Categorical Perception of Lexical Tones: Behavioral and Psychophysiological Study

ZHENG, Hongying

A Thesis Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

in

Electronic Engineering

The Chinese University of Hong Kong

April 2010

To my parents,

Tianhe ZHENG and Meiyu CHEN,

To my husband,

Jianyong, CHEN

for their love, consideration, and support.

ACKNOWLEDGEMENTS

This thesis is a synthesis of the research during my PhD study at both City University of Hong Kong (CityU) and The Chinese University of Hong Kong (CUHK). The most important person in my academic life is my supervisor, Prof. William Shi-Yuan WANG. All the words in the world do not suffice to thank him. His special wisdom led me into the beautiful realm of linguistics. His comprehensive lectures on speech perception inspired me to choose the topic on perception on lexical tones. He has guided me to learn how to appreciate the beauty and cherish the joy of doing research in basic science, and his encouragement has been all along supporting me to overcome various difficulties in research. He encouraged me to go into the new, exciting area of neuroscience. Without his recommendation and support, I would never have the chance to attend several international conferences and summer schools during the past four years or have the chance to visit the University of California, Berkeley during the summer in 2008.

I would like to extend my gratitude to Prof. Peter WM TSANG, who acted as my co-supervisor during my study in CityU, and sincerely supported me to continue my research when Prof. WANG leaves CityU. Especial thanks should be given to Prof. Richard IVRY and Prof. Paul KAY, who kindly hosted me when I visited the University of California, Berkeley. The working experience with them enriched my knowledge about psychology. More importantly, they helped me to raise the ability to do research work scientifically, critically, and creatively. I also want to thank Prof. Thomas Hun-Tak LEE, whose great knowledge about language acquisition and critical comments on my work are valuable for me to reconsider the experimental design.

I own a lot to the technical support and moral encouragement given by all the members of LEL. It was my great honor to collaborate with Dr. James W. MINETT and Dr. Gang PENG. Their rich experiences and critical challenge greatly inspired me to fulfill my research. I am grateful to Dr. Jin-Yun KE, Dr. Tao GONG, Dr. Francis Chun-Kit WONG, and Ms. Yao YAO to help me to improve my awkward English and to reorganize my paper. In addition, I am especially grateful to Dr. Tao GONG, Dr. Ching-Pong AU and Ms. Ruo-Xiao YANG who kindly helped me to collect some field work data presented in this thesis. Furthermore, I enjoyed a lot discussion with my other colleagues, Dr. Susan Lan SHUAI, Mr. Ying-Wai WONG, Mr. Raymond Wai-Man Ng, Mr. Guang-Ting MAI, Mr. Manson

FONG, Ms. Lin ZHOU, and Ms. Hui CHEN on various research topics. I am very fortunate to have these close amigos in my life.

In the end, I would like to thank the members in a bigger family, the Digital Signal Processing and Speech Technology Laboratory (DSPSTL). The leaders of this laboratory, Prof. Pak-Chung CHING and Prof. Tan LEE have generously supported me to perform my basic science research in this engineering laboratory. And I enjoyed working, growing, and improving my skills with the colleagues in DSPSTL, Dr. Meng YUAN, Ms. Yu-Jia LI, Dr. Neng-Heng ZHENG, Ms. Ning WANG, Dr. Feng TIAN, Ms. Hou-Wei CAO, Mr. Jia-Xian PAN, Mr. Li QIANG, Dr. Yvone Siu-Wa LI, and Ms. Natalie Li ZENG. Without these friends, my PhD life in CUHK would have not been so interesting.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Speech sounds vary across different conditions and subjects; nevertheless, listeners perceive the phonemes without difficulties. *Categorical perception* (CP) occurs when listeners map the varying speech sounds into discrete phonemic categories. In CP, to discriminate a pair of stimuli that cross a category boundary is much easier than those that lie within the same category, even though both pairs are separated by an equal physical difference. CP is one of the important properties essential for speech perception.

Pitch contour or its acoustic correlate, *fundamental frequency* (F0), distinguishes lexical meanings in tone languages. Two topics on CP of lexical tones were studied in the thesis: (1) the factors influencing CP, and (2) the temporal process of CP. These two topics were investigated through both behavioral and *event-related-potential* (ERP) methods on Cantonese and Mandarin tones.

Four factors were studied. They were (1) intrinsic acoustic properties of pitch contours by comparison between continua of level tones and contour tones; (2) positions of target syllables relative to context (without contextual sentence, at the beginning and at the end of the contextual sentence); (3) language backgrounds by comparison between listeners with different tone experiences; and (4) carrier syllables (real word, non word, and nonspeech). Three temporal stages were studied in the same experimental paradigm. They were (1) the preattentive stage investigated through the *mismatch negativity* (MMN); (2) the attentive stage investigated through the P300; and (3) the overt response stage investigated through the hit rate data.

All these four factors influence the degree of CP. In the discussion, both general auditory processing and language specific processing are suggested to be responsible for the various types of exhibition of CP, although they have different weights for different factors. Different patterns of CP were also observed in three temporal stages due to different weights of these two types of processing. In summary, a multistage model which includes both general auditory processing and language specific processing is proposed to explain the CP of lexical tones. This model improves previous models by proposing that the weights of these two types of processing in speech perception depend on the types of factors, and the temporal processing stages.

Finally, for the first time in the literature, the thesis also reported that even though a tone contrast (i.e., level vs. rising) is present in both tone systems, the same contrast is perceived differently by the two groups of subjects by virtue of their different language experiences.

摘要

同一句話，不同的人會發出不同的語音信號，甚至相同的人在不同場合下發出的語音信號也不一樣。盡管如此，聽衆總能夠毫不闲難地識別出這句話。語音的最小單位是音素。儅人們將所聽到的變化多端的語音信號對應到有限的幾個音素上時，一種稱爲範疇感知的現象出現了。範疇感知意味著兩种類別交界處的分辨率遠遠高於類別內部的分辨率。範疇感知是語音感知的一個重要特性。

在聲調語言中，聲音的音高或者基頻曲綫也可以用來區別語義，因此也是一種音素。本論文討論兩個關於聲調範疇感知的議題：(1) 影響聲調範疇感知的因素 (2) 聲調範疇感知的時間動態特性。這兩個議題將通過行爲和事件相關電位兩种實驗手段進行研究。研究的對象是普通話和廣東話（粵語）的聲調。

本論文研究了四种因素。它們分別是 (1) 基頻曲綫的內在聲學特性，包含了平聲調和曲折調；(2) 目標音節在上下文句子中的位置，包含孤立詞（沒有上下文），在句子開頭，和在句子結尾這三种情況；(3) 被試的聲調語言背景；和 (4) 載波音節的類型 (真字、 假字和非語音)。本論文還在同一种試驗範式下，研究了範疇感知在三种不同時間階段的表現。這三种時間階段是 (1) 前知覺階段，主要通過失匹配負波來探測；(2) 知覺階段，主要通過失 P300 來探測；(3) 外縣反映階段，主要通過命中率來探測。

研究結果表明所有測試因素都影響了範疇感知的程度。在論文討論部分，不同的範疇感知表現被認爲是因爲測試因素中兩种處理過程的不同加權的體現。這兩种過程是普遍聲音處理過程和語言特有的處理過程。範疇感知在三個不同時間階段，表現形式也不一樣。同理，這些不同也是因爲上述兩种過程的不同加權的體現。總之，本論文提出一種包括兩种過程的新的語音感知模型來解釋聲調的範疇感知問題。這種新的模型與以前的模型一樣，都包含了普遍聲音處理和語言特有的處理過程，但是這個模型首次提出這兩种過程在不同的狀態和不同的時間階段有不同的加權表現。

最後，本論文首次發現一種新的語言現象：即使兩种不同的聲調系統都包含同一對聲調對照物，因爲被試的母語有不同的聲調系統，因此他們對這一對相同的聲調對照有不同的感知模式。這種語言現象据我們所知從未被報道過。

CHAPTER 1 INTRODUCTION

1.1. Categorical Perception and Its Importance

Every day, we cluster things around us into different categories with or without being conscious of doing so. We know a table is a table, no matter whether it is round or square, large or small. We would not mistake a table for a chair or a stool. The ability to classify things and events into categories is one of the important cognitive abilities of human beings: cognition is categorization (Harnad, 1987). Categorization reduces the number and variety of the many objects, which need to be remembered. "Categorization occurs when we focus on important properties that are common to different objects and ignore irrelevant details (Repp, 1984)."

Categorization occurs in many domains, from a concrete object such as a cat or a dog to an abstract idea such as goodness or truth, from the visual domain such as different colors to the auditory domain such as different types of sounds. When some categories are formed, there is a relative perceptual change on the category boundary by enlarging the perceptual distance on the category boundary while shrinking the perceptual distance within a category. This phenomenon is called *Categorical Perception* (CP).

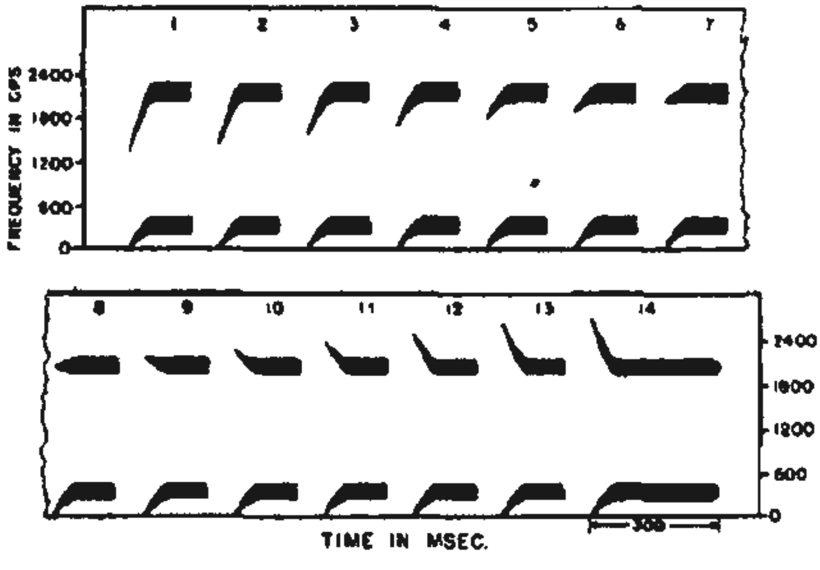


Figure 1.1.1 Illustrations of the spectrographic patterns for stimuli used in the experiment by Haskins Laboratories. Pattern 14, at the lower right, is complete in both time and spectral domains. (For discussion, see the text for details) [Taken from (Liberman, Harris, Hoffman, & Griffith, 1957)]

CP in speech sounds is first observed by Haskins Laboratories (Liberman, et al., 1957). In this first CP experiment, Liberman et al. constructed a series of syllables spanning the three categories */b/*, */d/*, and */g/* preceding a vowel */e/* by varying the transition of the *second formant* (F2). The first and second formants of this series of syllables are shown in Figure1.1.1. In the figure, fourteen 300 ms stimuli were constructed to represent the series. The results showed that the continuum can be reliably clustered into three rather sharply divided phonemic categories. Moreover, even though the physical difference along the continuum was the same, the subjects were better to discriminate between the sounds that lie on opposite sides of a phoneme boundary than between the sounds that fall within the same phoneme category. Furthermore, if the discrimination scores for adjacent pairs with the same step were connected together to form a curve along the continuum, there was a peak located on the phonemic boundary in the discrimination curve. Figure 1.1.2 (a) showed a stereotypical CP with two categories.

On the other hand, the other pattern of perception is called continuous perception, when the CP is absent. In the continuous perception mode, there is no sharp boundary to divide two categories. Rather, the proportion to perceive Figure 1.1.2 shows a schematic plot of identification and discrimination in stereotypical categorical perception mode and in stereotypical continuous perception mode (absence of categorical perception).

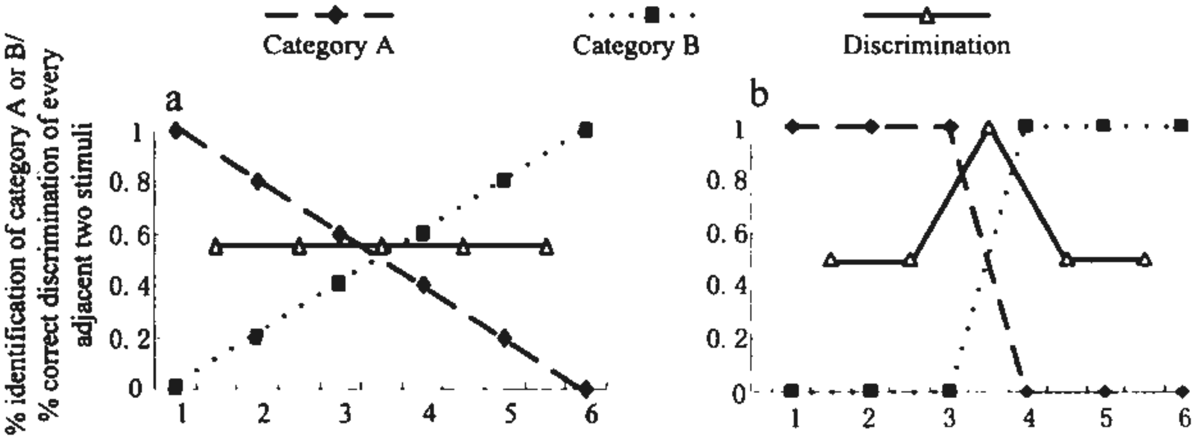


Figure 1.1.2 Plot of stereotypical categorical perception (a) and stereotypical continuous perception (b). The dashed line with a diamond and the dotted line with a rectangle represent identification scores of category A and B respectively. Solid line with an open triangle represents correct discrimination of every two adjacent stimuli pair.

The mechanisms of CP are not without controversy and are reviewed in Chapter 2. Nonetheless, the ability to place a range of phonetic features embedded within the quasi-continuous sound signal into discrete phonological categories is crucial for language processing (Baddeley & Wilson, 1993). CP is one manifestation of such ability and is associated with different types of dysfunction on language processing. (1) The children at risk of reading difficulties were less sensitive than not-at-risk children to changes between tokens that crossed the *voice-onset-time* (VOT) phonetic boundary (Breier, Fletcher, Denton, & Gray, 2004; Breier, et al., 2001). (2) Subjects in schizophrenia had shallower phonemic boundary slopes than the control normal subjects for *place of articulatory* (POA) (from */ba/* to */da/*) (Cienfuegos, March, Shelley, & Javitt, 1999). (3) Children with dyslexia showed an increased perceptibility of within-category differences on POA continuum (from */ba/* to */da/*), which is significantly different from the normal children. Moreover, the pattern is less clear in nonspeech condition (Serniclaes, Sprenger-Charolles, Carre, & Demonet, 2001). A followed-up experiment on investigation of CP in illiterate adults showed no difference of boundary effect between the illiterate adults and the normal adults, which further suggested that the CP anomalies displayed by dyslexics are indeed a cause rather than a consequence of their reading problems (Serniclaes, Ventura, Morais, & Kolinsky, 2005). (4) In non-alphabetic Chinese system, dyslexia also correlated with less accurate phonological awareness. For example, while performing similarly to reading-level controls, dyslexic children perceived tone and aspiration contrasts less categorically and less accurately than age-matched controls (Cheung, et al., 2009) .

Since CP is an important phenomenon associated with some language abilities, it attracts many research interests in either clinical application, or in the cognitive study, or in linguistic research.

1.2. Lexical Tone and Its CP Studies

Previous studies have demonstrated that CP is observed in both consonants and vowels, although more often in consonants than in vowels [See review in (Diehl, Lotto, & Holt, 2004)]. Both consonants and vowels are segments, which are defined in linguistics as "any discrete unit that can be identified, either physically or auditorily, in the stream of speech" (Crystal, 2003) (pp408-409). In contrast, the suprasegmental features usually extend over more than one segment such as tone, stress, and intonation. Among those, lexical tones also distinguish word meanings in tone language, in addition to consonants and vowels (Wang, 1967). Tone is superimposed mainly over the vowel of a syllable in Mandarin. The length of the tone is usually around several hundreds of milliseconds. So it is called suprasegmental information. Table 1.2.1 shows a syllable structure of a Chinese dialect.

Table 1.2.1 Syllable structure of a Chinese dialect. Elements enclosed in [ ] are optional [Taken from (Wang, 1973)]

Tone			
Initial	Final		
	[Medial]	Nucleus	[Ending]

Lexical tones are primarily determined by the fundamental frequency, or F0. The fundamental frequency is the number of glottal pulses per second that a speech signal contains. Each glottal pulse corresponds to a single vibration of the vocal folds inside the larynx (Wang & Peng, 2006). The tighter the vocal folds are pulled, the higher the vibration frequency is. This is how subjects change the F0 of their speech sounds. The typical number of cycles is around 100 Hz for male subjects. Women and children have higher frequencies of vocal folds vibration (around 300 Hz). Our perception of F0 is called pitch. If the pitch can be used to distinguish word meanings, it is called tone. A language is called a *tone language* if it has tones. Tone languages are found all around the world (Wang, 1973). Based on a very rough estimation, about 70 percent of the world's languages are tone languages (Yip, 2002) and over half the world's population speak a tone language (Fromkin, 1978).

Mandarin and Cantonese, the two Chinese dialects studied in the present thesis, are both tone languages. Mandarin has four tones (neutral tone, which does not distinguish lexical meaning, is excluded from the four): high level tone (tone 1), high rising tone (tone 2), dipping tone or falling rising tone (tone 3), and high falling tone (tone 4). Without consideration of duration, Cantonese has six tones[a]: high level tone (tone 1), high rising

---

[a] There are three entering tones with stop endings — tone 7 to tone 9 — share the same pitch patterns of level tones.

tone (tone 2), mid level tone (tone 3), low falling tone (tone 4), low rising tone (tone 5) and low level tone (tone 6).

An example of the tones and their relationship with the pitch patterns in Mandarin is shown in Figure 1.2.1. The right column shows the F0 of four monosyllabic Mandarin words, with the same syllable *ma*. When the F0 decreases, syllable *ma* means '*scold*'. When this syllable is produced with an increasing F0, it means '*hemp*'. A high level pattern of the F0 with syllable *ma* means '*mother*'. And this syllable means '*horse*' with a low dipping pattern of the F0. The pitch contour of Cantonese tones is shown in Figure 1.2.2.

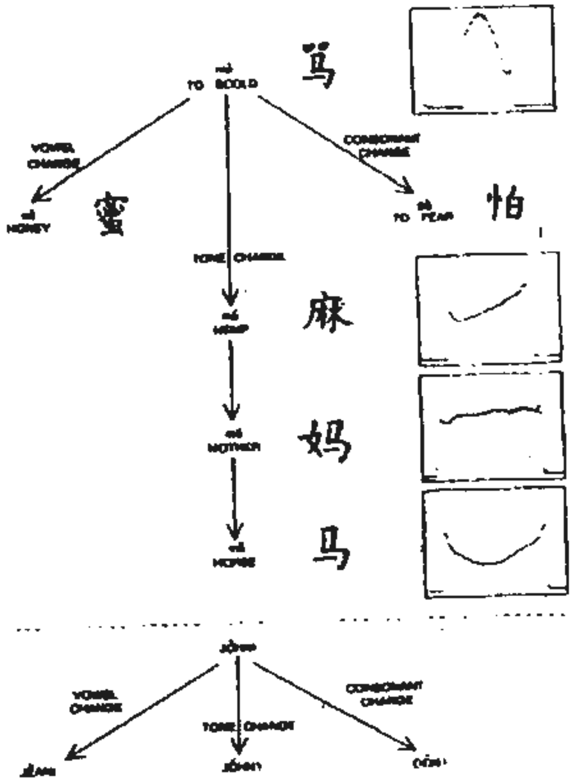


Figure 1.2.1 Mandarin tones. The right panel represents the pitch contours. The middle four characters in left panel represented the corresponding characters to the pitch contours. [Taken from (Wang, 1973)]

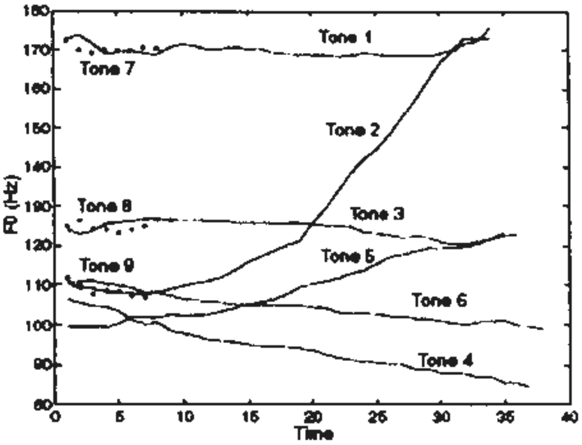


Figure 1.2.2 Pitch contours for Cantonese tones. Each tone is uttered by a male speaker. [Taken from (Peng, 2006)]

Lexical tone is an important constituent to distinguish the word meaning; however, its suprasegmental feature makes it different from phonetic segments. Are lexical tones also perceived categorically? Do they show the similar 'phonetic mode' as vowels and consonants? The previous CP studies in lexical tones showed inconsistent results from different tone languages and different paradigms. The continua varying along Thai level tones (Abramson, 1979) or varying along Cantonese level tones (Francis, Ciocca, & Ng, 2003) are perceived continuously if the tones are in citation forms. However, also in citation forms, the continuum varying along Mandarin level and rising tones is perceived categorically or quasi-categorically by native subjects (Hallè, Chang, & Best, 2004; Y. S. Xu, Gandour, & Francis, 2006). The presence of CP is also influenced by the listeners' linguistic backgrounds. The same Mandarin level to rising tone continuum was perceived continuously by non-tone language listeners (Hallè, et al., 2004; Y. S. Xu, Gandour, & Francis, 2006).

The limit studies on CP of lexical tones are not comparable to the numerous studies in consonants and vowels. In this thesis, the perception of Cantonese level tones as well as level and rising tones in both Mandarin and Cantonese tone systems was systematically investigated through different types of manipulation. In Chapter 3, behavioral experiments were conducted to show the effect of different factors on CP of Cantonese level tones. In Chapter 4, behavioral experiments were conducted to show the effect of different factors on CP of level vs. rising tones in both Mandarin and Cantonese tone systems.

1.3. Event-Related-Potential and Its Applications in CP Studies

The early CP studies were mainly behavioral studies through various types of manipulation. Although such manipulation is very useful to investigate *how* various factors influence CP, the inherent very low temporal resolution of the behavioral studies hinders their application in investigating *when* CP happens. Is it during the sensory and perceptual stage or is it during a later stage after the signal has been transferred onto a higher level? A technique called *electroencephalography* (EEG), which has a very high temporal resolution (Donchin, 2006), is suitable to investigate this *when* question.

The EEG, the electrical activities of the brain, can be reliably measured from the human scalp. The EEG reflects thousands of simultaneously ongoing brain processes. The brain response to a single stimulus or an event of interest, called *event-related-potential* (ERP), is not usually visible in the EEG recording of a single trial. Therefore, to see the brain response to the stimulus, the experimenter must conduct many trials and average the results together, causing random brain activity to be averaged out and the relevant ERP to be remained (Luck, 2005). A schematic illustration of the recording procedure is shown in Figure 1.3.1. In the left panel of the figure, EEG recording system is presented; while, in the right panel of the figure, the recorded brain wave synchronizing to each stimulus is shown. The continuous brain wave is segmented into different epochs aligning with the onset of a stimulus. The averaged wave of the same type of epochs is called an ERP. The procedure to obtain an ERP is shown in Figure 1.3.1.

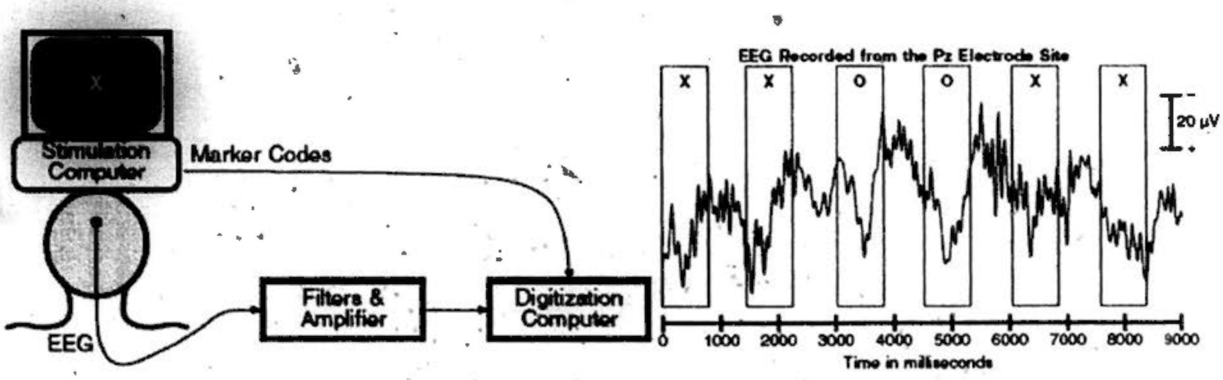


Figure 1.3.1 A schematic illustration of EEG recording procedure. See the text for details. [Taken from(Luck, 2005)]

Experimental psychologists and neuroscientists have discovered that many different stimuli can elicit reliable ERPs from participants. The ERPs may be described in terms of a series of positive and negative peaks or components that occur at different characteristic time. The timing of ERPs is thought to provide a measure of the timing of the brain's communication or the timing of information processing [reviewed in (Hillyard & Kutas, 1983)]. For example, an ERP component called P300 refers to a positive peak at around 300 ms after a rarely occurring event (Donchin, Tueting, Ritter, Kutas, & Heffley, 1975). Another ERP component called N400 refers to a negative peak at around 400 ms after a word violates the semantic meaning in a sentence (Kutas & Hillyard, 1984).

ERP is also used to study the temporal processing of CP and thus to investigate the neural mechanisms underlying CP in different stages. Since CP is a perception phenomena happened in phonological level, two typical ERP components are used to study CP: *mismatch-negativity* (MMN) [reviewed in (Näätänen, Paavilainen, Rinne, & Alho, 2007; Pulvermüller & Shtyrov, 2006)] and P300 (Maiste, Wiens, Hunt, Scherg, & Picton, 1995). The technical details and their applications in CP were introduced in Chapter 2.

1.4. Organization of the Thesis

CP is an important phenomenon in language and lexical tone is a component of language. Both are essential topics in studying language, and attract many research interests. There is intensive research on CP of other linguistic relevant features. However, only a little work on CP studies on lexical tones was available. Two topics on CP were studied in this thesis: (1) the factors influencing CP, and (2) the temporal process of CP. The work of the present thesis was done through both behavioral and *event-related-potential* (ERP) methods on Cantonese and Mandarin tones. It is organized as following.

Chapter 2 provided the background of CP studies in three sections. Selected mechanisms underlying CP and the factors influencing CP were introduced in the first section. The motor theory (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967) is the first proposed model to explain the mechanisms underlying CP. The motor theory emphasizes the influence of production on perception. Therefore, it can be called a single factor model. It gains many research interests recently because some new evidence from brain imaging studies confirms the involvement of some brain areas, which are traditionally responsible for speech production, in speech perception. However, some studies suggested that many other factors such as the general auditory processing, in addition to the production, also influence CP [reviewed in (Repp, 1984)]. Those other factors were also briefly introduced in this chapter. Integrate the general auditory processing into linguistic processing, some multi-factor models were proposed. A *multistore* model (Y. S. Xu, Gandour, & Francis, 2006), which introduces different types of memory systems to deal with these two types of processing, was one of multi-factor. However the *multistore* model did not consider the temporal dynamic of different types of processing on CP. Therefore, another multi-factor model called *multilevel* model (Phillips, 2001) was proposed to suggest the timing of influence from different factors.

In chapter 2, following the first section, which introduced the mechanisms underlying CP, the second section of this chapter focused on the typical analysis methods on behavioral CP studies. The criteria of the presence of CP and the measure of CP were detailed in this section to facilitate the experimental description in later chapters. Finally, the two ERP components to study CP, as well as the reasons to select these two components, were reviewed.

Chapter 3 described the behavioral experiments on CP of Cantonese level tones. A series of experiments were conducted to examine the perception behavior on Cantonese level tones, especially the effect of context position on CP and the influence of the linguistic background. Perception performance by native listeners in citation forms was studied first and followed by studies of contextual position effects on target's CP. In the same time, the influence of different language backgrounds was also studied by examining the performance from three groups of subjects: native subjects (Cantonese), subjects who speak another tone language (Mandarin), and subjects who do not speak a tone language (French). Furthermore, the perception of nonspeech sounds and non-word speech sounds, both of which mimic the pitch contours of the real Cantonese words, was also investigated to check the influence of carrier syllables. Through the investigation on these different types of factors, the mechanisms underlying CP was proposed.

Chapter 4 described both the behavioral and ERP experiments on CP of a continuum spanning from a level to a rising tone in both Mandarin and Cantonese tone systems. The aim of ERP experiments was to investigate the temporal dynamics of CP process. Three types of data (MMN, P300, and the hit rate) were measured to investigate CP under three temporal stages: the preattentive stage, the attentive stage, and the overt response stage. The behavioral experiments in this Chapter had two functions: (1) *post hoc* to verify CP behaviors, and (2) to look into the speech perception in the final cognitive stage of speech recognition—decision making. Both cross-language comparison and the manipulation of carrier syllables were also included in this chapter to show how long-term linguistic training influences the tone perception as a whole.

Chapter 5 was general discussion. CP experiments described in this thesis provided more information on how different types of factors influence the performance of tone perception. The results show that the (1) CP is not absolute but context dependent, (2) CP is influenced by linguistic training, and (3) CP is observed in nonspeech sounds in some conditions. Furthermore, the work also presented the time course of CP processing. The discussion on the obtained results led to better understanding of the complex behavior of CP, especially the underlying mechanisms of CP on different cognitive stages. A *multistage* model was proposed to explain the observed data. The model integrates the merits from both the *multistore* model (Y. S. Xu, Gandour, & Francis, 2006) and the *multilevel* model (Phillips, 2001). The key point of the proposed model in the present thesis is that two types of processing (general auditory processing and language specific processing) have different weights according to the different types of conditions and different temporal stages in speech perception. Moreover, major contributions, as well as some future work, were also summarized in this chapter.

CHAPTER 2 BACKGROUND

2.1. Mechanisms Underlying the Categorical Perception

2.1.1. Motor Theory

The analyses on speech sounds had shown that acoustic features of speech sounds vary a lot, even when the same word is uttered by the same people. Moreover, the variation of acoustic features highly depends on speech context (Liberman, et al., 1967). Nonetheless, humans map the complex acoustic signals into finite phonemic categories without difficulties. This invariant mapping process was illustrated by Liberman et al. (1967), who showed that sounds with different spectral patterns are perceived as the same consonant (*/d/*), when they are followed by different vowels. The invariance of speech perception can also be illustrated by CP on a sound continuum, which shows that the continuously and infinitely changing acoustic signals can only be mapped into a limited number of phonemes without ambiguities on phonemic boundaries.

To explain such invariance in perception, the motor theory was proposed to model the process of speech perception (Liberman, et al., 1967). In the motor theory, the complicated articulatory encoding is assumed to be decoded in the perception of speech by the same processes that are involved in production. In other words, people perceive spoken words by identifying the vocal tract gestures with which they are pronounced (Liberman & Mattingly, 1985, 1989). The discontinuity on the phonemic boundaries when humans perceive speech sounds, where CP happens, is mainly due to the realization of speech production. This hypothesis is consistent with early reported experimental evidence — the categorically perceived speech sounds are continua with manipulation of features that cannot be produced continuously [See reviewed in (Repp, 1984)]. These features include (1) VOT (Liberman, Harris, Eimas, Lisker, & Bastian, 1961; Liberman, Harris, Kinney, & Lane, 1961), (2) POA (Carden, Levitt, Jusczyk, & Walley, 1981; Johnson & Ralston, 1994; Liberman, et al., 1957), (3) manner of articulation (Bastian, Eimas, & Liberman, 1961; Fitch, Halwes, Erickson, & Liberman, 1980; Miller & Eimas, 1977) and (4) nasality of the consonant (Larkey, Wald, & Strange, 1978; Miller & Eimas, 1977). For some other phonemic features, such as vowels (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973) or lexical tones (Abramson, 1979), the continuous articulation between phonetic categories is a possible, so failure or less CP is observed.

Therefore, the motor theory proposed that both perception and production share a special "phonetic module", which is human-specific. [See reviewed in (Diehl, et al., 2004)]. The advocated specific phonetic module in the brain by the motor theory was illustrated by the duplex perception (Liberman & Mattingly, 1989), although a later study provided an alternative explanation for the duplex perception (Fowler & Rosenblum, 1990). The duplex perception refers to the phenomenon that the same acoustic information is used for both a nonspeech and a speech percept. In practice, a listener is presented a dichotic stimulus— two ears are presented with two different sounds simultaneously. One sound is an isolated third-formant transition sounding like a nonspeech chirp. The other sound is the base syllable consisting of the first two formants with complete formant transitions, and the third formant without a transition. The base syllable is perceived ambiguously as either */da/* or */ga/*. If the sound of third-formant transition is more */ga/* biased, the subject will unambiguously hear the stimulus as both chirp and the unambiguous */ga/*. Moreover, when the third-formant transition sound is presented in the isolation form, the perception mode is continuous. However, if it is presented in the dichotic stimulus, the perception mode is categorical (Liberman & Mattingly, 1989). The duplex perception, where whether the same sounds are perceived categorically depends on whether they are perceived as speech or nonspeech, is used as a strong evidence to show the existence of the speech module. Figure 2.1 shows the conceptual diagram of special speech module proposed in the motor theory.

