 6). The interaction between tonal type and stimulus type was not significant, (F (2, 52) = 1.03, MSE = 25769.53, $p = .78$, $\eta^2 = 0.05$).

Figure 5. Lateralization patterns (RT) across stimulus type in Experiment 1 A.

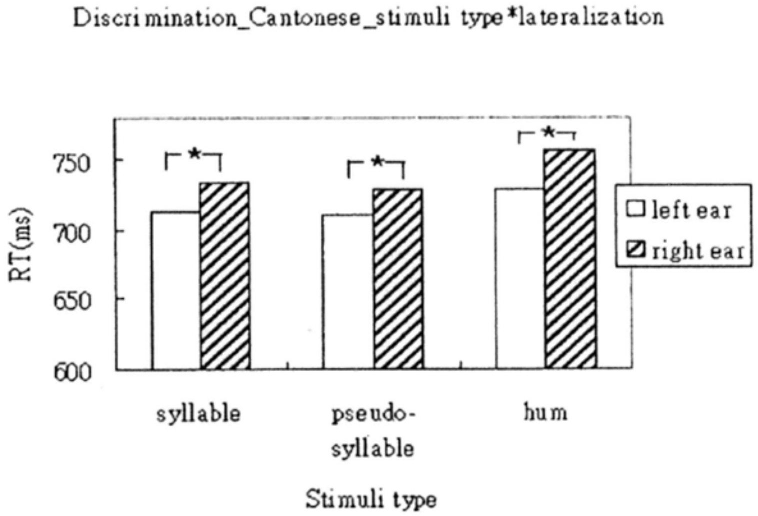
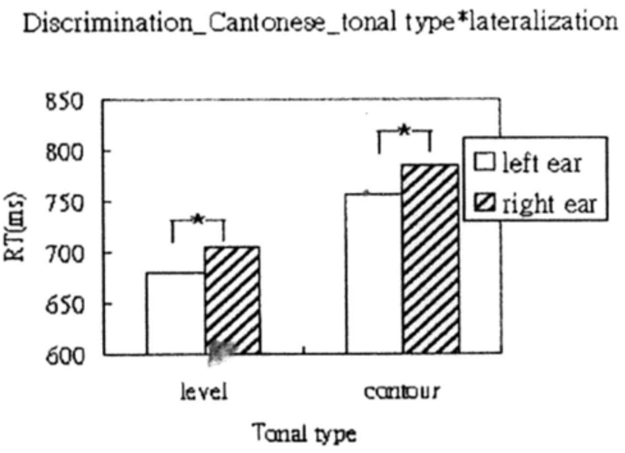


Figure 6. Lateralization patterns (RT) across tonal type in Experiment 1A.



Experiment 1B - Mandarin tone

Although the Cantonese and Mandarin share the same written system, but the speech of them is quite different. In terms of the lexical tone, Cantonese has six tones, while Mandarin has four tones. The acoustic features of them are different. Generally speaking, the acoustic resolution of the four Mandarin tones is better than that of the six Cantonese tones (Gandour, 1983). But both of them can be quickly processed during speech perception without effort.

Some studies were done on both of them, but the identification task was usually used (Smith, & Arreaga, 1973; Smith &Shand, 1974; Baudoin-Chial, 1986; Wang Jongman, & Sereno, 2001). In order to test hemispheric advantage of tonal processing in Mandarin using the discrimination task, the same procedure was done on Mandarin tone. Because there is only one level tone in Mandarin, we could not divide the four tones into level and contour groups like in Cantonese. Thus, all the four tones were studied together as in most of past studies on Mandarin tone. Therefore, there were two IVs (stimuli type and ear) in this Experiment 1B. The main effect of ear (lateralization) and stimuli type, and the modulation of stimuli type to lateralization were examined.

Participants

Thirty two (15 males, aged 18-24) undergraduates in the South China Normal University were recruited in this experiment. All of them were native Chinese speakers and speak Mandarin in everyday life, and none of them had more than three years of musical training. In addition, all of them were right handed, and none of them had hearing impairment. And an audiometric test was conducted to guarantee the hearing ability of the two ears was not different significantly. Ten RMB each hour were paid to them for rewards.

Materials and design

The speech stimuli in Experiment 1B also consisted of three kinds: four sets of syllables (/ma/, /yi/, /guo/, and /shi/ at four tones), four sets of pseudo-syllables (/bou/, /dv/, /biou/, and /diang/ at four tones), and hums synthesized from syllables. All the 16 real words were high frequency, and the word frequencies and NOH across condition were matched (*F*s ≦ 1.69 and *ps* ≧ .28). The recording and digitalizing methods of materials in this experiment was similar to that in Experiment 1A.

Table 2. Four sets of syllables used in Experiment 1B.

Tone	Set 1	Set 2	Set 3	Set 4
1	/ma1/ e.g., 媽 "mother"	/yi1/ e.g.,醫 "medicine"	/guo1/ e.g., 鍋 "pan"	/shi1/ e.g., 師 "teacher"
2	/ma2/ e.g., 麻 "herb"	/yi2 / e.g., 姨 "aunt"	/guo2/ e.g., 國 "country"	/shi2/ e.g., 十 "ten"
3	/ma3/ e.g., 馬 "horse"	/yi3/ e.g., 椅 "chair"	/guo3/ e.g., 果 "fruit "	/shi3/ e.g., 史 "history"
4	/ ma4/ e.g., 罵 "scold"	/yi4/ e.g., 意 "meaning"	/guo4/ e.g., 過 "pass"	/ shi4/ e.g., 是 "right"

Note. Following each syllable, the Chinese character example and the English equivalent of the character are given. The set 3 and set 4 are adopted from Wang Jongman and Sereno (2001).

Procedure

The procedure was the same to that in Experiment 1 A. There were 48 trials for syllable, pseudo-syllable, and hum respectively. The three kinds of materials were tested separately in different blocks, and the order of them was random between participants.

Results and discussion

For the RT, trials with incorrect response were eliminated from further analyses (9.08%). For both RT and accuracy, data exceed 3 SDs from means has been eliminated from further analysis (2.5 %). A correlation analysis was check before difference analysis to guarantee there was no trade-off of speed and accuracy.

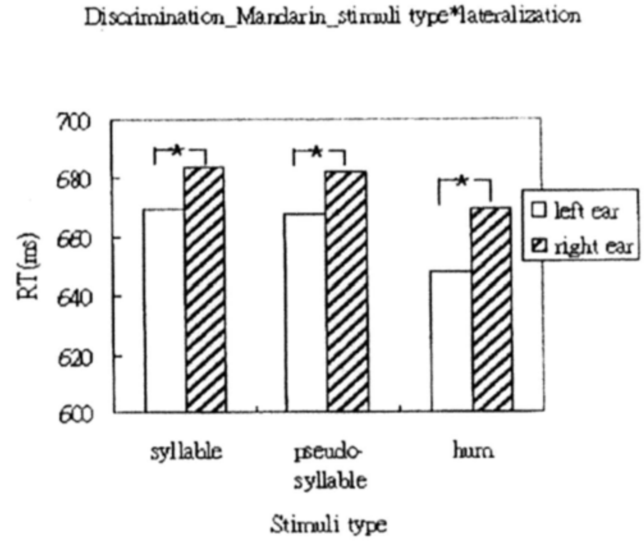
There were 2 IVs: stimuli type (syllable, pseudo-syllable, and hum), and ear (left and right), and two DVs (RT and accuracy). Two two-way (stimuli type and ear) ANOVAs were conducted on RT and accuracy to test the main effects of the two IVs, and how the stimuli type affects ear advantage. Greenhouse-Geisser correction was applied when appropriate to protect against inflation of Type 1 errors. There was no main effect and interaction effect on accuracy, thus, only the results of RT were reported.

RT

The ANOVA revealed a significant main effect of stimuli type, (F (2, 62) = 6.85, MSE = 42865.57, p <.05, η^2 = 0.13). The mean RT on syllable, pseudo-syllable, and hum were 677 ms, 675 ms, and 659 ms respectively. The response on syllable and pseudo-syllable was significantly slower that on hum (*ps* $< .05$), and the speed in syllable and pseudo-syllable were comparable.

And the main effect of lateralization was also significant, ($F(1, 31) = 6.28$, $MSE = 55802.60$, $p < .05$, $\eta^2 = 0.14$). Results revealed a LEA (662 ms) relative to the right ear (678 ms) performance. There was no interaction effect.

Figure 7. The lateralization patterns (RT) across stimuli type in Experiment 1B.



The results in the above revealed the LEA of tone discrimination in Mandarin. This advantage existed in syllable, pseudo-syllable, and hum. The performance on syllable was similar to that on pseudo-syllable. These results suggest that the LEA of Cantonese tone processing in Experiment 1B is reliable. In the following, the *Summary of Experiment 1* mainly talks about the Cantonese tone processing.

Summary of Experiment 1

Table 3. Main results of ear advantage in Experiment 1.

Language	Stimulus type	Tonal type	Ear advantage
Cantonese	Syllable	Level	LEA
		Contour	LEA
	Pseudo-syllable	Level	LEA
		Contour	LEA
	Hum	Level	LEA
		Contour	LEA
Mandarin	Syllable		LEA
	Pseudo-syllable		LEA
	Hum		LEA

Experiment 1A studied lateralization of tonal processing by using the discrimination task in Cantonese. In result, a LEA was found and this effect was not affected by the stimulus type. But he tonal type modulated lateralization in accuracy analysis, and the LEA only appeared on contour tone. The null effect on level tone might due to the ceiling effect, which presumption can be supported the LEA at both level and contour tone in RT analysis. In summary, a LEA of tonal processing was found in discrimination task, and this effect was not modulated by stimulus type and tonal type.

The evident LEA on hum is easy to understand, because on hum the segmental information was removed. And the f0 changing on hum should be processed as general auditory information that changes along frequency dimension. Studies (Zatorre & Belin; 2001; Zatorre, Belin, & Penhune, 2002) found that frequency varying information should be preferentially processed by the temporal lobes in the right hemisphere.

For Cantonese tone on syllable condition, two past studies (Benson, Smith, & Arreaga, 1973; Smith & Shand, 1974) failed to find ear advantage. On one hand, these two studies used the traditional DL tasks (identification), on the other hand, less participants (about 7) was recruited in these two studies. The results discrepancies may due to these two reasons. The LEA on syllable, pseudo-syllable, and hum in this study suggests the acoustic processing of tonal information in the discrimination task.

Secondly, the main effect of stimulus type was significant. The performance on syllable and pseudo-syllable was faster (around 20 ms) than that on hum. And the performance on syllable and pseudo-syllable was comparable. The performance advantage on syllable and pseudo-syllable relative to hum might due to linguistic experience. According to memory trace, both syllable and pseudo-syllable are naturally pronounced sounds because they have segmental information, but the hum does not. Thus, the linguistic memory also plays some role in tonal discrimination, although the LEA in this study reflects the dominant role of acoustic feature. The similar performance on syllable and pseudo-syllable suggests that the naturalness rather than the meaning of the sound is important.

Thirdly, performance at level tone was better than that at contour tone. On one hand, these results support the acoustic hypothesis. The f0 lines in the Figure 1 shows that level tones are distinguished from the initial section, however, contour tones are distinguished till from the center section. Thus, performance at level tone and contour tone is consistent with their intrinsic acoustic features. On the other hand, these results reflect the psychological reality of tonal dimension, and suggest that level tone and contour should be studied separately.

In summary, tonal discrimination is finished based on comparison of acoustic patterns, and it is mainly engaged by the right hemisphere. Meaning information is not accessed in this experiment. Acoustic processing does not exclude the role of language experience. The performance on naturally pronounced stimuli (syllable and pseudo-syllable) is much better than that of hum.

However, it is well established that lexical tone in tonal languages are phonological contrasts. And it is generally accepted that linguistic related information should be predominately processed by the left hemisphere. Furthermore, some studies have found the left hemispheric advantage of tonal processing with DL and neuroimaging methods (Wang, Jongman et al., 2001; Wang, Behne et al., 2004; Wong, Parsons et al., 2004). The substantial difference between those studies and the Experiment 1 was in task requirements. Thus, Experiment 2 further studied Cantonese tone processing by using the identification task (Bergen test).

Chapter 5

Tonal Processing in Identification DL

Experiment 2A- Cantonese tone

Experiment 1 found the LEA in tonal processing, which indicates an acoustic role. From the Introduction part, we know that with the DL paradigm some studies revealed the REA in tonal processing (Van Lancker & Fromkin, 1973 and 1978; Wang, Jongman et al., 2001; Wang, Behne et al., 2004), and most of them used the identification DL task. Thus, the result discrepancies of Experiment 1 and past studies might be caused by task requirement. In the Experiment 2, the Bergen test (Hugdahl, 1995), which was much similar to traditional DL tasks, was employed to testify the lateralization pattern of tonal processing.

The IVs of Experiment 2 were the same to that in Experiment 1. But for DV, only the report proportions of the left and right ear were recorded and compared in Experiment 2.

The linguistic involvement in the identification task is more than that in the discrimination task, which is mentioned in the Introduction part (the introduction of the DL paradigm). If that is the case, the REA would be found in this experiment, at least in syllable condition.