Even though the motor theory has evolved several versions (Liberman, et al., 1967; Liberman & Mattingly, 1985, 1989), the mechanisms of speech perception proposed in this theory remains controversy [See review in (Galantucci, Fowler, & Turvey, 2006; Hickok, Holt, & Lotto, 2009; Lane, 1965; Scott, McGettigan, & Eisner, 2009; Studdert-Kennedy, Liberman, Harris, & Cooper, 1970)]. In this thesis, I only focus on the controversy about one particular phenomenon of speech perception, the CP.

Two key mechanisms underlying CP suggested in the motor theory are the phonetic module and the discontinuity of speech production. Therefore, criticisms for the explanation are coming from these two aspects. (1) Whether there is a phonemic module, so that CP is dedicated to speech and human specific. (2) Whether CP relies on production. If CP is also observed in nonspeech sounds, or in other species than humans, or in other modalities, the hypothesis of point (1) is not valid [See review (Repp, 1984)]. The point (2) attracts lots of interests because of the explosion of brain imaging techniques' application on cognitive science, which shows speech perception involves access to the motor system (Burton & Small, 2006; D'Ausilio, et al., 2009; Ravizza, Henri, & Claire, 2005). However, the evidence from the brain imaging studies is not without controversial. Detailed evidence for these two points, which includes auditory modality only, was reviewed in the following sections.

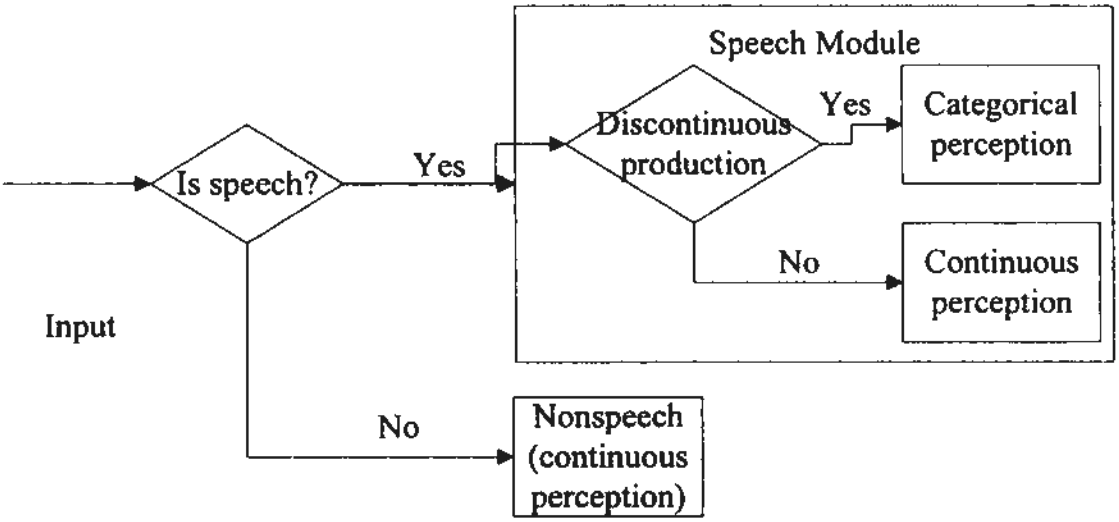


Figure 2.1.1 A conceptual diagram of the motor theory. A speech detector is presented at first to identify the incoming signal. If it is nonspeech, it will be perceived continuously. If it is speech signal, it is processed with "Speech Module". There are two pathways inside the "Speech Module". (1) If the continuous production is possible, it is perceived continuously. (2)If it is produced discontinuously, it is perceived categorically.

2.1.2. Some Issues in the Motor Theory

2.1.2.1. CP in nonspeech sounds.

As it was mentioned in section 2.1.1, many phonemic features are perceived categorically; however, later studies found that nonspeech sounds sharing the same acoustic features are also perceived categorically. For example, a set of nonspeech tonal stimuli that differ in the relative onset time of their components were employed (Pisoni, 1977) to investigate the perception mode on this onset time relation [also called *tone-onset time* (TOT)]. The result showed that the TOT, which is analogous with VOT, is also perceived categorically. Following this pioneer work by Pisoni, Holt et al. used nonspeech sounds, which mimick certain temporal properties of VOT stimuli, to train the subjects to categorize the continuum (Holt, Lotto, & Diehl, 2004). The subjects were divided into two groups: one group of listeners learned categories with a boundary coincident with the perceptual discontinuity, while another group learned categories defined such that the perceptual discontinuity fell within a category. Listeners in the latter group required significantly more experience to reach criterion categorization performance. These two studies by Pisoni and Holt et al. indicated that there is perceptual discontinuity for feature of onset time relationship. However, this perceptual discontinuity is not restricted to speech sounds.

Moreover, if the nonspeech sounds were constructed to share some critical transient acoustic cues of consonants (ramp of noise amplitude), the perception of these nonspeech sounds is categorical (Mirman, Holt, & McClelland, 2004). If the nonspeech sounds mimick the steady-state spectral cues of simplified synthetic vowels (filtered frequency bands of the noise), the perception was less categorical (or continuously) (Mirman, et al., 2004). This characteristic is very similar to the observation in consonants and vowels.

The above evidence showed that some acoustic features are also perceived categorically by humans. The features are not limited to speech sounds; rather, the general features are utilized in language. Languages make use of perceptual discontinuities to promote distinctiveness among sounds within a language inventory, instead of using a special phonetic module.

2.1.2.2. CP in animals.

CP is also observed in other species than humans. The most well-studied VOT feature is perceived categorically by chinchillas and the CP performance of chinchilla is similar to English adults (Kuhl & Miller, 1975). An interesting feature of VOT is that its perception is highly correlated with *first formant* (F1) onset frequency. Researchers found that when F1-onset frequency is lower, longer VOT is required for human listeners to perceive synthesized stop consonants as voiceless. This correlation effect was also observed in several types of animals: Japanese quails (Kluender & Lotto, 1994), budgerigars (Dooling, Okanoya, & Brown, 1989), chinchillas (Ohlemiller, Jones, Heidbreder, Clark, & Miller, 1999) and monkeys (Steinschneider, et al., 2005) [also see review (Belin, 2006)]. This boundary effect on VOT feature can be also reflected from neural responses. Intracortical responses recorded in human and monkey primary auditory cortex, which responded to VOT or TOT, showed similar patterns of the discontinuity on 20 ms (Steinschneider, et al., 2005), the critical threshold to distinguish English voiceless and aspiration sounds. The available evidence showed that encoding the VOT phonetic feature is not specific to human beings nor relies on production. The 20 ms threshold is not limited to VOT or TOT, but a more general temporal threshold in the auditory sensory system. For example, the house mice categorize the 50 kHz with various durations into two categories, which is sharply divided on 25~30 ms boundary. Two sounds are discriminated only when they differ at least 20~25 ms (Ehret, 1992).

POA or formant transition is another well studied phonetic feature. This feature is also perceived categorically by many types of animals. Japanese quails can be trained to distinguish */d/* from */b/* and */g/* preceding by different vowels (Kluender, Diehl, & Killeen, 1987). It was also found that rhesus monkeys perceive */bæ/-/dæ/-/gæ/* in a categorical manner, with different heart rates when they hear across-category pairs and within-category pairs (Morse & Snowdon, 1975) [also reviewed in (Repp, 1984)]. This category boundary effect in monkey was also observed behaviorally in a more recent study (Kuhl & Padden, 1983).

In addition to the phonetic features in speech sounds, some species related calls or songs are also perceived categorically by non-human species. (1) In a study, crickets' responses to some sound frequencies showed CP on the boundary of 16 kHz. The CP was measured through cricket's turning performance in both labeling and habituation-dishabituation paradigms (Wyttenbach, May, & Hoy, 1996). The 16 kHz boundary frequency reflects the frequency to identify the crickets' enemy: crickets call at 4 to 5 kHz; while, their enemy (bats) call at 25 to 80 kHz. (2) Position of an inflection from a rising to a falling frequency in acoustic sounds is a key feature for Japanese macaque communication signals. Therefore, the synthetic sounds, which resemble the critical feature of inflection position, can be clustered categorically by trained Japanese macaques (May, Moody, & Stebbins, 1989). (3) Note duration is a critical feature for swamp sparrow to distinguish two note categories with different roles in song construction. Swamp sparrow not only showed CP behaviorally (Nelson & Marler, 1989), the neurons in freely behaving swamp sparrows also expressed categorical auditory responses to changes in note duration (Prather, Nowicki, Anderson, Peters, & Mooney, 2009).

More studies on response of non-human species to speech and species-specific sound have been reviewed in some early papers (Ehret, 1987; Kuhl, 1986; Snowdon, 1979). All the above evidence showed that CP of auditory sounds is not necessarily processed by a specific speech module and is not limited to humans.

2.1.2.3. Broca's area and the motor cortex in speech perception.

The special "phonetic module" proposed in the motor theory is shared by perception and production. In other words, the discontinuity observed in CP is due to the discontinuity of vocal gestural. Some evidence showed that there is a production-perception link in the speech. For example, repeated presentation of */pa/* leads to less ambiguity of identifying */pa/* in a */ba/-/pa/* continuum (Cooper, 1979). Similarly, the same adaption effect is reflected in production. Since VOT is correlated with F1 onset frequency, the VOT for */pi/* is shorter than */ti/*. When the subjects adapt to repeated presentation of */pi/*, VOT is reduced when producing the */pi/-/ti/* continuum [(Cooper, 1979), see also in (Galantucci, et al., 2006)].

In addition to the behavioral evidence, the finding of mirror neurons and the involvement of Broca's area and the motor cortex in speech perception showed some neural evidence of the perception-production link. Both Broca's area and the motor cortex are important for speech production. The Broca's area is located at the posterior inferior frontal lobe. The importance of Broca's area in speech production is well-recognized now, since Paul Pierre Broca (1824-1880) reported impairment of speech production from a patient with this area damaged. The motor cortex includes three parts: the *primary motor cortex* (M1), *the premotor area* (PMA) and the *supplementary motor area* (SMA). It is most involved in controlling body's voluntary movements, including speech production.

The discovery of mirror neurons in monkey provides direct neural evidence for motor system (F5) involvement in perception (Pellegrino, Fadiga, Fogassi, Gallese, & Rizzolatti, 1992). A mirror neuron is a neuron that fires both when an animal acts and when the animal observes the same action performed by another (Rizzolatti & Craighero, 2004). Later finding showed that the mirror neuron also fired when the monkey heard the sounds of actions (Kohler, et al., 2002). However, it is normally impossible to study single neurons in the human brain, so scientists cannot be certain that humans have mirror neurons. By using *functional magnetic resonance imaging* (fMRI), researchers found that some brain regions (left inferior frontal cortex, including the Broca's area—homologue to Monkey's F5, and right superior parietal lobe) were active [(Chong, Cunnington, Williams, Kanwisher, & Mattingley, 2008; Welberg, 2008), although a contradict evidence was presented in (Dinstein, Gardner, Jazayeri, & Heeger, 2008)] both when a subject performed an action and when he saw another individual performing an action. It has been suggested that these brain regions contain mirror neurons, and they have been defined as the human mirror neuron system (Iacoboni, et al., 1999).

Some indirect measures were used to study the mirror neuron system in humans. For example, one measure is recording the size of a *motor evoked potential* (MEP) from muscles induced by *transcranial magnetic stimulation* (TMS) on motor cortex. When a person observed another person's action, his motor cortex became more excitable, so that the size of the MEP was enhanced (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995). It was found that the size of the MEP from tongue and lip muscles was enhanced when subjects listened to speech than when they listened to non-verbal sounds (Fadiga, Craighero, Buccino, & Rizzolatti, 2002; Watkins, Strafella, & Paus, 2003). Furthermore, by using the technique of combining *positron emission tomography* (PET) with TMS, the posterior part of the Broca's area was identified to modulate the excitability of the motor system during speech perception (Watkins & Paus, 2004). Recently, by using fMRI, distinct motor regions in the precentral gyrus, which are sparked by articulatory movements of the lips and tongue, were found differentially activated when subjects listened to the lip- or tongue-related phonemes (Pulvermüller, Huss, et al., 2006).

In CP study, deficits in the prefrontal region, including Broca's area, were found to be associated with impairment of categorical production. Some evidence showed that patients with lesions to Broca's area or who with Broca's aphasia are more likely to produce phonemes that are not good exemplars of a given target category. Moreover, this type of disorder is thought to be an impairment of motor planning but not weakness of the articulatory musculature [See review in (Ravizza, et al., 2005)]. More surprisingly, the prefrontal regions are also involved in categorical perception. Patients with frontal lesions showed difficulty in identifying phonemes with VOT and POA features and displayed more continuous function than normal subjects and patients with subcortical lesions (Ravizza, 2003).

Above evidence showed that "*it is possible the prefrontal regions maintain abstract motor representations that are used for speaking as well as comprehending*" (Ravizza, et al., 2005). However, the studies of involvement of Broca's area in speech perception did not show a consistent result. In a PET study, only overtly discrimination of the phonological features actived part of the Broca's area in the left hemisphere, but this area did not activate in passive listening (Zatorre, Evans, Meyer, & Gjedde, 1992). Ravizza (Ravizza, et al., 2005) proposed the function and the timing of the involvement of the motor cortex in the speech perception to reconcile the discrepancy. (1) The motor information is needed when the speech signal is attended to and is difficult for a clear categorical decision. So, the time sequence for different brain regions to be involved is different: the motor cortex is involved later than the temporal-parietal gyrus. In deed, some *magnetoencephalography* (MEG) results have shown that the activation of inferior frontocentral areas in speech perception for action words was around 15 ms later than the activation of superior temporal areas even during the passive listening (Pulvermüller & Shtyrov, 2006; Pulvermüller, Shtyrov, & Ilmoniemi, 2005). (2) Alternatively, the inconsistent engagement of Broca's area in passive listening may indicate that this region is for attention-based retrieval but not for motor-based phonological representations. However, this explanation is less possible because of some new evidence, which suggests Broca's area is involved in the representation of articulatory information and not just in the controlled retrieval of such information (Ravizza, et al., 2005). For example, experiments showing excitability of the motor system during speech perception in passive listening tasks with low attention demands (Pulvermüller, Huss, et al., 2006; Pulvermüller & Shtyrov, 2006; Watkins & Paus, 2004; Watkins, et al., 2003).

In summary, accumulating brain imaging studies revealed the involvement of the motor cortex and mirror systems in speech perception and some of their functions in CP. However, exact function and the timing of the involvement of the motor cortex and mirror systems need to be investigated more before we have a clear picture.

2.1.2.4. Comments on the motor theory

From the above review, CP is observed in nonspeech sounds, which cannot be produced by humans; in animals, who cannot produce human speech; and in very early infants before they can produce the sounds (Jusczyk, Rosner, Cutting, Foard, & Smith, 1977). Therefore, the explanation provided by the motor theory cannot explain all CP phenomena.

Even though there is increasing evidence to show the involvement of the motor cortex in CP, the role of the motor cortex in CP is unclear yet. Moreover, the motor cortex is not the only brain area responsible for CP. As Lotto, et al. input: "*It's very different to claim that motor area activity is present during perception than that motor activity plays a necessary part in perception, as proposed by MT (motor theory)* ." (Lotto, Hickok, & Holt, 2009)

Speech processing involves many resources and the procedure is not instantaneous, so there should be more than one mechanism underlying CP, including that proposed by the motor theory. Moreover, different mechanisms may work at different time during the process. The multiple mechanisms can be illustrated by the learning procedure. The above review has shown that there is a hardwired threshold in the brain for the VOT feature. However, the threshold can also be learnt and thus to be altered with more efforts later (Holt, et al., 2004).

Most languages use this innate VOT threshold for phonetic boundary, while some other languages use more complex boundaries in addition to the innate VOT threshold. The complex boundaries are trained by experiences (Keating, Mikoś, & Ganong III, 1981; Kessinger & Blumstein, 1997; Lisker & Abramson, 1964) [See review in (Lasky, Syrdallasky, & Klein, 1975)]. The learning ability was also exhibited in Japanese quails (Holt, Lotto, & Kluender, 2001). The VOT examples show that there are at least two mechanisms for CP: innate and learning. The learning procedure is more common on perception training than production training.

The learning procedure is not constrained to the VOT feature; it can also change the performance on other features, such as the POA feature and calls specific to specials. For example, the topographic pattern of evoked potentials, which were recorded from monkeys in response to the place feature, changed after training. The responses shifted from the right hemisphere to the left hemisphere after the monkeys were exposed to the contrast prenatally or postnatally, when they were compared with those from the control group (Molfese, et al., 1986). Moreover, the group with prenatal exposure only showed a response to across-category discrimination, while both postnatal exposure and control groups showed responses to both across-category and within-category discrimination. This evidence showed that the training period in monkeys is as important as in humans. Finally, calls specific to special can also be learnt through training. For example, swamp sparrow populations that learn different song dialects show different categorical perceptual boundaries, which are consistent with the boundary being learnt (Prather, et al., 2009) .

The learning effect described above showed that at least two procedures involved in CP processing: the innate non-linearity in general auditory systems and the perceptual reorganization through training. Moreover, the fact that the training effect is also present in nonhuman animals further opposes the idea of a special phonetic module in the brain. The behavioral and neurophysiological studies on the perception-production link also indicated that the timings of involvement for different brain areas in CP are different. Therefore, it is more reasonable to suggest that there are multiple stages along the process, which will be discussed in the section 2.1.4.

2.1.3. Factors Influencing Categorical Perception

As reviewed in the section 2.1.2, the motor theory cannot fully explain all empirical data on CP, so some other top-down information — lexical information, semantic information, and visual speech — influences CP as well (Hickok, et al., 2009). Moreover, since CP is observed in the general auditory processing, likewise, other factors than the top-down information — the non-linearity in the sensory system, the auditory memories, and the attention — also influence CP. In general auditory perception, there are two types of perception modes: categorical and continuous perception. When the non-linearity in the sensory system is not available for some acoustic features, the general auditory perception is continuous. On the contrary, when the non-linearity mapping is available, it is a categorical perception. The general auditory perception is irrelevant for linguistic experience. On the other hand, since the phonemic categories are discrete and limited, the perception mode for phonetic features is categorical. These two types of processing —linguistic relevant and general auditory processing — interact with each other. To investigate the various factors from these two types of processing, some manipulations on experimental design are done in practice (Repp, 1984).

First type of manipulation is on the auditory memory. (1) Adding a noise or an irrelevant sound between two trials interferes with the auditory memory. Through such manipulation, the phonetic procedure will be more prominent. Indeed, the vowel discrimination showed an increased categoricalness of CP through this manipulation

(Fujisaki & Kawashima, 1969; Lane, 1965), while consonant discrimination showed little effect with the same manipulation (Fujisaki & Kawashima, 1969). (2) Similarly, increasing the duration of an *inter-stimulus-interval* (ISI) decreases the auditory memory, while leaves the phonetic procedure intact. Indeed, experimental evidence showed that discrimination score for vowels was decreased when ISI increased, while ISI had little effect on consonant discrimination. Moreover, ISI affected discrimination of a within-category pair more than that of an across-category pair [(Pisoni, 1973), cited in (Repp, 1984)]. (3) In the interfering procedure, the acoustic properties and the context positions of interferential stimuli also matter. A more similar interference to the target introduces a greater interference effect. In addition, a following context interferes with auditory memory more than a preceding one (Repp, 1984). (4) Use some methods to increase the discrimination sensitivity. For example, rating scales and reaction times provide more information for fine discrimihation sensitivity. In color CP studies, the reaction times have been used as a measure to examine the hemispheric lateralization for CP, while there was no difference when regarding with the discrimination score (Gilbert, Regier, Kay, & Ivry, 2006). Different types of presentation paradigms also have different discrimination sensitivity. Some presentation paradigms have more discrimination sensitivity than others. For example, the AX paradigm is more sensitive in the discrimination task [See in (Repp, 1984)].

Second type of manipulation is on stimulus. As reviewed in section 2.1.2, the acoustic properties play an important role in CP: vowels are perceived less categorical than consonants. Some nonspeech sounds, which mimic some characteristics as vowels or consonants, share the similar degree of CP as the speech sounds. One of the acoustic properties to distinguish vowels and consonants is stimulus length. Therefore, a reduction in stimulus duration weakens the auditory trace and thus leads to clearer CP [See in (Repp, 1984)].

The last type of manipulation is on subjects. The cross-language studies are often used to investigate the linguistic influence on CP. However, the studies on human infants [e.g.(Eimas, Siquelan.Er, Jusczyk, & Vigorito, 1971)] and non-human animals show the general auditory processing underlying CP. The training experience and strategies of the subjects also influence CP. There are two alternative modes, phonemic and auditory, when subjects listen to the synthetic sounds. The neural circuits are different when subjects switch the mode (from the auditory processing to the speech processing) to listen to the same stimuli (Dehaene-Lambertz, et al., 2005). Several studies also showed that in the phonetic mode, listeners integrated all the relevant acoustic information into a phonetic perception, while in the auditory mode, they either selectively attended to individual auditory dimension or divided attention among several of them (Best, Morrongiello, & Robson, 1981).

All the above evidence showed that many factors, in addition to production gestures, influence CP.

2.1.4. Multilevel and Multistore Processing Models

The review from section 2.1.2 showed that there are at least two types of processing underlying CP: the phonemic processing and the auditory processing. This two-process model is also called dual-process model (Fujisaki & Kawashima, 1971). The dual-process model proposed that the phonemic mode is categorical, and the auditory short-term store is continuous. Although such model can explain the early behavioral CP results, which showed an increased discrimination performance on the phonetic category boundary, it cannot explain recent CP experiments on nonspeech and animals. Moreover, it does not provide clear information on how these two modes are processed in the time dimension. Therefore, two recent models on speech perception — *multilevel* and *multistore* models — are introduced in the following sections.

2.1.4.1. Multilevel processing model

A. INTRODUCTION.

To study the speech process in the temporal domain, neurophysiological indexes reflected by EEG or MEG are the best methods. During a word production, fine-grain spatiotemporal progression of lexical, grammatical, and phonological processing within Broca's area has been revealed by using the intracranial electrophysiology (Hagoort & Levelt, 2009; Sahin, Pinker, Cash, Schomer, & Halgren, 2009). The speech perception, similarly, is a complex processing containing multiple stages from acoustic input to syntactic process. Traditionally, serial model is proposed for the process of speech perception. Some neurophysiologic evidence also supports this model. For example, the acoustic information is processed at 20–200 ms ~~after a~~ stimulus' onset and is reflected by ERP component of P20–N100; the phonological information is processed at a delay of 100–400 ms, which is reflected by N100, N200; and, the lexical and semantic information is dealt with at around 400 ms reflected by N350 and N400 respectively. [Reviewed in (Pulvermüller & Shtyrov, 2006)]. However, more recent studies, by using MMN, show that psycholinguistic information—phonological, lexico-syntatic, and semantic information, as well as context integration—in speech perception is processed nearly simultaneously in this early time window (around 200 ms after information is available) [Reviewed in (Pulvermüller & Shtyrov, 2006; Pulvermüller, Shtyrov, & Hauk, 2009)].

These above models just roughly break down the stages of spoken sentence comprehension. Both the serial and near-parallel models of speech comprehension did not consider the fine-grain process from the acoustic signal to lexical presentations; neither did it explain the neurophysiological process of CP. The mapping from acoustic signals to lexical representations is mediated by a number of different levels of representation as shown in Figure 2.1.2, as Phillips stated: "*It is standard to distinguish at least the levels of acoustics, phonetics, and phonology in the representation of speech sounds.*" (Phillips, 2001)

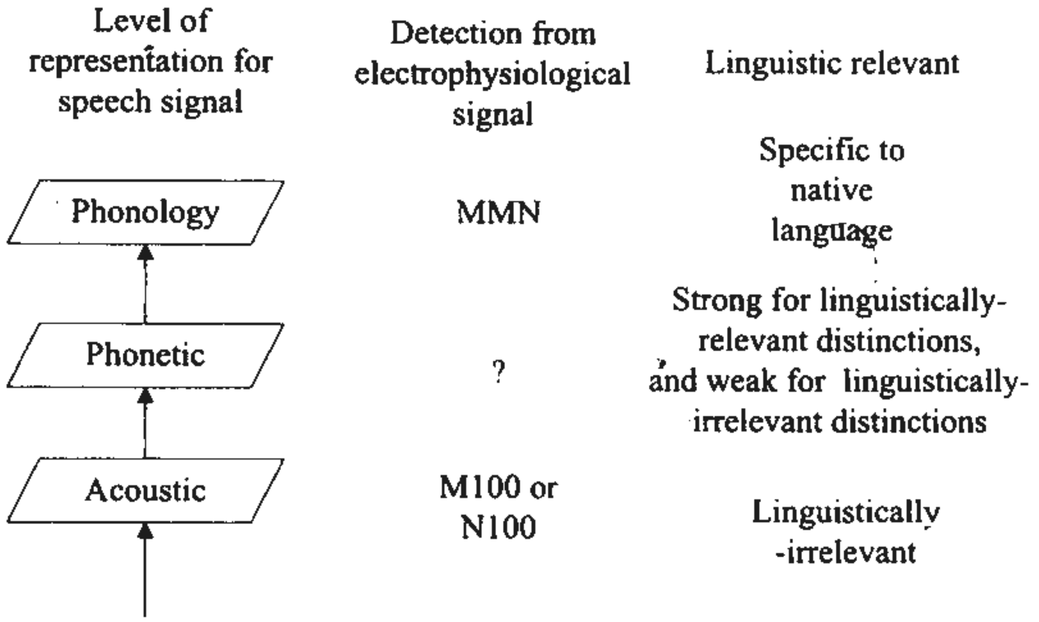


Figure 2.1.2 A multilevel model of CP. There are three levels of representation for mapping between acoustic signals and phonology categories. Electrophysiological signals to detect each level of representation are shown in the middle column. It is not clear what signal is for detections of phonetic level yet. The last column shows how each level of representation is relevant to language background. CP can be happened in all levels.

In this multilevel of presentation, the representation of the *acoustic* aspect of speech is believed not to be modified by exposure to specific languages, while the abstract *phonological* representations differ substantially across subjects of different languages. The intermediate *phonetic* stage may contain multiple representations, which present linguistically-relevant acoustic distinction more strongly, and present linguistically-irrelevant distinctions less strongly. The discontinuities of some acoustic features, whose CPs can also be observed in infants and nonspeech, are observed through N100 or M100. For example, the M100 latency varied with vowel category (*/a,i,u/*), specifically with vowel height (F1), but not with speaker (Poeppel, et al., 1997). The general 20 ms threshold on VOT feature is also reflected by N100, which is due to the inactivation period of the response (Sharma & Dorman, 1999). The phonological representation is uniform across the range of different phonological categories. One characteristic of such uniform, irrelevance of within-category variation, has been well documented by MMN (Kazanina, Phillips, & Idsardi, 2006; Phillips, et al., 2000). Nonetheless, other characteristics of uniform in phonological level (e.g. acoustic diversity of categories, neutralization of some sounds) do not gain convincing neurophysiological evidence yet. Although the MMN studies show evidence that phonological categories are available to the auditory cortex and in the early stage, they give no indication of how the categories are encoded.

B. COMMENTS.

The *multilevel* processing model, which is based on electrophysiological data, proposed a serial processing in speech perception. The first level is an acoustic representation which is linguistic irrelevant and the later two levels are linguistic relevant. Although in this model, CP can be present in all levels, the influence of linguistic background is only present on later two levels. Since phonology representation is discrete and happens in the last stage, the multilevel model predicts CP on all speech sounds. This prediction contradicts to what have been observed on vowel perception. Therefore, there is something missing in this model.

2.1.4.2. Multistore processing model

A. INTRODUCTION.

With the inter-linguistic behavioral study on speech and nonspeech, Xu et al. proposed a *multistore* processing model (in Figure 2.1.3.), which introduces a memory system, to explain speech perception. The memory system includes four types of memories: sensory memory trace, analyzed sensory memory, short-term categorical memory, and long-term categorical memory (Y. S. Xu, Gandour, & Francis, 2006).

According to the model, the sensory memory trace is derived from unanalyzed raw sensory data with around 300 ms integration time window. The analyzed sensory memory extracts the analyzed sensory codes, which include steady-state (e.g. pitch height), time-varying (e.g. pitch slope), and event-timing (e.g. VOT) etc. Both sensory memories deal with the information in a continuous way. The two memories differ in their life time and generators in the brain. The sensory memory trace only lasts around 300 ms, and the analyzed sensory memory lasts several seconds. Referring to Lü et al.'s MEG result (Lü, Williamson, & Kaufman, 1992), the sensory memory trace has a neural generator in the primary auditory cortex, and the analyzed sensory memory has a neural generator in the association auditory cortex. The relatively longer sensory store is necessarily required for context-coding.

The categorical processing is dealt with by both the short-term categorical memory and the long-term categorical memory. The short-term categorical memory captures only critical features of the stimuli that are used for perceptual categorization, so it is responsible for nonspeech CP and CP in animals. Short-term categorical representations can be permanently preserved in long-term memory pursuant to perceptual learning. This long-term categorical representation may serve as *templates* to be activated later by *bottom-up* matching of similar features. This long-term categorical representation may also provide *top-down* expectations in the encoding of short-term categorical memory that allow listeners to better direct selective attention to critical stimulus features or dimensions. Xu, et al. referred Luo et al.'s (Huan Luo, Husain, Horwitz, & Poeppel, 2005) study to show that the difference between short-term and long-term categorical memory lies in the auditory area. According to Luo et al.'s results, the long-term categorical memory is responsible for "natural" (long-term training) categorization for speech sounds, while the short-term categorical memory is responsible for "newly" learnt (short-term categorization) for nonspeech sounds.

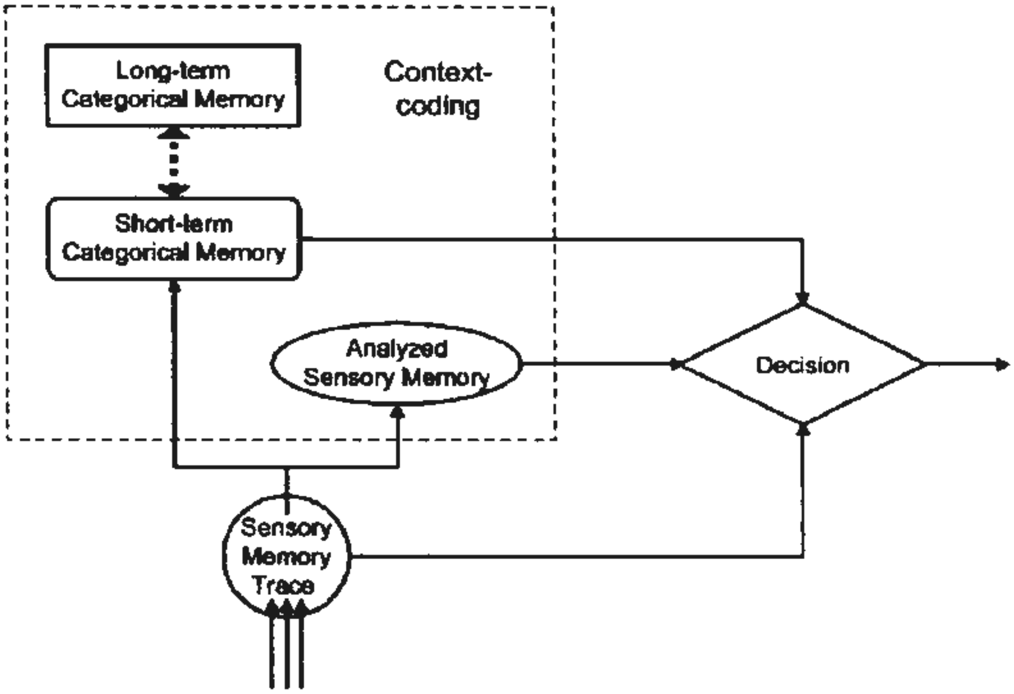


Figure 2.1.3 A multistore model of CP. It includes four memory stores: sensory memory trace, analyzed sensory memory, short-term categorical memory, and long-term categorical memory. Information is encoded in a hierarchical order but short-term categorical memory and analyzed sensory memory can be processed in parallel. All the sensory and short-term categorical components are subject to memory decay. The available memory traces after decay are input for decision-making. If long-term categorical memory is also available, it will interact with short-term categorical memory via both top-down and bottom-up mechanisms. All the memory components with relatively longer lifetime are involved in context-coding. [Taken from(Y. S. Xu, Gandour, & Francis, 2006)]

A. COMMENTS

In the model, the final decision comes from three ways: the sensory memory traces, or the analyzed sensory memory, or the short-term categorical memory which is modified by the long-term categorical memory. The *multistore* model can be viewed as an extension of the dual-process model, but it did not provide clear information on the temporal course of the speech processing. In addition to the missing temporal information in the *multistore* model, the information about how to do the final decision from these three memories is not clear too. Is it an on/off switch in each type of memory to reach the final decision, or is it an optimal judgment based on weights from different types of memory? In Chapter 5, I will return to the *multilevel* and *multistore* models and propose a combination model called *multistage* model to fill in the missing information.

2.2. Behavioral Methods to Study Categorical Perception

To study CP behaviorally, identification (labeling) and discrimination tasks are two typical methods. The testing objects used in both tasks are a continuum of stimuli varying their physical features in one or some dimensions with two end points unambiguously perceived as two categories. In an identification task, these stimuli are repeatedly presented to subjects in a random order for classification into one or another category. In contradict to the rather fixed paradigm for the identification task, discrimination task has various paradigms. Typically, an ABX or an AX paradigm is used In an ABX paradigm, stimulus A always differs from B, while, the X is either the same as A or as B. The subject is asked to discriminate stimulus X to be A or to be B. In an AX paradigm, the subject is asked to discriminate whether X is different from A.

The ideal CP has the following characteristics (Repp, 1984): (1) In the labeling function, there is a sharp boundary between two categories; (2) In the discrimination function, there is a peak at the category boundary; (3) In the discrimination function, within a category, the discrimination performance is at or near chance level; (4) The discrimination function can be fully predicted from the identification function.

However, in the practice, there is no ideal CP. Therefore, it is necessary to measure the degree of categoricalness, especially when we compare the performance from different sets of continua or when we compare the performance from different groups of subjects. In the below two sections, the analytic methods on the degree of categoricalness for identification function and discrimination function will be introduced.

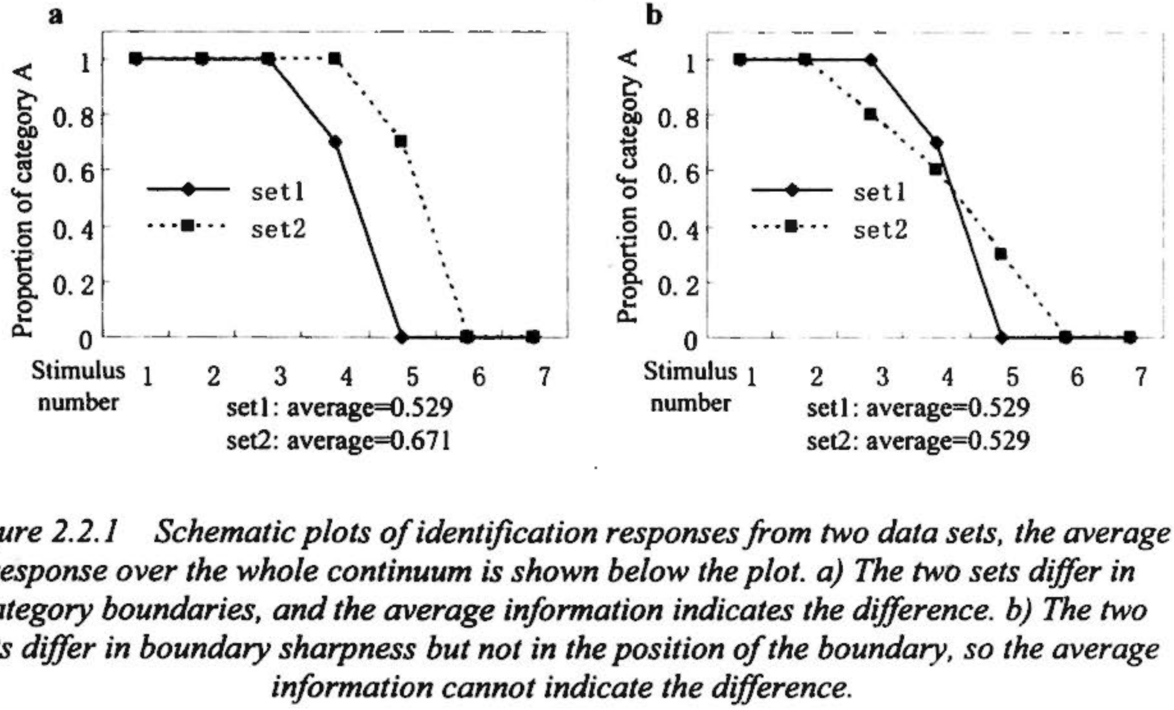
2.2.1. Identification Task

2.2.1.1. Response to each stimulus

To investigate whether a sound continuum is perceived categorically, three stimuli are required in a continuum to check the location of identification boundary. Therefore, response to each stimulus is a natural measurement. When there are two or more continua tested and each subject participates in more than one condition, the *repeated measure* (RM) *analysis of variance* (ANOVA) is preferred, with two within-subject factors: continuum and stimulus. On the other hand, when the data comes from two groups of subjects, a mix model is used where there's a between-subjects (grouping) factor and a within-subject factor (stimulus) (Keating, 2006; Max & Onghena, 1999).

2.2.1.2. Boundary steepness and boundary position

The statistical test on response to each stimulus is whether different labels are provided for different stimuli. There is no information about the degree of CP on different continua or conditions. Researchers generally have interests on the category boundary (the 50% crossover point), and the boundary steepness (the gradience, or the boundary width, or the range of stimuli spanned by the 25% to 75% responses). Instead of analyzing response to each stimulus, different methods to obtain the boundary steepness and position are introduced in this section.



set1: average=0.529
set2: average=0.671

set1: average=0.529
set2: average=0.529

Figure 2.2.1 Schematic plots of identification responses from two data sets, the average response over the whole continuum is shown below the plot. a) The two sets differ in category boundaries, and the average information indicates the difference. b) The two sets differ in boundary sharpness but not in the position of the boundary, so the average information cannot indicate the difference.

A. AVERAGE ALL RESPONSES OF A DATA SET.

First method is to average all the responses over different stimuli of one data set together, and then to compare the average responses between different data sets (Repp, Liberman, Eccardt, & Pesetsky, 1978). This method is effective when two data sets differ in the category boundary (Figure2.2.1a). However, in this method, there is no information about sharpness of category boundary, which is one of the central interests in the analysis of speech perception. If two data sets do not differ in the category boundary but in the boundary sharpness (Figure2.2.1b), the averaging method cannot distinguish them.

B. TRANSFORM TO PREDICTED DISCRIMINATION.

Second method is to transform the identification data into the predicted discrimination data by equation (2.1). Then, the comparison of sharpness of a category boundary can be done by the comparison of peakness on predicted discrimination data (Y. S. Xu, Gandour, & Francis, 2006).

Predicted discrimination:

$$P_{pred.disc(i,j)} = (1 + (P_{Ai} - P_{Aj})^2)/2 = (1 + (P_{Bi} - P_{Bj})^2)/2 \qquad (2.1)$$

Where, $P_{pred.disc(i,j)}$ is the predicted discrimination score on stimulus pair of (i, j). P_A and P_B are identification scores for category A or B respectively. P_{Ai} represents the proportion of stimulus i labeled as category A. For example, to calculate predicted discrimination score on stimulus pair 1-3, we can obtain $P_{pred.disc(1,3)} = 0.5 + (P_{A1} - P_{A3})^2/2 = 0.5 + (P_{B1} - P_{B3})^2/2$. After the score of predicted discrimination is calculated, the value and the position of the peak can be obtained from the curve.

C. CURVE FITTING: LOGISTIC.

The last method is curve fitting, which can be used to obtain both information of category boundary and category sharpness. Categorization data are generally not quite linear, nor exponential; they are generally more S-shaped. Logistic has this shape and can be constrained to fall between 0 and 1, so, it is the best theoretical match to the expected shape of the data (Keating, 2006). Figure2.2.2 shows the logistic functions presenting different degrees of gradience (Frisch, Pierrehumbert, & Broe, 2004). After the raw data has been transformed into logistic function, the two features: category boundary and boundary steepness can be obtained by the following equations (Frisch, et al., 2004; Keating, 2006):

$$P_{\log} = \frac{e^{b_0+b_1x}}{1+e^{b_0+b_1x}} \tag{2.2}$$

$$y = b_0 + b_1x \tag{2.3}$$

In equation (2.2), $P_{\log}$ is the smoothed version of raw data (or the proportion of one category), which has an S-shape, x is the input parameter, here refers to stimulus number. In equation (2.3), y is a linear regression function, and b_0 and b_1 are two parameters estimated from the raw data through "maximum likelihood".

In the estimation, the highest data value should be less than 1, and lowest data value should be greater than 0. So in practice, change "0" to 0.001 (or similar value), and correspondingly, "1" to 0.999 (or similar value). The **sharpness** information can directly be obtained from b_1 (Kutner *et al.*, 2005, p. 567)(Y. S. Xu, Gandour, & Francis, 2006). Even though b_1 is not the actual slope of data, it is related to the slope, with higher values reflecting shallower curves. The boundary information can be calculated through x from the equation (2.3) when $f(x)$=0.5. Therefore, the **position of CP boundary** can be obtained from equation (2.4)

$$\textbf{Position of CP boundary}: x_{cb} = \frac{-b_0}{b_1} \tag{2.4}$$

D. CURVE FITTING: PROBIT.

Another curve fitting approach is more in accord with general statistical practice: If data cannot be fitted by a line, try transforming the data to make them look linear, then fit a line. Z-scores has been used to transform the data to fit a line by linear regression and the boundary is where z=0 [(Eimas, Cooper, & Corbit, 1973; Miller & Liberman, 1979), also refers to (Keating, 2006)]

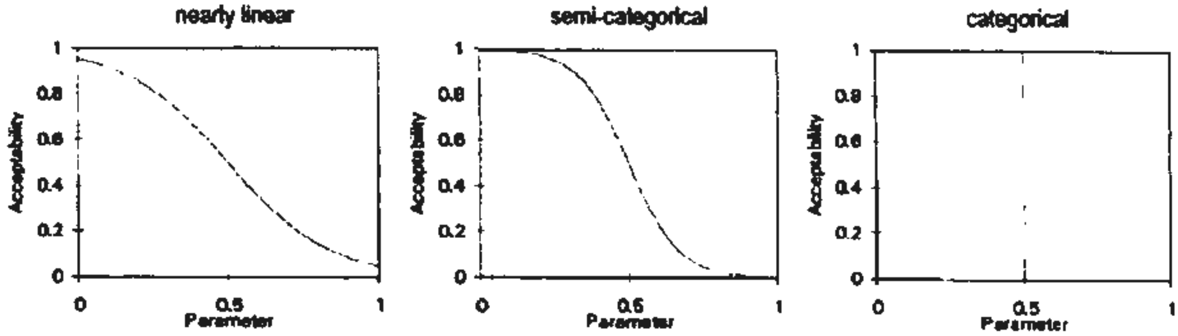


Figure 2.2.2 Three constraints with different degrees of gradience representing possible phonotactic constraints. [Taken from (Frisch, et al., 2004)]

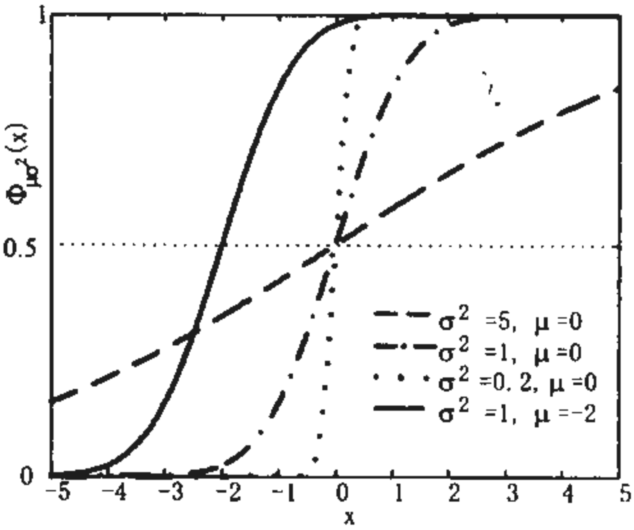


Figure 2.2.3 CDF function $\Phi_{\mu\sigma^2}(x)$ with different combinations of σ^2 and μ values.

Recently, a very popular curve fitting approach is *probit* transformation. It is very similar to logistic curve but has a steeper slope than a logistic curve, and is preferable for fitting very categorical responses. The probit model use inverse *cumulative distribution function* (CDF) associated with standard normal distribution. The CDF with normal distribution with *variance* σ^2 and *mean* μ can be written in (2.5)

$$\Phi_{\mu\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{\frac{-(t-\mu)^2}{2\sigma^2}} dt \tag{2.5}$$

The σ^2 determines the steepness of CDF function, and μ determines the center of the CDF function. See Figure 2.2.2 for the comparison of CDF function with different μ and σ^2.

When σ^2=1 and μ=0, the distribution is called standard normal distribution. Figure2.2.3 shows the CDF with different combinations of the parameters.

If the distribution is not the standard normal distribution, the linear function (2.6) is used to transform into a standard normal distribution and then the probit equation becomes (2.7). Rewrite the equation (2.6) by changing the parameters of σ^2 and μ into b_0 and b_1, the linear transformed function y is obtained in (2.8). CP boundary is x when $P_{probit} = 0.5$ in equation (2.7), or $y = 0$ in equation (2.8). Therefore, the boundary x_{cb} can be obtained from equation (2.9), the same as (2.4).

$$y = (x - \mu) / \sigma \qquad (2.6)$$

$$P_{probit} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y} e^{\frac{-u^2}{2}} du \qquad (2.7)$$

$$y = \frac{1}{\sigma} x - \frac{\mu}{\sigma} = b_0 + b_1 x \qquad (2.8)$$

Position of CP boundary: $x_{cb} = \dfrac{-b_0}{b_1}$ (2.9)

To avoid the negative quantities when P_{probit} is less than 50%, in practice, the equation (2.7) can also be rewritten as (2.10), and similarly, equation (2.8) can be rewritten in format of (2.11) and the position of CP boundary is in (2.12) [P23, (Finney, 1971)].

$$P_{probit} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y-5} e^{\frac{-u^2}{2}} du \qquad (2.10)$$

$$y = \frac{1}{\sigma} x - \frac{\mu}{\sigma} + 5 = b_0 + b_1 x \qquad (2.11)$$

Position of CP boundary: $x_{cb} = \dfrac{5 - b_0}{b_1}$ (2.12)

The boundary slope, or sharpness, can be obtained by boundary width, which is the span of x, when the value of P_{probit} between 25% to 75% in (2.7) or (2.10). From the equation (2.13), similar to the case of logistic approximation, b_1 is associated with the slope information. A higher the b_1 is associated with a steeper slope or a narrower boundary width.

Boundary width or sharpness: $$w = x_{p=0.75} - x_{p=0.25} = \frac{1.35}{b_1} \qquad (2.13)$$

2.2.2. Discrimination Task

Discrimination data can be shown by correctness of an AX or ABX response. A discrimination curve is obtained by connecting the responses to different pairs in order of the stimulus number. If a continuum is perceived categorically, there is a peak on the curve, which corresponds to the discrimination between categories.

In an ABX task, the correctness of a response is obtained by averaging all correct responses of ABA and ABB together. In an AX task, each unit is comprised of all the trials in four types of comparisons (AB, BA, AA, and BB). Adjacent comparison units contain overlapping AA or BB trials, (e.g., the 3-3 pairs were included in both 1-3 and 3-5 units in a two-step discrimination). Discrimination response D_{obt} for each comparison unit is defined by equation (2.14) [See also in (Y. S. Xu, Gandour, & Francis, 2006)].

$$D_{obt} = P('S'/S) \bullet P(S) + P('D'/D) \bullet P(D) \qquad (2.14)$$

The percentages of "same" ('S') and "different" ('D') responses of all the same (S) and different (D) trials (i.e., the correct responses) in each comparison unit are represented by two conditional probabilities, *P('S'/S)* and *P('D'/D)* respectively. *P(S)* and *P(D)* are the probabilities of *S* (AA or BB) and *D* (AB or BA) trials in each unit.

After the discrimination curve is obtained, there are some different methods to analyze the discrimination data and to do the comparison between different data sets. The first type of analysis on discrimination data is on **discrimination peak**: the location of the peak and the height of the peak. The location of the peak indicates the category boundary, while the height of the peak indicates the resolution of discrimination on the category boundary. The second type of analysis is the **comparison between across- and within-category**

discrimination. In practice, to improve the signal to noise ratio, the responses to all within-category pairs are averaged, and likewise to all across-category pairs, and these averages are compared (e.g. by paired t-test) (Pisoni, 1973; Y. S. Xu, Gandour, & Francis, 2006). The comparison of discrimination between within- and across-category provides information for the degree of categoricalness, with a greater difference indicating a higher degree of categoricalness. The third type of analysis is the **comparison between obtained discrimination vs. predicted discrimination** (Best, et al., 1981; Y. S. Xu, Gandour, & Francis, 2006). The comparison between predicted and obtained discrimination is rare and the results usually showed that the discrimination score on within-category is better in obtained function than that in predicted function (Keating, 2006; Liberman, et al., 1957).

2.3. ERP Components to Study Categorical Perception

2.3.1. ERP Components Correlating with Categorical Perception

Due to its high temporal resolution (millisecond), the ERP is a good method to investigate the temporal course of CP. The ERP, which directly reflects the brain activities, is often correlated with some cognitive functions. For example, N400 is a negative component correlated with easiness of semantic integration (Chope, Metzlutz, Wioland, Rumbach, & Kurtz, 1994; Kutas & Hillyard, 1980; Lau, Phillips, & Poeppel, 2008).

Similarly, there are some different ERP components correlated with CP. For example, in a CP study on facial emotional expression, three ERP components — bilateral occipito-temporal negativities (N170), the vertex positive potential (P150), and P3b wave — are found to be modulated by the types of deviants' category (Campanella, Quinet, Bruyer, Crommelinck, & Guerit, 2002). Reduction of N1m (the magnetic equivalent to N1 wave of EEG), which is measured by MEG, is observed when two values of VOT are across a category boundary than within a category (Simos, et al., 1998). The amplitude of N1-P2 complex is increased as perception improved after training. This change in waveform morphology are thought to reflect an increase in neural synchrony, as well as strengthened neural connections associated with improved speech perception (Tremblay, et al., 2001).

Nonetheless, some of the above described ERP components are modality dependent, such as the N170; and thus cannot be used to study the auditory processing. Furthermore, it is difficult to use them to study higher cognitive processing like language function because the language function associated with them is not clear yet. For example, although N1 is frequently studied in auditory experiment, the auditory N1 does not directly reflect some higher language functions. It is often referred to as an exogenous response, which is evoked by any acoustic stimulus with a well-defined onset, regardless of the listener's task or attentional state (Näätänen & Picton, 1987; Phillips, 2001). Moreover, the inter-cultural effect, which is very important in studies of speech perception, on the early ERP components is not clear or consistent. For example, although ERP components between 40 ms and 550 ms are influenced by the melodic structure type of the music phrase, only P3-like component is affected by cultural familiarity (Nan, Knösche, & Friederici, 2009).

In particular, some mid and late ERP components, such as MMN, N400, P600 etc., are found to be affected by language influence or to be correlated with some linguistic functions. Typically, the present thesis uses two ERP components (MMN and P300) under different types of conditions to investigate how different linguistically relevant factors modulate these two components; thus, the temporal dynamic of brain activities affected by these factors can be revealed. Both ERP components are elicited by a rare change in the auditory sequence, but they occur at different stages of speech processing: MMN occurs at the early speech processing without attention, and P300 occurs at the late speech processing with attention. More details about these two components are reviewed in the next two sections.

2.3.2. MMN

2.3.2.1. Characteristics of MMN

MMN is especially used to study brain activities in pre-attentive stage. The '*traditional*' MMN is generated by the brain's automatic response to any change in auditory stimulation exceeding a certain limit, which roughly corresponds to the behavioral discrimination threshold (Näätänen, et al., 2007). The MMN can be obtained experimentally by *passive oddball paradigm* without subject's attention. In an oddball paradigm, there are at least two types of stimuli presented in a block, where one type of stimulus occurs ('standard') more frequently than the others ('deviants'). To distract their attention, subjects are often guided to focus on another modality such as movies or books when presented with the auditory input. By subtracting ERP waves of deviants from those of standards in the passive oddball paradigm, a difference wave called MMN is obtained. In addition to the experimental paradigm, the topographic distribution and the temporal information also distinguish the MMN from other ERP components. The MMN has a frontocentral and central distribution (relative to a mastoid or nose reference electrode) and often peaks (negatively) at 100–250 ms after the change onset [Reviewed in (Kujala, Tervaniemi, & Schröger, 2007; Näätänen, et al., 2007)].

2.3.2.2. Mechanisms and generators of MMN

The MMN is typically recorded without subjects' focal attention. It can even be recorded during non-attentive states such as sleep (Atienza, Cantero, & Dominguez-Marin, 2002), or in coma (Fischer, Morlet, & Giard, 2000). This demonstrates that brain can respond to the change automatically (Garrido, Kilner, Stephan, & Friston, 2009).

Such early and automatic response leads to the proposal that MMN is the differential activation of the afferent N1 transient detectors for repetitive *'standard'* stimuli and '*deviant*' stimuli. The recent studies using *identity* or *control* MMN reveal that the traditional oddball MMN reflects both a non-comparison sensory processing and a comparison based cognitive processing (Maess, Jacobsen, Schröger, & Friederici, 2007; Näätänen, Jacobsen, & Winkler, 2005; Näätänen, et al., 2007). The *identity* MMN is obtained by subtracting ERP of deviants in the oddball block from that in a control block. The control block is a separated block containing the same number of standard stimuli and deviant stimuli. Moreover, the number of deviants in the control block is the same as the number of deviants in the oddball block. Consequently, by subtracting ERP of deviants in the oddball block from that in a control block, the refractory effects of the afferent neurons, which are introduced by the probability difference between the standard and the deviant stimuli in the oddball block, are extracted from the identity MMN. The *identity* MMN indicates that the elicitation of an MMN requires the presence of a memory representation of the standard. Therefore, two components — an early components (N1) and a late component (identity or control MMN) — are presented in the oddball MMN. The early component (N1) is induced by refractory effects and the later component (identity or control MMN) is induced by the comparison procedure. The neural generators for these two components are separated in the time domain but overlapped in the spatial domain. Both are located in the auditory cortex.

There is another neural generator of MMN located in the frontal lobe. Although the MMN is seldom affected by attention, the frontal generator of MMN is associated with an involuntary attention switching process, an automatic orienting response to an acoustic change (Escera, Alho, Schröger, & Winkler, 2000). The role of the prefrontal generators is supported by studies of patients. The patients with prefrontal lesions showed diminished temporal MMN amplitudes (Alain, Woods, & Knight, 1998).

Research also reveals that an MMN can be elicited not only by changes in physical properties, but also by changes in abstract rules, such as grammatical rule (Pulvermüller, Shtyrov, Hasting, & Carlyon, 2008; Shtyrov, Pulvermüller, Naatanen, & Ilmoniemi, 2003), stimulus patterns (Tervaniemi, Saarinen, Paavilainen, Danilova, & Naatanen, 1994), and cross-modality integration (Colin, et al., 2002) etc. [Reviewed in (Kujala, et al., 2007; Näätänen, et al., 2007; Pulvermüller & Shtyrov, 2006)]. In these so-called 'abstract feature' MMN studies, there is no physically identical, repetitive standard stimulus but rather a class of several physically different 'standard' stimuli. The invariant 'abstract' feature, uniting the various exemplars of the standard stimuli, is based on some common rules that they all obey (Näätänen, et al., 2007). With the help of MEG, the source of MMN in prefrontal or premotor area is found to have an 20–40 ms delay than that in the auditory cortex, when MMN is elicited by intelligent speech sounds [Reviewed in (Pulvermüller & Shtyrov, 2006)]. These studies suggest that the prefrontal source provides some top-down information to facilitate the automatic detection. In addition to the sources located in the prefrontal lobe and the temporal lobe, intracranial MMN recordings in animals suggest that at least in some species, MMN subcomponents also may be generated in the thalamus and the hippocampus (Alho, 1995).

In summary, the MMN is caused by at least two underlying functional processes: a sensory memory mechanism related to generators from the temporal lobe, and an automatic attention-switching process related to the generators from the frontal lobe. The sources in the temporal areas may be involved in processing changes of the sounds' physical properties or abstract rules; whereas, the sources on the frontal area may reflect reorientation of attention, which provides top-down prior expectations (Garrido, et al., 2009).

2.3.2.3. Applications of MMN

Although the MMN is mainly studied in auditory modality, it can be elicited from other sensory modalities such as the somatosensory modality (Akatsuka, et al., 2005), the olfactory modality [Reviewed in (Pause & Krauel, 2000)], and the visual modality [Reviewed in (Pazo-Alvarez, Cadaveira, & Amenedo, 2003)].

MMN reflects the accuracy of auditory discrimination, and it can be obtained even from inattentive subjects. Therefore, it becomes an especially attractive tool for studying various central auditory-system dysfunctions both in adults and children [Reviewed in (Näätänen, 2000, 2002, 2003)]. The most promising clinical application of MMN is in schizophrenia patients. Different types of abnormalities of MMN elicited from schizophrenia patients have been found to be associated with different types of dysfunctions. Therefore, MMN can be used as a diagnosis measure [Reviewed in (Näätänen & Kähkönen, 2008), also in (Garrido, et al., 2009)]. Another important clinical application is in the field of dyslexia. Dyslexia is currently thought, in the majority of the cases, to be result either from a dysfunction of the phonological system or a more general auditory deficit (Kujala & Näätänen, 2001). MMN can be elicited either by changes of abstract phonological rules or changes of acoustic properties. Therefore, through different types of manipulation, the MMN is useful to determine which aspects of auditory information are deficiently processed in dyslexia [Reviewed in (Kujala & Näätänen, 2001), also in (Garrido, et al., 2009)].

MMN is also a hot topic in the linguistic area, because it can be used to index the memory traces for phonemes and to probe the permanent language-specific speech-sound memory traces (Näätänen, et al., 1997). Different levels of linguistic processes are found to affect MMN, which indicates the early and automatic access of these linguistic processes and supports a near-parallel linguistic processing (Pulvermüller, Shtyrov, Ilmoniemi, & Marslen-Wilson, 2006).

These different levels include lexical access and selection, semantic processes, and syntactic processes. (1) At the lexical level, familiar sounds of one's native language and unfamiliar sounds can be distinguished through the size of MMN. For example, a same vowel elicited a larger size MMN by subjects whose native phonemic system has this vowel than by subjects whose native phonemic system doesn't have this vowel (Näätänen & Alho, 1997). (2) At the semantic level, unattended word stimuli elicited an activation sequence starting in the superior-temporal cortex and rapidly progressing to the left-inferior-frontal lobe (Pulvermüller, et al., 2005). More importantly, the spatio-temporal patterns of cortical activation depended on lexical and semantic properties of word stems and affixes, thus indicating that the MMN can give clues about lexico-semantic information processing stored in long term memory. (3) At the syntactic level, the size of MMN was found to reflect whether a word string conforms to abstract grammatical rules (Pulvermüller, et al., 2008; Pulvermüller, Shtyrov, et al., 2006). These results suggested that lexical, semantic and syntactic information can be processed by the central nervous system outside the focus of attention in a largely automatic manner, which favor the payroll processing of language. [More examples on application of MMN in linguistic information can be found from the following review paper: (Näätänen, et al., 2007; Pulvermüller & Shtyrov, 2006; Pulvermüller, et al., 2009)]

Moreover, since neither explicit responses nor attention is required to elicit an MMN, it is particularly suitable to study the language development in infants. For example, in a Finnish-Estonian cross-linguistic design, the obtained evidence suggested that the language-specific speech-sound memory traces develop between 6 and 12 months of age (Cheour, et al., 1998).

2.3.3. P300

2.3.3.1. Characteristics of P300

P300 is a positive ERP component, which peaks at 300 ms or more (up to 900 ms) after a stimulus. P300 is also called P3 or *late positive component* (LPC) in the literature. In addition to the latency, P300 also has a characteristic parielocentral scalp distribution (referencing to nose or mastoid). Although there are some other paradigms to elicit P300, the *oddball* paradigm, the same paradigm used in recording MMN, is a typical one (Polich, 2007). However, in contradict to MMN, which is recorded without subjects' focal attention and is irrelevant of tasks, the P300 wave only occurs if the subject is actively engaged in the task of detecting the targets. Therefore, it depends very much on the processing of the stimulus context and levels of attention and arousal (Linden, 2005; Polich & Kok, 1995). The P300 was first observed to be correlated with stimulus uncertainty (Sutton, Braren, Zubin, & John, 1965). Accumulating research work makes it clearer now that P300 reflects the discriminability of detection of change (Polich, 2007).

2.3.3.2. Mechanisms and generators of P300

P300 reflects brain activities that update the information from the memory when subjects pay attention to the stimuli (Donchin, 1981). The amplitude of P300 increases when the probability of the target is lower or when the target is easier to be discriminated. P300 amplitude indicates the amount of central nervous system activities, which are related to incoming information processing. Figure2.3.1 shows a *context-updating* model to explain the processing of P300 (Polich, 2003b). In the model, the sensory input is evaluated with attention. If the incoming stimulus is not different from what has been stored in the working memory, the current mental model is maintained, so that only N100, P200 and N200 are evoked. However, if a change is detected, the attentional process governs the updating of the stimulus representation in the memory, so that P300 is evoked (Polich, 2003b, 2007). The latency of P300 increases when targets are harder to be discriminated from standards, but not when response times increase for other reasons (Linden, 2005). In other words, the latency of P300 is correlated with reaction time only when the accuracy is emphasized.

P300 has two subcomponents—P3a and P3b. P3b, which has a parietocentral topographic distribution and is a bit late in time, is elicited by task-relevant deviant stimuli that are attended to. Whereas P3a, which has a frontocentral scalp topographic distribution and occurs a bit early than P3b, is elicited by stimuli more salient than the target without subjects' attention (task irrelevant) (Polich, 2003b, 2007). The P3a sometimes is also called "novelty" P300, which is traditionally elicited by inserting unique and highly salient (or novel) stimuli in the trains of repeated standard and target stimuli (Spencer, Dien, & Donchin, 1999). Since P3a is generally largest over the anterior and central recording sites, has a comparatively short peak latency, and rapidly habituates, it is thought to reflect frontal lobe function (Polich, 2004).

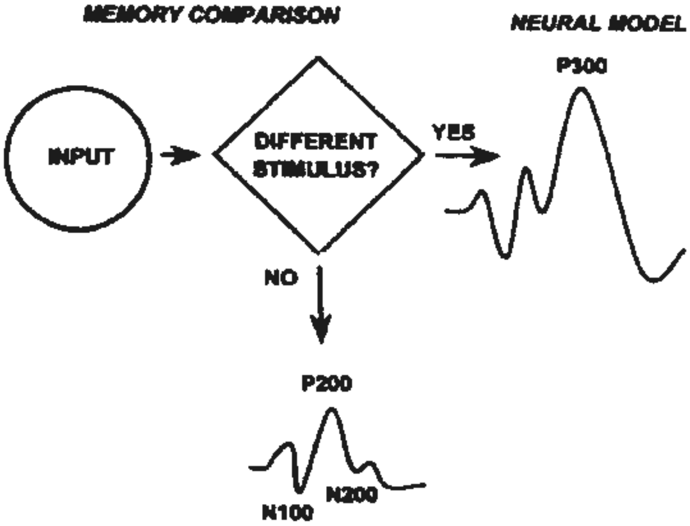


Figure 2.3.1 Schematic illustration of the P300 context-updating model. [Taken from (Polich, 2003b)]

The neural generators of P300 are not precisely determined even after three decades of research on this field (Linden, 2005). "*The P300 may have multiple intracerebral generators, with the hippocampus and various association areas of the neocortex all contributing to the scalp-recorded potential*" [(Picton, 1992), see also in (Linden, 2005)]. The intracranial recording studies revealed that multiple generators in neocoritical and limbic areas showed responses to rare stimuli when subjects paid attention to them and did the categorization (Linden, 2005). The areas include the hippocampus (located in medial temporal lobe) (McCarthy, Wood, Williamson, & Spencer, 1989), the inferior frontal sulcus (Baudena, Halgren, Heit, & Clarke, 1995), the inferior parietal lobe (Smith, et al., 1990), the superior parietal lobe (Halgren, et al., 1995), the temporal lobe (Halgren, et al., 1995), and the anterior cingulated cortex (Baudena, et al., 1995).

Furthermore, the neural generator for P3a and P3b may be different. The lesion studies revealed that "*the target-P300 seems to be mainly affected by temporoparietal junction lesions, whereas novelty responses are compromised in patients with a wide range of lesion sites, including the medial temporal, frontal, and parietal lobes*" (Linden, 2005). By using the event-related fMRI, the areas around the Sylvian fissure activate to infrequent targets (Linden, et al., 1999; McCarthy, Luby, Gore, & Goldman-Rakic, 1997). Although fMRI studies also found some different neural sources other than the Sylvian fissure for novel P300 than the target P300, the studies are yet to be confirmed [See also in (Linden, 2005)]. The EEG/MEG source localization combining with fMRI showed that the neural generator of P3a is more anterior than that of P3b, which are located in anterior superior temporal gyrus (Alho, et al., 1998), prefrontal cortex (He, Lian, Spencer, Dien, & Donchin, 2001), and anterior cingulate or supplementary motor area (Dien, Spencer, & Donchin, 2003). In summary, although it's still unclear about the neural generator for P3a and P3b, the converging data still "*provide some strong evidence that the inferior parietal lobe and the temporoparietal junction, particularly the supramarginal gyrus, in the generation of the P3a and P3b, whereas lateral prefrontal areas seem to contribute to the P3a only, in particular, that evoked by novel stimuli. More superior and posterior parietal areas showed higher target than distractor responses.*" (Linden, 2005)

2.3.3.3. Applications of P300

P300 latency reflects stimulus classification speed (Kutas, McCarthy, & Donchin, 1977), which is generally unrelated to the methods of overt response (Magliero, Bashore, Coles, & Donchin, 1984; McCarthy & Donchin, 1981). Moreover, P300 latency is independent of behavioral reaction time especially when the reaction time is affected by factors more than stimulus difficulties [(Verleger, 1997), see also in (Polich, 2004)]. Therefore, it can be used as a motor-free measure of cognitive function. The peak latency of P300 is shown to be correlated with age. Picton (1992) summarized the relationship between P300 peak latency and age. The function of this relationship has a "V" shape. Latency gets a minimal value in the middle of the age axis (around 18 years) and increases towards two end points (either with the subjects getting younger in the range 5-18 years or with the subjects getting older) [See also in (Polich, 2004)].