Participants

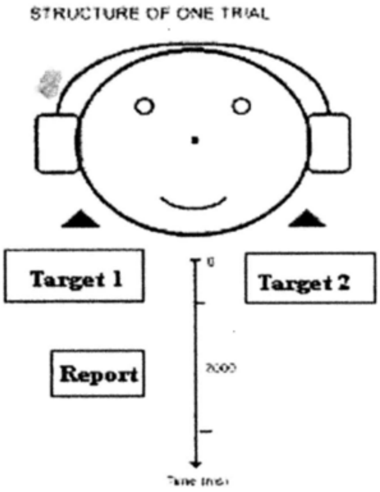
Thirty four native Cantonese speakers (13 males, aged 18-23) participated in this task to get course credits. All of them were right handed, and none of them had more than three years of musical training or had hearing impairment. Audiometric test was conducted to choose participants whose listening ability of the two ears was not different significantly.

Materials and design

Stimuli were adopted from Experiment 1. The three level tones constituted 9 dichotic-pairs, including six different dichotic-pairs (1-3, 1-6, 3-6, 3-1, 6-1, and 6-3) and three same dichotic-pairs (1-1, 3-3, and 6-6). The four syllable sets (/fu/, /yi/, /si/, and /se/) made up 36 trials for level tones, and 36 similar trials for contour tones. The level trials and contour trials were studied separately in two blocks with random order. In addition, pseudo-syllable (/bi/, /bu/, /di/, /du/) and hum also had one level and one contour tone blocks. The test order of stimulus type was random in participants.

Procedure

Figure 8. The diagram of the Bergen test in one trial.



In the Bergen test (Hugdahl, 1995), two stimuli (one dichotic pair) were presented to the two ears simultaneously. Participants were required to report the first one they heard. In past studies, report manner could be oral report, key-press report, or putting down the answer on an answer sheet. But for Cantonese tone, it is impossible to judge participant's oral responses exactly based on individual vocal answer, given the tone 1, tone 3, and tone 6 are level tones which should be identified in relative to other tones or in context. Thus, the key-press report manner was employed. Numbers 1-6 in the numeric keyboard were used to denote tone 1-6. Participants needed to report their answer by pressing a number key.

Each block had 12 same dichotic-pairs and 24 different dichotic-pairs. For the same dichotic-pairs, no matter which ear were report, the answer would be the same. Thus, only the different dichotic-pairs were counted in analysis. In each trial, participants could report the left or the right ear. Report proportions of the left and right ears were recorded and analyzed. For example, there were 24 valid trials for the level tone block in syllable condition. And one participant reported 16 times from the left ear and 6 times from the right ear. The proportions of left and right ear report were 66.7 % and 33.3 % respectively, which shows a left ear advantage in this condition.

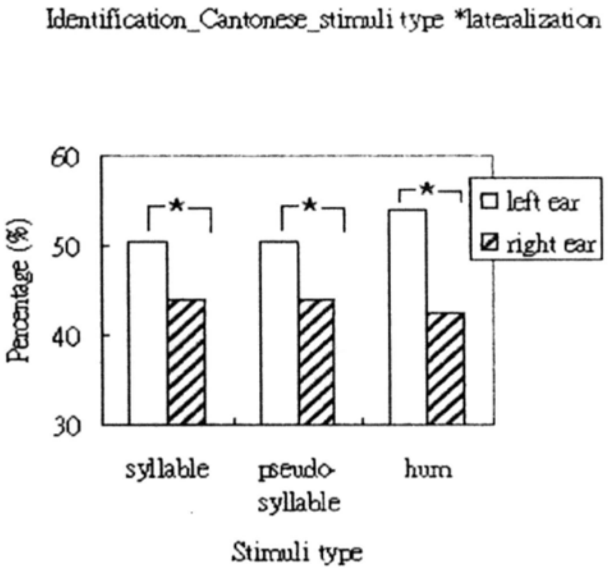
Results and discussion

Similar data trimming methods were done before data analysis and 5.5 % of data was discarded. One three-way ANOVA with the factors of ear, stimulus type, and tonal type was done on report proportion. The analysis focused on the main effect of the three factors, and how stimulus type and tonal type modulate lateralization.

Greenhouse-Geisser correction was applied when appropriate to protect against inflation of Type 1 errors.

The ANOVA showed a significant lateralization effect (*F* (1, 33) = 8.52, *MSE* = 0.054, $p < .05$, $\eta^2 = 0.23$), which came from the performance advantage of the left ear (52 %) relative to the right ear (43 %). Results also showed a significant main effect of tonal type (*F* (1, 33) = 6.25, *MSE* = 0.065, $p < .05$, $\eta^2 = 0.25$). The identification accuracy at contour tones (M = 48 %) was better than that at level tone (47 %).

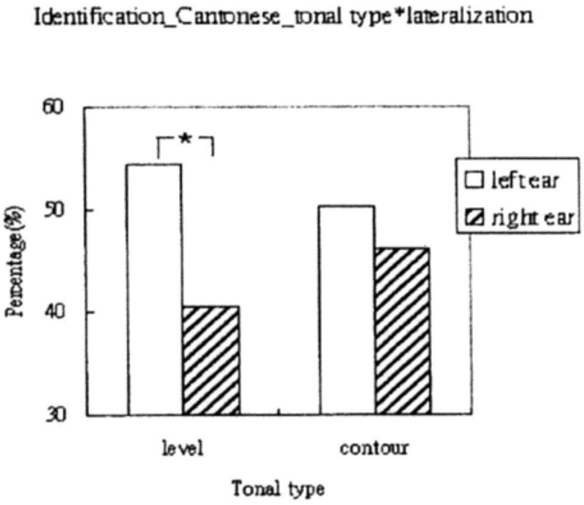
Figure 9. Lateralization patterns across stimulus type in Experiment 2A.



The LEA in hum (12 %) seemed greater than that in syllable (5 %) and pseudo-syllable (5 %), but the interaction between stimulus type and lateralization was not statistically significant (Figure 8). The interaction between tonal type and lateralization (Figure 9) was marginally significant (Figure 9) (*F* (2, 66) = 6.08, *MSE* = 0.055, $p < .05$, $\eta^2 = 0.19$). At level tones (t (33) = 3.81, $p < .05$), the performance of the left ear (54.5 %) was better than that of the right ear (40.6 %). At contour tones (*t*

(33) = 1.89, *p* = *.067*), the performance of the left ear (50.3 %) was not significantly better than that of the right ear (46.3 %).

Figure 10. Lateralization patterns across tonal type in Experiment 2A.



Experiment 2B- Mandarin tone

Experiment 2A failed to find the REA on Cantonese tone processing[9], which seemed strange to past studies on Mandarin tone. Given that scarce identification DL study has been conducted on Cantonese tone, it is not easy to compare the results of Experiment 2A with similar studies on Cantonese tone. However, identification DL studies on Mandarin tone were relatively more. Thus, Experiment 2B testified the lateralization of Mandarin tone processing with Bergen test (both oral report and key-press report), by which, I wanted to testify the validity of the Bergen test, and in turn to examine the results of Experiment 2 A.

---

[9] The lateralization pattern in the syllable condition might be affected by that in the pseudo-syllable and hum conditions, given that the syllable blocks were embedded in the pseudo-syllable and hum blocks. In order to remove this confounding, I analyzed the data from the first syllable block of each participants, but the results still showed the same pattern of LEA (*F* (1, 33) = 3.95, *MSE* = 0.034, $p < .05$).

Participants

Twenty six (10 males, aged 18-24) undergraduates in the South China Normal University (SCNU) were recruited in this experiment. All of them were native Chinese speakers and speak Mandarin in everyday life, and none of them had more than three years of musical training. In addition, all of them were right handed, and none of them had hearing disability. And an audiometric test was conducted to guarantee the hearing ability of the two ears was not different significantly. Ten RMB each hour was paid to them for rewards.

Materials and design

Syllable /yi/ at four tones (mean "medicine", "aunt", "chair", and "meaning" respectively) and /ma/ at four tones (mean "mother", "herb", "horse", and "scold" respectively) were used as materials (refer to Table 2). Four /yi/ syllables constituted 12 different dichotic-pairs and 4 same dichotic-pairs. The two sets of syllables constituted 32 dichotic-pairs. Similarly, only the 24 different pairs were valid trials.

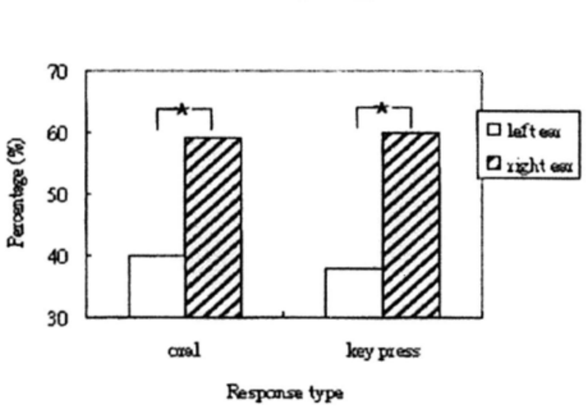
Procedure

The Bergen test was adopted. In the key press report manner, number keys 1-4 on the numeric keyboard were used to denote tone 1-4. Participants needed to press a key to indicate their answer. In the oral report task, participants were required to report the first syllable they heard as quickly as possible. Experimenter marked participants' answers on the answer sheet during the experiment. And their vocal responses were recorded using a voice recorder, which was used as backup.

Results and discussion

Two t-tests were conducted to compare the left and right ear performance on oral report and key press report respectively. The t-test on oral report showed significant REA (*t* (25) = 2.39, p < .05, Cohen' *d* = 0.41[10]), the performance of the right ear (59 ± 20 %) was significantly higher than that of the left ear (40 ± 19 %). The t-test on key press report task also showed a significant REA (*t* (25) = 2.37, p < .05, Cohen' *d* = 0.44), the performance of the right ear (60 ± 23 %) was significantly higher than that of the left ear (38 ± 23 %). For the oral report manner, there were 16, 5, and 4 participants showing the REA, LEA, and no advantage ($\chi2$ = 5.63, p < .05). And for the key-press report manner, there were 16, 6, and 4 participants showing the REA, LEA, and no advantage ($\chi2$ = 4.87, p < .05). Thus, the report manner does not matter.

Figure 11. Lateralization patterns in two report manners in Experiment 2B.



---

[10] The Cohen' *d* is an appropriate measure of effect size in T-test. And the effect sizes of about 0.2, 0.5, and 0.8 are termed as small, medium, and large, respectively (Cohen, 1988).

Summary of Experiment 2

Table 4. Main results of ear advantage in Experiment 2.

Language	Stimulus type	Tonal type	Ear advantage
Cantonese	Syllable	Level	LEA
		Contour	
	Pseudo-syllable	Level	LEA
		Contour	
	Hum	Level	LEA
		Contour	
Mandarin	Syllable		REA

Experiment 2A studied lateralization of Cantonese processing in identification DL task and found a LEA in general. The LEA was not affected by stimulus type, although the LEA on syllable and pseudo-syllables seemed weaker than that on hum. However, the LEA was modulated by tonal type, and the LEA only appeared on level tone.

It seems strange that the LEA was found on Cantonese tone processing in the identification task, given that a REA was usually found in past studies on Mandarin tone using identification tasks. In order to testify the validity of the Bergen test in this study, the similar procedure was done on Mandarin tones in Experiment 2B. In result, the REA was found, which repeated those similar studies on Mandarin tone identification (Wang, Jongman et al., 2001; Wang, Behne et al., 2004). Thus, the Bergen test in this study is valid.

Categorical perception studies revealed that Mandarin tones are perceived categorically (Wang, 1976). And for Cantonese tones, different results were found at level and contour tones. The contour tones rather than the level tones are processed categorically (Francis, Ciocca, Ng, 2003). Thus, Mandarin tone and contour tone in

Cantonese are categorically represented in the brain. And in the in the identification DL, a REA is more likely to be found. But for the level tone in Cantonese, it is continuously presented in the brain as frequency varying continuum. And the LEA should be found in DL studies, irrespective of the task requirement.

Chapter 6

Tonal processing in Passive Oddball Task

Experiments 3 and 4 further studied the time course and lateralization of Cantonese tone processing by using the ERP technique, and aimed at 1) testing the time courses of the effects of tonal dimensions and stimulus type, 2) testing the time point of segment and tone combination, and 3) retesting the lateralization effect.

In Experiments 3 and 4, three IVs were manipulated: the f0 direction (level and contour), f0 height-deviation magnitude (large and small), and stimulus type (syllable and hum). In Experiment 3, the passive oddball task was employed, in which the MMN and P3a were evoked. The latencies and amplitudes of the MMN and P3a were measured as the DVs.

If the functional model is correct, both f0 direction and f0 height-deviation magnitude would not affect Cantonese tone processing, because all the six tones are equally linguistic contrasts. But the stimulus type would affect tonal processing, the left hemisphere advantage would be found on syllable. If the acoustic model is correct, both the f0 direction and f0 height-deviation magnitude would affect tonal processing. For instance, the level tone would be processed faster than the contour tone according to the pitch lines (Figure 1).

For the relation of segment and tone, if they are processed separately, the processing manner would not be affected by stimulus type. If they are interactive, the processed speed of tone would be affected by stimulus type.

Participants

Eighteen undergraduates (8 males, aged 18-23) in the CUHK participated in this experiment. All of them were right handed, and none of them had hearing difficulty. They were native Cantonese speakers and lived in HK for at least 10 years. They participated in this experiment to get course credits.