In addition to age (Walhovd & Fjell, 2001), P300 is also used as an indicator for general cognitive efficiency (i.e., how well an individual patient can process incoming information). It indexes some cognitive process that is related to the general mental ability measured on *intelligent quotient* (IQ) test (McGarry-Roberts, Stelmack, & Campbell, 1992). P300 is also correlated with many types of mental dysfunctions, such as schizophrenia, autism, dementia, alcoholism, Korsakoff's syndrome, etc. [See (Picton, 1992) for a detailed review]. However, since it is hard to determine how brain disorders affect the fundamental cognitive operations of attention allocation and immediate memory and thus influence P300 amplitude or latency, the application of P300 in the clinical measure should be careful (Picton, 1992; Polich, 2004).

Another important application of P300 is in the area of *brain-computer-interface* (BCI) (Birbaumer & Cohen, 2007; Farwell & Donchin, 1988; Sellers, Kübler, & Donchin, 2006). The BCI systems are especially developed to help patients, who lack of motor movements (such as paralyzed people), to communicate with others through brain waves. In the P300 based speller, as long as the subjects are able to pay attention to a specific target presented among a metric of selection, the P300 to the target will be able to be detected without the subject's motor movement. Therefore, the target of a subject's in mind can be identified by other people without the subject's overt movement.

Although P300 is not language specific, since P300 reflects the general cognitive stage including attention and memory, it can and has been used to investigate psycholinguistic issues (Kutas & Van Petten, 1994). (1) The P300 indicates social awareness. For example, vocal sounds directly reflect a communicative intent, so that intensity change of vocal but not nonvocal sounds is socially relevant. P300 showed that when subjects listened to a sequence of vocal or nonvocal sounds that contained rare deviants which differed from standards in sound intensity, the vocal sounds recruited enhanced sensory and attentional resources than nonvocal sounds (Schirmer, Simpson, & Escoffier, 2007). (2) The conceptual integration of text to the sentence can be indexed by P300 (Yang, Perfetti, & Schmalhofer, 2007). (3) Stroop effect, a psychology effect when subjects are asked to report the color in which a word is displayed, subjects are influenced by word meaning even though it is irrelevant, is a demonstration of automatic semantic interference. Although the reaction time is delayed in incongruent condition, the P300 latency remains stable. This suggests that the response-related processes are the primary source of Stroop interference (Ila & Polich, 1999; Szűcs, Soltész, & White, 2009). In summary, the combination methods of reaction time and the P300 enrich the understanding on some linguistic phenomena.

2.3.4. Comments on the MMN and P300

Both MMN and P300 are elicited from the oddball paradigm, although they are in different attentive stages and different time windows. The MMN reflects the automatically detection of violation in the pre-attentive stage, which characterizes the implicit, intuitive knowledge; while the P300, reflects the conscious detection of violation to the standard in perceptual stage, which characterizes the explicit knowledge (van Zuijen, Simoens, Paavilainen, Naatanen, & Tervaniemi, 2006). The same experimental paradigm but different attentional processes make MMN and P300 often to be studied together to examine the contrast between the automatic and controlled processing (Sussman, et al., 2004).

2.3.5. Neurophysiological Bases for Speech Perception

MMN and P300 are often examined together to study the model of speech perception. On one hand, the early studies suggested a three-factor (acoustic, phonetic and phonemic) hierarchy for speech perception (Pegg & Werker, 1997; Werker & Logan, 1985). Based on these studies, Dalebout and Stack (1999) suggested that MMN, P300, and behavioral responses could be used to test the proposition that there are different processing levels in the perception of speech in addition to attentional conditions (Dalebout & Stack, 1999). In this serial model, the auditory level of processing can be reflected by MMN. Secondly, the P300 indexes the phonetic/phonemic level, which is responsible for phonemic contrasts that exists in response to the native languages or phonetic contrasts of any language. Thirdly, there exists a higher order linguistic level wherein lexical, syntactic, and semantic knowledge mediates perception. This level is only shown in behavioral performance.

On the other hand, accumulating evidence has shown that the language influence has been reflected as early as in the time window of MMN (see section 2.3.2). Studies on a phonological task revealed both acoustic properties and phonological information influence the magnitude of P300 (Tampas, Harkrider, & Hedrick, 2005). These recent studies suggested the parallel or near parallel processing of acoustic and phonetic information.

CHAPTER 3 BEHAVIORAL EXPERIMENTS ON CANTONESE LEVEL TONES

3.0. Purposes of the Experiments

Three level tones which only exist in the Cantonese tone system were investigated by behavioral experiments. All the stimuli used here were synthesized by *pitch-synchronous overlap and add* (PSOLA) through a speech processing software — Praat (Boersma & Weenink, 2005). The PSOLA changes the pitch contour of a sound, while keeps the voice quality unchanged (Lemmetty, 1999).

There were four experiments in this chapter. The purposes of this series of experiments were to examine (1) whether Cantonese level tones are perceived categorically; (2) how the contextual sentences influence the perception of Cantonese tones; (3) how linguistic experiences of subjects influence CP; and (4) whether the degree of CP on the same pitch contours depends on the types of carrier syllables (i.e., real word, non-word, nonspeech)

3.1. Experiment I: to Examine the Context Effect

Only an identification task was carried out in Experiment I on Cantonese level tones. Native listeners participated in Experiment I to check whether Cantonese level tones are perceived categorically by native subjects. In this experiment, the position of the target tonal syllable was manipulated to have three types: in *isolation* (ISO) without a contextual sentence, at the beginning of a contextual sentence (*right context*, RC), and at the end of a contextual sentence (*left context*, LC). When the target syllables were in isolation form, they were tested in an absolute manner. That means the pitch contour of the target syllable was manipulated so that the target syllable is perceived as one of the three Cantonese level tones. When the target syllable was embedded in a sentence, it was tested in a relative manner. That means the target syllable can be perceived as one of the three Cantonese level tones depending on the pitch contrast between the target and the contextual sentence. This pitch contrast was obtained by changing the pitch contours of the sentence only (the target syllable was unchanged). Experiment I was conducted to examine the first two points proposed in section 3.0. The two points were repeated here: (1) whether Cantonese level tones are perceived categorically; and (2) how the contextual sentences influences CP of Cantonese level tones.

3.1.1. Method

3.1.1.1. Subjects

Thirteen native Cantonese speakers (7 male and 6 female, age: mean=23 years, SD=4.5 years)[b] volunteered to participate in the experiment. All of them are university students from the City University of Hong Kong. None of them reported a history of speaking or hearing disability. All of them are right handed according to self-report. Since this behavioral experiment was not aimed to test the lateralization pattern, no formal handedness test was done.

3.1.1.2. Stimuli

A. ISOLATION FORM.

Table 3.1.1 F0 value of stimuli in Experiment Ia (isolation form). Anchor syllables, which were used as filler, were naturally uttered speech. Synthetic syllables were based on stim#8, the base syllable.

	Mean F0 (Hz)			
Stimu#	Synthetic syllable		Anchor syllable	
1	189	(HighLevel)	189	(High-Level)
2	182			
3	175			
4	168			
5	161			
6	154			
7	147			
8	140	(Mid-Level, base syllable)		
9	133			
10	126			
11	119	(Low-Level)	119	(natural Low-level)

---

[b] To estimate the sample size is not an easy job. Different estimation methods are required for different types of statistical analysis methods (Eng, 2003). For *repeated measures analysis of variance* (RMANOVA), there is not a mature method to predict the sample size. Therefore, the number of subjects was chosen according to rule of thumb. In behavioral experiment, the number of subjects is between 10 and 20. The more the better, however, subjects are expensive, we should balance between the cost and the reliability of the results. In practice, we started from 10 and analyzed on the data. If there are some effects with low power effect but near significant, we will increase the number of subjects and test again to check the robustness of the significant effect.

Cantonese three level tone syllables [分 *fen*^{55} (to divide), 瞓 *fen*^{33}(to sleep), and 份 *fen*^{22} (to share)] were uttered by a native Cantonese subjects five times. The best utterance of each tone syllable, which was measured by two independent researchers, was chosen as the anchor syllable. Among the three anchor syllables, the mid-level tone was chosen as the base syllable. Eleven stimuli were synthesized by varying pitch contours from low-level to high-level based on the base syllable (mid-level syllable). The pitch value's of two end points (#1, high-level; and #11, low-level) were determined by two anchor syllable (naturally uttered high-level and low-level tones). Therefore, all the 11 stimuli differed only on pitch contours but they were the same in all other acoustic properties. The duration of each target syllable was normalized to 500 ms. Mean pitch values of the 11 stimuli, with 7 Hz step, were shown on Table.3.1.1.

B. WITH CONTEXTUAL SENTENCE.

The *target syllables* (TSs) [分 *fan*^{1} (to divide), 瞓 *fan*^{3} (to sleep) and 份 *fan*^{6} (to share)][c] were embedded in two types of contexts. One was the *left context* sentence (LC)—the target is at the end of the sentence) and the other was the *right context* sentence (RC)—the target is at the beginning of the sentence). The LC sentence was 呢個字係 *fan* (*ni*^{1} *go*^{3} *zi*^{6} *hai*^{6} *TS*, This word is *fan*) and the RC sentence was *fan* 係乜意思 (*fan hai*^{6} *mat*^{1} *ji*^{3} *si*^{1}, *fan* is what meaning). Durations of the sentences were 1.11 s for LC, and 1.15 s for RC respectively. The duration of the targets was 450 ms for LC and 350 ms for RC.

LC and RC stimuli were constructed and presented with the same procedure. Similar to those in ISO form, the three LC and RC sentences, which differed in target syllable but had the same contextual sentences, were uttered five times by a native male speaker. The best utterance of each sentence, which was measured by two independent researchers, was chosen as the anchor sentence. The sentence with the target of mid-level tone [瞓 *fan*^{3} (to sleep)] was chosen as the base sentence. Other stimuli were synthesized by varying the pitch contour of context syllables. The F0 difference (F0 dist.) between the carrier syllables and the target syllables determined what the target was perceived. Therefore, instead of using the pitch contours of the target as the anchor point, the mean F0 dist. between the

---

[c] All the superscript value represents for the tonal category in Cantonese. For example, *TS*^{1} is tone 1 target syllable with tone

contextual sentence and the target syllable was chosen as the anchor. The two anchors were from naturally uttered high-level and low-level sentences.

The difference in mean F_0 between the [*hai*^{6}] and [*fan*^{1}] was measured, which we called anchor 1. The values of the difference were 44Hz and 46Hz in LC and RC sentences respectively. Similarly, anchor 2 was obtained from the difference in mean F_0 between the [*hai*^{6}] and [*fan*^{6}]. They were -6 Hz and -4Hz in LC and RC sentences respectively. The values of two anchors were slightly modified according to the rule of distance between 55-33-22. In other words, the difference of F0 distance between the high-level sentence and the mid-level sentence was twice times than that between the mid-level and the low-level sentence.

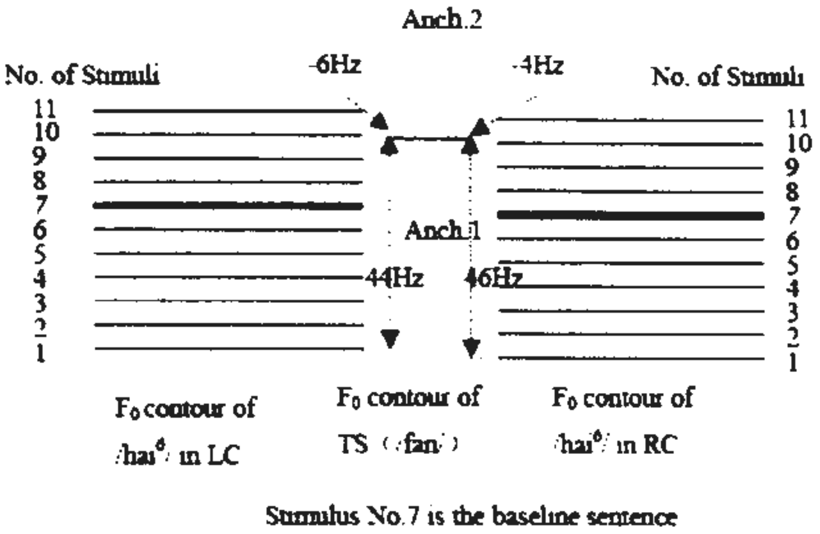
Table 3.1.2 Mean F0 value (Hz) of stimuli in Experiment Ib (with contextual sentence). Synthetic syllables were based on stim#7, the base syllable

Stim#	LC context	LC target	LC F0 dist.		RC target	RC context	RC F0 dist.
1	83		44	(High-Level)		90	46
2	88		39			95	41
3	93		34			100	36
4	98		29			105	31
5	103		24			110	26
6	108		19			115	21
7	113	127	14	(Mid-level, base)	136	120	16
8	118		9			125	11
9	123		4			130	6
10	128		-1			135	1
11	133		-6	(Low-Level)		140	-4

The difference in F_0 between anchor 1 and the baseline was equally divided at 5 Hz step. Hereby, we obtained 7 points as reference (No.1-7 in Tab. 3.1.1). Similarly, four reference points (No.8-11) were obtained between anchor 2 and the baseline. Totally there were 11 points including 2 endpoints. No.7 was the baseline point. Different stimuli or testing sentence were differed only by pitch contours of the carrier sentences, and all the other acoustic properties were kept the same. A schematic diagram of pitch contours was shown in Figure 3.1.1. Real pitch contours for LC and RC continua were shown in Figure3.1.2. The value of the pitch contours was listed in Table. 3.1.2. The F0 dist. is obtained by the equation (3.1):

$$\text{F0 dist.} = \text{F0}\ (TS)\text{-F0 (context)} \tag{3.1}$$

Where, the 'context' here referred to *hai*^{6}.



Stimulus No.7 is the baseline sentence

Figure 3.1.1 Stimuli continua structure of Experiment I. The step in the continua was 5Hz. We anticipated TS in stimulus No.1 to be perceived as [fan^1*] and TS in stimulus No.11 to be perceived as [fan*^{6}*].*

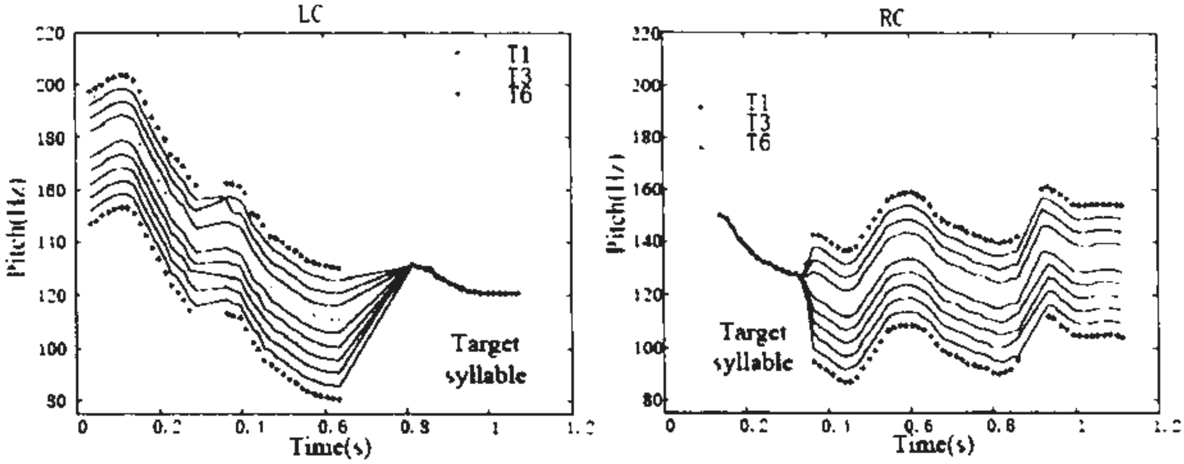


Figure 3.1.2 Pitch contours of LC (left context) and RC (right context) in Experiment I. We anticipated TS (target syllable) in stimulus No.1 (blue line) to be perceived as [fan^1*]. TS in stimulus No.7(green line) to be perceived as [fan*^{3}*], and TS in stimulus No.7(red line) to be perceived as [fan*^{6}*] .T1: tone 1; T3: tone 3; T6: tone 6. The dashed vertical lines in each plot are the boundaries of each word.*

3.1.1.3. Identification task

Every trial was constructed by two repeated identical stimuli, which were separated by a 500 ms *inter-stimulus-interval* (ISI). There was a 4 s *inter-trial-interval* (ITI) after every trial. The stimuli were presented to listeners via a loudspeaker. The presentation paradigm was shown in Figure3.1.3. Subjects were required to identify the target after a trial as one of the three target words (分 *fan*^{1}, 瞓 *fan*^{3}, or 份 *fan*^{6}) by circling one out of the three, even if they were not sure about the answer. There were13 types of trials in *Isolation* form including 11 synthetic stimuli constructed from the base syllable and two anchor stimuli, which were natural uttered high-level and low-level tone syllables as the fillers; while, there were 11 types of trials in context form. The subjects who failed to respond to anchor points in isolation form correctly were rejected from further data analysis. Each type of trial was repeated twice and all were presented to the subjects in a random order. Each of three conditions (Isolation, RC, and LC) contains five such blocks. There was a one minute's break after each block and a three-minute's break after each condition. Three conditions are counterbalanced among subjects. The whole experiments took 30 minutes. A sample sheet is attached in appendix 2.

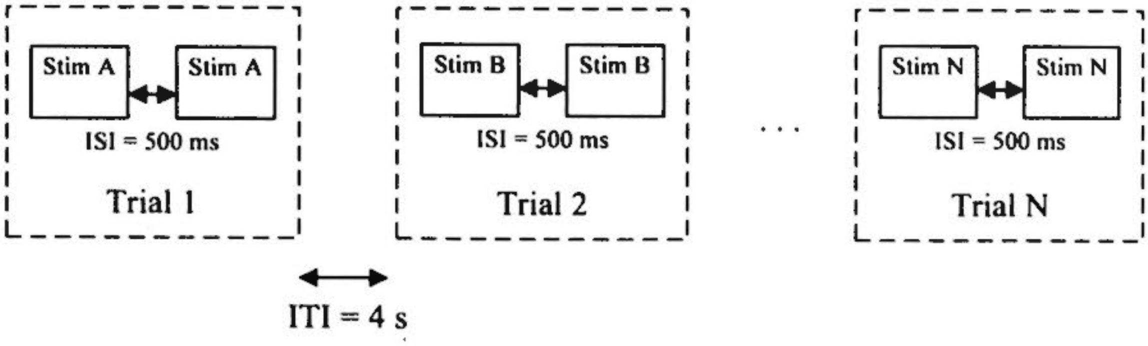


Figure 3.1.3 Presentation paradigm in identification task for Cantonese level tones.

3.1.2. Data Analysis

All thirteen subjects were included in the data analysis, because none of them failed to reach at least 75 % of correctness on the two anchor points in each continuum. Since the three continua with different target positions were perceived by the same subjects, all the statistical analyses were either *repeated measures analysis of variance* (RMANOVA) or paired *t*-test. The analyses were to test whether the three continua were perceived differently by the subjects, especially whether they showed different CP. All the statistical analyses were conducted through the software called SPSS (http://www.spss.com/) (Brace, Kemp, & Snelgar, 2003). When conducting RMANOVA, Greenhouse-Geisser adjustment was used when the assumption of sphericity was violated and the significant level was set at 0.05. To correct the type I error by multiple tests, the threshold value p was corrected by Bonferroni adjustment in pairwise t-test.

3.1.2.1. Identification curves.

The identification curve was obtained by sorting the proportion of response to each category in an order of the stimulus number. Smooth of the identification curve indicates how confident and consistent the subjects categorize the stimuli. The deviant of the first derivation of the identification curve was such an indicator. The equation (3.2) showed how to obtain the first derivation of the identification curve.

$$\Delta p_i = \left[p_{T1}(i+1) - p_{T1}(i)\right] + \left[p_{T6}(i) - p_{T6}(i+1)\right] \qquad i=1\sim10 \tag{3.2}$$

Here, only T1 and T6 were calculated, since the T3 is not monotonic increased and moreover, T3 can be uniquely determined by T1 and T6. i here, represented stimulus number. Since T1 and T6 had reverse direction of increasing along the continuum, they contributed different signs to the derivation. Since there were 11 stimuli on each continuum, the i was in the range of 1 to 10.

Larger Δp_i indicated change of adjacent identification responses more abruptly. Moreover, the larger *variance of* Δp_i over the continuum or subjects indicated less smooth of the identification responses.

3.1.2.2. Predicted discrimination.

Equation of (2.1) only deals with two categories. To deal with three category data in this experiment, the (2.1) was modified according to that described in (Francis, et al., 2003) and rewritten as equation (3.3).

$$\begin{aligned} p(pred.disc_{ij}) = 0.5 + 0.25\{&[p_{T1}(i) - p_{T1}(j)]^2 \\ &+ [p_{T3}(i) - p_{T3}(j)]^2 + [p_{T6}(i) - p_{T6}(j)]^2\} \end{aligned} \tag{3.3}$$

According to the crossover points in identification curves, the predicted discrimination value was grouped into across- and within- category results. Then the comparison was on these across- and within- category groups.

3.1.3. Results

3.1.3.1. Identification curves.

Identification responses averaged across listeners according to three target positions were presented in Figure 3.1.4. The figures showed that listeners exhibited crossovers in their identification function. Three categories were obtained along the continuum. The crossovers corresponded roughly to the expected location of boundaries between three tone categories (at the 3-5 and 9-11 pairs in both the isolation form and the LC continuum and at 4-6 and 7-9 pairs in the RC continuum[d]). Category boundaries in RC continuum are sharper than those in LC continuum and isolation form.

The identification curve in the isolation form was messier than that in the LC or in the RC continuum. A two-way RMANOVA was conducted on the delta value of identification responses (Δp_i) [3 target position (POSITION): Isolation, LC and RC × 10 *i*, the delta steps (STEP)] . There was no any main effect for POSITION or for STEP. However, there was an interaction effect [POSITION × STEP: F(2,18)=3.234, $p<0.001$], which revealed that patterns of the identification score between an adjacent pair were different for these three continua. Figure 3.1.5 showed the Δp_i of these three target positions. The *post hoc* analyses on each *i* over three continua by RMANOVA showed that when *i*=5 and *i*=6, the Δp_i significantly differed in three continua [*i*=5: F(2,24)=11.877, $p<0.001$; *i*=6: F(2,24)=5.165, $p=0.014$]. To correct the type I error by multiple tests, the threshold value *p* was divided by 11, so $p<0.005$ was the significant value. After the correction, only when *i*=5, the Δp_i significantly differed in three continua.

Another method to examine the smoothness of the identification curve was to check the *deviants of* Δp_i over the whole continuum for three types of positions. Therefore, there were three such values corresponding to each continuum for each subject. A RMANOVA

---

[d] Only two major crossovers (tone1and tone3, tone3 and tone6) were considered

on the factor of POSITION showed *variance of* Δp_i significantly differed in three continua [F(2,24)=5.714, *p*=0.020]. However, the *post hoc* pairwise comparison showed that none of them significantly differed from each other, although there is a trend that *deviants of* Δp_i in Isolation form is larger than in RC or LC condition [mean of *deviants of* Δp_i: ISO=0.617; RC=0.405; LC=0.382].

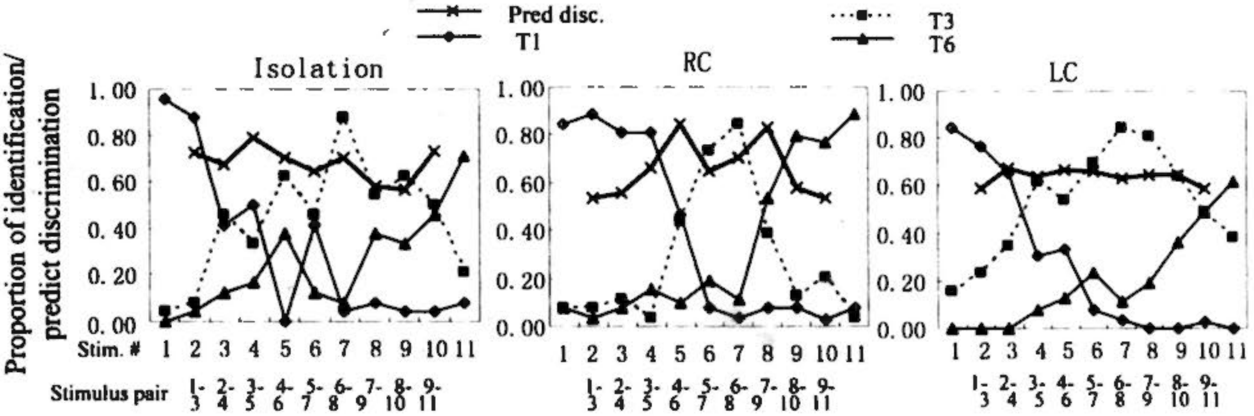


Figure 3.1.4 Results of natural speech based Cantonese level tones perceived by native subjects in Experiment I. The target syllables had three types of positions: in the isolation form (Isolation), at the beginning of the sentence (RC), at the end of the sentence (LC). A solid line with a cross represented predicted discrimination (Pred disc.) calculated from the identification results. A solid line with a circle, a dotted line with a square, and a solid line with a rectangle represented the proportion of the stimuli perceived as tone 1 (T1), tone 3 (T3) and tone 6 (T6) respectively.

Figure 3.1.5 The delta value of identification responses (Δp_i) in three continua with different positions of target in Experiment I. The solid line with a rectangle, the dash line with a diamond, and the dotted line with a square represented the Δp_i when it was in isolation form, in RC continuum and in LC continuum respectively.

3.1.2.2. Predicted discrimination.

A two way [3 POSITION ×9 stimulus pair (PAIR)] RMANOVA showed there were both a significant interaction effect and a PAIR main effect, but no POSITION main effect [POSITION × PAIR: F(16,192)=3.219, $p<0.001$; POSITION: F(2,192)=2.346, $p=0.117$; PAIR: F(8,192)=2.711, $p=0.01$]. The interaction effect indicated that predicted discrimination score over the nine pairs did not show the same pattern for three POSITION continua. A one-way RMANOVA on the factor of PAIR for the predicted discrimination showed that only RC continuum had a significant PAIR effect, while Isolation and LC continua had none [RC: (F(8,96) =7.311, $p<0.001$; LC: F(8,96)=0.422, $p=0.831$; Isolation: F(8,96)=2.094, $p=0.108$].

Table 3.1.3 Grouping information, grouped mean value, and statistical analysis of predicted discrimination scores in three continua of Experiment I. ** : $p<0.003$; * : $p<0.017$. (significant level $p<0.017$ after correction)

	RC		LC		ISO	
	mean	pairs	mean	pairs	mean	pairs
AC	0.838	4-6, 7-9	0.623	2-4, 9-11	0.702	2-4, 3-5, 9-11
WC	0.595	others	0.634	other	0.676	others
AC vs. WC paired t-test	t(12)=4.818, p<0.001***		t(12)=-0.328, p=0.748		t(12)=0.563, p=0.583	
AC	RC vs. LC	t(12)=4.420, p=0.001**				
	RC vs. ISO	t(12)=1.899, p=0.082				
	LC vs. ISO	t(12)=-1.567, p=0.143				
WC	RC vs. LC	t(12)=-1.718, p=0.111				
	RC vs. ISO	t(12)=-4.210, p=0.001**				
	LC vs. ISO	t(12)=-1.573, p=0.142				

To improve the signal to noise ratio, the nine stimulus pairs for each continuum were grouped into *within-category* (WC) and *across-category* (AC) pairs according to the crossover of identification curves. Therefore, the *p(pred.disc)* values of WC and AC were averaged as in Table 3.1.3 and plotted in Figure3.1.6. The RMANOVA of *p(pred.disc)* on two within-subject factors (3 POSITION × 2 CATEGORY: AC and WC) revealed a significant interaction effect and two significant main effects [POSITION: F(2,24)=5.152, $p=0.014$; CATEGORY: F(1,24)=18,087, $p<0.001$; POSITION × CATEGORY: F(2,24)=8.178, $p=0.002$]. The observed interaction effect indicated that patterns of the difference between AC and WC categories in three continua were different. To further check what kind of patterns differ between AC and WC in each continuum, three paired t-tests on the difference between AC and WC for each continuum were conducted. The results (listed in Table 3.1.3) revealed that only RC continuum had a significant effect on CATEGORY by showing a higher *p(pred.disc)* in AC than that in WC even after correction. The paired t-tests confirmed that CP was only observed in the RC continuum and absent in both the LC continuum and the Isolation continuum. Moreover, six paired t-tests revealed that RC continuum had a higher predicted discrimination than LC continuum only in across-category pairs but not in within-category pairs; on the contrary, RC continuum had a higher predicted discrimination than ISO continuum only in within-category pairs but not in across-category pairs. These results further supported that the RC continuum was perceived more categorically than the LC continuum, and the identification curve from the ISO continuum was less smooth than that from both LC and RC continua.

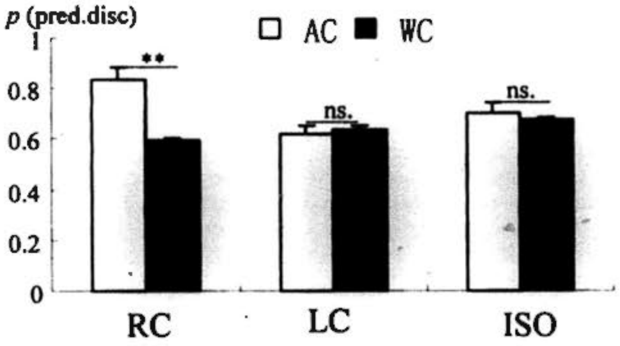


*Figure 3.1.6 Grouped value of predicted discrimination according to AC (Across-category) and WC(within-category) in three continua of Experiment I. **: p<0.01; ns. Non significant, RC: right context, LC: left context; ISO: isolation. Error bar represented standard error.*

3.1.4. Summary and Discussion

The result obtained from Experiment I confirmed our hypothesis that perception of Cantonese level tones is affected by the context. In this experiment, target syllables were kept unchanged in both the RC and LC continua, therefore, tone categories of the target syllables were totally determined by the context. Nonetheless, the subjects successfully categorized the continua into three tone categories. Moreover, the identification function was smoother when the targets were embedded in a sentence than when they were isolated. This smoothness was illustrated by a smaller deviance of Δp_i over the whole continuum in RC and LC conditions than in ISO condition. Furthermore, the larger deviance in ISO condition was not due to the sharper categorical boundary, since no CP was observed. The results showed that pitch changing in the adjacent syllable influences the perception of the TS. Therefore, categorization of tonal categories not only depends on the absolute pitch value of the TS but also on the pitch distance between the TS and adjacent syllables. The result in Experiment I is consistent with the findings from a previous study on Mandarin tones (Lin & Wang, 1984).

The context effect observed in Experiment I was not only restricted to the presence of a contextual sentence. The effect of the target position in a sentence was also observed. CP was only observed in RC continuum, while no CP was observed in LC or ISO continuum. These results confirmed the early findings (Francis, et al., 2003) that noncategorical perception of Cantonese level tones in isolation but categorical perception of them with a context. However, our finding extended that finding by further illustrated that CP was only observed in RC but not in LC condition.

The results obtained from Experiment I provided some suggestions to two questions raised at the beginning (1) whether Cantonese level tones are perceived categorically; (2) and (2) how a contextual sentence influences CP of Cantonese level tones. The results suggested that (1) the perception of Cantonese tones highly depends on the context; (2) the Cantonese level tones are perceived categorically only when they are with a contextual sentence, especially when the contextual sentence follows the target level tone; and (3) the position of contextual sentence also influences the degree of CP on Cantonese level tones, with increment of CP only in RC condition but not in LC condition.

However, there was a limitation in Experiment I. In this experiment, the stimuli were synthesized based on naturally uttered speech, therefore, mean F_0 of TSs in LC and RC continua were different. The difference of F_0 between the /hai^6/ and TS was also different for each corresponding stimulus in LC and RC sentences. Therefore, it was hard to tell whether the difference of categoricalness was due to context position or due to a higher mean F_0 value in RC continuum. Experiment II, where TS and F_0 distance were kept strictly equal, was carried out to rule out the influence of this latter factor.

3.2. Experiment II: to Reexamine the Context Effect

Experiment II was conducted to investigate further the mechanisms underlying the position effect observed in Experiment I, especially to investigate whether the position effect was due to the artifact of naturally uttered stimuli.

3.2.1. Method

3.2.1.1. Subjects

Twenty three native Cantonese subjects (12 male and 11 female, age: mean=22 years, SD=4.2 years)[e] volunteered to participate in this experiment. All are students from the City University of Hong Kong, with no reported history of speaking or hearing disability. All of them are right handed according to self-report. Since this behavioral experiment was not aimed to test the lateralization pattern, no formal handedness test was done.

3.2.1.2. Stimuli

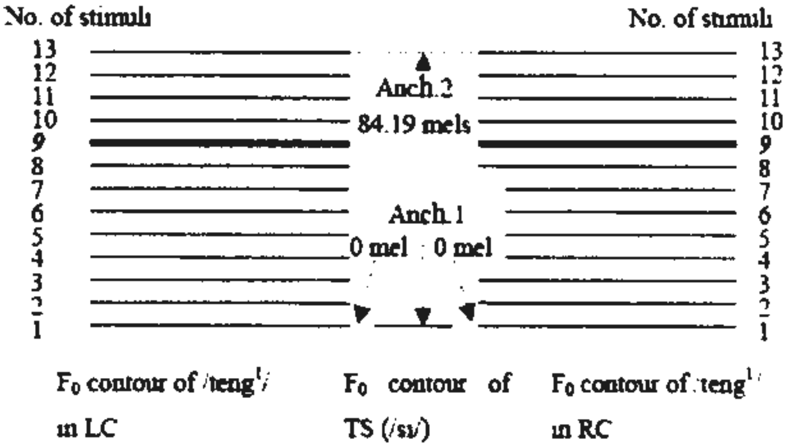
Table 3.2.1 Mean F0 value and F0 dist. of stimuli in Experiment II (with contextual sentence). Synthetic sentences were based on stim#9, the base syllable

Stim#	Context [*teng*^{1}] (Hz)	TS (Hz)	Step (Hz)		F0 dist. (Hz)	F0 dist. (*mel*)
1	133		-5.2	(High-Level)	0	0
2	138.27		-5.2		-7	-5.2
3	143.4		-5.3		-14	-10.4
4	148.7		-5.3		-21.1	-15.7
5	154		-5.3		-28.1	-21
6	159.3		-5.4		-35.1	-26.3
7	164.7		-5.4		-42.1	-31.7
8	170.1		-5.4		-49.1	-37.1
9	**175.5**	**133**	**0**	**(Mid-level, base)**	**-56.1**	**-42.5**
10	181		5.5		-63.1	-48
11	186.5		5.5		-70.2	-53.5
12	192		5.5		-77.2	-59
13	197.6		5.6	(Low-Level)	-84.2	-64.6

---

[e] We first enrolled 10 subjects in the experiment, however, the preliminary result show a marginal effect with low power effect. We later added more subjects to increase the power effect and reanalyzed the data.

Stimuli of Experiment II were resynthesized sentences based synthetic male speech generated from CUTALK (http://dsp.ee.cuhk.edu.hk/speech/cutalk) (a text-to-speech engine). Only two contextual sentences (RC and LC) were used in this experiment. The LC was presented by the sentence: [*tau*^{4} *sin*^{1} *nei*^{5} *teng*^{1} *TS* (You have just heard the word TS)]. The RC was presented by the sentence: [*TS teng*^{1} *dak*^{1} *hou*^{2} *hou*^{2} (The word of TS is heard quite clearly)]. The TSs were: [詩 *si*^{1} (poem), 嗜 *si*^{3} (hobby), and 事 *si*^{6} (thing)]. Six (three for LC and RC each) sentences were generated from the CUTALK. Three sentences differed in target syllables but had the same of contextual sentences. The engine generated the identical context sound of [*teng*^{1}], and three TSs. The three TSs differed only in F0 value and were identical in both LC and RC conditions. The duration was 280 ms for target syllable, 1.317 s for RC sentence and 1.344 s for LC sentence.



Stimulus No.9 is the baseline sentence

Figure 3.2.1 Stimuli continua structure of Experiment II. The step in the continua was 7.0 mels. We anticipated TS in stimulus No.1 to be perceived as /si^1*/ and TS in stimulus No.13 to be perceived as /si*^{6}*/. Stimulus No.9 was the base sentence, based on which the continua were synthesized.*

Following the similar procedure described in section 3.1.1.2, thirteen stimuli were synthesized for each continuum. Stimulus No.9 was the base sentence. Two anchors in LC and RC conditions were obtained by measuring the F0 dist. between the context ([*teng*^{1}]) and TSs. The distance between the two anchors was larger than that in Experiment I, so *mel* scale was chosen for step unit. The F0 difference (F0 distance) between the carrier syllables and the target syllables determined what the target was perceived. The calculation of F0 dist. was the same as in equation (3.1), and $F_0(TS)$=133Hz was kept constant. The F0 dist. step between adjacent stimuli was 7 *mels*. The information of pitch distance was listed in Table

3.2.1. Figure 3.2.1 and Figure 3.2.2 showed the schematic and real pitch contours respectively.

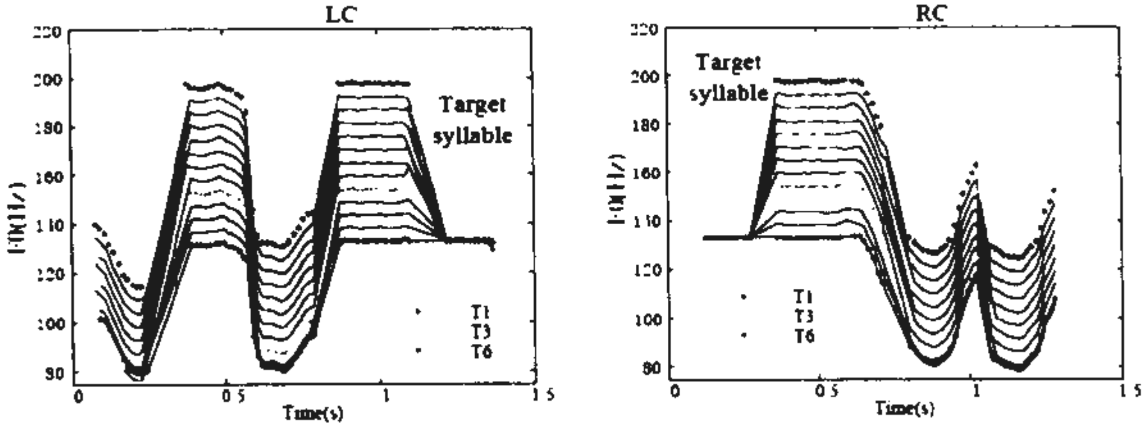


Figure3.2.2 Pitch contours of LC (left context) and RC (right context) in Experiment II. We anticipated TS (target syllable) in stimulus No.1 (blue line) to be perceived as [si^1], TS in stimulus No.9(green line) to be perceived as [si^3], and TS in stimulus No.9(red line) to be perceived as [si^6] .T1: tone 1; T3: tone 3; T6: tone 6.

3.2.1.3. Identification task

Thirteen tokens in each of the RC and LC continua were randomized into a block with 1.5s ITI. There were nine such blocks for each continuum. In other words, each trial was repeated nine times. There was a one-minute's break between blocks and a three-minute's break between conditions. The first block was a practice block, which was not counted into the response. Two sets of continua were counterbalanced among subjects to reduce the order effect. In contradict to Experiment I, one trial in Experiment II contained only one stimulus without a repetition as shown in Figure 3.2.3. Subjects heard the stimuli over a SONY headphone (MDR CD-777) and conducted the identification task following the same procedure as in Experiment I described in section 3.1.1.3B. The whole duration was 30 minutes.

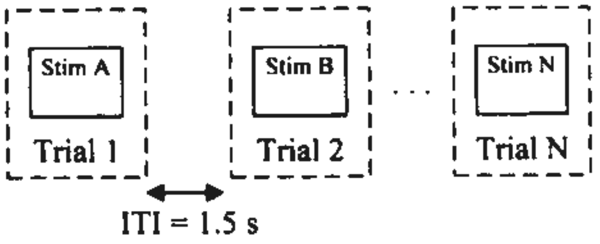


Figure 3.2.3 Presentation paradigm in identification task for Cantonese level tones (synthetic based stimuli).

3.2.2. Data Analysis

Similar to Experiment I, all the statistical analyses were either RMANOVA or paired *t*-test. When conducting RMANOVA, Greenhouse-Geisser adjustment was used when the assumption of sphericity was violated and the significant level was set at 0.05. To correct the type I error by multiple tests, the threshold value *p* was corrected by Bonferroni adjustment in pairwise *t*-test.

The analysis procedures were similar as in Experiment 1. The procedure includes (1) whether three categories were obtained in each continuum; (2) whether the identification curves had equal smoothness in LC and RC, which indicates the consistency of the category classification; and (3) whether the degree of CP was different in LC and RC conditions. The dependant variables for the above three points were identification scores (identification curves), Δp_i (or deviance of Δp_i), and the predicted discrimination scores respectively.

3.2.3. Results

3.2.3.1. Identification curves.

Identification responses averaged across listeners according to three target positions were presented in Figure 3.2.4. The figures showed that listeners exhibited crossovers in their identification function. Three categories were obtained along each continuum. The crossovers corresponded roughly to the expected location of boundaries between the three tonal categories (at the 3-4 and 9-10 pairs in LC continuum and at 1-2 and 6-8 pairs in RC continuum). Category boundaries in RC continuum were sharper than those in LC continuum and isolation form.

Similar to Experiment I, a two-way RMANOVA was conducted on the delta value of identification responses (Δp_i) [2 target position (POSITION): LC and RC × 12 *i*, the delta steps (STEP)] . There was a significant STEP main effect [F(11,242)=3.568, $p<0.001$] but no significant POSITION main effect [F(11,242)=0.037, $p=0.848$]. The interaction effect was also significant [POSITION×STEP: F(11,242)=3.075, $p=0.001$], which revealed that the patterns of the identification score between the adjacent pair were different for these two continua. Figure 3.2.5 showed the Δp_i from both continua with different target positions. The *post hoc* analyses on each i over these two continua by paired t-test showed that when i=1, 5, 6 and 10, the Δp_i significantly differed in two continua [i=1: t(22)=2.463, p=0.022; i=2: t(22)=-3.022, p=0.006; t(22)=2.736, p=0.012; t(22)=-2.498, p=0.020]. To correct the type I error by multiple tests, the threshold value p was divided by 13, so p<0.0017 was the significant value. After the correction, no significant difference of Δp_i between these two continua was obtained.

The second method to examine the smoothness of the identification curve was to check the *variance of* Δp_i over the whole continuum in the two position types. A paired t-test between two continua showed *variance of* Δp_i was not significantly differed for two continua [t(22)=1.31, p=0.204].

Both types of analysis on identification curves did not show evidence that there was difference of smoothness between these two continua.

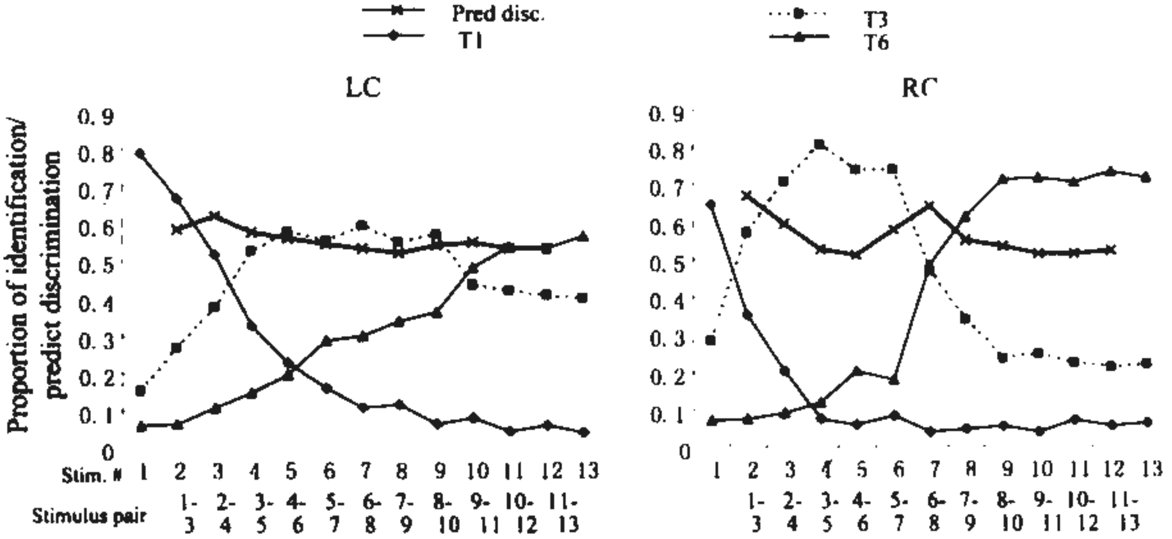


Figure 3.2.4 Results of synthetic based Cantonese level tones perceived by native subjects in Experiment II. The target syllables had two types of positions: at the beginning of the sentence (RC), at the end of the sentence (LC). A solid line with a cross represented predicted discrimination (Pred disc.) calculated from the identification results. A solid line with a circle, a dotted line with a square, and a solid line with a rectangle represented the proportion of the stimuli perceived as tone 1 (T1), tone 3 (T3) and tone 6 (T6) respectively.

Figure 3.2.5 The delta value of identification responses (Δp_i) in two continua with different positions of target in Experiment II. The solid line with a diamond and the dash line with a rectangle represented the Δp_i in RC continuum and in LC continuum respectively.

3.2.3.2. Predicted discrimination.

The predicted discrimination scores for 11 stimulus pairs from each continuum were grouped into *within-category* (WC) and *across-category* (AC) pairs according to crossovers of identification curves. The *p(pred.disc)* of WC and AC were averaged according to the groups described in Table 3.2.2 and plotted in Figure 3.2.6. The RMANOVA of *p(pred.disc)* on two within-subject factors (2 POSITION × 2 CATEGORY: AC and WC) revealed a significant interaction effect and two significant main effects [POSITION: F(1,22)=8.007, p=0.01; CATEGORY: F(1,22)=16,551, p=0.001; POSITION × CATEGORY: F(1,22)=13.645, p=0.001]. The results indicated that the overall predicted discrimination from AC is better than WC. In other words, there was an abrupt boundary in the identification curve. The observed interaction effect suggested that the patterns of difference between AC and WC categories in both continua were different.

Table 3.2.2 Grouping information, grouped mean value, and statistical analysis of predicted discrimination scores in two continua of Experiment II. **: $p<0.005$;*: $p<0.025$ (significant level $p<0.025$ after correction).

	RC		LC	
	mean	pairs	mean	pairs
AC	0.657	1-3, 6-8	0.573	2-4, 3-5, 8-10, 9-11
WC	0.539	others	0.545	other
AC vs. WC paired t-test	t(22)=4.26, $p<0.001$**		t(22)=2.137 p=0.044	
AC	RC vs. LC	t(22)=3.288, p=0.003**		
WC	RC vs. LC	t(22)=-1.293, p=0.210		

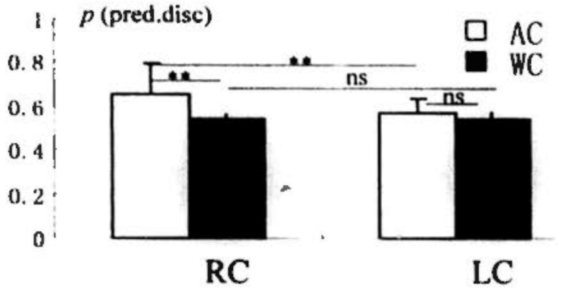


*Figure 3.2.6 Grouped value of predicted discrimination according to AC (Across-category) and WC (within-category) from LC(left context) and RC(right context) continua in Experiment II. **: p<0.005; *: p<0.025; ns. Not significant. Error bar represented standard error.*

To further look into what kind of difference, a paired *t*-test on the difference value between AC and WC for each continuum was conducted. To correct the type I error due to the multiple t-test, significant level is set at $p<0.025$. The results (listed in Table 3.2.2) revealed that only RC continuum had significant effects on CATEGORY by showing a higher *p(pred.disc)* in AC than that in WC. Moreover, RC and LC continua differed in AC pairs only. The paired *t*-test confirmed that CP was only observed in RC continuum and absent in LC continuum.

3.2.4. Summary and Discussion

The comparison of the categorical boundaries showed that they were sharper in RC than in LC continuum. These results replicated the position effect observed in Experiment I. Experiment II improved Experiment I with better control of stimuli by using synthetic speech. Therefore, the target syllable and the immediate neighbor syllable were the same in LC and RC continua. With such control, the observed categoricalness difference between LC and RC could not be attributed to the acoustic difference of target stimuli or F0 height in Experiment I.

One of the major differences between the LC and RC continua in Experiment I and II was the timing of perceiving the target tone. This timing difference was related to the auditory memory as reviewed in the chapter 2. When the target was presented in LC continuum, the reference of the pitch range has been set up, while the reference has not been set up in RC continuum. Therefore, in LC condition, the listeners could immediately identify which tone they had perceived, therefore, they could pay more attention to the acoustic details. On the contrary, in RC condition, the listeners could only identify the tone category of the target after the later occurring context. In the later case (RC condition), the auditory memory decayed more than that in the LC condition, especially when there was an interference sound, the context, before the subject made the judgment. The auditory memory theory predicts the same asymmetric effect will be observed in nonspeech syllables.

The second major difference between LC and RC continua in both Experiment I and II was the naturalness of the transition of pitch contours between the target and the context. According to the study of tone production, pitch contours of the context affect the pitch contours of the target syllables. Therefore, depending on the tonal types of the context, the same target will have different pitch contours on the transition point. Furthermore, this influence is more prominent when the context precedes the target than when the context follows the target. This is also referred as that the anticipatory effect is smaller than carryover effect in tone production (Y. W. Wong, 2006; Y. Xu, 1994, 1997). Consequently, the degree of naturalness of LC continuum should be less than that of RC continuum, when the synthetic procedure was only simply to shift the context upwards and downwards in both Experiment I and II.

According to different types of perception theories, there are two contradicting anticipations of effects of the pitch transition. In the motor theory, the speech perception is the perception of the intention of articulation gestures or the perception of neural command. In other words, perception is invariant and irrespective of the actual acoustic features. This is why the motor theory can be used to explain the invariant perception in coarticulation conditions (Liberman & Mattingly, 1985). The motor theory considered the link between production and perception is innate but not learnt. Consequently, no matter how the surface acoustic feature changes, as long as the changing is consistent with the production constraints, the perception is invariant (Liberman & Mattingly, 1985). These production constraints are primary due to coarticulation with different types of contexts. Therefore, the position of the context or the content of the context should not affect CP, since the neural command is the same in both conditions, even though surface acoustic features due to the realization are different in these two conditions. In other words, the motor theory anticipates a symmetric CP between LC and RC with a natural transition.

On the other hand, the *multistore* model suggests the language experience influencing CP. The linguistic experience makes the pitch contour of the target in LC more similar to that of the context than that in RC. Therefore, listeners are more ambiguous on identifying the target when it is in LC than when it is in RC condition. When the pitch contour of the target is less similar to the target, more reliable and consistent tonal category will be obtained. Under this model, we will predict that more natural pitch transition between the target and the context will reduce CP in LC. However, since the transition is not so prominent in RC, CP will not be influenced too much by it. In other words, the *multistore* model predicts a more natural transition will induce more asymmetric CPs between LC and RC. Moreover, the model also predicts that such asymmetric CP will be less when the listener without a corresponding linguistic experience.

In the following two experiments (III and IV), the above hypotheses were tested. Experiment III was designed to investigate how the pitch transition affects the asymmetric effects of CP due to the position of contexts. The influence from carrier syllables was also studied in Experiment III. There were three types of carrier syllables tested in Experiment III: real words, non-words, and nonspeech sounds. The exposure or the experience of the native listeners on the three syllables is decreased accordingly. Through such manipulation, language influence can be investigated.

Experiment IV was designed to further study the language effect by a cross-language study. Three groups of subjects participated in the experiment: native subjects, subjects who speak another tone language, and subjects who do not speak a tone language. Similar to the effect of the carrier syllable, the language effect on these three groups of subjects is decreased accordingly. Nonspeech carriers were also employed in Experiment IV. We anticipated subjects without a tone language experience would be less affected by the carrier syllable than the native Cantonese subjects.

Comparison between Experiment I and Experiment II on stimuli construction and the number of subject indicated that more consistent category boundaries were obtained and smaller number of subjects was needed with the same power effect in Experiment I. Moreover, the following experiments were conducted to investigate how the naturalness influences CP. Therefore, all the stimuli in Experiment III and IV were synthesized from the base sentences in Experiment I.

3.3. Experiment III: to Check the Effect of Syllable and Pitch Transition

3.3.1. Method

3.3.1.1. Subjects

Twelve native Cantonese subjects (seven male and five female, age: mean=21 years, SD=2.6 years), were paid to participate in the experiment. All are students from the City University of Hong Kong, with no reported history of speaking or hearing disability. All of them are right handed according to self-report. Since this behavioral experiment was not aimed to test the lateralization pattern, no formal handedness test was done.

3.3.1.2. Stimuli

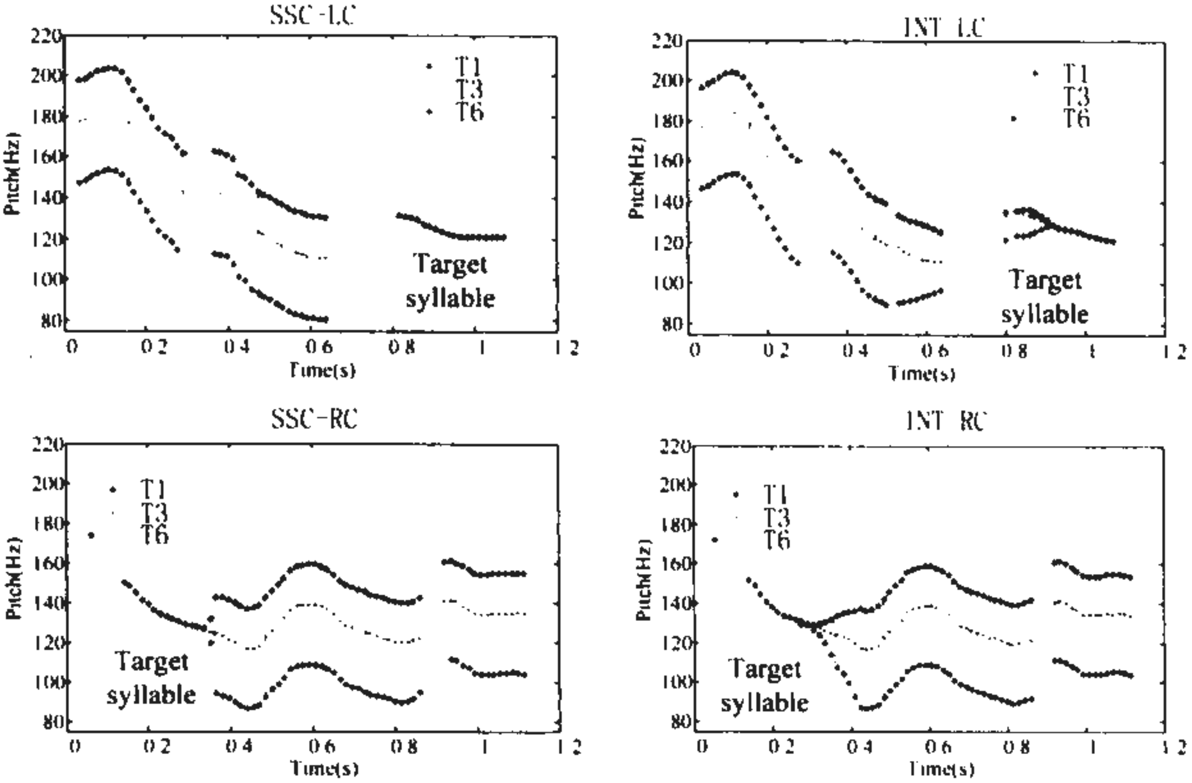


Figure 3.3.1 Pitch contours used in Experiment IIIa. SSC: simply shifted continuum; INT: interpolation continuum; LC: left context; RC: right context; T1: tone 1; T3: tone 3; T6: tone 6. The dashed vertical lines in each plot are the boundaries of each word.

A. MANIPULATION OF PITCH TRANSITION

To study the effect of the pitch transitions, two sets of stimuli were needed. One set of stimuli, simply shifted (SSC) continuum, was the same set used in Experiment I. The second set of stimuli, interpolation (INT) continuum, was synthesized by the following four major steps. Firstly, the three uttered sentences with three targets were normalized on pitch contours to let the pitch contours of contexts in three sentences be similar and on the same height level. Secondly, shift the naturally uttered tone-1 and tone-6 sentences upwards or downwards to make the pitch height of the target syllable in these two sentences similar as that in the tone-3 sentence. Thirdly, two anchor sentences were obtained by shifting the context of the tone-3 base sentence upwards and downwards. After shifting, the pitch contours of the altered contexts was similar to that in the shifted version of tone-1 and tone-6 sentences, while the main pitch contour of the target was kept unchanged. Then, the pitch transition between the context and the target was modified to be analogous to those in shifted tone-1 and tone-6 sentences. Finally, other stimuli were obtained by interpolating the pitch contours between the two anchor sentences. Figure 3.3.1 showed the pitch contours of 11 stimuli used in Experiment IIIa.

B. DIFFERENT TYPES OF CARRIER SYLLABLES

In addition to the real words [*fan*] (SSC set), a nonword sound [*fe*] (NWS set) and a nonspeech sound (saw wave complex tones, CPT set) were also used in Experiment III to investigate the effect of carrier syllables. Three sets of stimuli only differed in the spectral information of the carrier syllables, while all other aspects (i.e., the context, the pitch contours, the duration, and the amplitude) were kept the same. The nonword sound has a correct Cantonese phonological rule, but does not represent any semantic meanings. The nonword sound was synthesized by concatenating the consonant portion from the syllable [*fan*] and the vowel portion from a naturally uttered syllable [*se*] (Eng. to write down). The nonspeech sound was synthesized by concatenating the consonant portion from the syllable [*fan*] and a saw wave. Both the nonword and the nonspeech sounds were normalized on the amplitude profile and duration. Therefore, except for spectral information, these two sets of stimuli were kept as similar to the set of real words as possible.

3.3.1.3. Identification task

Identification task was only applied when the pitch transition was investigated (SSC and INT sets). The procedure was the same as that described in Experiment II. Four subjects (three male and one female) were excluded from further analyses due to their failure to reach identify the category of two endpoints on the practice round in either set.

3.3.1.4. Discrimination task

After the identification task, discrimination task was applied on all the four sets (SSC, INT, NWS, CPT) in Experiment III. The four sets were presented to subjects in a counterbalanced order and tested in two days. Stimuli were presented to listeners over a SONY headphone (MDR CD-777). In each set, the stimuli were consisted of all pairwise combinations of individual sentences separated by zero or two tokens along the continuum, with an ISI of 500 ms. There were a total of 29 such pairs for each of LC and RC continua. Twenty-nine pairs repeated each three times were distributed into 15 blocks randomly. The subjects were instructed to select 'yes' or 'no' on paper to indicate whether two TSs in a pair were the same or different. Each block contained 6 pairs (with 6s ITI), except for the last block which only contained 3 pairs. There was a one-minute's break between blocks and a three-minute's break after 15 blocks. The duration for one experiment was one hour. A presentation paradigm for discrimination task was shown in Figure 3.3.2.

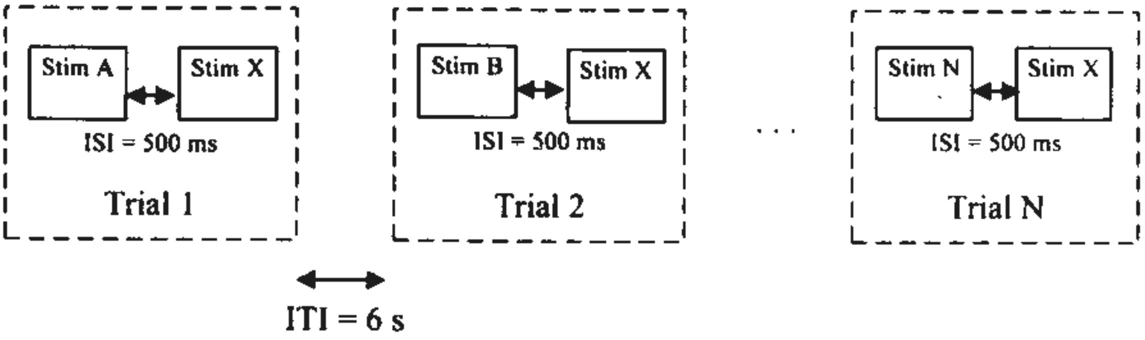


Figure 3.3.2 Presentation paradigm for discrimination task.

3.3.2. Data Analysis

3.3.2.1. Predicted discrimination.

The analysis on predicted discrimination was the same as that described in Experiment I and II. Since in Experiment I and II, the Δp_i did not provide more information than the predicted discrimination, only analysis on predicted discrimination was considered in Experiment III.

3.3.2.2. Obtained discrimination.

The score of discriminations for each pair was calculated following descriptions in previous study (Francis, et al., 2003). Obtained discrimination scores were calculated by equation (2.14), which was described in chapter 2. The equation (2.14) was repeated here.

$$D_{obt} = P('S'/S) \bullet P(S) + P('D'/D) \bullet P(D) \tag{2.14}$$

3.3.3. Results

3.3.3.1. Identification results and predicted discrimination.

Identification responses averaged across listeners according to three target positions were presented in Figure 3.3.3. The figures showed that listeners exhibited crossovers in their identification function. Three categories were obtained along each continuum. The crossovers corresponded roughly to the expected location of boundaries between the three tonal categories (at the 4-5 and 8-9 pairs in INT-LC continuum and at 4-5 and 7-8 pairs in other three continua).

Similar to Experiment I and II, the predicted discrimination scores were grouped into *within-category* (WC) and *across-category* (AC) pairs according to the crossovers from identification curves. The *p(pred.disc)* of WC and AC was averaged as shown in Table 3.3.1 and plotted in Figure 3.3.4. The RMANOVA of *p(pred.disc)* on three within-subject factors (2 POSITION × 2 CATEGORY × 2 CONTINUUM: SSC and INT) revealed no significant three-way interaction effect, but one significant two-way interaction and two significant main effects [POSITION: F(1,7)=11.91, p=0.011; CATEGORY: F(1,7)=22.25, p=0.002; POSITION × CATEGORY: F(1,7)=7.304, p=0.031]. The main CATEGORY effect indicated that the overall predicted discrimination from AC was better than WC, which suggested that there was an abrupt boundary in the identification curve. The observed POSITION × CATEGORY interaction effect suggested that the degree of CP (difference between AC and WC) was depending on the target position (LC or RC). No three-way interaction effect suggested that two types of continua (SSC and INT) had a similar pattern regarding on the effect of target position on the degree of CP.

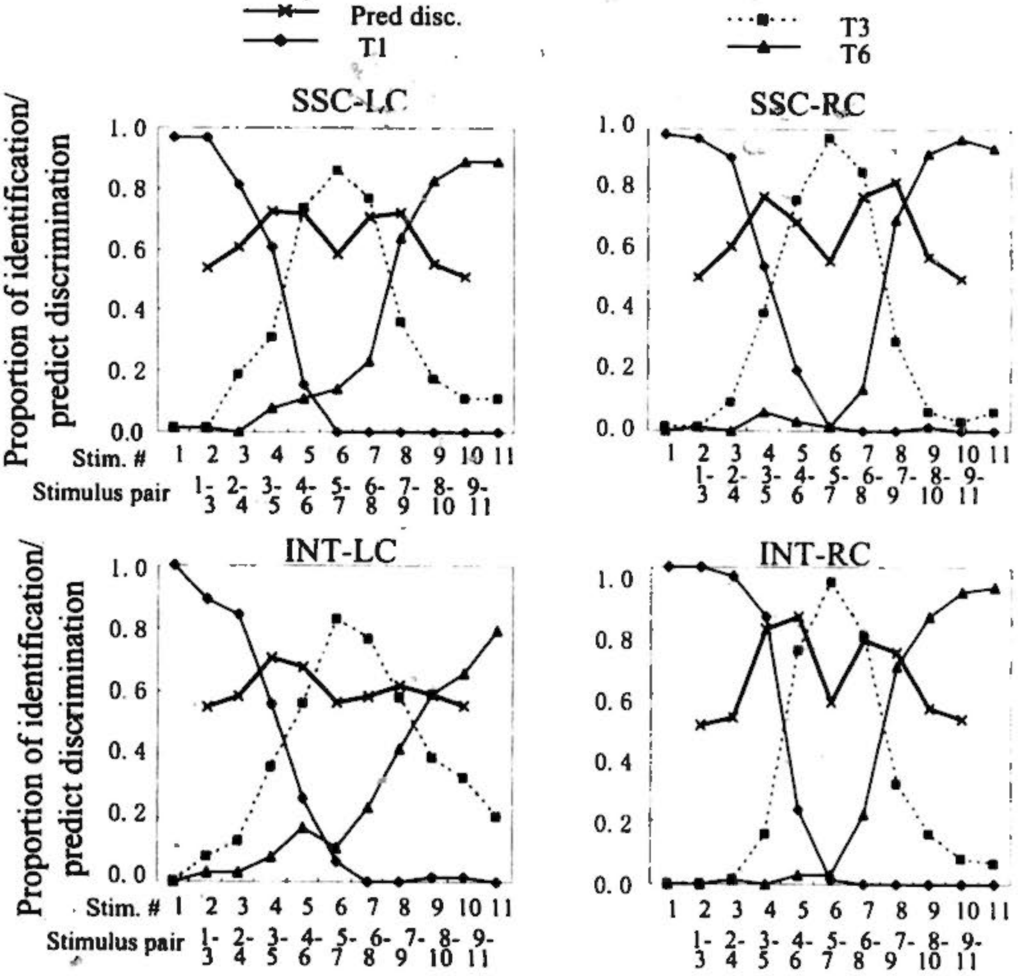


Figure 3.3.3 Results of two sets of Cantonese level tones perceived by native subjects in Experiment IIIa. SSC: simply shifted continuum; INT: interpolation continuum. The target syllables had two types of positions: at the beginning of the sentence (RC), at the end of the sentence (LC). A solid line with a cross represented predicted discrimination (Pred disc.) calculated from the identification results. A solid line with a circle, a dotted line with a square, and a solid line with a rectangle represented the proportion of the stimuli perceived as tone 1 (T1), tone 3 (T3) and tone 6 (T6) respectively.