Material and design

Cantonese syllable */yi/* pronounced at six tones were used as stimuli materials (please refer to Table 1 in Experiment 1). These stimuli were produced by a female native speaker, and were normalized to 400 ms in duration (including the 15 ms rise and fall ramps) and 60dB in peak intensity. Additionally, the six hums that were synthesized from the six syllables were also tested.

Table 5. Descriptive features of the six tones that were quantized from recorded syllables (syllable /yi/ at six tones). The tonal patterns of the synthesized hums were the same as the corresponding syllables.

Tone	TP (ms)	f0 (Hz)			
		Onset	Offset	TP	Average
Yi1	No	270	260	NA	265
Yi2	200	155	260	150	187
Yi3	No	255	210	NA	215
Yi4	No	180	110	NA	140
Yi5	210	160	200	150	175
Yi6	No	185	170	NA	177

Note: TP was the abbreviation of turning point.

The variations in f0 over time for the six syllables are illustrated in Figure 1 and Table 2. From Figure 1 and Table 2, we can know that yi6, yi1 and yi3 are level tones and yi4, yi2, and yi5 are contour tones. Thus, yi6/yi1 (standard/deviant) and yi6/yi3 were oddball pairs of level tones, yi4/yi2 and yi4/yi5 were oddball pairs of contour tones. For yi6/yi1 and yi6/yi3, the f0 height-deviation magnitude of yi6/yi1 was larger than that of yi6/yi3, thus, they were the level-large (yi6/yi1) and level-small (yi6/yi3) conditions respectively. Similarly, yi4/yi2 and yi4/yi5 denoted the contour-large and contour-small conditions respectively.

Procedure

In an electromagnetically-shielded and sound-attenuated room, the participants were seated comfortably about 1 meter from a computer screen. The oddball pairs were presented to them by a headphone when they were watching a self-chosen, closed-caption silent movie.

There were two level conditions and two contour conditions on both syllable and hum. And the level and contour conditions were studied separately in two blocks. For the level block on syllable, sound sequence that contained 640 /yi6/, /80 /yi1/, and 80 /yi3/ (S/D = 8/1/1) were presented pseudo-randomly with the restriction that the number of consecutive standards did not exceed the range 6 to 10. Similarly, for the contour block on syllable, which contained 640 /yi4/, / 80 /yi2/, and 80 /yi5/ (S/D = 8/1/1), and all of them were presented randomly with similar constrain. Same manipulation was done on hum. The order of syllable and hum was counterbalanced in participants. And the order of level and contour blocks within syllable and hum was also random. Totally, there were 3200 trials that were divided into 4 blocks, each trial consisted of a 400 ms stimuli sound and a 400 ms blank, and there were a 3-minute rest between two blocks. The whole experiment lasted for about 50 minutes.

Data recording

EEGs was recorded by a 64-channel (Ag-AgCl) NeuroScan system. Electrodes were positioned following the 10-20 system. One electrode placed on the medial frontal aspect served as online reference and data were referenced to the electrode on the nose in offline analysis. The left and right mastoids were also recorded. Vertical electrooculogram (EOG) was recorded supra- and infra-orbitally from the left eye. The horizontal EOG was recorded as the left versus right orbital rim. The impedance of all electrodes was kept below 5 kΩ. EEG and EOG were digitized online at 500 Hz and bandpass filtered from AC 0.05 to 100 Hz.

Data analysis

The resulting data were offline filtered from 1 to 30 Hz. Eye movement artifacts (mean EOG voltage exceeding ± 80 μV) and trials containing amplifier clipping and peak-to-peak deflection exceeding ± 80 μV were discarded. The remaining artifact-free trials were used to compute the ERPs evoked by the standard (only the pre-deviant standards were used) and deviant sounds in each condition. The epoch for analysis was 600 ms, including a 100 ms pre-stimulus baseline.

The standard-related ERP was subtracted from the deviant-related ERP to extract the difference waves in each condition. As shown in Figure 11 and Figure 13, the MMN and a following P3a was elicited in all conditions. Both past studies (Näätänen, 2001; Polich, 2007) and the topographies in this experiment showed that the MMN and P3a were largest on the frontal scalp. Therefore, we measured the two components from the following 9 points: F3, Fz, F4, FC3, FCz, FC4, C3, Cz, and C4. The time windows of measuring the MMN and P3a were defined according to the onset and offset time points of the two components. We measured the peak amplitudes and peak latencies of MMN at the most negative points between 100 – 300 ms after stimulus onset. Similarly, the peak amplitudes and peak latencies of P3a were measured at the most positive point between 250 – 400 ms after stimuli onset. In addition, the ground waves showed different patterns on onset latencies and peak latencies of MMN. Thus, the MMN onset latencies were also measured. Table 3 and Table 4 show the mean amplitudes and latencies of each index in each condition.

ANOVAs were separately conducted on the amplitudes and latencies of the MMN and the P3a to test the difference across conditions. The results were reported in three parts. Firstly, for the syllable, five 3-way ANOVAs [f0 direction: level or contour, f0 height-deviation magnitude: large or small, and lateralization: left (F3, Fc3, C3), center (Fz, FCz, Cz), and right (F4, FC4, C4)] were conducted on each DV (MMN onset latency, MMN peak latency, MMN amplitude, P3a peak latency, and P3a amplitude); secondly, for the hum, the same five ANOVAs were done on each DV; and thirdly, ignore f0 height-deviation-small conditions (based on the results of the above two parts) and conducted five three-way ANOVAs [stimulus type (syllable, hum), f0 direction (level, contour), and lateralization (left, center, and right)] on the five DVs. Greenhouse-Geisser correction was applied when appropriate to protect against inflation of Type 1 errors.

The first and second parts wanted to analysis the effects of f0 direction and f0 height-deviation magnitude on tonal processing in syllable and hum respectively. And the third part focused on the effects of stimulus type and f0 direction. The three parts were reported one by one in the following.

Results and discussion

Part 1: Syllable

For the MMN onset latency, the main effect of tonal type (f0 direction) was significant (*F* (1, 17) = 38.33, $p < .001$), the onset latencies of level tone (112 ms) were much earlier than that of the contour tone (161 ms). The main effect of the f0 height-deviation magnitude was marginally significant (*F* (1, 17) = 3.78, $p = .059$), large deviation (132 ms) elicited earlier MMNs than small deviation did (141 ms). The interaction of tonal type and f0 height-deviation magnitude was significant (*F* (1, 17) = 7.17, $p < .05$), further analysis revealed that larger deviation (104 ms) elicited earlier MMN than smalle deviation (120 ms) only at level tone (*t* (17) = 5.56, $p < .05$).

Table 6. Latencies and amplitudes (standard deviations in parentheses) of MMN and P3a on syllable across condition in Experiment 3.

	MMN			P3a	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
Large-level	104 (7)	162 (8)	- 4.72 (.43)	304 (10)	2.21 (0.23)
Small-level	120 (7)	232 (7)	- 3.13 (0.38)	314 (8)	2.23 (0.29)
Large-contour	160 (8)	232 (10)	- 3.41 (0.45)	362 (8)	2.14 (0.38)
Small-contour	162 (7)	240 (9)	- 2.74(0.33)	360 (10)	2.44 (0.41)

For the MMN peak latency, the main effect of tonal type was significant (*F* (1, 17) = 23.01, p < .001), the latencies of level tone (197 ms) were much earlier than that of the contour tone (236 ms). The main effect of the f0 height-deviation magnitude was also significant (*F* (1, 17) = 23.78, $p <.001$), large deviation (197 ms) elicited earlier MMNs than small deviation did(236 ms). The interaction of tonal type and f0 height-deviation magnitude was significant (*F* (1, 17) = 6.58, $p < .05$), further analysis revealed that large deviation elicited earlier MMN than smalle deviation did only at level tone, and level tone elicited earlier MMN than contour tone did only when deviation magnitude was large, which means that the level-large condition elicited earlier MMN than the other three conditions, and the MMNs in other three conditions were comparable.

For the MMN amplitude, the main effect of tonal type was significant (F (1, 17) = 4.32, $p < .05$), the amplitude of level tone (-3.92 μV) were larger than that of the contour tone (-3.07 μV). The main effect of the f0 height-deviation magnitude was also significant (F (1, 17) = 4.98, p = < .05), large deviation (-4.06μV) elicited larger MMNs than small deviation did (-2.93μV). There was no interaction effect.

For the P3a peak latency, the main effect of tonal type was significant (F (1, 17) = 52.56, $p < .001$), the latencies of level tone (309 ms) were much earlier than that of the contour tone (361 ms). No effect was found on P3a amplitude.

The results showed no lateralization effect and interaction effect between lateralization and other IVs (Figure 13).

Figure12. Grand average waveforms on syllable across condition in Experiment 3.

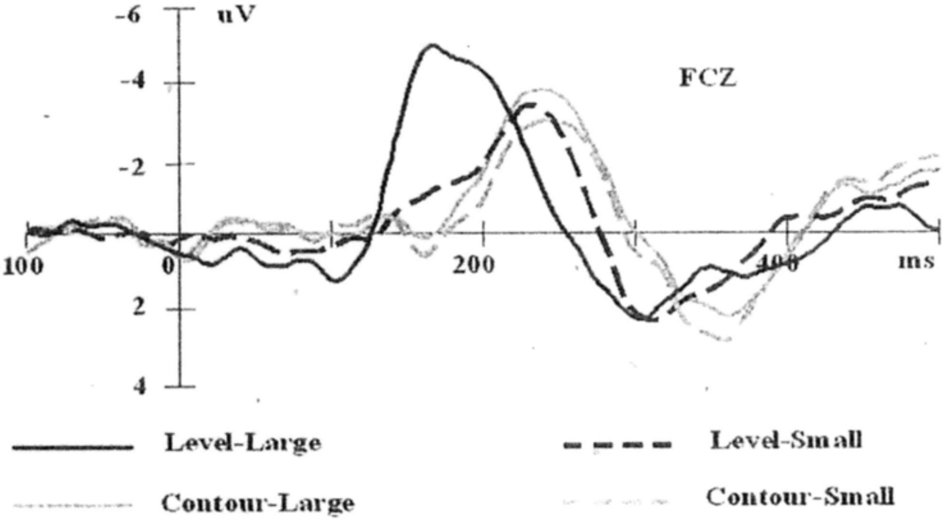
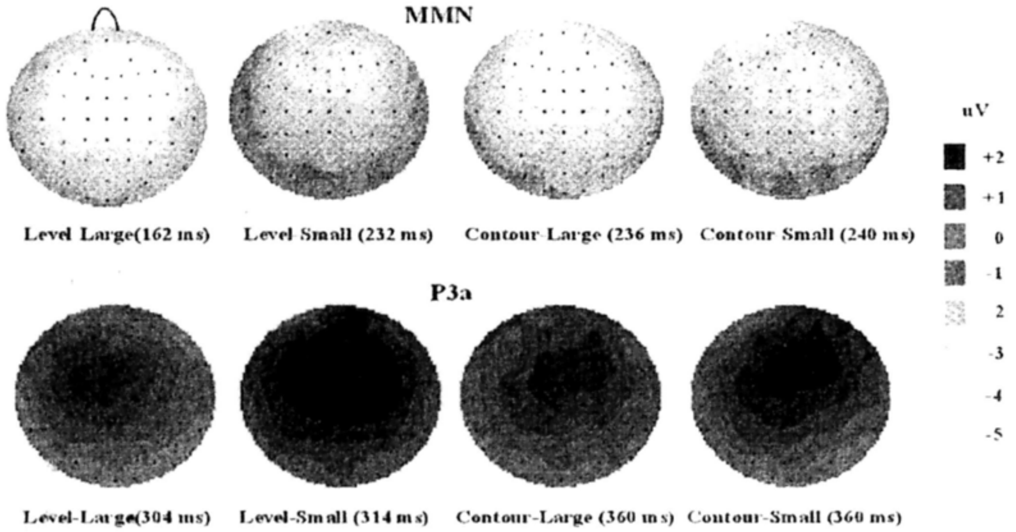


Figure 13. Topographies of MMN and P3a on syllable in the four conditions. Those maps showed the time points at which the largest effect appeared. Actually, all topographies in the following showed the time points that had largest effect.



Part 2: Hum

Table 7. Latencies and amplitudes (standard deviations in parentheses) of MMN and P3a on hum across four condition in Experiment 3.

	MMN			P3a	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
Large-level	95 (6)	180 (7)	- 2.92(.25)	305 (7)	1.52 (0.18)
Small-level	120 (7)	228 (6)	- 1.81 (0.34)	295 (8)	1.41 (0.25)
Large-contour	170 (7)	250 (7)	- 1.63 (0.52)	365 (10)	0.93 (0.36)
Small-contour	165 (10)	245 (8)	- 1.31 (0.36)	360 (9)	1.04 (0.30)

For the MMN onset latency, the main effect of tonal type was significant (F (1, 17) = 58.84, p < .001), the onset latencies of level tone (107.5 ms) were much earlier than that of the contour tone (167.5 ms). The main effect of the f0 height-deviation magnitude was marginally significant (F (1, 17) = 3.95, p = .055), large deviation (132.5 ms) elicited earlier MMNs than small deviation did (142.5 ms). The interaction between tonal type and f0 height-deviation was significant (F (1, 17) = 5.91, p < .05), further analysis revealed that large deviation (95 ms) elicited earlier MMN than smalle deviation (170 ms) did only at level tone ($t(17) = 16.23, p < .01$).