However, two *post hoc* analyses on each continuum (SSC and INT) repetitively revealed that the POSITION×CATEGORY interaction effect was only presented in INT continuum but not in SSC continuum [SSC: F(1,7)=0.597, *p*=0.465; INT: F(1,7)=26.773, *p*=0.001]. Furthermore, four *post hoc* paired *t*-tests on CP (difference between AC and WC) revealed that there was a significant CP on RC but not on LC in both SSC and INT continua, which further confirmed the three-way interaction effect obtained from RMANOVA.

Another four *post hoc* paired t-tests on position effect (difference between RC and LC) revealed that the position effect was only presented in across-category paired in INT continuum (listed in Table 3.3.1).

In summary, the results in identification performance replicated the position effect (a greater CP in RC than in LC) in general. Although the effect of transition is weak, there was a trend that the pitch transition modulated the position effect by showing a more natural transition would induce a more asymmetric CP when positions of the contextual sentence were different.

Table 3.3.1 Grouping information, grouped mean value, and statistical analysis of predicted discrimination scores in two continua (SSC and INT) of Experiment III. **: $p<0.0025$; *: $p<0.0125$ (significant level $p<0.0125$ after correction).

	SSC-RC		SSC-LC		INT-RC		INT-LC	
	mean	pairs	mean	pairs	mean	pairs	mean	pairs
AC	0.769	3-5, 4-6, 6-8, 7-9	0.719	3-5, 4-6, 6-8, 7-9	0.787	3-5, 4-6, 6-8, 7-9	0.648	3-5, 4-6, 7-9, 8-10
WC	0.552	others	0.560	other	0.535	others	0.568	other
AC vs. WC paired t-test	t(7)=5.05, p=0.001**		t(7)=2.551, p=0.038		t(7)=4.784, p=0.002**		t(7)=2.455, p=0.044	
AC RC vs. LC	t(7)=1.016, p=0.344				t(7)=5.105, p=0.001**			
WC RC vs. LC	t(7)=-0.269, p=0.796				t(7)=-2.966, p=0.021			

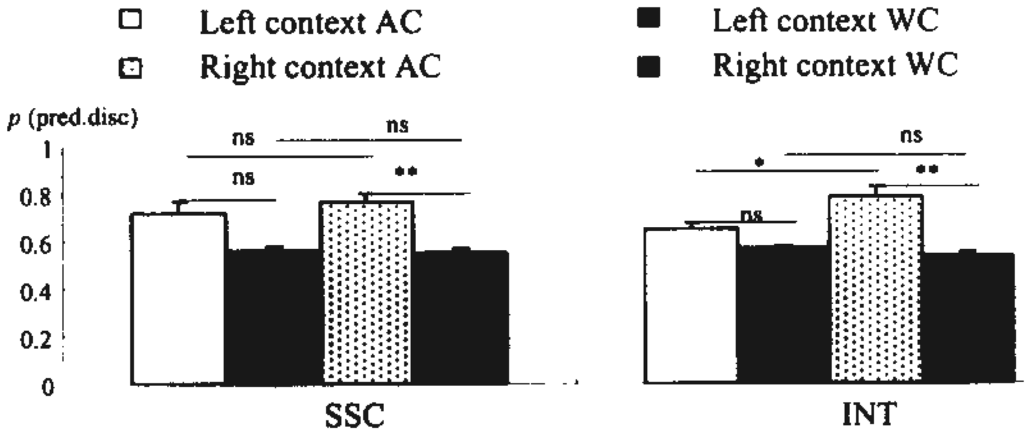


*Figure 3.3.4 Grouped value of predicted discrimination according to AC (Across-category) and WC (within-category) from left context and right context continua in Experiment III. SSC: simply shifted continuum; INT: interpolation continuum; **: p<0.0025; *: p<0.0125; ns. Not significant. Error bar represented standard error.*

3.3.3.2. Obtained discrimination.

Table 3.3.2 One-way RMANOVA on effect of stimulus in Experiment III. **: $p<0.001$; *: $p<0.05$.

	LC	RC
SSC	F(8,56)=4.302, $p<0.001$**	F(8,56)=11.661, $p<0.001$**
INT	F(8,56)=2.553, $p=0.019$*	F(8,56)=6.792, $p<0.001$**
NWS	F(8,56)=5.091, $p<0.001$**	F(8,56)=5.166, $p<0.001$**
CPT	F(8,56)=1.503, $p=0.177$	F(8,56)=2.293, $p=0.034$*

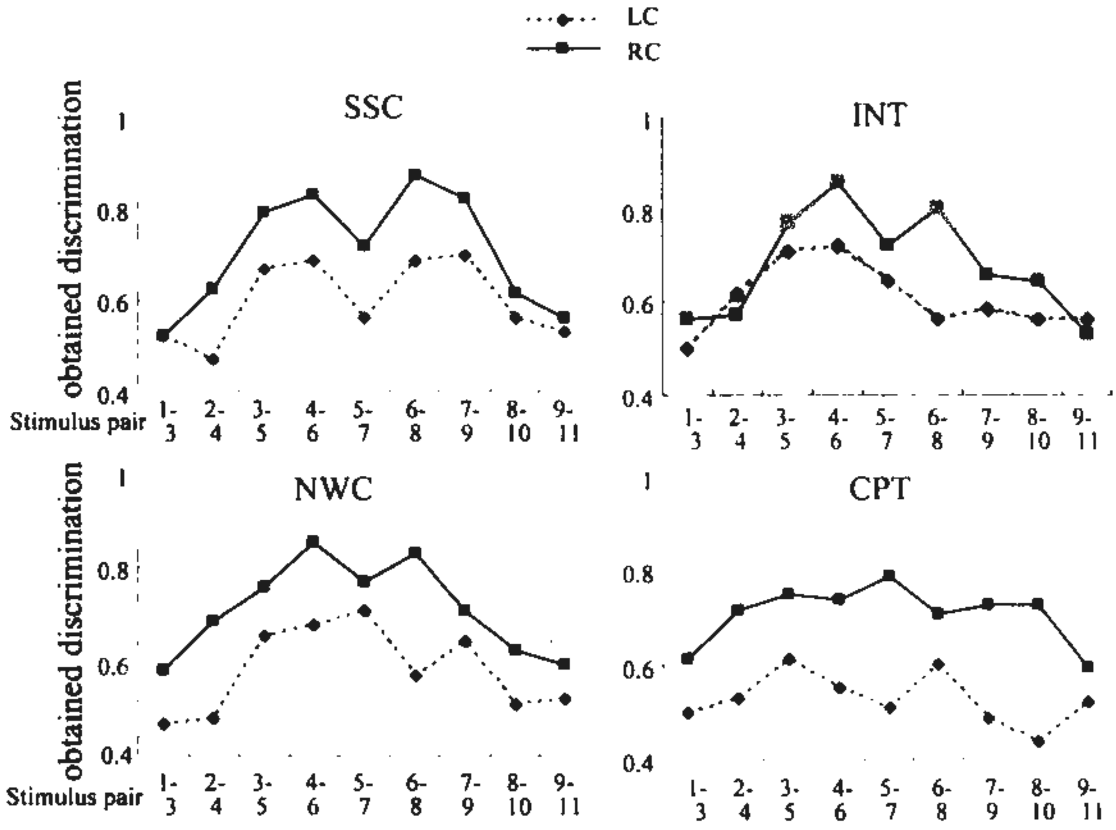


Figure 3.3.5 Obtained discrimination scores of LC and RC in Experiment III. SSC: simply-shifted continuum; INT: interpolation continuum; NWS: non-word sounds; CPT: complex tones

In total, there were 8 continua [4 CONTINUUM (SSC, INT, NWS, and CPT) ×2 POSITION (LC and RC)] of obtained discrimination scores, which were showed in Figure 3.3.5. Eight one-way RMANOVA on each continuum were carried out to check whether there were peaks and valleys on the discrimination curves. The within-subject factor was

STIMULUS with 9 steps. The results showed that except for CPT-LC continuum, all other seven continua had peaks and valleys on the discrimination curves (Table 3.7). According to results from the pairwise t-test and the grand average discrimination curves (in Figure 3.3.5), the 9 steps were grouped into *within-category* (WC) pair and *across-category* (AC) pairs (listed in table 3.3.3.). Further analyses were on these grouped discrimination scores. Results of statistical analyses were detailed in appendix 3.

A. EFFECT OF PITCH TRAINSITION

To investigate the effect of pitch transition on CP, comparison was done on SSC and INT. Following the same steps as those in predicted discrimination scores, the D_{obt} of WC and AC were averaged as shown in Table 3.3.3 and plotted in Figure3.3.6. The RMANOVA of D_{obt} on three within-subject factors (2 POSITION×2 CATEGORY ×2 CONTINUUM: SSC and INT) revealed no significant three-way interaction effect, but one significant two-way interaction and two significant main effects [POSITION: F(1,7)=30.389, p=0.001; CATEGORY: F(1,7)=64.009, p<0.001; POSITION × CATEGORY: F(1,7)=16.699, p=0.005]. The main CATEGORY effect indicated that the overall predicted discrimination from AC was better than WC, which suggested the presence of a CP. The observed POSITION×CATEGORY interaction effect suggested the degree of CP was depending on the target position (LC or RC). No three-way interaction effect suggested that two types of continua (SSC and INT) had a similar pattern regarding on the effect of target position on the degree of CP, which is consistent with the predicted discrimination results.

However, two *post hoc* analyses on each continuum (SSC and INT) repetitively revealed that the POSITION×CATEGORY interaction effect was only presented in INT continuum but not in SSC continuum [SSC: F(1,7)=2.64, p=0.148; INT: F(1,7)=16.701, p=0.005]. Furthermore, four *post hoc* paired t-tests on position effect (difference between RC and LC) revealed that the position effect was only presented in across-category of INT continuum. Another four *post hoc* paired *t*-tests on CP revealed that all four sets had a significant CP (listed in Table 3.3.3).

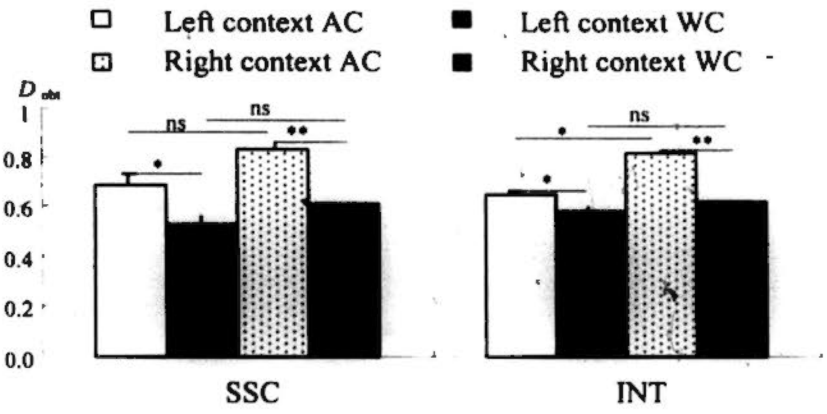


*Figure 3.3.6 Grouped value of obtained discrimination according to AC (Across-category) and WC (within-category) from left context and right context continua in Experiment III to study the effect of pitch transition. SSC: simply shifted continuum; INT: interpolation continuum; **: p<0.0025; *: p<0.0125; ns. Not significant. Error bar represented standard error.*

In summary, the results in obtained discrimination replicated the results from identification task in following two aspects. (1) The asymmetric CP due to target position is obtained — the degree of CP is higher in Right Context (RC) continuum than that in Left Context (LC) continuum. (2) When the transition is more natural (INT), RC has a higher CP than LC does; while, there is no such significant asymmetric effect observed when the transition is less natural (SSC).

Table 3.3.3 Grouping information, grouped mean value, and statistical analysis of obtained discrimination scores in two continua (SSC and INT) of Experiment III. **: p<0.0025; *: p<0.0125 (significant level p<0.0125 after correction).

	SSC-RC		SSC-LC		INT-RC		INT-LC	
	mean	pairs	mean	pairs	mean	pairs	mean	pairs
AC	0.831	3-5, 4-6, 6-8, 7-9	0.685	3-5, 4-6, 6-8, 7-9	0.809	3-5, 4-6, 6-8	0.643	3-5, 4-6, 7-9, 8-10
WC	0.608	others	0.529	other	0.615	others	0.577	other
AC vs. WC paired t-test	t(7)=5.631, p=0.001**		t(7)=4.503, p=0.003**		t(7)=8.335, p<0.001**		t(7)=3.654 p=0.008*	
AC RC vs. LC	t(7)=2.626, p=0.034				t(7)=7.388, p<0.001**			
WC RC vs. LC	t(7)=2.668, p=0.032				t(7)=1.665, p=0.140			

B. *EFFECT OF CARRIER SYLLABLE*

To investigate the effect of carrier syllable on CP, comparison was done among SSC (Real word), NWS (Nonword) and CPT (Complex tone). Similarly, the D_{obt} of WC and AC were averaged as in Table 3.3.4. and plotted in Figure 3.3.7. The RMANOVA of D_{obt} on three within-subject factors (2 POSITION × 2 CATEGORY × 3 CONTINUUM: SSC, INT, and CPT) revealed no significant three-way interaction effect, but one significant two-way interaction and two significant main effects [POSITION: F(1,7)=26.004, p=0.001; CATEGORY: F(1,7)=52.296, p<0.001; CATEGORY × CONTINUUM: F(1,7)=4.322, p=0.035]. The CATEGORY × CONTINUUM interaction effect indicated that CP was different on these three continua, although the main CATEGORY effect indicated that the overall discrimination in AC pair was better than that in WC. The main POSITION effect suggested that the overall discrimination performance was different regarding on the target position.

However, three *post hoc* analysis on each continua (SSC, NWS, and CPT) repetitively revealed that none of the continua showed a POSITION × CATEGORY interaction effect. This suggested no evidence of the effect of target position on CP in these three continua. Furthermore, six *post hoc* paired t-tests on position effect (difference between RC and LC) revealed that discrimination performance from both category pairs in NWS continuum was affected by target position by showing better discrimination in RC than in LC in both category pairs. However, no evidence of such target position effect was presented on SSC or on CPT continuum. Another six *post hoc* paired *t*-tests on CP (difference between AC and WC) revealed that except for CPT-LC and RC continua, all the other four sets had a significant CP (listed in Table 3.3.4.), though CP in CPT continua was also marginally significant.

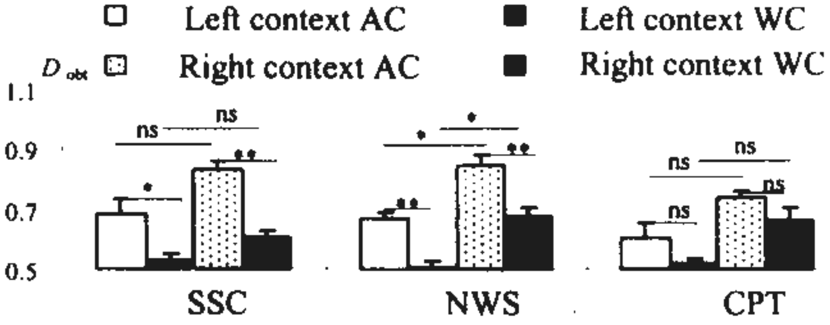


*Figure 3.3.7 Grouped value of obtained discrimination according to AC (Across-category) and WC (within-category) from left context and right context continua in Experiment III to study the effect of carrier syllables. SSC: simply shifted continuum; NWS: non-word sounds; CPT: complex tones; **: p<0.008; *: p<0.002; ns. Not significant. Error bar represented standard error.*

Table 3.3.4 Grouping information, grouped mean value, and statistical analysis of obtained discrimination scores in two continua (SSC, NWS, CPT) of Experiment III. **: $p<0.0025$; *: $p<0.0125$ (significant level $p<0.0125$ after correction).

	SSC-RC		SSC-LC		NWS-RC		NWS-LC		CPT-RC		CPT-LC	
	mean	pairs	mean	pairs	mean	pairs	mean	pairs	mean	pairs	mean	pairs
AC	0.83	3-5, 4-6, 6-8, 7-9	0.68	3-5, 4-6, 5-7, 7-9	0.84	3-5, 6-8	0.67	3-5, 6-8	0.74	3-5, 4-6, 6-8	0.61	3-5, 4-6, 7-9, 8-10
WC	0.61	others	0.53	others	0.68	others	0.51	others	0.66	others	0.51	1-3, 2-4, 6-8, 9-11
AC vs. WC paired t-test	t(7)=5.631, p=0.001**		t(7)=4.503, p=0.003**		t(7)=5.53 p=0.001**		t(7)=9.69, p<0.001*		t(7)=1.886, p=0.101		t(7)=2.797, p=0.028	
AC RC vs. LC	t(7)=2.626, p=0.034				t(7)=4.17, p=0.004*				t(7)=2.08, p=0.076			
WC RC vs. LC	t(7)=2.668, p=0.032				t(7)=4.12, p=0.004*				t(7)=2.91, p=0.023			

In summary, the results of comparison on different types of carrier syllables showed that CP was presented in continua with real word and non-word carrier syllables. This difference between speech and nonspeech suggested that CP is influenced by linguistic experience. Although no evidence showed a significant CP on complex tones, the trend of discrimination of AC pair better than WC pair did not reverse. This same trend suggested the generalization of tone experience on non speech sounds. Moreover, the main POSITION effect but no POSITION ×CONTINUUM effect suggested that discrimination performance of RC was better than LC and this trend was regardless of the carrier syllables as long as the pitch contours were the same. In other words, the POSITION effect was affected by pitch contours but not by carrier syllables.

3.3.4. Summary and Discussion

Experiment III employed both the identification and the discrimination tasks to investigate whether the asymmetric CP regarding the position of the target was due to the naturalness of speech sounds. The results from identification and discrimination tasks consistently showed that a more natural pitch transition induced a more asymmetric CP. Since a symmetric CP in natural speech was anticipated from the motor theory (discussed in section 3.2.4), the observation favored the *multisore* account of CP in that other mechanisms than the motor theory also take effect.

The *multistore* model considers effects from both general auditory processing and linguistic influence. The observed asymmetric CP regarding on the target position can be explained by both accounts. However, the results from experiment on manipulation of pitch contours suggested the linguistic influence is more responsible for the asymmetric CP. Due to the asymmetric coarticulation effect — the preceding context influences the target more than the following context — the pitch contours of the target vary more in LC than in RC (Y. W. Wong, 2006; Y. Xu, 1994, 1997). Therefore, listeners are more ambiguous when identifying the target when it is in LC than when it is in RC in natural speech (the INT continuum). When the pitch contour of the target is less similar to the target as in SSC continuum, more reliable and consistent tonal category will be obtained.

Furthermore, linguistic experience in *multistore* model (Xu et al., 2006) predicts less CP and asymmetric effect when the target with less experience. The comparison on different types of carrier syllables partially confirmed the prediction. When the carrier syllable was nonspeech sound, even the pitch contour was the same as that in real word, neither obvious CP nor obvious asymmetric effect was observed. The results that similar CP was presented in non-word sound and in real word suggested that CP observed in the experiment was a more phonological processing than the semantic processing. In Experiment IV, the influence of linguistic experience on CP was further investigated by a cross-language study.

On the other hand, the general auditory account suggested by the *multistore* model also gained supports from the results by showing that the discrimination performance in RC was better than that in LC in general, which was regardless of the types of carrier syllables. Nonetheless, better discrimination in RC than in LC in general might also be due to the artifact during the construction of the stimuli, which needs to be investigated in the future.

3.4. Experiment IV: to Check the Effect of Linguistic Background

3.4.1. Method

3.4.1.1. Subjects

In addition to eight valid Cantonese subjects, who participated in Experiment III, 17 more subjects from different language backgrounds participated in the experiment: nine Mandarin subjects (7 male and 2 female, aged 18-26) from universities in Fuzhou, and eight French subjects (4 male and 4 female, aged 18-23) from Université Lumière lyon 2. All the subjects were paid to participate. None of them had reported history of speaking or hearing disability. Speakers from the other two language (Mandarin and French) have no experience on Cantonese. None of the subjects have experiences of musical training.

3.4.1.2. Stimuli

Two sets of stimuli were used: SSC (simply shifted continuum) and CPT (complex tone continuum). These two sets of stimuli were the same as described in Experiment I and III. Both sets of stimuli included LC and RC conditions.

3.4.1.3. Discrimination task

Due to the lack of corresponding phonemic labels of level tones in Mandarin, French and nonspeech, only the discrimination task was carried out in the experiment. The experimental procedure was the same as that described in Experiment III. The data of Cantonese was the same as in Experiment III and repeated here to investigate the influence of language backgrounds.

The experimental settings were listed below: (1) Cantonese subjects listened to the simply shifted real words (CS: Cantonese + Speech); (2) Cantonese subjects listened to the complex tones (CN: Cantonese + nonspeech); (3) Mandarin subjects listened to the simply shifted real word (MS: Mandarin + Speech); (4) French subjects listened to the simply shifted real words (FS: French + Speech); (5) French subjects listened to the complex tones (FN: French + nonspeech).

3.4.2. Data Analysis

Obtained discrimination was analyzed with the same method as that described in Experiment III.

3.4.2. Results

In total, there were 10 continua [5 CONTINUUM (CS, CN, MS, FS, FN) × 2 POSITION (LC and RC)] of obtained discrimination scores shown in Figure 3.4.1. Similar to Experiment III, ten one-way RMANOVAs with factor of STIMULUS were carried out to check whether there were peaks and valleys on the discrimination curves. The results showed that peaks and valleys were significantly obtained from both LC and RC continua in CS condition. MS and CN had a marginally significant STIMULUS effect on RC continuum. The remaining continua did not show a significant STIMULUS effect (Table 3.4.1). According to results from the pairwise t-test and the grand-averaged discrimination curves (in Figure 3.3.5), the 9 steps were grouped into *within-category* (WC) pair and *across-category* (AC) pairs (listed in Table 3.4.2). Further analyses were on these grouped discrimination scores. They were analyzed from different aspects: performance on speech set, performance on nonspeech set, and performance on interaction between speech and nonspeech sets. Results of statistical analyses were detailed in appendix 4.

Table 3.4.1 One-way RMANOVA on the effect of stimulus in Experiment IV **: $p<0.001$; *: $p<0.05$.

	LC	RC
CS	F(8, 56)=4.302, $p<0.001$**	F(8, 56)=11.661, $p<0.001$**
MS	F(8, 64)=0.688, $p=0.700$	F(8, 64)=2.157, $p=0.043$ *
FS	F(8, 56)=0.998, $p=0.448$	F(8, 56)=1.544, $p=0.1631$
CN	F(8, 56)=1.503, $p=0.177$	F(8, 56)=2.293, $p=0.034$ *
FN	F(8, 56)=0.848, $p=0.565$	F(8, 56)=1.717, $p=0.115$

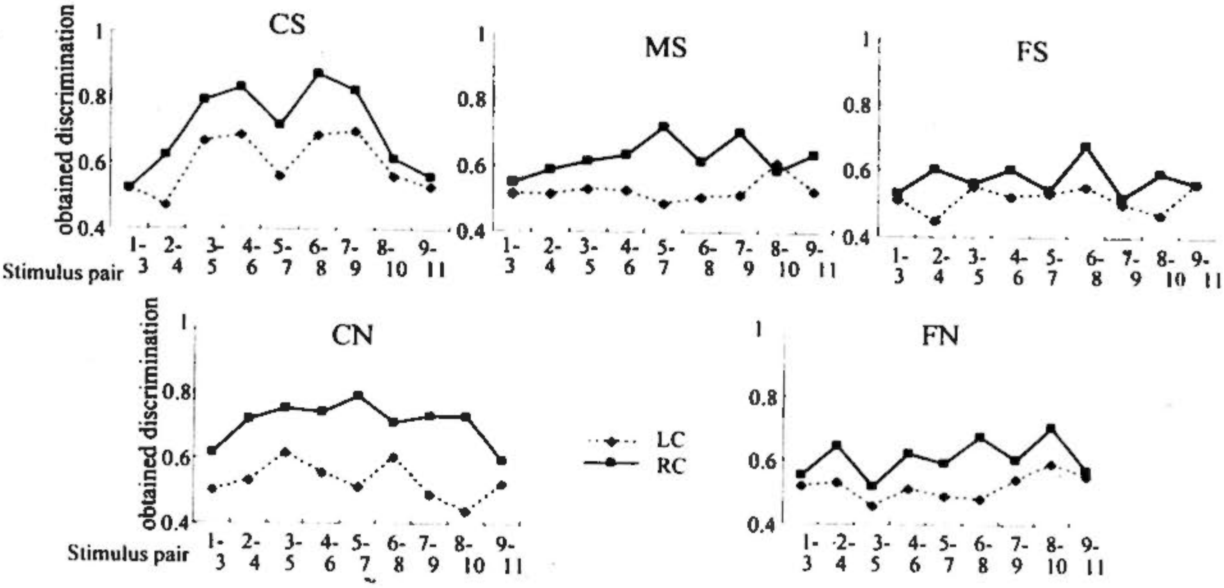


Figure 3.4.1 Obtained discrimination scores of LC and RC in Experiment IV. CS: Cantonese + Speech; CN: Cantonese + nonspeech; MS: Mandarin + Speech; FS: French + Speech; FN: French + nonspeech.

A. *PERFOMANCE ON SPEECH SET*

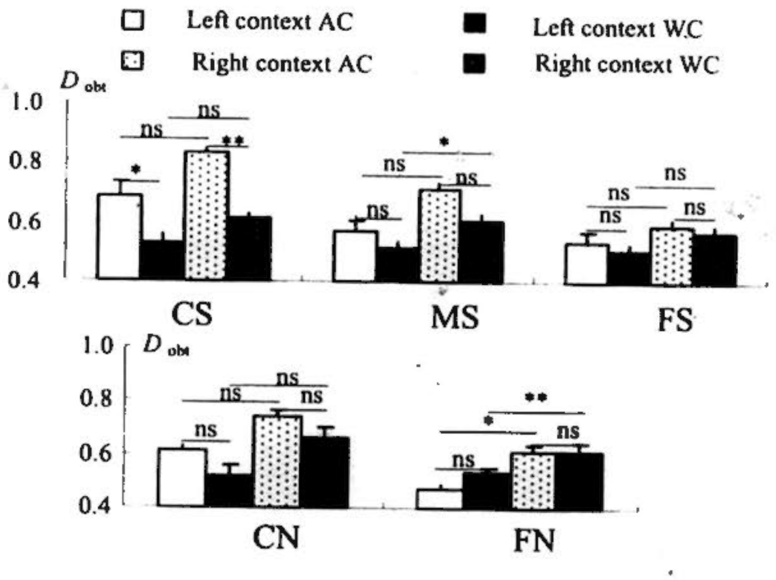


*Figure 3.4.2 Grouped value of obtained discrimination according to AC and WC from left context and right context continua in Experiment IV. CS: Cantonese+speech; MS: Mandarin+speech; FS:French+speech; CN:Cantonese+nonspeech; FN: French+nonspeech; **: $p<0.001$; *: $p<0.008$; ns. Not significant. Error bar represented standard error.*

The D_{obt} of WC and AC in SSC from three languages were averaged as shown in Table 3.4.2. and plotted in Figure3.4.2. The mix design of D_{obt} had two within-subject factors (2 POSITION × 2 CATEGORY) and one between-subject factor (3 LANGUAGE: Cantonese, Mandarin, and French). The LANGUAGE effect reached a significant level [F(2,22)=10.141, $p=0.001$]. The *post hoc* pairwise comparison revealed that performance by Cantonese subjects was significantly better than that by Mandarin and French subjects, while no significant difference was obtained between Mandarin and French subjects. The mix design also showed a two-way interaction effect between CATEGORY and LANGUAGE, and two main within-subject effects as well [POSITION: F(1,22)=25.451, $p<0.001$; CATEGORY: F(1,22)=48.903, $p<0.001$; CATEGORY × LANGUAGE: F(2,22)=11.007, $p<0.001$]. The main CATEGORY effect indicated that the overall predicted discrimination from AC was better than WC, which suggested the presence of CP. The main POSITION effect but no POSITION × LANGUAGE effect indicated that discrimination in RC was better than in LC overall. The observed LANUAGE × CATEGORY interaction effect suggested that different language background had different degrees of CP.

Table 3.4.2 Grouping information, grouped mean value, and statistical analysis of obtained discrimination scores in three conditions (CS, MS and FS) of Experiment IV. **: $p<0.0001$; *: $p<0.008$ (significant level $p<0.008$ after correction).

	CS-RC		CS-LC		MS-RC		MS-LC		FS-RC		FS-LC	
	mean	pairs	mean	pairs	mean	pairs	mean	pairs	mean	pairs	mean	pairs
AC	0.83	3-5, 4-6, 6-8, 7-9	0.68	3-5, 4-6, 5-7, 7-9	0.71	5-7, 7-9	0.57	3-5, 8-10	0.59	3-5, 4-6, 6-8	0.53	3-5, 4-6, 7-9, 8-10
WC	0.61	others	0.53	others	0.60	others	0.51	others	0.57	others	0.50	1-3, 2-4, 6-8, 9-11
AC vs. WC paired t-test	t(7)=5.631, $p=0.001$**		t(7)=4.503, $p=0.003$*		t(7)=2.773 $p=0.024$		t(7)=2.544, $p=0.034$		t(7)=0.826, $p=0.436$		t(7)=1.048, $p=0.329$	
AC RC vs. LC	t(7)=2.626, $p=0.034$				t(7)=2.892, $p=0.020$				t(7)=1.775, $p=0.119$			
WC RC vs. LC	t(7)=2.668, $p=0.032$				t(7)=5.361, $p=0.001$*				t(7)=1.555, $p=0.164$			

Three *post hoc* analyses on different language groups (Cantonese, Mandarin and French) revealed that the CATEGORY effect was presented in both Cantonese and Mandarin on real word continua [CS: F(1,7)=37.345, $p<0.001$; MS:F(1,8)=12.131, $p=0.008$; FS: F(1,7)=2.305, $p=0.173$]. Furthermore, six *post hoc* paired t-tests on position effect (difference discrimination score or CP between RC and LC) revealed that the position effect was only presented in within-category paired in MS continuum. Another six *post hoc* paired *t*-tests on CP (difference between AC and WC) revealed that only two continua in CS reached a significant level, while the effect in MS continua was marginally significant and was in FS was not significant (listed in Table 3.4.2).

In summary, CP depends on language experience: the native Cantonese subjects showed a strong CP; the subjects with another tone experience showed a marginally CP; and the subjects without any tone experience did not show a CP. In general, the discrimination from RC continuum was better than that from LC continuum and the difference was more prominent when the subjects with tone experiences than when the subjects without any tone experiences.

Table 3.4.3 Grouping information, grouped mean value, and statistical analysis of obtained discrimination scores in two conditions (CN and FN) of Experiment IV. **: $p<0.0025$; *: $p<0.0125$ (significant level $p<0.0125$ after correction).

	CN-RC		CN-LC		FN-RC		FN-LC	
	mean	pairs	mean	pairs	mean	pairs	mean	pairs
AC	0.74	3-5, 4-6, 6-8	0.61	3-5, 4-6, 7-9, 8-10	0.61	3-5, 4-6, 6-8	0.47	3-5, 4-6, 7-9, 8-10
WC	0.66	others	0.51	1-3, 2-4, 6-8, 9-11	0.61	others	0.53	1-3, 2-4, 6-8, 9-11
AC vs. WC paired t-test	t(7)=1.886, p=0.101		t(7)=2.91, p=0.023		t(7)=-1.909 p=0.098		t(7)=0.104, p=0.92	
AC RC vs. LC	t(7)=2.08, p=0.076				t(7)=6.137, p<0.001**			
WC RC vs. LC	t(7)=2.03, p=0.039				t(7)=3.487, p=0.010*			

B. PERFOMANCE ON NONSPEECH SET

The D_{obt} of WC and AC in CPT from two languages were averaged as shown in Table 3.4.3 and plotted in Figure3.4.2. There were two within-subject factors (2 POSITION × 2 CATEGORY) and one between-subject factor (2 LANGUAGE: Cantonese, and French) in the mix design of D_{obt}. The LANGUAGE effect reached a significant level [F(1,14)=19.498, p=0.001], which indicated that the performance on the nonspeech set was significantly better by Cantonese subjects than by French subjects. The mix design also showed a three-way interaction effect, a two-way interaction effect, and a main within-subject effect

[POSITION: F(1,22)=18.815, p=0.001; CATEGORY × LANGUAGE: F(2,22)=7.302 p=0.017; CATEGORY×POSITION×LANGUAGE: F(2,22)=5.211, p=0.039]. The three-way interaction effect indicated that the language experience modulated CP in LC and RC continua differently. The main POSITION effect but no POSITION ×LANGUAGE effect indicated that discrimination in RC was better than in LC on nonspeech continua overall,. The observed LANUAGE×CATEGORY interaction effect suggested different language backgrounds had different degrees of CP.