For the MMN peak latency, the main effect of tonal type was significant ($F(1, 17) = 30.09, p < .001$), the MMN latencies of level tone (204 ms) were much earlier than that of the contour tone (247.5 ms). The main effect of the f0 height-deviation magnitude was also significant ($F(1, 17) = 10.45, p < .05$), large deviation (215 ms) elicited earlier MMNs than small deviation (236.5 ms) did. The interaction of tonal type and f0 height-deviation was significant ($F(1, 17) = 5.83, p < .05$), further analysis revealed that large deviation elicited earlier MMN than smalle deviation did only at level tones, and level tone elicited earlier MMN than contour tone did only when deviation magnitude was large, which means that the level-large condition elicited earlier MMN than all other three conditions, and the MMNs in other conditions were comparable.

For the MMN amplitude, the main effect of tonal type was significant ($F(1, 17) = 4.01, p < .05$), the amplitude of level tone (-2.35 μV) were larger than that of the contour tone (-1.45 μV). The main effect of f0 height-deviation was also marginally significant ($F(1, 17) = 3.94, p = .065$), large deviation (-2.25 μV) elicited larger MMNs than small deviation (-1.55 μV) did. There was no interaction effect.

For the P3a peak latency, the main effect of tonal type was significant ($F(1, 17) = 53.56, p < .001$), the MMN latencies of level tone (300 ms) were much earlier than that of the contour tone (362.5 ms).

For the P3a amplitude, level tones (1.45μV) elicited larger P3a than contour tones (0.95μV) did ($F(1, 17) = 4.34, p < .005$).

Results showed no lateralization effect and no interaction between lateralization and other IVs (Figure 15).

Figure 14. Ground average waveforms on hum of the four conditions in Experiment 3.

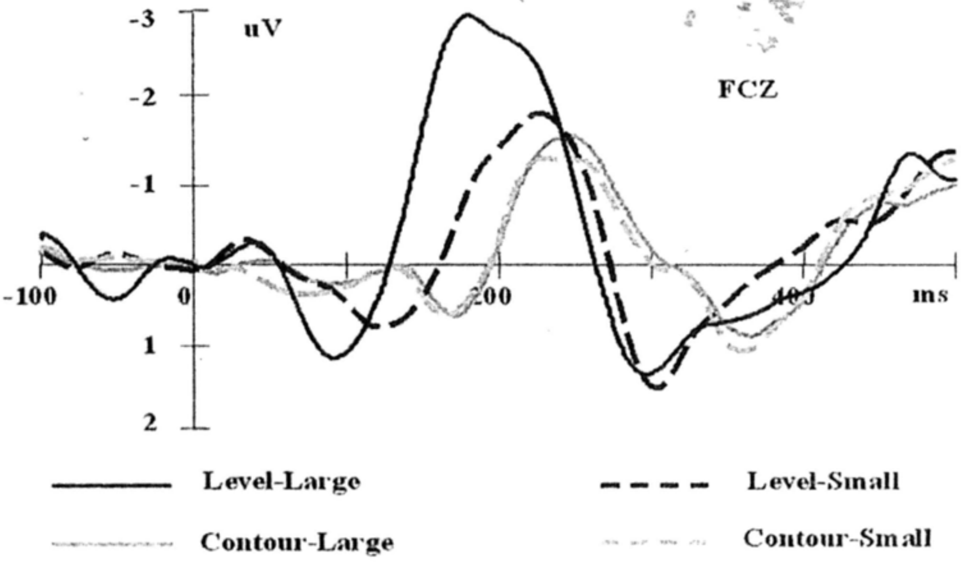
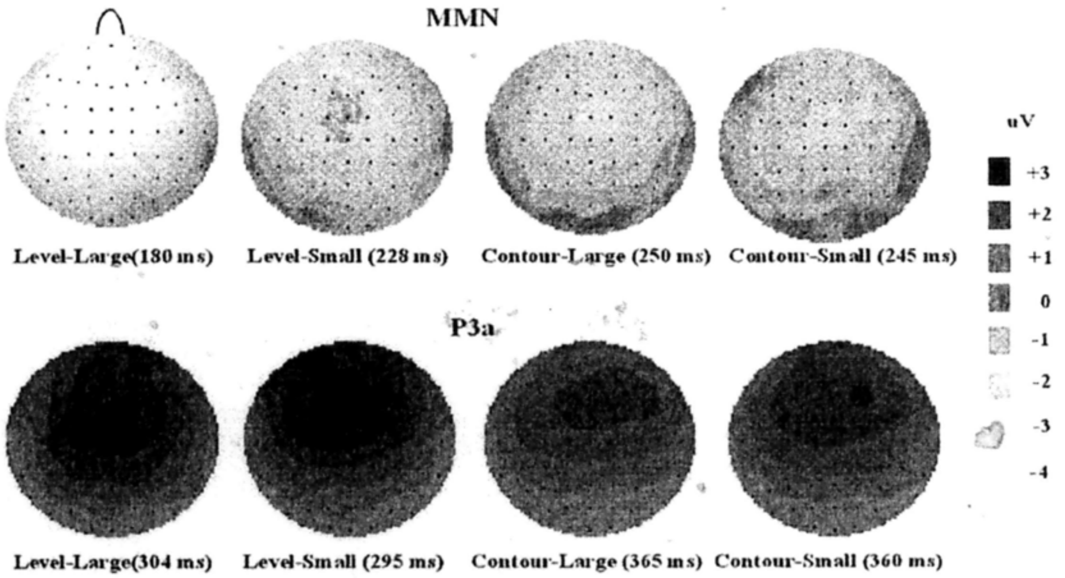


Figure 15. Topographies on hum of MMN and P3a in four conditions.



*Part3 Stimulus type * f0 direction*

The above two parts revealed the effect of tonal type on all indices, but the effect of f0 height-deviation magnitude was not as evident as that of the tonal type.

The f0 height-deviation magnitude had almost no effect on contour tone, and the f0 height-deviation effect on level tone mainly displayed on the MMN but not on P3a. Thus, in order to focus on the effect of stimulus type, the analysis in the third parts was done only in large-deviation conditions. Five ANOVAs with three factors (stimulus type, f0 direction, and lateralization) were conducted on the five DVs.

Table 8. MMN and P3a values in syllable and hum (only large-deviant condition).

	MMN			P3a	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
Syllable-level	104 (7)	162 (8)	- 4.72 (0.43)	304 (10)	2.21 (0.23)
Syllable-contour	160 (8)	232 (10)	- 3.41 (0.45)	362 (8)	2.14 (0.38)
Hum-level	95 (6)	180 (7)	- 2.92 (.25)	305 (7)	1.52 (0.18)
Hum-contour	170 (7)	250 (7)	- 1.63 (0.52)	365 (10)	0.93 (0.36)

For the MMN onset latency, the main effect of tonal type was significant (F (1, 17) = 80.01, p < .001), the onset latencies of level tone (99.5 ms) were much earlier than that of the contour tone (165 ms).

For the MMN peak latency, the main effect of the stimulus type was significant (F (1, 17) = 10.45, p <.05), syllable (197 ms) elicited earlier MMNs than hum (215 ms) did. The main effect of tonal type was significant (F (1, 17) = 60.01, p < .001), the latencies of level tone (171 ms) were much earlier than that of the contour tone (241 ms).

For the MMN amplitude, the main effect of the stimulus type was significant (F (1, 17) = 5.14, p <.05), syllable (- 4.07 μV) elicited larger MMNs than hum (- 2.27 μV) did. The main effect of tonal type was significant (F(1, 17) = 4.01, p < .05), the amplitude of level tone (- 3.82 μV) were larger than that of the contour tone (-2.52

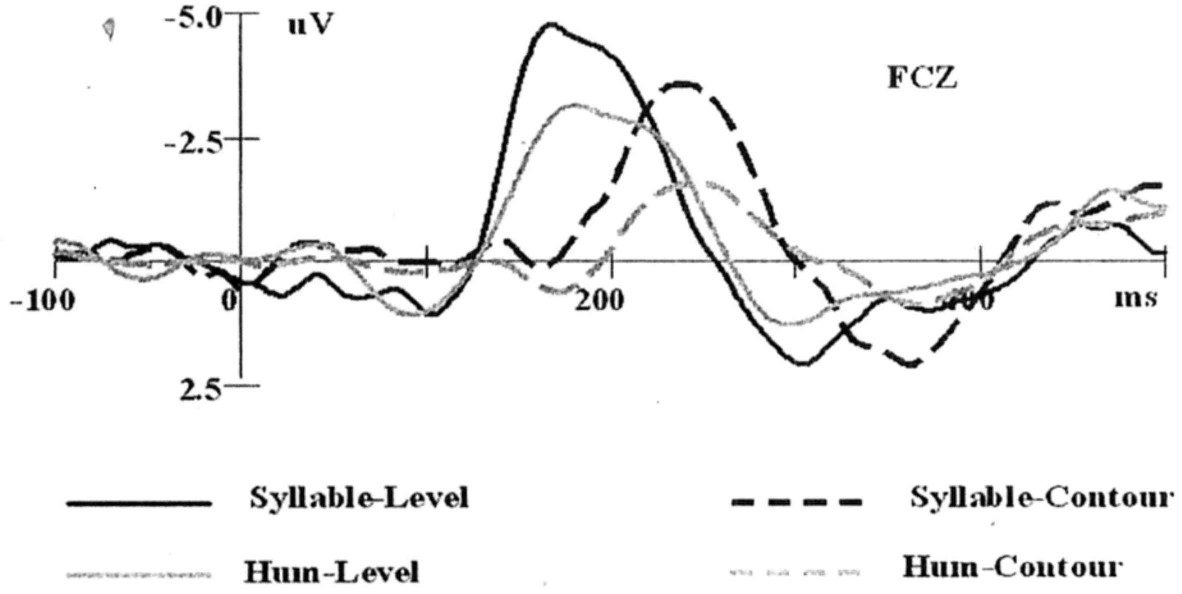
μV).

For the P3a peak latency, the main effect of tonal type was significant (*F* (1, 17) = 50.78, p < .001), the latencies of level tone (304.5 ms) were much earlier than that of the contour tone (363.5 ms).

For the P3a amplitude, the main effect of the stimulus type was significant (*F* (1, 17) = 3.89 p <.05), syllable (2.18 μV) elicited larger MMNs than hum (1.25 μV) did.

The results showed no effect of lateralization and no interaction between lateralization and other IVs (figure 4).

Figure 16. The ERPs of level and contour tones in syllable and hum (all were large f0 height-deviation conditions). For topographics, please refer to Figure 13 and Figure 15.



Summary of Experiment 3

Table 9. Main results in Experiment 3. " √ " means a significant effect was found.

IV	MMN			P3a	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
F0 direction	√	√	√	√	√ (hum)
F0 height		√	√ (syllable)		
Stimulus type		√	√		√
Lateralization					
Direction × height	√	√			

Note: Two effects were only significant on syllable or hum, which were marked. Except there two, all other effects were significant on both syllable and hum.

Experiment 3 studied the brain responses to Cantonese tone processing by using the passive oddball paradigm, in which, tonal dimensions and stimulus type were manipulated, and MMN and P3a were measured as DVs.

ANOVAs showed that the tonal type (f0 direction) effect appeared at MMN onset stage and lasted till to the P3a stage. The level tone elicited earlier MMN and P3a, and larger MMN than contour tone did. The effect coincides with the acoustic features in Figure 1 and Table 2, which suggests an acoustic processing of Cantonese tone.

The effect of f0 height-deviation was modulated by tonal type. The brain responses to large deviation and small deviation at level tone were separated at about 100 ms (MMN onset) and merged at about 250 ms (MMN peak), which means that tone 1 and tone 3 are distinguished at about 100 ms and the processes of level tone is finished at about 250 ms. These results coincided with the pitch lines in Figure 1. But the deviation magnitude could not affect brain responses at contour tone. The MMN and P3a to large and small deviation contour tone were almost coincided, which means that our brain can not discriminate tone 2 and tone 5 easily, at least not easily in this discrimination task. The acoustic distinction of tone 2 and tone 5 did not elicit different brain response. These results are consistent with the notion that the contour tones rather than the level tones are categorically perceived in Cantonese (Francis, Ciocca, & Ng, 2003).

In addition, this experiment compared the tonal processing on syllable and on hum. Results showed that stimulus type effect appeared at about 250 ms (the MMN peak) and lasted till to P3a stage. Syllable elicited larger MMN and P3a than hum did, which was consistent with the better performance on syllable relative to hum in the DL experiment. These results indicate that segment can modulate tone processing from 250 ms to about 350 ms after stimuli onset. Segment and tone are interactive rather than independent during speech perception.

But this experiment failed to find any lateralization effect and interaction between lateralization and other factors. It is possible that in the discrimination task, the acoustic processing and lexical processing are mixed together from beginning to the end of the task, no one can override the other.

Chapter 7

Tonal Processing in Active Oddball Task

Experiment 4 was designed to study tonal processing in the active oddball task, which could complement results in Experiment 3. The stimuli in Experiment 4 were the same as that in Experiment 3. But the tasks of them were different. In Experiment 4, participants needed to do clarification task, in which they needed to press one key for target sounds (standard) and press another key for non-target sounds (non-standard).