Two *post hoc* RMANOVAs with two within-subject factors of CATEGORY and POSITION were conducted on performance from Cantonese and French subjects repetitively. Cantonese subjects reached a marginal significant CATEGORY and POSITION main effect [CATEGORY: F(1,7)=6.574, p=0.037; POSITION:F(1,7)=6.648, p=0.037], while French subjects reached a interaction and a POSITION main effect [CATEGORY × POSITION: F(1,7)=10.545, p=0.014; POSITION:F(1,7)=29.920, p=0.001]. Furthermore, four *post hoc* paired t-tests on position effect (difference between RC and LC) revealed that the position effect was presented in French subjects. Another four *post hoc* paired *t*-tests on CP (difference between AC and WC) revealed that none of the comparisons reached a significant level (listed in Table 3.4.3.).

In summary, CP depends on language experience: the Cantonese subjects perceive the nonspeech in a weak categorical mode, while the French subjects did not show any CP. Similarly to the performance in real word continua, in general, the discrimination from RC continuum was better than that from LC continuum in nonspeech condition. However, contradicting to the real word condition, the difference in nonspeech was more prominent for subjects without tone experience.

C. PERFOMANCE ON INTERACTION OF SPEECH AND NONSPEECH

The last type of analysis was to check how the language experience influenced the subjects' performance on real word and nonspeech sounds. A mix design of D_{obt} on both types of sounds was carried out. The mix design had three within-subject factors (2 POSITION×2 CATEGORY×2 CONTINUUM: SSC and CPT) and one between-subject factor (2 LANGUAGE: Cantonese and French). The LANGUAGE effect reached a significant level [F(1,14)=55.811, p<0.001]. The mix design showed a LANGUAGE×

CATEGORY interaction effect [F(1,14)=20.242, $p<0.001$]. This result indicated that different language experiences induced different degrees of CP.

Break down the data into two language groups and conduct two three-way RMANOVA (2 POSITION × 2 CATEGORY × 2 CONTINUUM: SSC and CPT). There was a marginal significant interaction effect between POSITION and CONTINUUM [CATEGORY × CONTINUUM: F(1,7)=5.255, $p=0.056$], and a significant POSITION main effect [F(1,7)=14.034, $p=0.007$] from French subjects. The results showed no significant CP from French subjects. Moreover, the performance from them indicated that although the discrimination resolution in RC continuum was higher than that in LC continuum in general, the trend was stronger in nonspeech continua than that in real speech continua. The Cantonese subjects showed the same results as in Experiment III. The detailed statistical results were shown in Appendix 3 and in Experiment III, therefore they were not replicated here. The performance from Cantonese subjects indicated that CP was more prominent in real word continua than in nonspeech continua. Moreover, the discrimination resolution in RC continuum was higher than that in LC continuum in general and this trend was independent of carrier syllables.

3.4.4. Summary and Discussion

The cross-language study in this section was to investigate how different linguistic backgrounds influence CP and the position effect. The results from the real word continuum indicated that CP was present in native listeners. A weaker CP was also present in listeners with experience of another tone language, the Mandarin. However, only one obvious category boundary (or two close boundaries) was observed in Mandarin listeners, in contradict to two obvious separated category boundaries were observed in Cantonese listeners. The two level tone categories may be corresponding to the high level tone and the low level tone (the tone 3 in continuous speech) (Y. Xu, 1994) in Mandarin speech. This suggested that Mandarin subjects may extend their experience to an unknown speech to facilitate the discrimination. Finally, there was no CP observed from subjects without any tone experience (French subjects). The results on real speech continua from subjects of three language backgrounds showed that CP was modulated by the language experiences, with a more corresponding experience inducing a stronger CP.

Furthermore, the language modulation effect was also present on a comparison of CP on different types of carrier syllables. The native subjects had a greater CP in speech continua than in nonspeech continua, while, French subjects did not show different degrees of CP regarding the carrier syllables. This indicated that native subjects had a greater sensitivity to the difference between speech and nonspeech, and CP can be used as an indicator to distinguish speech from nonspeech. The results that the language experience modulated CP on types of carrier syllables further support the linguistic factor on CP.

The position effect by showing a higher discrimination in RC than in LC continuum was observed in general. The factor of language experience had little effect on the position effect. However, the particular greater target effect in nonspeech than in real speech by French subjects suggested the involvement of an auditory processing for the asymmetric effect.

The results from CP and position effect further confirmed the influence from two factors — linguistic experience and auditory processing proposed — in the multistage model.

3.5. Summary of chapter 3

Four experiments on investigation of CP performance on the continuum spanning across Cantonese three level tones were reported in this chapter. Experiment I was designed to examine the context effect through the manipulation of target position in a sentence. Stimuli used in Experiment I were based on natural speech and the targets were presented in three positions within a sentence: isolation (without a contextual sentence), LC (left context, at the end of a sentence), and RC (right context, at the beginning of a sentence). The results indicated that the presence of a contextual sentence improved CP of Cantonese level tones. Moreover, the results also showed that CP was more prominent in RC condition than in LC condition, which showed an *asymmetric position effect of context* (APEC) on CP.

To examine whether the observed APEC on CP was due to the artifact of the stimuli, Experiment II was carried out on synthetic speech with the similar manipulation of context position. The results replicated the APEC on CP, which indicated that the effect was not only due to the artifact of construction of stimuli but some other factors also contributed to the APEC on CP.

Experiment III was designed to examine whether the APEC effect observed in Experiment II was linguistically relevant or not, especially whether it was relevant to articulation gestures as proposed in the motor theory. Two types of manipulations were conducted in Experiment III: (1) on pitch transition between the context and the target; and (2) on the carrier syllables. The results from Experiment III showed that CP of Cantonese level tones and APEC on CP were affected by both linguistic processing and general auditory processing, thus the results favored the multistage model more than the motor theory model.

Experiment IV was designed to further investigate how different linguistic backgrounds influenced CP and the APEC through a discrimination task by both a cross-language study and a comparison between speech and nonspeech conditions. Different language groups performed differently on the tasks in two aspects: the discrimination scores for the level tones and the CP. (1) Native listeners (Cantonese) had a higher discrimination score for level tones than non native listeners did (Mandarin and French). (2) CP was found from two groups of tone language listeners (Mandarin and Cantonese) especially in speech condition. However, the position of CP boundary was affected by subjects' tone experience from their native language. The perception of nonspeech pitches from native subjects was generalized from the linguistic experience, but had a smaller degree of CP. Non-tone language listeners (French) did not show CP either from the speech or from the nonspeech condition. However, non-tone language listeners discriminated between the across-category pair better than the within-category pair in speech condition, but a reverse pattern was obtained in the nonspeech condition. There was no language effect on APEC; actually, APEC effect was present from all language groups and in both speech and nonspeech conditions, by showing that discrimination from RC was higher than from LC. Moreover, no APEC effect was observed on CP in this experiment. The results further confirmed the influence from two factors—linguistic experience and auditory processing.

CHAPTER 4 BEHAVIORAL AND ERP EXPERIMENTS ON MANDARIN AND CANTONESE TONES

4.0. Purposes of the Experiments

Although experiments in Chapter 3, which were designed to investigate the perception of Cantonese level tones, have shown that both linguistic experiences and some aspects from general auditory processing affect CP, there was no information on the temporal dynamics of these influences on CP. Moreover, although the experiments in Chapter 3 compared the performance from tone and non-tone language groups (Chinese and French), and also compared the performance from two tone language groups (Mandarin and Cantonese) with different tone inventories, the tone contrast investigated in those experiments was native to only one tone language system (Cantonese) but absent in the other tone language system (Mandarin) and in the non-tone language system (French). Therefore, the experiments conducted in Chapter 3 were not enough to look into the influence of different tone inventories on perception as a whole. Whether different phonemic inventories affect the perception on the same native contrast has never been investigated yet. The experiments in this chapter were conducted to investigate the above two questions further—the temporal dynamics of CP and the influence of different tone inventories on the perception of the same native tone contrast.

CP between the high level and the high rising tones, which exist in both Mandarin and Cantonese tone systems, was investigated by ERP and behavioral experiments. Similar to chapter 3, all stimuli used here were synthesized by PSOLA using Praat.

There were three experiments in this chapter to investigate three temporal cognitive stages: MMN for the preattentive stage, P300 for perceptual stage, and behavioral for decision stage. In all three experiments, two types of stimuli were used: speech and nonspeech sounds. The comparison between speech and nonspeech better illustrated the influence of linguistic experience. In P300 and behavioral experiments, a cross-language comparison was carried out to further look into the influence from different language backgrounds.

4.1. Experiment I: MMN

Experiment I was an MMN study on speech and nonspeech sounds by native Mandarin subjects. A *post hoc* behavioral task was supplemented to verify the category boundary on the continuum.

4.1.1. Method

4.1.1.1. Subjects

Nineteen right-handed[f] (Laterality Quotient=75.5±20.8) (Oldfield, 1971) native Mandarin subjects (11 females, 8 males, aged 23–31 years) participated in both MMN and behavioral discrimination tasks in two separated days with payment. Ten of them, as well as another eight subjects, participated in a behavioral identification task to supplement the procedures after a few months. All of them were students from the Chinese university of Hong Kong. All of them were from mainland China. None of them can speak Cantonese. All are with normal hearing. Informed written consent was obtained from each subject. Ethics approval was obtained from Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong.

4.1.1.2. Stimuli

Based on the Mandarin syllable *yi*^{1} recorded by a male speaker, we synthesized a set of eleven 500 ms stimuli by varying the pitch contours from level to rising (Fig 4.1.1a). The pitch contours were bilinear approximations of the level and rising tones as those used in (Wang, 1976). Although the stimuli are not natural tones, they were labeled almost perfectly in the behavioral task (ref. Wang, 1976 and section 4.1.3.1). We synthesized 11 nonspeech counterparts using pure tones (frequency-modulated sine waves) with the same pitch contours. To make the loudness of nonspeech comparable to the speech, the speech was delivered at 65dB and nonspeech at 85dB, according to subjective judgment by two independent researchers. Intensity envelopes were kept unchanged in each continuum and closely matched between speech and nonspeech contexts (Figure4.1.1b and c.) According

---

[f] In ERP experiment, since lateralization pattern is one of the testing factors, especially when data from different electrodes should be grouped together, the handedness is formally tested. Similarly to what has done in behavioral task (Chapter 3), the number of subjects enrolled in ERP experiments is determined according to the rule of thumb. A typical number is 15 to 20. For some small effect component, large number of subjects is preferred.

to previous results on identification (Wang, 1976), one standard (#8) and two types of deviants (across-category-deviant, #6; and within-category-deviant, #10) were chosen in ERP study.

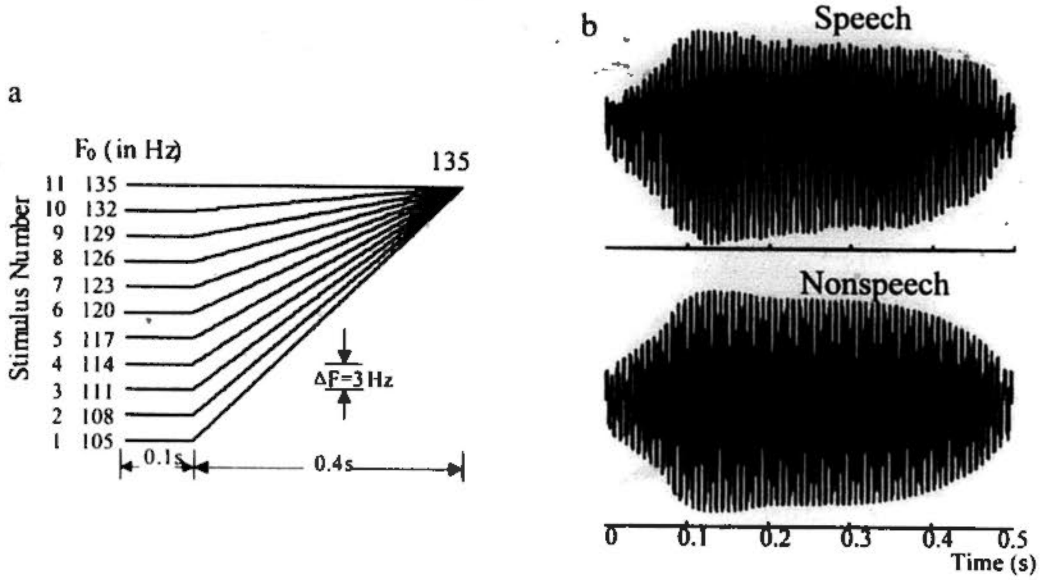


Figure 4.1.1 a) Schematic diagram for pitch contours of 11 stimuli. Stimulus #7 was the category boundary (following Wang, 1976). Stimulus #10, #8 and #6 were used as within-category-deviant, standard and across-category-deviant, respectively. b) The amplitude profiles of speech and nonspeech.

4.1.1.3. EEG recording

Before the EEG recording, subjects were instructed to watch a silent movie with subtitles and to ignore the input auditory sounds. They were instructed that the input auditory sounds were distractor and their task was to focus on the movie and answer relevant questions after the experiment. Failure to answer the question will affect their final payment.

An oddball passive listening paradigm for auditory input (our actual target) was used in the ERP experiment. The subjects were distracted from input audio stimuli by watching a silent movie with subtitles. Sound sequences were presented in six blocks (three blocks for each speech context) via Philip headphones (SHG5000) in a random order. Each block contained one standard and two deviants to make up 360 trials delivered at a rate of 1Hz. In other words, the interval between successive trials was 500ms. The standard (#8) occurred with probability of 7/9 while each of the two deviants (across-category-deviant, #6; and within-category-deviant, #10) occurred with probability of 1/9 in a block. At least two standards precede a deviant. The standard and the deviants were selected based on earlier results (Wang, 1976). The presentation paradigm was shown in Figure 4.1.2a.

Raw EEG data were recorded continuously with 128-channel Electrical Geodesics, Inc. (EGI) system (referenced to the vertex, sampling rate at 250 Hz, amplifier passband at 0.1–100 Hz) (Tucker, 1993). Impedances were maintained below 50 kΩ. The EEG data were offline digitally band-pass filtered at 1–30 Hz and then segmented into 1000 ms epochs (100 ms pre-stimulus). Eye blinks and eye movements were screened by vertical and horizontal EOG channels. Epochs with voltage changes exceeding 55 μV from EOG channels and 100 μV from the other channels were automatically removed from further analysis.

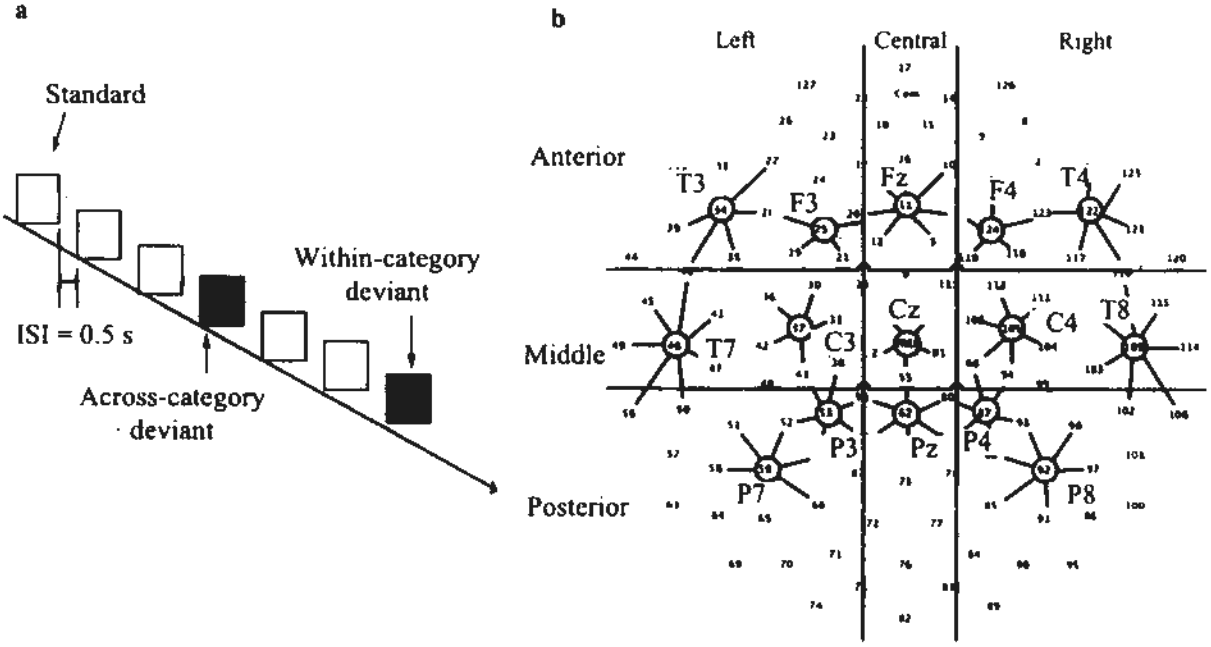


Figure 4.1.2 a) Schematic diagram for the oddball paradigm. b) EGI's 128-channel Channel layout and the clustered channels that correspond to the international 10/20 system and nine spatial locations for the mean amplitude and statistical analysis of the ERP and MMN waves. EGI: Electrical Geodesics, Inc.

4.1.1.4. Behavioral task

A. DISCRIMINATION TASK.

Subjects discriminated from a pair of stimuli in a behavioral same/different test. The stimulus pairs included both two-step and zero-step pairs (e.g. 6-6, 8-8, 6-8 and 8-6) in both contexts. Eight repetitions of each pair were presented to the subjects in different blocks. ISI (inter-stimulus-interval) was set at 500ms and ITI (inter-trial-internal) was 4s. One extra practice block (data was excluded from the analysis) was conducted to familiarize subjects with the procedure. The paradigm was similar as that used in Chapter 3 Experiment III and IV and shown in Figure 3.3.2 with different ITI. The discrimination task lasted 15 minutes after the EEG experiment.

B. IDENTIFICATION TASK.

Subjects labeled the randomly presented speech stimulus #4, #6, #8 and #10, each of which was repeated 12 times, as either Mandarin syllable *yi*^{1} or *yi*^{2}. ITI was set at 4s. Since the tone categories are defined within linguistic contexts, only speech sounds were tested. The paradigm is similar as that used in Chapter 3 Experiment II and shown in Figure 3.2.3 with different ITI.

4.1.2. Data Analysis

4.1.2.1. ERP data

RMANOVA was used for statistical analysis, and Greenhouse-Geisser correction was used for violation of Mauchly's test of sphericity in within-subject effect. When there were any interaction effects, post hoc paired-samples t-tests were carried out.

A. TO OBTAIN MMN

To make the number of trials comparable between the deviants and the standard, only the trials followed by at least two trials of standard were used. After artifact detection, the remaining trials were sorted by stimulus types and averaged to compute the ERPs. On average, there were 56 (SD=±16) trials for each condition. Averaged ERPs were baseline corrected. They were re-referenced to the averaged mastoid. The subject-averaged ERPs were then averaged together to produce a grand-average ERP for each condition. In each condition, MMN was calculated from a difference wave by subtracting the standard ERP from the deviant ERP within the 100–350 ms after a stimulus onset. The ERP and difference waves by electrodes were mapped into 15 clusters based on the 10/20 system.

The 15 electrode-clusters were further grouped into 3×3 (REGION×HEMISPHERE) spatial locations (Figure 4.1.2b).

B. TO VERIFY MMN

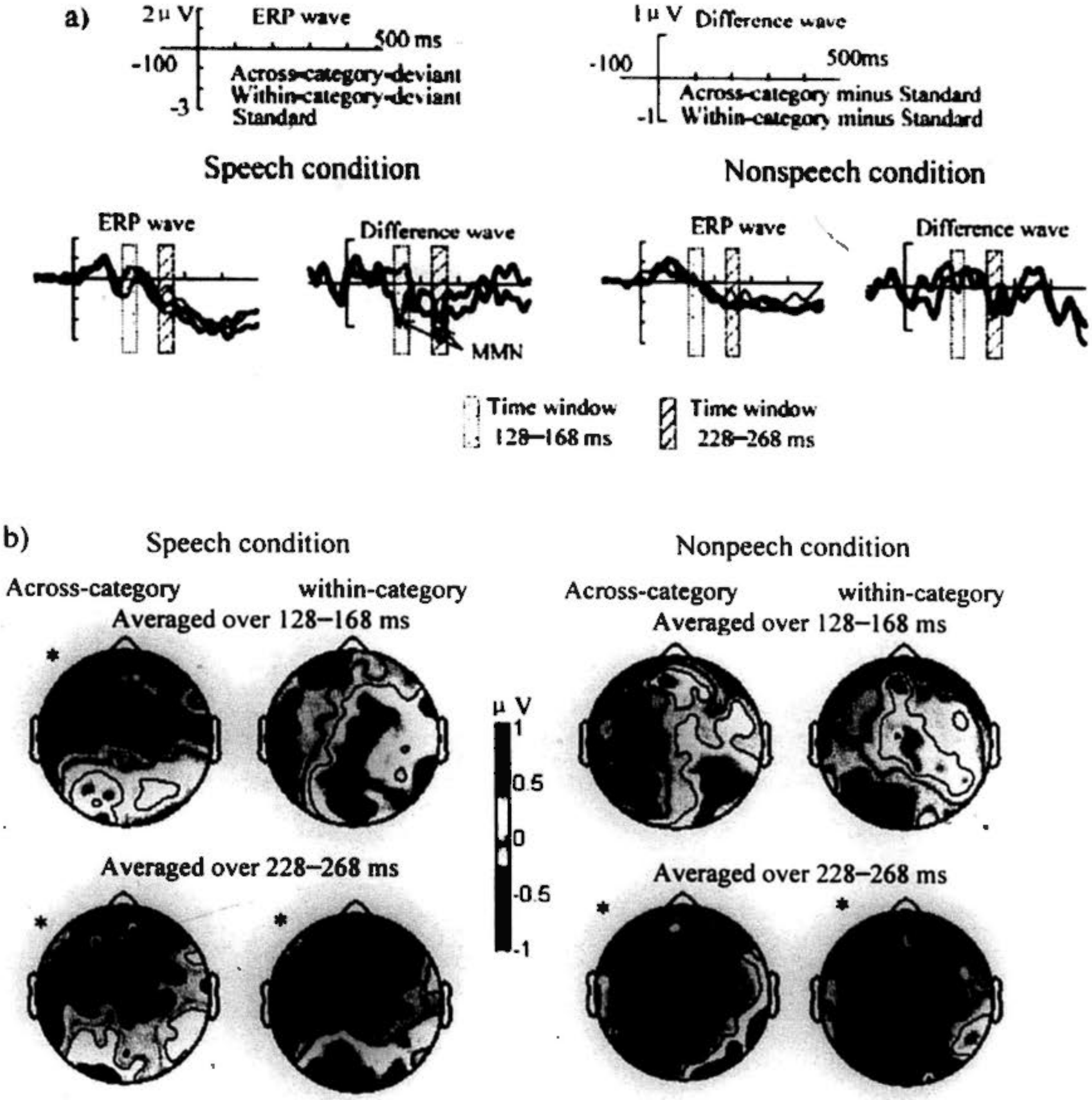


*Figure 4.1.3 a) Grand averaged, from 19 subjects, ERP and difference waves for speech and nonspeech conditions from Fz cluster. Dot-shadowed and slash-shadowed areas represent two time regions (128–168 ms and 228–268 ms) of MMNs. b) Top view of a topographic map for two types of difference waves (across-category minus standard; within-category minus standard) in speech and nonspeech conditions over the two time windows. * : p<.05 on at least one spatial location (details are shown in text).*

MMNs were identified by the negative peaks from grand-average difference waves over all conditions within the determined time window at Fz (Luck, 2005). There were four types of MMNs: within- and across-category deviants in speech context, and within- and across-category deviants in nonspeech context. Each MMN was further verified by the comparison of mean ERP voltage (over the window ±20 ms with the center at the MMN peak) between deviants and standard. The selection of time window followed the steps from previous studies (Luck, 2005; Hao Luo, et al., 2006). Four 2×3×3 three-way within-subject Analysis-of-Variances (ANOVA) for factors of STIMULUS (deviant and standard), REGION (anterior, middle, and posterior) and HEMISPHERE (left, central, and right) were carried out on the four types of deviants respectively. The dependent variables were mean voltages of ERPs.

C. TO ANALYZE MMN

To investigate the difference between across- and within-category deviants, two 2×3×3 three-way within-subject ANOVAs were carried out on speech and nonspeech contexts respectively for factors of CATEGORY (within-category and across-category), REGION and HEMISPHERE. To investigate the difference between speech and nonspeech, two 2×3×3 three-way within-subject ANOVAs were carried out on within- and across-category pairs respectively for factors of CONTEXT (speech and nonspeech), REGION and HEMISPHERE. The dependent variables in these two types of analyses were mean voltages from difference waves.

4.1.2.2. Behavioral data

To obtain the score of discrimination of each pair, calculation was done following the descriptions in (Y. S. Xu, Gandour, & Francis, 2006). In order to compare the behavioral performance with the electrophysiological signature, responses from the across-category pair (6, 8) and the within-category pair (8, 10) were selected for statistical analysis. A 2×2 within-subject ANOVA for factors of CONTEXT and CATEGORY was carried out with the obtained discrimination scores as the dependent variables.

4.1.3. Results

4.1.3.1. ERP results

Two negative peaks (MMN) were identified from grand-averaged difference waves on Fz location: the first one peaked at 148 ms and the second peaked at 248 ms (Figure 4.1.3a).

The high-density EEG recording is good at showing the spatial distribution of ERP. We plotted topographic maps of difference waves over the scalp surface by EEGLAB (Delorme & Makeig., 2004) in Figure 4.1.3b. Two time windows were averaged over 128–168 ms and 228–268 ms respectively for these plots and statistical analyses.

A. FIRST TIME WINDOW (128–168 MS)

(1) Deviant vs. standard.

A significant three-way (STIMULUS×REGION×HEMISPHERE) interaction effect, was found in the across-category pair in speech context [F(4, 72)=3.252, *p*=0.045]. Post Hoc paired-samples t-tests between the across-category-deviant and the standard in speech context revealed that the STIMULUS effect reach a significant level in left-anterior and left-middle locations (Figure 4.1.3b)[$M_{left\text{-}anterior}$ = -1.65 μV, t(18)=-2.58, *p*=.019; $M_{left\text{-}middle}$=-1.14 μV, t(18)=-3.33, *p*=.004]. A significant two-way (STIMULUS×HEMISPHERE) interaction effect was also found in the across-category pair in nonspeech context [F(2, 36)=4.659, *p*=.037]. However, post hoc paired-samples t-tests failed to find a significant effect at any of the nine spatial locations. There was no significant STIMULUS main or interaction effects observed in the other two types of deviants. The results showed that only across-category-deviant in speech context elicited significant MMNs in this time window.

(2) Across- vs. within-category.

A significant main CATEGORY effect was obtained from speech context but not from nonspeech context through the three-way (CATEGORY×REGION×HEMISPHERE) within-subject ANOVA [speech: F(1,18)=6.08, *p*=.024, Macross=-0.608μV, Mwithin=0.318μV; nonspeech: F(1,18)=0.246, *p*=.626, Macross=-0.041μV, Mwithin=0.109μV]. The result verified that CP was observed in speech context only.

(3) Speech vs. nonspeech context.

A significant three-way (CONTEXT×REGION×HEMISPHERE) interaction effect was obtained from across-category stimuli [F(2, 72)=3.766, *p*=.023]. Post Hoc paired-samples t-test between speech and nonspeech on different locations revealed a significant difference at the right-middle location [t(18)=2.19, p =.042]. No significant CONTEXT main or interaction effect was obtained from within-category stimuli. It showed that speech and nonspeech differed in across-category only.

B. SECOND TIME WINDOW (228–268 MS)

(1) Deviant vs. standard.

In across-category-deviant of speech context, no significant STIMULUS main effect or STIMULUS interaction effect was observed. However, a significant two-way (STIMULUS×REGION) interaction effect was found from within-category-deviant. Post Hoc paired-samples t-tests between within-category-deviant and standard in speech context revealed that STIMULUS effect reach a significant level in Fz and left-middle locations [M_{Fz} = -1.06 μV, t(18)=-2.12, *p*=.048; $M_{left-middl}$e=-1.26 μV, t(18)=-2.35, *p*=.03]. In nonspeech context, significant STIMULUS main effects were observed in both category types: within-category F(1,18)=5.46, *p*=.031; and across-category, F(1,18)=4.922, *p*=.04. No significant interaction effect of STIMULUS was obtained in nonspeech context, which showed that the MMN was a whole head distribution. The results showed that in this time window, except for across-category-deviant in speech context, all other three types of deviants elicited significant MMNs. Moreover, MMNs from speech and nonspeech contexts showed different spatial distributions.

(2) Across- vs. within-category and speech vs. nonspeech.

There was no significant CATEGORY main or interaction effect in the comparison between across- and within-category stimuli, or any significant CONTEXT main or interaction effect in the comparison between speech and nonspeech context.

4.1.3.1. Behavioral results.

A. CATEGORY BOUNDARY

Results of obtained discrimination and identification scores were shown in Figure.4.1.4. The discrimination peak in the speech context was located at stimulus pair (6, 8) and consistent with previous results (Wang, 1976). In the identification task, stimulus #4 and #6 were correctly labeled as yi2 at 100% (SD = ±2%) and 93% (SD = ±10%) respectively; #8 and #10 were correctly labeled as yi1 at 91% (SD = ±20%) and 100% (SD = ±0%) respectively. Therefore, linguistic category was located between stimulus #6 and #8.

However, the discrimination peak in the nonspeech context was slightly different from that in the speech context: two pairs of stimulus (4, 6) and (5, 7) got exactly the same mean scores at the peak. Thus, the category boundary in nonspeech context was less clear and was different from that in speech context.

B. ACROSS- VS. WITHIN-CATEGORY

A CONTEXT×CATEGORY within-subject ANOVA on obtained discrimination scores showed no significant SPEECH main effect, but a significant CATEGORY main effect [F(1,18)=42.509, $p<.001$], and a two-way interaction effect [F(1,18)=9.108, $p=.007$]. Post Hoc paired-samples t-tests revealed that discrimination of the across-category pair was easier than that of the within-category pair in both speech and nonspeech contexts [speech: t(1,18)=7.97, $p<.001$, $M_{across}=0.728$, $M_{within}=0.580$; nonspeech: t(1,18)=2.713, $p=.014$, $M_{across}=0.658$, $M_{within}=0.593$]. The significant difference between across- and within-category pairs indicated the presence of CP for lexical tones in both speech contexts. The interaction effect indicated that CP is more salient in speech context.

C. SPEECH VS. NONSPEECH CONTEXT

Two paired-samples *t*-tests on discrimination scores between speech and nonspeech contexts were carried out in both across- and within-category pairs. A significant difference was observed in the across-category pair [$t(1,18)=2.476$, $p=.023$], but not in the within-category pair [$t(1,18)=-0.426$, $p=.625$]. The observation indicated that discrimination difference between speech and nonspeech is more prominent in across-category stimuli.

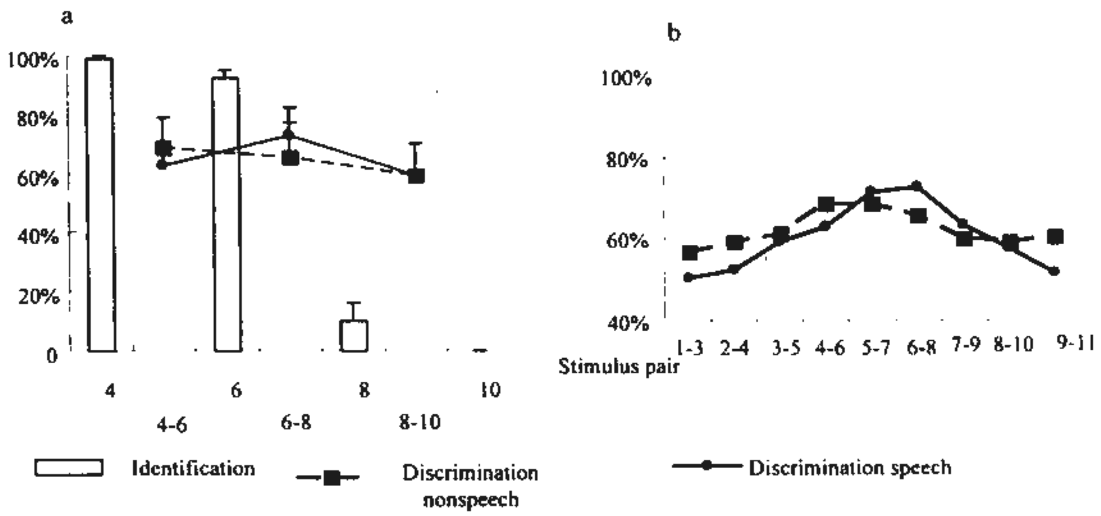


Figure 4.1.4 a) Identification and part of discrimination scores in speech and nonspeech conditions. b) Discrimination scores along the completed continuum in speech and nonspeech conditions. Error bars represented standard error.