There were three IVs: f0 direction, f0 height, and stimulus type. The latencies and amplitude of N2 and P3b were measured as DVs on both syllable and hum (indeed, the N1 and P2 were also elicited on syllable, and the latency and amplitude of N1 and P2 were also analyzed on syllable).

In the active oddball task, subjects were required to process lexical tone actively. Thus, the processing level in Experiment 4 was deeper than that in Experiment 3. And the lexical involvement in Experiment 4 would be more than that in Experiment 3. Therefore, Experiment 4 hoped to find more obvious effect of linguistic memory on tonal processing. For instance, the main effect of stimulus type would be found, ERP patterns elicited by syllable and hum would be different. And the interaction between tonal dimension and stimulus type was hoped to be found, the effects of tonal dimensions on syllable and hum would be different to some extent.

Participants

Nineteen native Cantonese speakers (8 males, aged from 18 to24) who were undergraduate students in the CUHK were recruited in the experiment. All of them were right handed with normal hearing ability. And none of them had more than three years of musical training. They participated in this experiment to get rewards of 50 HK$ each hour.

Materials and design

Stimuli materials and design the same to that in Experiment 3.

Procedure

In each trial of Experiment 4, a sound was presented for 400 ms, after that was a blank that was 2000 ms in maximum. Participants were requirement to judge the sound was target or non-target and press a corresponding key after the sound onset immediately. If the response was made during sound onset, the blank would not appear, if the response was made during blank onset, the response would terminate the blank. And then, a feedback of this trial was presented for 1000 ms. And after each block, the feedback of the whole block was given.

Before each block, a training stage was done to guarantee participants understand the rule. During the training stage, the target (deviant) in this block was given to participant and they were required to press one key to target and press another key to non-target.

Each trial lasted for about 3 seconds, and the whole experiment (3200 trials) lasted for about two and a half hours, which was finished in two times by the same participant.

Data recording

EEG was recorded by a 64-channel (Ag-AgCl) NeuroScan system. Electrodes were positioned following the 10-20 system convention. The left and right mastoids were also recorded. Onemedial frontal electrode was used as inline reference and the data was referred to the linked mastoids offline. VEOG was recorded supra- and infra-orbitally from the left eye. HEOG was recorded at the left versus right orbital rim. The impedance of all electrodes was kept below 5 kΩ. EEG and EOG were digitized online at 500 Hz and bandpass filtered from 0.05 to 100 Hz.

Data analysis

The resulting data were offline filtered from 1 to 30 Hz. Eye movement artifacts (mean EOG voltage exceeding ± 80 μV) and trials containing amplifier clipping and peak-to-peak deflection exceeding ± 80 μV were discarded. The remaining artifact-free trials were used to compute the ERPs evoked by the standard (only the pre-deviation one) and deviant sounds in each condition. The epoch for analysis was 800 ms, including a 100 ms pre-stimulus baseline.

Subtract the standard-related ERP from the deviation-related ERP to get purely deviant-related components for each condition. From the Figure 17 and Figure 19 we could know that the components on syllable and hum were different, therefore, syllable and hum were analyzed separately. For the syllable part, as shown in Figure 17, deviant stimuli elicited N1, P2, N2, and P3b, the peak latencies and amplitudes were measured for the four components respectively. According to the grand average ERPs and topographies, and also referring to past studies (see Polich, 2007), time window of N1, P2, N2, and P3b were defined at 100-200 ms, 150-250 ms, 200-350 ms, and 300-500 ms respectively. And those components were measured from the following nine points: FC3, Cz, FC4, C3, Cz, C4, CP3, CPz, and CP4.

For the hum part, N2 and P3b were elicited. The peak latencies and amplitudes of M2 and P3b were measured from the same 9 point as in syllable. The time window of N2 and P3b were 150-350 ms and 250-600 ms respectively. Also, the onset latencies of the N2 were measured.

For the syllable part, 8 three-way ANOVAs were done on the latencies and amplitudes of N1, P2, N2, and P3b. The three factors were: f0 direction, f0 height deviant magnitude, and lateralization. And for the hum part, 5 ANOVAs were done with the same three factors on latencies and amplitudes of N2 and P3b, and on the onset latencies of N2. Greenhouse-Geisser correction was applied when appropriate to protect against inflation of Type 1 errors.

Results and discussion

Part 1: syllable

Table 10. Latencies and amplitudes (SDs in parentheses) of N1, P2, N2 and P3b on syllable in four conditions in Experiment 4.

	N1		P2		N2		P3b	
	Latency	Amplitude	Latency	Amplitude	Latency	Amplitude	Latency	Amplitude
Large-level	134	- 2.78	166	-2.30	232	- 5.4	360	5.93
	(6)	(0.36)	(6)	(0.35)	(7)	(0.54)	(9)	(0.63)
Small-level	150	- 0.50	176	-0.48	284	- 3.7	470	2.11
	(6)	(0.05)	(6)	(0.01)	(8)	(0.38)	(10)	(0.34)
Large-contour	132	- 1.92	186	1.61	274	- 5.02	470	2.15
	(6)	(0.21)	(5)	(0.05)	(8)	(0.49)	(10)	(0.36)
Small-contour	142	- 1.32	190	0.55	295	- 5.05	460	3.23
	(7)	(0.26)	(7)	(0.01)	(8)	(0.51)	(12)	(0.23)

Note: in this table, all latencies are peak latencies.

For the N1, latency data showed the significant main effect of f0 height-deviation magnitude (F (1, 18) = 4.04, p <.05), large deviation (133 ms) elicited earlier N1 than small deviation (146 ms) did. And the amplitude data also showed significant main effect of f0 height-deviation magnitude, (*F* (1, 18) = 5.24, *p* <.05), large deviation (-2.35 μV) elicited earlier N1 than small deviation (-0.91μV) did.

For the P2, both latency and amplitude showed significant tonal type effect. Level tone (171 ms, -1.4μV) elicited earlier (*F* (1, 18) = 3.84, *p* <.05) and smaller (*F* (1, 18) = 5.36, *p* <.05) P2 than contour tone (188 ms, 1.1 μV) did.

For the N2, the latency showed that level tone (252 ms) elicited earlier N2 than contour tone did (290ms), (*F* (1, 18) = 28.65, *p* <.001), and large deviation (257 ms) also elicited earlier N2 than small deviation (284 ms) did, (*F* (1, 18) = 12.36, *p* <.001).

The amplitude showed marginally significant effect of deviation magnitude (*F* (1, 18) = 3.69, *p* = .056). Large deviation (- 5.2 μV) elicited larger N2 than small deviation (-4.3 μV) did. And the interaction between tonal type and f0 height-deviation magnitude (*F* (1, 18) = 5.01, *p* <.05). The large deviation (-5.4μV) elicited larger N2 relative to the small deviation (-3.7μV) only at level tone.

For the P3b, the latency showed significant main effect of tonal type (*F* (1, 18) = 39.62, *p* <.001), P3b of level tone (415 ms) was earlier than that of contour tone (465 ms). Latency also showed significant main effect of f0 height-deviation magnitude (*F* (1, 18) = 40.02, *p* <.001), the large deviation (435 ms) elicited earlier P3b than small deviation (465 ms) did. The interaction between them was significant (*F* (1, 18) = 5.36, *p* <.05), further analysis showed that the P3b in Large-level condition (360 ms) was earlier than all other three conditions (470 ms, 470 ms, 460 ms), and the other three conditions were comparable.

The P3b amplitude showed significant main effect of tonal type (*F* (1, 18) = 4.23, $p <.05$), P3b of level tone (4.05 μV) was larger than that of contour tone (2.66μV). And amplitude also showed significant main effect of f0 height-deviation magnitude (*F* (1, 18) = 4.56, p <.05), the large deviation (4.10 μV) elicited larger P3b than small deviation (2.60 μV) did. The interaction between them was significant (*F* (1, 18) = 4.68, $p <.05$), further analysis showed that the P3b in Large-level condition (5.93μV) was larger than that in all other conditions (2.11μV, 2.15μV, 3.23μV), and the other three conditions were comparable.

The lateralization effect analysis showed no difference between left and right sides and no interaction between lateralization and other IVs (Figure 18).

Figure 17. Grand average waveforms on syllable across condition in Experiment 4.

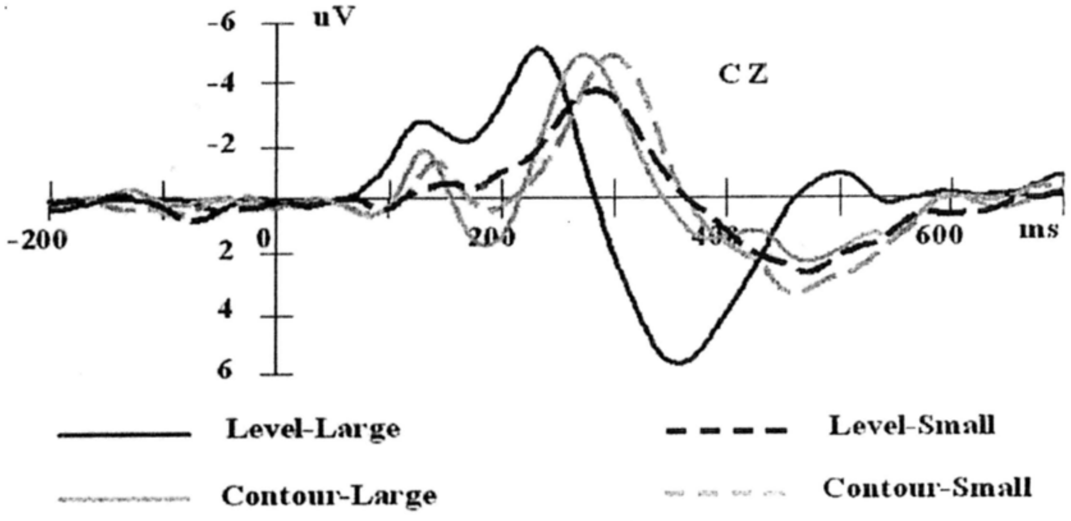
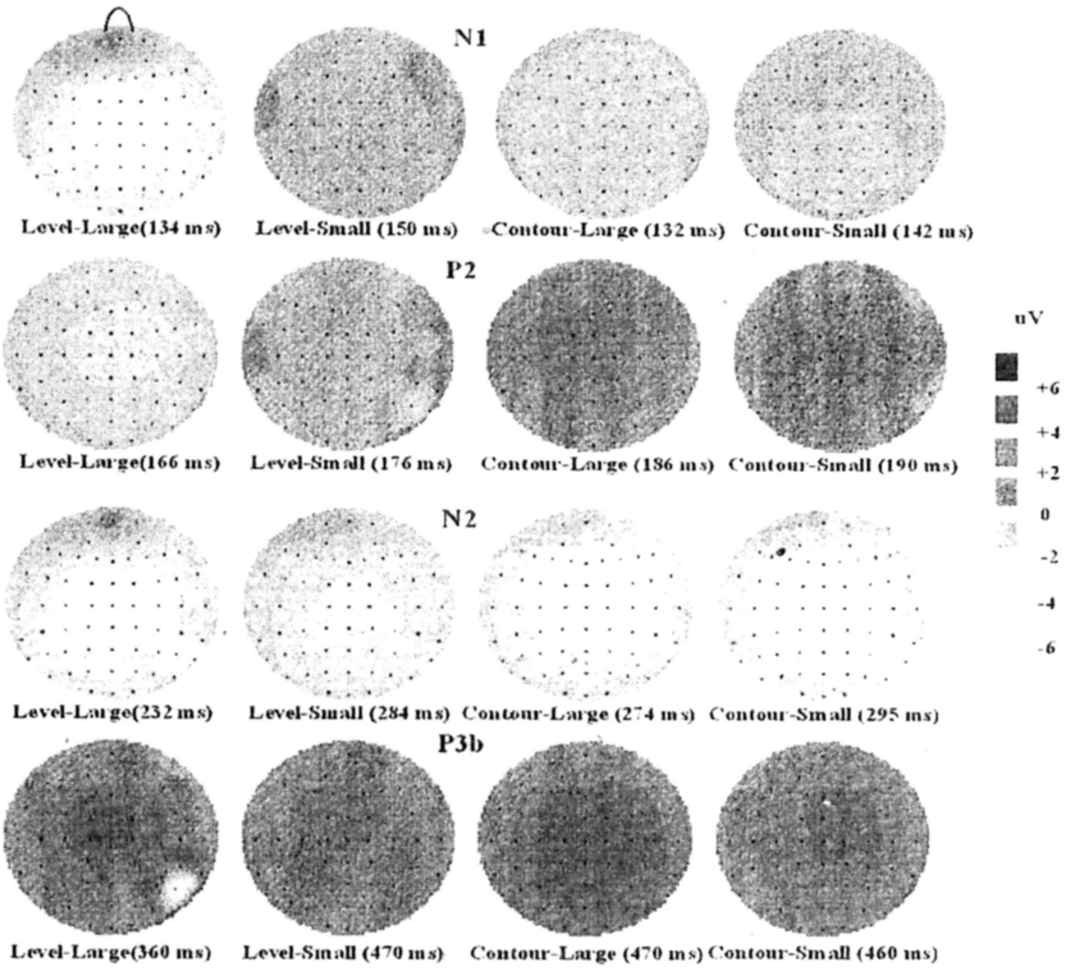


Figure18. Topographies of N1, P2, N2, and P3b on syllable in each condition in Experiment 4.



Part2: hum

Table 11. Latencies and amplitudes of N2 and P3b on hum in each condition in Experiment 4.

	N2			P3b	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
Large-level	70 (5)	218 (9)	- 4.54 (0.54)	350 (11)	5.22 (0.56)
Small-level	80 (6)	250 (9)	- 3.24 (0.48)	402 (11)	1.73 (0.23)
Large-contour	180 (8)	285 (9)	- 3.33 (0.35)	525 (12)	2.42 (0.34)
Small-contour	190 (8)	275 (9)	- 1.75 (0.30)	550 (12)	1.03 (0.21)

For the N2, the onset latency revealed significant main effect of tonal type, (F (1, 18) = 88.6, p <.001), level tone (78 ms) elicited earlier MMN than contour tone did (187 ms).

The N2 peak latency reveled similar main effect of tonal type, (*F* (1, 18) = 36.98, *p* <.001), level tone (234 ms) elicited earlier N2 than contour tone did (280 ms). The main effect of f0 height-deviation was also significant (*F* (1, 18) = 4.68, *p* <.05), large deviation (251.5 ms) elicited earlier N2 than small deviation (262.5 ms) did. The interaction between tonal type and f0 height-deviation magnitude was significant (*F* (1, 18) = 4.01, *p* <.05), and the large deviation (218 ms) elicited earlier N2 relative to small deviation (250 ms) only at level tone.

The N2 amplitude revealed significant main effect of both tonal type (*F* (1, 18) = 4.34, *p* <.05) and f0 height-deviation magnitude (*F* (1, 18) = 4.22, *p* <.05). Level tone (3.93μV) elicited larger N2 than contour tone (2.50μV) did, and large deviation (3.95μV) elicited larger N2 than small deviation (2.51 μV) did.

For the P3b, latency revealed significant main effect of both tonal type (*F* (1, 18) = 108.34, p <.001), and f0 height-deviation magnitude (*F* (1, 18) = 23.75, p <.001). Level tone (375 ms) elicited earlier N2 than contour tone (536 ms) did, and large deviation (435 ms) elicited earlier N2 than small deviation (475 ms) did.

Amplitude revealed significant main effect of both tonal type (*F* (1, 18) = 4.69, p <.05), and f0 height-deviation magnitude (*F* (1, 18) = 8.36, p <.001). Level tone (3.50μV) elicited larger N2 than contour tone (1.70μV) did, and large deviation (3.81μV) elicited larger N2 than small deviation (1.35 μV) did.

No lateralization effect or interaction between other variables and lateralization was found.

Figure 19. Grand average waves on hum in the four conditions in Experiment 4.

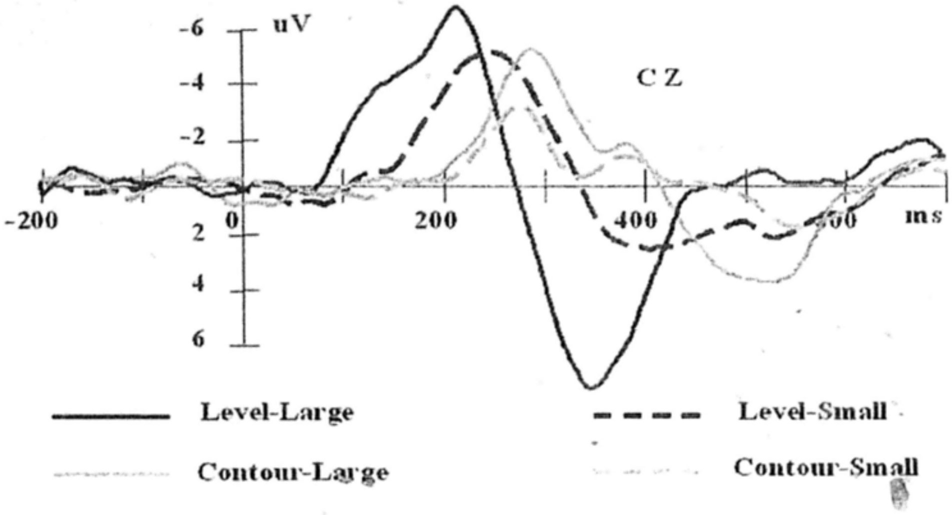
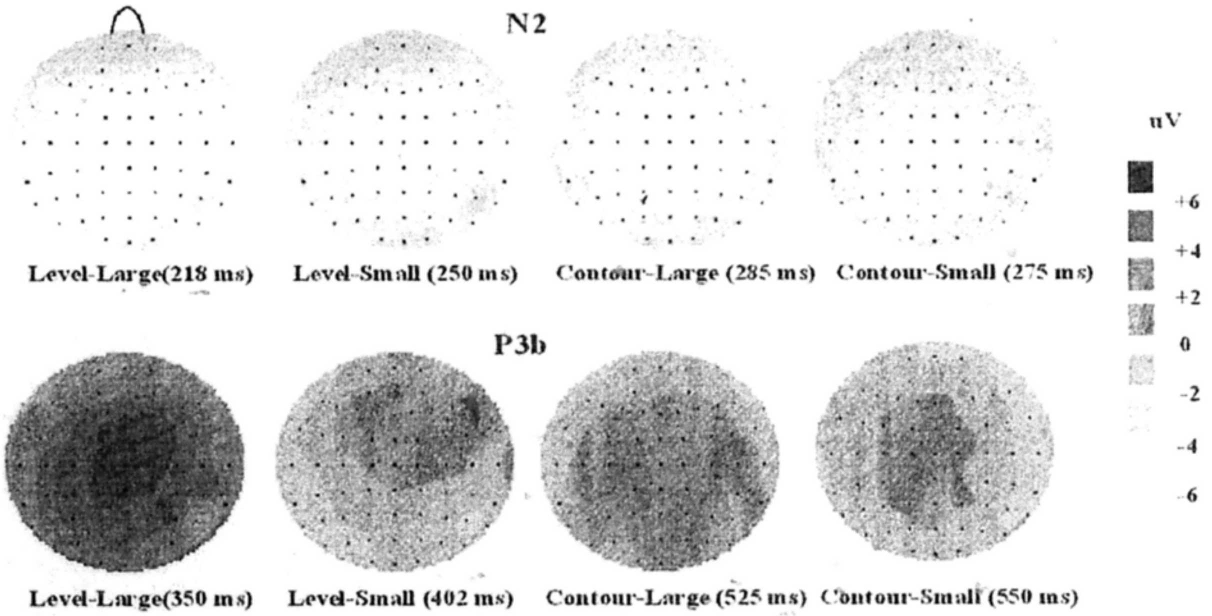


Figure 20. Topographies of N2 and P3b on hum in each condition in experiment 4.



Summary of Experiment 4

Table 12. Main results in Experiment 4 (syllable). " √ " means a significant effect was found.

IV	N1		P2		N2		P3b	
	latency	amplitude	latency	amplitude	latency	amplitude	latency	amplitude
F0 direction			√	√	√		√	√
F0 height	√	√			√		√	√
Lateralization								
Direction × height						√	√	√

Table 13. Main results in Experiment 4 (hum). " √ " means a significant effect was found.

IV	N2			P3b	
	Onset latency	Peak latency	Amplitude	Peak latency	Amplitude
F0 direction	√	√	√	√	√
F0 height		√	√	√	√ ✓
Lateralization					
Direction × height		√			

In the active oddball paradigm, results on syllable and hum were different. On hum, the effect of f0 direction appeared from MMN onset to P3b, and the effect of f0 height appeared from MMN peak stage to P3a stage. These results were similar to that in passive oddball paradigm on both syllable and hum.

But on syllable, the results were quite different. The effect of f0 direction began at the P2 (170 ms) and continued to N2 and P3b, the level tone elicited earlier and larger P2 and N2 than contour tone did, which was consistent with the acoustic lines in Figure 1. The effect of f0 height began from N1 (130 ms) to N2 and P3b, but not on P2. Large deviant elicited larger and earlier N1 and N2 than small deviant. In Figure 1, the tone 2 and tone 5 are distinguished at initial section (corresponding to N1) and are mixed at about 120 ms (corresponding to P2) and then separate again from turning point (corresponding to N2). Thus, the brain responses to contour tone perfectly coincide with their pitch line, which denotes the effect of acoustic feature.

Compare the results on syllable to that on hum, we can know that segment can facilitate the processing of lexical tone. Lexical tones are processed earlier on syllable than on hum, especially for the f0 height dimension. On one hand, these results indicate that the linguistic memory is important to tonal processing. On the other hand, these results again suggest that tone and segment are interactive rather than separate in speech perception. And the integration of tone and segment begins from about 130 ms in this experiment.

Chapter 8

General Discussion

In order to test how tonal dimension, stimulus type, and task requirement affect tonal processing, four experiments were conducted on Cantonese tone. Experiments 1 and 2 were DL experiments. They manipulated the same IVs: 1) f0 direction (level or contour) and 2) stimulus type (syllable, pseudo-syllable, and hum), but used different tasks (a discrimination task in Experiment 1and an identification task in Experiment 2); Experiments 3 and 4 were ERP experiments. They manipulated the same IVs: 1) f0 direction (level or contour), 2) f0 height-deviation (large or small), and 3) stimulus type (syllable or hum), but the tasks in them were different (passive oddball task in Experiment 3 and active oddball task in Experiment 4).

In the following sections, the main results of this study, including the effects of tonal dimension, stimulus type, lateralization, and task requirement are presented one by one (the main results of the four experiments are also summed up in Table 3 (Experiment 1), Table 4 (Experiment 2), Table 9 (Experiment 3), Table 12 (Experiment 4), and Table 13 (Experiment 4)). Following that, the implications of these results with regard to the theoretical issues are presented. Finally, some potential directions for future research are introduced.

8.1 Main results

Firstly, tonal dimension showed main effects in both the DL and the ERP experiments. In the DL experiments, the performance for level tones was much better than that for contour tones, which were indicated by the higher accuracy (both discrimination and identification tasks) and shorter RT (discrimination task) for level tones relative to those of contour tones. The effect of tonal dimension was also demonstrated in the interaction between the tonal type (f0 direction) and the lateralization effect. In the discrimination task, the accuracy analysis indicated that the LEA appeared for contour tones only. The null lateralization effect for level tones might have been caused by a ceiling effect. The significant LEA for level tones in the RT analysis also supported this conjecture. Additionally, in the identification task, the LEA appeared only for level tones, which suggests that the level tones are more easily received according to their acoustic properties than the contour tones.

In the ERP experiments, both tonal dimensions (f0 direction and f0 height) were tested. Extensive effects were found for both of them. For *f0 direction*, in the passive task, level tones elicited earlier and larger MMN and earlier P3a than contour tones did for both syllables and hums. In the active task for syllables, level tones elicited earlier and larger P2 than contour tones did. The P3b for level tones was larger and earlier than that of contour tones only in the f0 large-deviantion condition. In the active task for hums, level tones elicited earlier and larger N2 and P3b than contour tones did. For *f0 height,* in the passive task, a significant effect only existed for level tones, such that the large f0 height-deviation condition elicited larger and earlier MMN than the small f0 height-deviation condition did. However, the two contour-related waveforms nearly coincided with each other for both syllables and hums. In the active task for syllables, the large f0 height-deviation elicited larger and earlier N1 than the small f0 height-deviation did. Also, the large f0 height-deviation elicited larger and earlier P3b than the small f0 height-deviation did only for level tones. For hums, the large f0 height-deviation elicited larger and earlier N2 and P3b than the small f0 height-deviation did.

Secondly, an effect for stimulus type was found. In the discrimination DL, performance on syllables and pseudo-syllables was comparable, and performance on both of them was better than that on hums. In the passive oddball paradigm of the ERP experiment, MMN and P3a elicited in the syllable condition was larger than that of the hum condition. In the active oddball paradigm, N1 and P2 were elicited only in the syllable condition, indicating that tones were processed faster in the syllable condition than in the hum condition. The results of the DL experiment and ERP experiment were generally consistent.

Thirdly, the lateralization effect was only found in the DL experiments. In the discrimination DL, the LEA was significant on both accuracy (null results at level tones might have been due to ceiling effects) and RT. Moreover, the LEA was not modulated by the stimulus type or the tonal type. In the identification DL, an LEA was reflected by a larger report proportion in the left ear than in the right ear, and the LEA for hums seemed greater than that for syllables and pseudo-syllables, though it was not statistically significant. However, the LEA was modulated by the tonal type in the identification of DL, and the LEA only appeared for level tones.

The LEA of Cantonese tone processing in the discrimination tasks was consistent with the LEA in pure tones in similar tasks (Brancucci, Babiloni et al., 2005; Brancucci, Anselmo et al., 2008). These results suggest that the acoustic feature is crucial for tonal discrimination. In the identification DL, the LEA only existed for level tones but not for contour tones, which can be explained by the Categorical Perception study on Cantonese tone (Francis, Ciocca, & Ng, 2003).

Task requirement also affected tonal processing. In the DL experiments, both the discrimination and identification tasks were used. The results showed that task requirement modulated the role of the tonal type. The tonal type influenced the ear advantage only in the identification DL. In the ERP experiments, both the active (discrimination) and passive (null task) oddball paradigms were employed. The task effect was not significant for hums. However, for syllables, N1 and P2 were elicited in the active rather than in the passive task. In summary, task requirement influences tonal processing. Moreover, in the ERP experiments, the task-requirement effect was more important for syllables than for hums.

Finally, this study suggests that different techniques provide different data. In order to understand a research question more sufficiently, multiple paradigms and/or techniques are necessary. In this study, the lateralization patterns were examined by both DL and ERP methods. Results showed that the DL seems more sensitive than the ERP method in lateralization testing. On the other side, the ERP method has higher temporal resolution and can provide latency and amplitude information through ERP components, which the DL cannot. Combining the advantages of different paradigms and technique is necessary in future studies.

The main findings are listed in above paragraphs. In the following sections, these findings are discussed in terms of the theoretical issues to determine whether this study can help us understand them. Five topics are discussed: 1) the effects of tonal dimension, 2) the debate centering on the acoustic versus the functional model, 3) the ACS framework, 4) the relationship between segment and tone, and 5) possible future directions.

8.2 The effects of tonal dimensions

Originally, tonal dimension was a research topic in phonetic studies, the aim of which was to use simplified dimensions to distinguish tones of all tonal languages. Gandour (1983) studied the tonal dimensions of Thai, Cantonese, Mandarin, and Taiwanese, and revealed the tonal dimensions of f0 height and f0 direction that underline 19 tonal patterns in these four languages. In a recent study, Khouw and Ciocca (2007) found that f0 height and f0 direction can be used to distinguish all tones in Cantonese.

In addition, one ERP study was done to explore the dimensional processing of Mandarin tone (Chandrasekaran, Gandour, & Krishnan, 2007). By using the passive oddball paradigm (standard/deviant), level-contour pairs (T1/T2, T1/T3) and contour-contour pairs (T2/T3) were presented to native Mandarin speakers and English speakers who were Chinese naïve. The MMN were analyzed using the multidimensional scaling method (MDS). F0 height and f0 direction were revealed, and the f0 direction was more important to native Mandarin speakers than the f0 height dimension.

However, the two dimensions have usually been mixed together in past studies[11]. This study manipulated the two dimensions in Cantonese, studying the effects of them on Cantonese tone processing using DL and ERP methods. The DL experiments showed that the main effect of f0 direction was significant, such that performance at level tones was better than that at contour tones. This performance differences can be confirmed by the effect of f0 direction in the ERP experiments, in which level tones elicited earlier and larger MMN, P3a, and N2 than contour tones did. These results were also consistent with past studies that found that contour tones were more difficult to distinguish than level tones in Cantonese (Gandour, 1983; Khouw & Ciocca, 2007).

In the passive oddball paradigm, f0 height dimension affected the processing of level tones but not contour tones. The large height-deviation elicited earlier and larger MMN than the small height-deviation for level tones only. The two contour waveforms nearly coincided with each other. However, in the active oddball paradigm, for syllables, the effect of f0 height appeared as early as from the N1 stage. The large height-deviation elicited larger and earlier N1 than the small height-deviation did. Additionally, results indicated that the effect of f0 height was modulated by task requirement and stimulus type. In the active task, for syllables, the effect of f0 height (N1) emerged earlier than the effect of f0 direction (P2) did, though the effect of f0 height was relatively later and smaller than the effect of f0 direction in the passive oddball paradigm and in the active task for hums.

---

[11] In another ERP experiment (Tsang, Jia, Huang, & Chen, 2011), level and contour tones in Cantonese were studied separately. The results showed that the size and latency of MMN were sensitive to the level tones, and the latency of P3a was sensitive to the contour tones.

In short, this study (especially the ERP experiments) revealed the psychological reality of tonal dimensions. Different dimensions have different effects on tonal perception. The processing time courses of them are different. Stimulus type and task requirement affect the processing of two dimensions differently. In tonal processing studies, the two dimensions should be manipulated separately.

8.3 Functional versus acoustic models

Model debates in speech perception mainly stem from the debates over the *functional* hypothesis vs. the *acoustic* hypothesis. Given that tonal information processing was primarily studied in the field of speech perception, theoretical disputations of tonal processing also primarily stem from the debates over the functional vs. the acoustic models.

For tonal information processing, the predictions of the two models mainly diverge at two aspects. Firstly, which information is most important for determining tonal processing, the function of lexical tone or the acoustic features of lexical tone? The functional model (Liberman & Mattingly, 1985; Liberman, 1996; Liberman & Whalen, 2000) posits that it is the linguistic function rather than the acoustic feature that determines how tone is processed in the brain. By contrast, the acoustic model (Zatorre, Belin, & Penhune, 2002; Poeppel, 2003) claims that it is not the function but rather the acoustic features that are crucial. Secondly, the two models predict different neural mechanisms for tonal processing. According to the functional model, lexical tone is preferentially processed by the language-related cortices in the left hemisphere because tone is phonetically contrastive. However, according to the acoustic model, lexical tone is predominately processed by the supra-temporal cortices in the right hemisphere because tone is a fine frequency-varying signal.

Therefore, resolving the debates over the two models can be focused on the two divergences of their predictions. For instance, the variation of acoustic features and linguistic-involvement levels can be manipulated. Thus, testing the two models can be translated into testing whether the acoustic features or the linguistic involvement modulates tonal processing more. If the acoustic features play a more important role than linguistic involvement does, the acoustic model would be supported, and vice versa. Likewise, testing the two models can also be translated into testing the lateralization of tonal processing. If the left hemisphere advantage is found, it means that the functional model is correct in that condition. By contrast, if the right hemisphere advantage is found, the acoustic model is acceptable.

In this study, the lateralization pattern of tonal processing was tested with two commonly used techniques of perceptual lateralization studies: the DL task and the ERP method. A null lateralization effect was found in the ERP experiments, which was reflected by both the ANOVA results and the topographies. In the DL experiments, the LEA was found for Cantonese tone processing in both discrimination and identification tasks, which were consistent with the acoustic hypothesis in this aspect.

For the second divergence, the study provided ample evidence to support both models. For the functional model, in the DL experiments, the main effect of lexical level was significant, such that performance on syllables and pseudo-syllables was much better than that on hums, which means that, although the acoustic features determined lateralization in the DL of Cantonese tone processing, the language experience also has an important effect. In the passive oddball ERP experiment, MMN and P3a elicited by syllables were much larger than those elicited by hums, which was consistent with the better performance on syllables than hums in the DL task. Furthermore, in the active oddball task, an evident N1 and P2 appeared in the syllable condition rather than in the hum condition. Together with the MMN effect of stimulus type in Experiment 3, it seems that the effect of lexical memory in tonal processing begins at about 100- 200 ms after syllable onset. For the acoustic model, level tones generally elicited earlier and larger MMN and N2 than contour tones did, which coincides with the pitch lines in Figure 1. Additionally, the larger height-deviation evoked larger and earlier MMN and N2 than the small height-deviation did, which suggest a role for acoustic features in tonal processing

8.4 The ACS framework

The above section discusses the debates over the functional vs. the acoustic models, revealing that both models can be supported by selective evidence, but neither one can explain all of the findings. Actually, there are a large body of studies, which were described in the introduction, that have selectively supported the two models (for a review, please refer to Zatorre & Gandout 2007; Wong, 2002). It seems that both acoustic features and the linguistic memory in the brain are substantially important to tonal processing.

Recently, scholars have proposed that speech perception and general auditory information processing are unnecessarily studied separately. On one hand, speech perception must begin with auditory information processing, although most current models of speech perception assume that speech perception deals with phonemes rather than sound waves (Holt & Lotto, 2008). On the other hand, speech may be represented as stable category memory in the brain, which can be used as typical sample in general audition studies. Therefore, the cross-fertilization between the two fields provides new approaches to studying speech perception and general audition (Holt & Lotto, 2008; Diehl et al., 2004). One new approach is termed the ACS framework (similar to the General Approach in Diehl, 2004).

According to the ACS framework, speech perception and general auditory information processing involve similar processing mechanisms and neural substrates. Both acoustic features and linguistic memory are important to auditory processing. This framework emphasizes the interactivity of ear-inputted acoustic signals and memory of the signals in the brain. In terms of tone, the processing of tone is the interactivity between f0 patterns from the cochlea and the memory of tonal patterns in the brain. Some factors (task requirement, lexical context) can modulate the effects of acoustic features and lexical memory.

In this study, in order to get a complete depiction of tonal processing, both tonal dimensions and stimulus type were manipulated. The influence of lexical memory and acoustic properties on tonal processing was found. These results cannot be explained by the functional or acoustic models individually, but are in line with the ACS framework.

Furthermore, the ACS explains language phenomena, especially the phenomenon of categorical perception, by using the *learning effect*. According to the learning effect, it is the learning procedure (intense practice) rather than the learning object (language) that determines the manner of perception. Learning procedure can be broadly divided into two stages: novice and expertise. In the novice stage, a signal is received based on its acoustic features, and tone is processed as a continuous f0 pattern by the right hemisphere. In the expertise stage, sound signals are processed as categories. In this stage, tone is processed as lexical contrast in the left hemisphere. Native speakers are experts of their native language. Thus, the left hemisphere advantage is usually found in language processing.

The learning effect has been supported by a study on pure tone perception (Brancucci, di Nuzzoa, & Tommasia, 2009). In that study, using the DL paradigm, a pitch identification task was conducted on two types of participants: a group with absolute pitch (AP) and a group without absolute pitch (NAP). Results showed opposite hemispheric patterns for them. The AP group showed a left hemispheric advantage, whereas the NAP group showed a right hemispheric advantage. The authors explained the results in terms of the learning effect[12]. The AP and NAP participants were at the expertise and novice stages, respectively Thus, their respective means of processing were different. These results suggest that, for auditory information processing, the processing manner is determined by the learning

---

[12] These results can also be explained by the familiarity effect (Brancucci, di Nuzzoa, & Tommasia, 2009).

procedure rather than by the processing object.

8.5 Segment and Tone

Speech perception begins with identifying the fundamental elements in speech sounds, including segmental information (consonant and vowel) and supra-segmental information (tone) in tonal languages. Those fundamental elements are integrated together to yield the meaning of syllables (Diehl, Lotto, & Holt, 2004). Thus, except for the processing manner of lexical tones themselves, the relation of tone and segment, especially how tone and segment are combined during speech perception, is also important in tonal languages.

One viewpoint is that tone and segment are separately processed (Ye & Connine, 1999), which is consistent with the view from both linguistics (Selkirk, 1986) and psycholinguistic (Speer, Shi, & Slowiaczek, 1989) studies. Similarly, in speech production, the *autosegmental* view proposes that tone has its unique representation separate from segmental units (Goldsmith, 1976).

However, two recent studies have shown that tone and segment are not separate; rather, they integrated during speech perception. Using the Garner speeded classification paradigm (Garner, 1976), researchers examined the reciprocal interferences between tone and segment in target dimension (tone, consonant, or vowel) classification tasks. Results indicated that segment and tone interfered with each other in the classification task (Repp & Lin1990), with the segmental information interfering more with tone classification than the reverse (Tong, Francis,

& Gandour, 2008). Moreover, the integrality of vowel and tone was stronger than that of the other pairs. These results suggest an interactive relationship between tone and segment.

In this study, the effect of segment on tonal processing was also examined. In the DL tasks, tonal patterns were conveyed by syllables, pseudo-syllables, and hums. In both syllables and pseudo-syllables, tone is realized by pronounceable sounds, which sound naturally according to phonetic memory. For hums, in which all segmental information was removed, these sounds do not exist in the memory and sound unnaturally. Results showed that performance on syllables and pseudo-syllables was better than that on hums, which means that language experience influenced tonal processing. Thus, the lexical memory of segment can facilitate tonal processing.

In the ERP experiments, tones realized on syllables evoked larger MMN and P3a than tones realized on hums in the passive oddball paradigm, indicating that brain activity was larger when processing tone-segment combinations than when processing tone alone (Luck, 2005). The more mental resource involvement in the syllable condition can explain the better performance in the syllable condition than in the hum conditions (e.g., experiment 1). In the active oddball task, except for the N2 and P3b, which were also evoked in the hum condition, the N1 and P2were also evoked in the syllable condition, which means that the tones realized on syllables can be earlier processed than tones realized on hums.

In addition, segmental information modulated the effects of the two tonal dimensions. The effect of f0 direction appeared earlier than that of f0 height for hums, which means that the average f0 level is processed later than is the f0 direction in hum condition. However, for syllables, the effect of f0 height and f0 direction began at the N1 (~130 ms) and P2 stages (~170 ms), respectively, which means that the f0 height was processed earlier than the f0 direction when tones were conveyed by syllables.

In summary, in speech perception of tonal languages, tone and segment are processed interactively rather than independently. Segmental information can facilitate tonal information processing in a discrimination DL task (Fox & Unkefer, 1985) and in both passive and active oddball ERP tasks. In the active oddball paradigm, for syllables, the interaction of segment and tone begins at about 130 ms.

8.6 Future directions

Firstly, one interesting finding in this study is the relationship between segment and tone. Results showed that segment can facilitate tone perception in both DL and ERP tasks, with the facilitation effect displayed in both behavioral performance and brain responses. In future studies, two issues can be further tested. The first one concerns the two kinds of segments, the consonant and the vowel, which one modulates tone perception more? The consonant is the initial part of a syllable and appears earlier than the vowel does. However, tone is mainly realized on the vowel. Thus, how the two kinds of segments affect tone perception, and which one modulates tone processing more than the other could be tested in the future. The second question concerns the effect of tone on segment processing. How and when tone affect segmental information processing is still not clear. In future studies, a complete map of the relation of tone and segment during speech perception should be constructed.

Secondly, the materials in this study were either mono-syllables or hums. Results showed effects of both acoustic features and linguistic memory on tone perception in both DL and ERP experiments but failed to find a left hemisphere advantage for Cantonese tone processing. The Introduction introduced the notion that processing speed and phonological role can be modulated by the lexical context and the task requirement. In order to further test how tone helps to extract meaning from speech, bisyllables or idioms could be used in the future. In addition, tasks that are more like natural speech perception should be developed.

Thirdly, in the discussion of the ACS framework, the learning effect was introduced, which can be used to explain some findings on speech perception and general auditory information processing. In order to examine the ACS framework, the learning effect can be tested in speech learning, musical sound learning, and general audition learning, to determine whether the learning stage or the learning object is crucial to the manner of processing.

8.6 Conclusions

Using two DL experiments and two ERP experiments, this paper documents a study on the effects of tonal dimension and stimulus type on lexical tone processing, specifically on Cantonese tone processing.

For the tonal dimension of the f0 direction, results showed that responses to level tones was faster than that to contour tones, and level tones elicited earlier and larger ERP components than contour tones did. For the tonal dimension of the f0 height, the results showed that a large f0 height-deviation elicited earlier and larger ERP components than a small f0 height-deviation did. The effect of the f0 direction appeared earlier than the effect of the f0 height. Thus, tonal dimensions play an important role in tonal processing. They should be considered in future studies.

For the stimulus type, in DL experiments on Cantonese tone, performance on syllables and pseudo-syllables was better than that on hums. In Experiment 3, syllables elicited larger MMN and P3a than hums did. In Experiment 4, the N2 and P3b were elicited by both syllables and hums, but the N1 and P2 appeared only in the syllable condition. Thus, lexical status can modulate tonal processing from N1 stage.

The effect of tonal dimensions indicates the role of acoustic features on tonal processing, and the effect of stimulus type reflects the role of linguistic memory on tonal processing. In addition, the LEA in the DL experiment on Cantonese tone also suggests the crucial role of acoustic features in tonal processing. Thus, both low-level acoustic features and high-level memory are important to tonal processing. Neither the *acoustic model*, which emphasizes the acoustic features only, nor the *functional model*, which emphasizes the linguistic function only, can explain all of the results in this study. The *ACS framework*, which emphasizes the equal roles of acoustic features and linguistic function, is more appropriate in this study.

In terms of the relation of tone and segment, this study suggests that tone and segment are interactively, rather than independently, processed during speech perception. The integration of the two appears within 100-200 ms after syllable onset.

References

Abramson, A. S. (1979). Thai tones as a reference system. In W. Gething, G. Harris, & P. Kullavanijaya (Eds.), *Tai linguistics in honor of Fang-Kuei Li.* 1-12. Bangkok: Chulalongkorn University Press.

Baudoin-Chial, S. (1986). Hemispheric lateralization of Modern standard Chinese tone processing. *Journal of Neurolinguistics, 2,* 189–199.

Benson, P. J., Smith, T. S., & Arreaga, L. (1973). Dichotic perception pf lexical tone by speakers of Cantonese and English. *The Journal of the Acoustical Society of America 53 (1),* 346-346.

Brancucci, A., Babiloni, C., Babiloni, F., Galderisi, S., Mucci, A., & Tecchio, F., et al. (2004). Inhibition of auditory cortical responses to ipsilateral stimuli during dichotic listening: Evidence from magnetoencephalography. *European Journal of Neuroscience, 19(8),* 2329–2336.

Brancucci, A., Babiloni, C., Rossini, P. M., & Romani, G. L. (2005). Right hemisphere specialization for intensity discrimination of musical and speech sounds. *Neuropsychologia, 43,* 1916–1923.

Brancucci, A., Anselmo, AD., Martello, F., & Tommasi, T. (2008). Left hemisphere specialization for duration discrimination of musical and speech sounds. *Neuropsychologia, 46,* 2013–2019.

Brancucci, A., di Nuzzoa, M., & Tommasia, L. (2009). Opposite hemispheric asymmetries for pitch identification in absolute pitch and non-absolute pitch musicians. *Neuropsychologia, 47 (13),* 2937-2941

Broadbent, D. E. (1954). The role of auditory localization in attention and memory span. *Journal of Experimental Psychology, 47,* 191–196.

Bryden, M. P., (1986). Dichotic listening performance, cognitive ability, and cerebral organization. *Canadian Journal of Psychology, 40,* 445-456.

Chandrasekaran, B., Gandour, J. T., & Krishnan, A. (2007). Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity. *Restorative Neurology and Neuroscience, 25,* 195-210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*, 148-156.

Chen, H. C., & Zhou, X. (1999). Processing East Asian languages: An introduction. *Language and Cognitive Processes, 14*, 425-428.

Chen, H. C. (2006). Language processing: Implications from the study of Chinese. In Q.C. Jing, M.R. Rosenweig, G. d'Ydewalle, H.C. Zhang, H. C. Chen, & K. Zhang (Eds.), *Progress in psychological science around the world: Neural, cognitive, and developmental issues*, 367-381. London: Psychology Press.

Chinese Character Database: With Word-formations. (2003). Retrieved on September 20, 2010, from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale, NJ: Erlbaum.

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

Cutler, A. (1986). Forbear is a homophone: lexical prosody does not constrain lexical access. *Language and Speech, 29,* 201-220.

Diehl, R. L., Lotto, A. J. & Holt, L. L. (2004). Speech perception. *Annual Review of Psychology, 55,* 149–179.

Fok, C. Y. Y. (1974). *A perceptual study of tones in Cantonese (Occasional Papers and Monographs No. 18).* Hong Kong: University of Hong Kong, Centre of Asian Studies.

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics, 13,* 69–90.

Francis, A.L., Ciocca, V.C., & Ng, B.K.C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65(6),* 1029-1044.

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

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of Phonetics, 11,* 149–175.

Gandour, J., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y. & Li, X. (2002). A crosslinguistic FMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience. 14,* 1076–1087.

Gandour, J., Tong, Y., Wong, D., Talavage, T., Dzemidzic, M., Xu, Y., Li, X., & Lowe, M. (2004). Hemispheric roles in the perception of speech prosody. *Neuroimage, 23 (1),* 344-57.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of*

Experimental Psychology: Human Perception and Performance, 6, 110–125.

Garner, W. R. (1976). Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology, 8,* 98-123.

Goldsmith, J. (1976). An overview of autosegmental phonology. *Linguistic Analysis, 2,* 23-68.

Goodglass, H., & Calderon, M. (1977). Parallel processing of verbal and musical stimuli in right and left hemispheres. *Neuropsychologia, 15,* 397.

Holt, L. L., & Lotto, A. J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Sciences, 17,* 42-46.

Hugdahl K. (1995). Dichotic listening: probing temporal lobe functional integrity, in: Davidson R. J., Hugdahl K. (Eds.), *Brain Asymmetry,* MIT Press, Cambridge, MA, 123– 156.

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

Kimura, D. (1967). Functional asymmetry of the brain in dichotic listening, *Cortex, 3,* 163– 168.

Ladefoged, P. (2001). *A course in phonetics,* 4th edition, Orlando: Harcourt Brace.

Liberman A. M., Harris K. S., Hoffman H. S., Griffith B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology. 54,* 358–68

Liberman, A. M., & Mattingly, I. G.. (1985). The motor theory of speech perception revised. *Cognition 21,* 1–36.

Liberman, A. M. (1996). *Speech: A Special Code.* Cambridge, MA: The MIT Press.

Liberman, A. M., & Whalen, D. H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences, 4 (5),* 1471-6.

Liu, S., & Samuel, A.G. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes, 22,* 566-594.

Luck S. J. (2005). *An Introduction to the Event-Related Potential Technique.* Cambridge, Mass.: The MIT Press.

Luo, H., Ni J-T., Li Z-H., Li X-O., Zhang D-R., Zeng F-G., & Chen L. (2006). Opposite patterns of hemisphere dominance for early auditory processing of lexical tones and consonants. *Proceedings of the National Academy of Sciences. 103 (51),* 19558–19563

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62(4),* 407-420.

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22,* 1376–1392.

McClelland, J. L., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology, 23*, 1-44.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology:*

Human Perception and Performance, 17, 433–443.

Moen, I. (1993). Functional lateralization of the perception of Norwegian word tones—Evidence from a dichotic listening experiment. *Brain and Language, 44,* 400–413.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, .M., Huotilainen, M., Iivonen, A., et al., (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature 385,* 432–434.

Näätänen, R., (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology. 38 (1),* 1 – 21

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9,* 97–113.

Piazza, D.M., (1980). The influence of sex and handedness in the hemispheric specialization of verbal and nonverbal tasks, *Neuropsychologia. 18,* 163–176.

Pike, K. (1948). *Tone languages.* Ann Harbour: University of Michigan Press.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118(10),* 2128-2148.

Poeppel, D. (2003). The analysis of speech in different temporal integration windows: cerebral lateralization as 'asymmetric sampling in time'. *Speech Communication. 41,* 245–255.

Ren, G. Q., Yang, Y. F., Li, X. Q. (2009). Early cortical processing of linguistic pitch patterns as revealed by the mismatch negativity. *Neuroscience, 162,* 87-95.

Repp, B. H., & Lin, H. B. (1990). Integration of segmental and tonal information in speech perception: A cross-linguistic study. *Journal of Phonetics, 18(4),* 481-495.

Schirmer, A., Tang, S. L., Penney, T. B., Gunter, T. C., & Chen, H. C. (2005). Brain responses to segmentally and tonally induced semantic violations in Cantonese. *Journal of Cognitive Neuroscience, 17,* 1–12.

Selkirk, E. O. (1986). *Phonology and syntax*. MIT Press, Cambridge, MA.

Smith, T. S., & Shand, M. (1974). Dichotic perception of initial consonants and lexical tone in Cantonese. *The Journal of the Acoustical Society of America. 55 (s1).* 11-11.

Speer, S., Shi, C. & Slowiaczek, M. L. (1989). Prosodic structure in language understanding: Evidence from tone sandhi in Mandarin. *Language and Speech 32,* 337-354.

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. In H. C. Chen, & O. Tzeng (Eds.) *Language processing in Chinese.* Amsterdam: North-Holland/Elsevier

Tong, Y., Francis, A., & Gandour, J.T. (2008). Processing dependencies between segmental and suprasegmental features of Mandarin Chinese. *Language and Cognitive Processes, 23(5),* 689-708.

Tsang Y. K., Jia S., Huang J., & Chen H. C. (2011). ERP correlates of pre-attentive processing of Cantonese lexical tones: The effects of pitch contour and pitch height. *Neuroscience Letters. 487 (3),* 268-272.

Van Lancker, D., & Fromkin, V. A. (1973). Hemispheric specialization for pitch and "tone": Evidence from Thai. *Journal of Phonetics, 1,* 101–109.

Van Lancker, D. & Fromkin, V. A. (1978). Cerebral dominance for pitch contrasts in tone language speakers and in musically untrained and trained English speakers. *Journal of Phonetics, 6,* 19–23.

Van Lancker, D. (1980). Cerebral lateralization of pitch cues in the linguistic signal. *Papers in Linguistics: International Journal of Human Communication, 13 (2),* 201-277.

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

Wang, W. (1976). Language change. In S. R. Harnad, H. D. Steklis, & J. Lancaster (Eds.), *Origins and evolution of language and speech (Annals of the New York Academy of Sciences, 280,* 61-72. New York: New York Academy of Sciences.

Wang. Y., Jongman, A., & Sereno, J. A. (2001) Dichotic perception of Mandarine Tones by Chinese and American Listeners. *Brain and Language ,78:* 332-348.

Wang, Y., Behne, D., Jongman, A., and Sereno, J.A. (2004). The role of linguistic experience in the hemispheric processing of lexical tone. *Journal of Applied Psycholinguistics, 25,* 449-466.

Wong, P. C. (2002). Hemispheric specialization of linguistic pitch patterns. *Brain Research Bulletin, 59,* 83–95.

Wong, P. C., Parsons, L. M., Martinez, M. & Diehl, R. L. (2004). The role of the insular cortex in pitch pattern perception: the effect of linguistic contexts. *The Journal of Neuroscience, 24,* 9153–9160.

Xi, J., Zhang, L., Shu, H., Zhang, Y., & Li, P. (2010). Categorical Perception of Lexical Tones in Chinese Revealed by Mismatch Negativity (MMN), *Neuroscience, 170 (1),* 223-231.

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

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes, 14*, 609-630.

Yip, M. (2003). *Tone.* Cambridge University Press, New York.

Zatorre, R. J., & Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cerebral Cortex, 11,* 946–953.

Zatorre, R. J., Belin, P. & Penhune, V. B. (2002). Structure and function of auditory cortex: music and speech. *Trends in Cognitive Science. 6,* 37–46.

Zatorre, R.J. & Gandour, J.T. (2007). Neural specializations for speech and pitch: moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B,* 1-17.