4.1.4. Discussion

4.1.4.1. Cognitive stages of categorical perception of lexical tones

In the preattentive stage, CP in speech context is exhibited by the difference of MMN between across- and within-category deviants. The MMN components include both a non-comparison, sensory stage and a comparison, cognitive stage (Maess, et al., 2007); but our MMN data cannot tease apart these two components — which is not our aim here in any case. In contrast, our experiment investigates whether the comparison procedure is only based on specific acoustical features, or also based on tonal category information. The latter view is supported from the results.

We obtained two MMNs peaking at different time windows, which were presumably elicited by two change points of our stimuli (an early 100 ms level portion and a later 400 ms rising portion). The across- and within-category deviants differ from the standard with an equal physical distance at the level portion. Nonetheless, we observed a difference of MMNs in speech context between across- and within-category deviants only in the first time window, which corresponds to the level portion. Thus, the observed MMN difference cannot be explained by acoustic differences. The difference of tonal category is the more plausible explanation. The perceptual distance between stimuli across a category is expanded, because they represent different phonemic categories; while the perceptual distance between stimuli within a category is reduced, because they represent the same phonemic category.

Studies on vowels and consonants demonstrate that MMN reflects both high-order linguistic process and CP (Kasai, et al., 2001; Kazanina, et al., 2006; Nenonen, Shestakova, Huotilainen, & Näätänen, 2005). Cross-language studies on lexical tones also show that processing of lexical tones is influenced by linguistic experience in preattentive stage (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007b). Moreover, the experience influence is observed not only in cortical level but also at the earlier brainstem level (Krishnan, Xu, Gandour, & Cariani, 2005). The training may not be limited to linguistic experience; rather, any relevant experience (e.g. musical training) also shapes the brain in both subcortical (P. C. M. Wong, Skoe, Russo, Dees, & Kraus, 2007) and cortical levels (Chandrasekaran, Krishnan, & Gandour, 2009).

The rising portion is a representative feature of Mandarin Tone 2. However, no significant CATEGORY effect was found from the rising portion in the study, which may be due to the following two reasons.

(1) The level portion of the stimulus fully predicts the rising portion, so that due to the effect of *time window integration* (TWI), the MMN of rising portion is reduced if there is a significant MMN of the level portion. The TWI predicts the reduction of a later MMN if two events are always paired together and occur within 170 ms (Näätänen et al., 2007). In our study, the onsets of level and rising portion are separated only by 100 ms, which is shorter than 170 ms. Therefore, since the change has been automatically detected by the level portion, the rising portion would elicit a smaller MMN or even would not elicit a new MMN at all from the across-category-deviant in speech context. However, we do not exclude the possibility that the rising portion is also involved in a comparison procedure based on the linguistic influence.

(2) The rising portion itself may not be able to elicit a significant linguistic representation of tone in the preattentive stage. For example, MMN elicited by linear approximation of tones (linear ascending ramps) did not show a group difference between Mandarin and English subjects, while the MMN elicited by curvilinear approximation did (Chandrasekaran, Krishnan, & Gandour, 2007a). The cross-language study shows that experience-dependent neural plasticity in the preattentive stage "is sensitive to naturally

occurring pitch dimensions". The curvilinear approximation of tones includes an early level portion in the shape, so the early level portion may form an important component to make the tones sound more natural. This result is also consistent with the production studies: the level portion part is consistently present when it is uttered in isolation (Peng, 2006; Y. Xu & Wang, 2001).

The effect of different shapes of tone approximations was studied in the brainstem stage by the Frequency Following Response (FFR) (Krishnan, Gandour, Bidelman, & Swaminathan, 2009). The tracking accuracy of FFR from the curvilinear approximation showed a group difference based on language experiences. Neither trilinear nor linear approximation of tones showed a group difference in the study. Their result does not contradict our hypothesis, because different stages of brain processing and different shapes of approximation were studied between our study and Krishnan et al.'s (2009). Multiple features are integrated to activate some abstract rules in the MMN stage (Näätänen, et al., 2007), while single feature processing is more prominent in the brainstem stage before the integration. Therefore, while the brainstem at the earlier stage responds more to natural occurring pitch shapes, the MMN at the later stage responds to more abstract pitch shapes as the stimulus becomes more integrated when it travels to higher levels. This is a possible hypothesis but further studies need to be done before we fully understand it.

Furthermore, the shapes of approximation may also play important roles in the brainstem stage. The approximation in Krishnan et al.'s study emphasizes fitting two ends of the pitch contours: no matter what kinds of shapes of approximation, the ends fit best. However, our bilinear approximation does not emphasize the ends but the slopes and angles around the turning points, which may make the sounds in our study more natural than the trilinear or linear approximation used in Krishnan et al.'s. We hope further study will clarify this point.

4.1.4.2. Representation of lexical tones in the brain

When the same pitch pattern occurs in different contexts, the perception is different. The difference is exhibited in both preattentive and postperceptual stages: (1) in preattentive stage, CP is observed in the speech context but not in the nonspeech context; (2) in postperceptual stage, CP is more prominent in the speech context than in the nonspeech context. This difference indicates that the tone perception relies on the base syllable.

These two types of base syllable, vowel and pure tones, differ in both linguistic roles and acoustic features. The difference in acoustic features — harmonic richness and spectral complexity, is not enough to explain the difference of perception. In our stimuli, the across- and within-category stimuli differ from the standard stimuli with the same physical distance at the level portion. If the harmonic richness is enough to make a difference between speech and nonspeech contexts in the across-category stimuli, the same facilitation should also be observed in the within-category stimuli, which is not true in our data. Although a study (Tervaniemi, et al., 2000) found that harmonic partials facilitate pitch height discrimination in the preattentive stage; it did not find an asymmetric effect of the facilitation about the direction of deviants to the standard. That means the facilitations are the same as long as the degrees of difference are the same, no matter whether the pitch height of deviants is higher or lower than the standard.

Moreover, topographic distributions of MMNs from speech and nonspeech contexts were different in our study: MMNs from speech context had an anterior and middle distribution, while MMNs from nonspeech context had a whole head distribution. However, enhancement of harmonic complexity does not change the spatial topographies of MMNs (Zion-Golumbic, Deouell, Whalen, & Bentin, 2007). Even though there is no one-to-one mapping between spatial distributions of scalp recording and location of neural sources, and the function correlated with the difference is unclear, this topographic difference still indicates that their MMN neural sources are different in some aspects. For example, a difference in topographic distribution of the MMN was correlated with a particular brain injury in an aphasia study (Becker & Reinvang, 2007). Because of the limitation of spatial resolution in EEG studies, the difference of source generators should be further investigated by techniques with higher spatial resolution such as MEG or fMRI.

In the preattentive stage, the same pitch pattern in nonspeech context does not show a similar CP as that in speech context. Furthermore, the acoustic complexity is not enough to explain that dissimilarity. Therefore, we conclude the linguistic roles carried by the base syllable contribute to tonal representation. In this stage, both base syllable and pitch patterns are integrated together to trigger neural-networks specific to linguistic experience.

In the postperceptual stage, a similar but less prominent difference between across- vs. within-category stimuli was obtained in nonspeech context when compared with speech context. This finding is consistent with the multistore model proposed by (Y. S. Xu, Gandour, & Francis, 2006): in the later stage, listeners turn their attention to the representative features of lexical tones to do the discrimination, and therefore, rely less on the base syllables.

4.1.4.3. Double MMNs

Double MMNs or MMNs from different time windows in one experiment have been observed in some early experiments (Summarized on Table 4.1.1). For example, a study investigated MMN elicited by complex tones, naturally spoken words, and pseudowords (Korpilahti, Krause, Holopainen, & Lang, 2001). Their results showed that tones elicited a bifurcated MMN, with early MMN (peaking at 150–200 ms) being more dominant. On the other hand, words elicited a strong late MMN, peaking at about 400–450 ms after the stimulus onset, and moreover, it was significantly weaker for pseudowords than for words. The late MMN wave, especially for word differences, was found to reflect summating MMN generators and memory trace formation on gestalt bases. Another study compared the MMN elicited by dyslexia and normal subjects on both tone and speech stimuli. The tone stimuli yielded two MMN components, but no group differences. Three components were found for the speech stimuli and significant differences were obtained between dyslexics and controls in two of the three time windows. The results suggested that phonological deficit is reflected on the later MMN window (Schulte-Körne, Deimel, Bartling, & Remschmidt, 2001).

Although the above two studies both reported linguistic related information is reflected at the later time window of MMN, it doesn't mean that all language related information (i.e. linguistic experience) is only reflected at a later time window. A study showed that language related information is accessed earlier than a general auditory difference (Horev, Most, & Pratt, 2007). In their study, ERPs for Hebrew VOT and *frequency onset time* (FOT) were compared. The results showed that VOT elicit different in N1, P2, but FOT in P2. Moreover, the voicing boundaries corresponded to Hebrew VOT values of production, suggesting that voicing perception in Hebrew is mediated mainly by linguistic experience rather than by innate temporal sensitivity. Therefore, ERP data differed to VOT versus FOT stimuli as early as N1, indicating that brain processing of the temporal aspects of speech and nonspeech signals differ from their early stages.

Horev et al.'s result on the comparison between speech and nonspeech is consistent with our results on that the speech and nonspeech difference was observed in the early MMN. Since the semantic and phonological information stands out in a later time window, the early difference between speech and nonspeech observed from our study may correspond to linguistic experience but not explicit phonological knowledge *per se*.

Table 4.1.1 Summary of major results in MMN experiment

	Preattentive Stage		Postperceptual Stage (Behavioral)
	TW-1	TW-2	
Speech	CP a-MMN –	– – w-MMN	CP
Nonspeech	– – –	– a-MMN w-MMN	CP*

TW-1: First time window (128-168 ms)

TW-2: Second time window (228-268 ms)

CP : Across-category > Within-category, and identification boundary is also consistent with the discrimination peak

* : The discrimination peak is not at the identification boundary defined in the speech context

a-MMN: MMN from across category (Across-category<standard)

w-MMN: MMN from within category (Within-category<standard)

4.1.5. Summary

Previous behavioral experience has shown that long-term linguistic learning induces CP of lexical tones. However, these tasks could not separate different stages of processing, because the responses are the merged product of attention and memory, which these behavioral tasks considerably demanded. The combination of electrophysiological and behavioral experiments reported here is the first one, to our knowledge, to reveal how CP of lexical tones takes place in the time dimension (major results are recapitulated in Table 4.1.1). CP of lexical tones was observed both from MMN waves in the preattentive stage and from behavioral tasks where subjects paid attention to stimuli and made overt discrimination. CP from these two stages may be due to different mechanisms. Difference between across- and within-category stimuli in the preattentive stage reveals that relevant experience influences the perception of lexical tones. In the same preattentive stage, the lack of category effect in nonspeech stimuli with the same pitch contour suggests that the categorical perception is influenced by experience with natural sounding speech. While CP of lexical tones occurs at the preattentive stage for speech, it occurs later for nonspeech at the postperceptual stage. This indicates that the pitch pattern for speech, has been factored away from its base syllable, and has been generalized to the processing of nonspeech materials. In conclusion, the results reported here lead us to understand that categorical perception is a complex multi-stage process which relies on both long-term memory traces and active verbal labeling.

4.2. Experiment II: P300

Experiment II was a P300 study on speech and nonspeech sounds by native Mandarin and Cantonese subjects. The oddball paradigm is similar to what used in MMN study. However, they reflect different cognitive stages. The P300 is used to investigate the cognitive stage when subjects pay attention to the stimuli, while MMN is used to investigate the cognitive stage when subjects' attention is distracted. Moreover, according to previous study, P300 is a more reliable phonemic indicator than other ERP components (Frenck-Mestre, Meunier, Espesser, Daffner, & Holcomb, 2005). Therefore, P300 and MMN are used to investigate different stages of speech perception.

Since the P300 requires the subjects' attention, behavioral responses were also recorded simultaneously. Simultaneous recording ensures the mental state of subjects and the task are the same in both behavioral and ERP data. Therefore, the difference between these two types of methods may be contributed by the different timing of the responses and different sources of the responses (behavioral responses reflect the brain controlled muscle activities and P300 directly reflects the brain activities) and thus provides more information about the temporal dynamics within one process of speech perception.

The impact of language or culture background on cognition has been reported in earlier studies. Specifically, some late ERP components are modulated by culture differences. For example, a study showed that distinct event-related brain potentials (ERPs) were elicited from participants with different cultural backgrounds in response to presentation of affective images (Hot, Saito, Mandai, Kobayashi, & Sequeira, 2006). In their study, culture-modulated differences in the ERPs arose from 170 ms after the stimulus onset, and were most evident for the parietal LPC (late positive component) within the 255–455 ms time window. In another line of research, the ERPs elicited during music-related tasks revealed significant culture-modulated differences in the perception of culturally distinct musical styles (Nan, Knösche, & Friederici, 2006; Nan, et al., 2009) and instruments (Arikan, et al., 1999; Zhu, et al., 2008). Zhu et al. (2008) found that music played to Chinese participants with the *guqin*, an instrument of Chinese origin, elicited a stronger amplitude P300 than piano music did. The authors suggested that participants' differential responses to the *guqin* and piano may be due to the acoustical similarity of the musical tones produced by the *guqin* and the lexical tones of spoken Chinese (Zhu, et al., 2008). These studies showed that the familiar feature will elicit a larger P3 response. The P3 can also be used to index the influence of linguistic experience on speech perception. For example, Japanese listeners do not distinguish */r/* from */l/*, while English listeners do. Such difference is reflected in P3 and behavioral performance by showing a deficient or absent discrimination between these two phonemes from Japanese listener than from native English listener (Buchwald, Guthrie, Schwafel, Erwin, & Vanlancker, 1994).

In the study of tone perception, although the influence of linguistic experience has been reported, most of the studies were focused on the difference between tone and non-tone language [Mandarin vs. English (Bent, Bradlow, & Wright, 2006; Klein, Zatorre, Milner, & Zhao, 2001; Mattock & Burnham, 2006; Wang, 1976; Y. S. Xu, Gandour, & Francis, 2006); Mandarin vs. French (Hallè, et al., 2004); Thai vs. English (Gandour, Wong, & Hutchins, 1998)]. However, 70% languages are tone languages (Yip, 2002) and different tone languages have different tone inventories. For example, Mandarin has four tones and each tone has a distinct pitch pattern (level, rising, falling rising, and falling); Cantonese has six tones and some of the tones share the same pitch patterns (three level tones and two rising tones); Thai has five tones including two level and two falling tones.

Quantity comparison between two tone systems (Mandarin and Cantonese) was conducted on acoustic features (Peng, 2006). In that study, large scale of production by native Mandarin and Cantonese speakers was analyzed in terms of both F0 height and F0 slope for pitch contours. The high level tone and high rising tone are acoustically similar in Mandarin and Cantonese, their speaker-normalized F0 height and slope being almost identical. The tone system of Cantonese is acoustically denser than that of Mandarin. Peng found that the four Mandarin tones tend to be produced distinctly from each other, allowing the Mandarin listener to discriminate them readily. However, he observed significant overlap in the values of F0 height and slope for the Cantonese tones. In particular, Tones 3 and 6 have the same F0 slope and only slightly different F0 height. Furthermore, Tones 2 and 5 have similar F0 height and only slightly different F0 slope. Although Wang (1971) comments that "a fundamental principle is that the sounds of a language tend to maximize the phonetic distance from each other", Peng's (2006) observations appear to provide a counterexample to this principle, and imply that the Cantonese listener might be required to make finer distinctions in perception of both F0 height and slope in order to discriminate certain tones than the Mandarin listener.

Cross-linguistic studies have shown that long-term language experience can influence how tones are perceived (Hallé et al., 2004; Wang, 1976; Xu et al., 2006). Gandour (1983) compared the performance of tone perception in five language groups (four tone language groups: Mandarin, Taiwaness, Thai, and Cantonese; one non-tone language group: English) by INDSCAL multidimensional scaling model. He found that not only tone and non-tone language groups had different weights on the 'height' and 'direction' dimensions, but also tone language groups systematically differed in the relative weights of these two dimensions. Moreover, some studies showed that tone language listeners discriminated between native linguistic tones better than between non native linguistic tones (Y. S. Lee, Vakoch, & Wurm, 1996; Y. S. Xu, Gandour, Talavage, et al., 2006). Nonetheless, none of the studies have investigated the linguistic influence on the perception of the same native tone contrasts existing in both tone systems (for example, rising and level tones existed in both Cantonese and Mandarin systems). The question therefore arises as to how the tone systems of Mandarin and Cantonese influence the perception of F0 contours that occur in both systems.

4.2.1. Method

4.2.1.1. Subjects

Twenty eight right-handed subjects from the Chinese university of Hong Kong, with normal hearing and no reported history of neurological illness, participated in the experiment with payment. Nineteen of them were native Mandarin subjects [7 female and 7 male, mean age = 23.8 ± 3.8; Laterality Quotient = 75.9 ± 25.7 (Oldfield, 1971)] and 19 of them were native Hong Kong Cantonese subjects [8 female and 6 male; age = 20.9 ± 1.6; Laterality Quotient = 69.1 ± 32]. Informed written consent was obtained from each subject. Ethics approval was obtained from Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong. Additional ten subjects, five for each group, also participated in the experiment, but their data was excluded from the analysis due to failure to meet some criteria described in section 4.2.1.5.

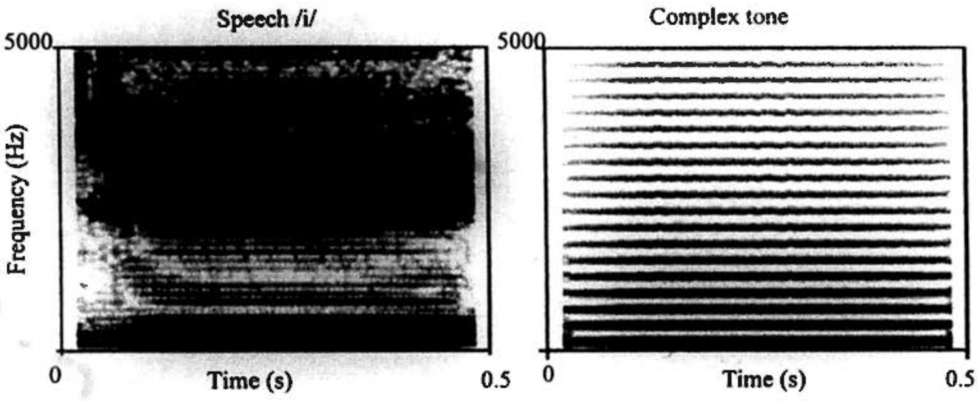


Figure 4.2.1 The spectrogram information of speech and nonspeech (complex tone) in Experiment II.

4.2.1.2. Stimuli

The 500 ms speech stimuli were the same as those used in Experiment I and shown in Figure 4.1.1. In contradict to the pure tone used in Experiment I, the nonspeech used here was based on a complex tone (saw wave), while the other aspects were the same as those in Experiment I. Figure 4.2.1 showed the spectral information for the speech and nonspeech. To make the loudness of nonspeech comparable to that of speech, the speech was delivered at 65dB and the nonspeech at 75dB, according to subjective judgment by two independent researchers.

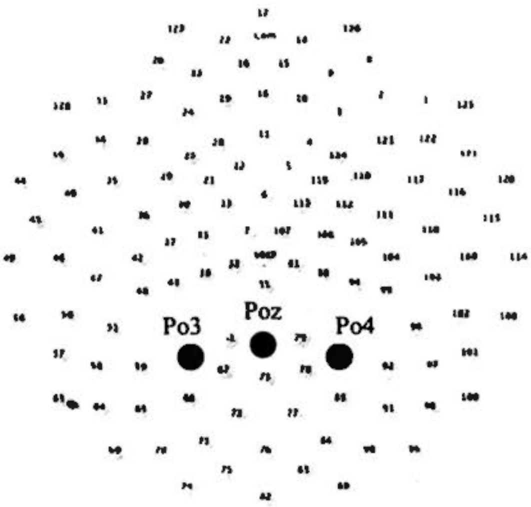


Figure 4.2.2 The location information of 129 channels. Three channels (Poz, Po3 and Po4) were picked up for P300 analysis.

4.2.1.3. Procedure during EEG recording

Four [2 sets (rising and level) × 2 syllable types (speech and nonspeech)] sets of stimuli were presented in the oddball paradigm, which was similar to that in Experiment I and shown in Figure 4.1.2. These four sets of stimuli were presented binaurally to the subjects through a pair of magnetic shielded headphones in a counterbalance order. Each set of stimuli contained one standard and two types of deviants [across-category (AC) deviant and within-category (WC) deviant]. The subjects were asked to press the mouse with two thumbs, whenever they heard a deviant.

In the rising set, the standard was the stimulus #4, the AC deviant was the stimulus #7, and the WC deviant was the stimulus #1; while in the level set, the standard was the stimulus #7, the AC deviant was the stimulus #4, and the WC deviant was the stimulus #10. The standard and the deviants were selected based on earlier results (Wang, 1976). There were 400 trials in each set. Seventy-five percent of them were standard and each type of deviants occupied 12.5% of the stimuli. The trials in a set were presented in a pseudo random order with at least one standard preceded a deviant. The ISI between two trials was jittered in the range of 800–1000 ms. In other words, the *stimulus onset asynchrony* (SOA) was 1300–1500 ms. Every set of stimuli was presented in three blocks with a 2-minute break between two adjacent blocks. After each set of presentation, there was a 5-minute break. Before the recording, a demo of three types of stimuli (standard and two deviants) from the tested set and a practice run were presented to the subjects for the familiar purpose.

4.2.1.4. Procedure during behavioral task

After the EEG recording, the subjects were asked to identify the category of the stimuli (#1, #4, #7 and #10). Demo of the sound categories for two end points (#1 and #10) was shown to the subjects before the task. To make the procedure comparable in speech and nonspeech stimuli, the category was demoed as “sound 1” and “sound 2” without linguistic labeling in speech set. Speech and nonspeech were presented in separated blocks with 15 repetitions for each stimulus in a random order.

4.2.1.5. EEG recording

Electroencephalograph (EEG) data were recorded using a GES 250 system (Electrical Geodesics, Inc., Eugene, OR, USA) with 128-channel Ag/AgCl electrode arrays.

Recordings were carried out in NetStation 4.1 (also supplied by Electrical Geodesics, Inc.). Data were recorded at a rate of 1000 Hz, referenced to the vertex, filtered with an analogue band-pass filter (0.1 Hz to 400 Hz), and digitized using a 16-bit A/D converter (Tucker, 1993). The EEG signals were low-pass filtered at 30 Hz. Eye movements and blinks were monitored with electrodes placed on the supra- and infraorbital ridges of each eye (vertical eye movement), and near the outer canthus of each eye (horizontal eye movement). Epochs with voltage changes exceeding 55 μV from EOG channels and 100 μV from the other channels were automatically removed from further analysis. The electrode impedances were generally kept below 50kΩ, following the recommendation of the manufacturer.

EEG segments were extracted from 100 ms before the stimulus onset to 900 ms after the stimulus onset. The mean voltage in the 100 ms interval prior to the stimulus onset was used as the baseline in the subsequent ERP derivation. If more than 40% of the deviant trials for a particular participant were contaminated by artifacts, which means at least 30 artifact free trials remains for each condition, the entire EEG recording for that participant was excluded. As a result, the EEG recordings obtained from seven participants (five Mandarin and two Cantonese) were excluded from further analysis. Another three participants were also excluded from analysis because they are early bilingual with Cantonese and Mandarin. After artifact detection, the remaining segments were sorted by stimulus types and averaged to compute the ERPs. Averaged ERPs were baseline corrected. They were re-referenced offline against average-mastoid reference. The subject-averaged ERPs were then averaged together to produce a grand-average ERP for each condition.

4.2.2. Data Analysis

4.2.2.1. ERP data.

RMANOVA and mix designs were used for statistical analysis, and Greenhouse-Geisser correction was used for violation of Mauchly's test of sphericity in within-subject effect. To correct the type I error by multiple tests, the threshold value p was corrected by Bonferroni adjustment in pairwise t-test.

A. TO OBTAIN P300.

In total, there were 24 sets of grand-average of ERPs [2 sets (rising and level) ×2 syllable types (speech and nonspeech) ×3 stimuli (1 standard, 2 deviants) ×2 languages (Mandarin and Cantonese)]. Since the comparison was conducted between different types of deviants, difference waves were used for future analysis. Difference wave was obtained by subtracting the standard ERPs from the deviant ERPs. Since P300 has a central-parietal topographic distribution and the topographic maps also showed that the channel 'Poz' had a maximum or a near maximum value, the P300 from Poz was chosen for statistical analysis. The central-parietal distribution and the positive polarity of the difference wave were further verified by the comparison between the mean value (over the window ±30 ms with the center at the peak on channel of Poz) to the "0" value through 24 independent t-tests. All the data from Poz were significantly larger than "0". Since we need to calculate peak latency and to compare it with reaction time, adaptive mean amplitude was used here. This method was provided by EGI Netstation 4.1, which is also adapted by other researchers (Blau, Maurer, Tottenham, & McCandliss, 2007; Electrical Geodesics, 2006; Scerif, Worden, Davidson, Seiger, & Casey, 2006).

B. TO ANALYZE P300.

Analyses of P300 were on the mean value (with 60 ms window centered at individual peak) and the peak latency of the difference wave (deviant-minus-standard) on channel of Poz. On each type of independent variable (mean value and peak latency), a main test was a four-way mix design with three within-subject factors (2 SET: level and rising; 2 CATEGORY: AC and WC; 2 SYLLABLE: speech and nonspeech) and one between-subject factor (2 LANUGAGE: Mandarin and Cantonese).

The SET effect was significant in the four-way interaction effect [SET × SYLLABLE × CATEGORY × LANGUAGE: F(1,26)=14.422, p=0.001]. More importantly, the presence of a two-way interaction effect between SET and CATEGORY [SET × CATEGORY: F(1,26)= 23.42, p<0.001] indicated that the categories were defined differently in the rising and level set. Consequently, level and rising sets were analyzed separately. On each set, a three-way mix design test was carried out. The test had two within-subject factors (2 CATEGORY; 2 SYLLABLE) and one between-subject factor (2 LANGUAGUE).

C. TO ANALYZE LATERALITY P300.

Laterality was tested through between the Po3 and Po4 channels (shown in Figure 4.2.2). Similarly, two independent variables were mean value (with 60 ms window centered at individual peak) and the peak latency of the difference wave. The data were also analyzed into different sets. On each type of independent variable (mean value and peak latency), a main test was a four-way mix design with three within-subject factors (2 CATEGORY: across-category and within-category; 2 SYLLABLE: speech and nonspeech; 2 HEMISPHERE: left and right) and one between-subject factor (2 LANUGAGE: Mandarin and Cantonese).

4.2.2.2. Behavioral data.

There were two types of behavioral results. One was the detection responses, which was recorded during the EEG recording. The analysis of this type of behavioral data was to supplement the P300 data. Since the subjects were pressing the key simultaneously during the recording, the correctness of hitting the key and the reaction time of the response were recorded. Although the reaction time was also recorded during the experiment, the data was not analyzed since we did not have a well control on that. Therefore, only the hit rate was considered in the first type of behavioral results.

When analyzing the data of detection responses, the rate of false alarm was analyzed first, because the discrimination sensitivity is reflected from both false alarm and hit rate. A main test for the rate of false alarm was a three-way mix design. The test had two within-subject factors (2 SET and 2 SYLLABLE) and one between-subject factor (2 LANGUAGE). There was no significant main effect nor any significant interaction effect observed. Therefore, only the hit rate on deviants was considered in the later analysis.

The main test on hit rate was a four-way mix design with three within-subject factors (2 SYLLABLE; 2 CATEGORY; 2 SET) and one between-subject factor (2 LANUGAGE). There was a four-way interaction effect [F(1,24)=5.518, p=0.027] and more important a significant two-way interaction effect between SET and CATEGORY [F(1,26)=47.979, p<0.001]. The strong two-way interaction effect indicated that CP was different in level and rising set. Therefore, the level set and rising set were analyzed separately. On each set, a three-way mix design with two within-subject factors (2 SYLLABLE; 2 CATEGORY) and one between-subject factor (2 LANUGAGE) was carried out as the main test.

The second type of behavioral data was the identification data, which were used to check how the subjects categorize the stimuli into two categories.

4.2.3. Results

4.2.3.1. Rising set (#1,#4,and #7)

A. ERP RESULTS ON POZ

(1) Mean value of difference wave

ERP and difference waves on the Poz were presented on Figure 4.2.3. The topographic maps of difference waves were plotted on Figure 4.2.4. The mean value from deviant#7 was larger than that from deviant #1 (Showed in Table 4.2.1). The results were consistent with our expectation and the results from the post hoc labeling task (described in 4.2.3.2). In the later description, across-category (AC) deviant referred to #4, and within-category (WC) deviant referred to #7 in the rising set.

The three-way mix design revealed a significant three-way interaction effect [CATEGORY × SYLLABLE × LANGUAGE: F(1,26)=4.431, p=0.045], a significant two-way interaction effect [CATEGORY × SYLLABLE: F(1,26)=4.953, p=0.035], and a significant main CATEGORY effect [F(1,26)=15.292, p=0.001]. The main CATEGORY effect indicated the presence of CP (difference between AC and WC). The two-way interaction effect indicated that CPs were different between speech and nonspeech conditions; and moreover, the three-way interaction effect indicated that different language experience contributed to the different degree of CP in speech and nonspeech conditions.

To test how the different carriers modulated CP in different language groups further, two *post hoc* two-way (2 SYLLALBE × 2 CATEGORY) RMANOVA tests were carried out for Mandarin and Cantonese groups respectively. There was a significant two-way interaction effect in Cantonese group but not in Mandarin group [Cantonese: F(1,13)=6.349, p=0.026; Mandarin: F(1,13)=0.014, p=0.908], although both groups had significant main

CATEGORY effects [Cantonese: F(1,13)=10.469, *p*=0.007; Mandarin: F(1,13)=4.932, *p*=0.045]. The results indicated that Cantonese subjects had a different CP between speech and nonspeech conditions, while Mandarin subjects had the similar CP regardless the carrier syllables. Since there was a two-way (CATEGORY and SYLLABLE) interaction effect in Cantonese group, two paired t-tests were carried out on Cantonese data. The results showed a significant CP in speech condition [t(13)=3.488, *p*=0.004], but not in nonspeech condition [t(13)=1.875, *p*=0.083]. Table 4.2.1 listed the mean value of the difference wave. In overall, the mean value from the AC deviant was higher than that from the WC deviant.

The second type of *post hoc* test was carried out to examine the how the language experience affected the speech and nonspeech continua respectively. Two two-way mix designs, with CATEGORY as the within-subject factor and LANGUAGE as the between-subject factor, were carried out. There was a significant two-way interaction effect in speech condition only [CATEGORY × LANGUAGE : F(1,26)=5.257, *p*=0.03], but not in nonspeech condition [CATEGORY × LANGUAGE :F(1,26)=0.198, *p*=0.660], although both conditions showed significant main CATEGORY effect [nonspeech: F(1,26)=6.418, *p*=0.018; speech: F(1,26)=16.344, *p*<0.001]. This *post hoc* test showed that the language experience modulated CP in the speech condition but not in the nonspeech condition, although CP is present in both conditions.

The last type of *post hoc* test was carried out to examine the how the language experience affected the speech vs. nonspeech difference in two categories. Two two-way mixed design with SYLLABLE as the within-category factor and LANGUAGE as the between-subject factor were carried out on across-category stimulus and within-category stimulus respectively. There was no significant two-way interaction effect in either condition, and moreover, the speech vs. nonspeech difference was only observed in across-category deviant [AC: F(1,26)=6.896, *p*=0.014; WC: F(1,26)=0.189, *p*=0.668]. This *post hoc* test showed that subjects performed speech and nonspeech differently only in the across-category stimulus.

In summary, the ERP results showed that CP can be reflected by P300. Moreover, both factors (Language experience and carrier syllables) affected the degree of CP which was reflected by P300 (shown in Figure 4.2.5a and 4.2.6a). The factor of language experience was demonstrated by different sensitivities to the CP difference between speech and nonspeech by different groups. Cantonese subjects had a greater CP in speech condition than that in nonspeech condition; while, Mandarin subjects had a similar CP in speech and nonspeech conditions. The factor of the carrier syllable was demonstrated by effect of language experience was different in speech and nonspeech conditions. Language experience affected CP in speech condition more than in nonspeech condition.

(2) Peak latency of difference wave

The three-way mix design on peak latency showed no significant main or interaction effect. Therefore, no further analysis was done.

Table 4.2.1 Mean value and latency of difference wave (deviant-minus-standard) and hit rate for deviants in rising set (#1, #4, and #7). *: $p<0.0125$ (significant level $p<0.0125$ after correction).

		Mandarin		Cantonese	
		Speech	Nonspeech	Speech	Nonspeech
Mean value (μV)	AC (#7)	4.17	3.69	7.16	4.54
	WC (#1)	3.02	2.61	3.01	3.00
Peak latency (ms)	AC (#7)	510.93	527.29	564.64	524.57
	WC (#1)	481.86	491.57	519.43	495.86
Hit rate	AC (#7)	0.88	0.86	0.85	0.85
	WC (#1)	0.46	0.62	0.36	0.55
paired t-test on mean value					
AC vs. WC paired t-test		p=0.057	p=0.109	p=0.004*	p=0.083
AC	speech vs. nonspeech	p=0.392		p=0.027	
WC	speech vs. nonspeech	p=0.609		p=0.991	
paired t-test on hit rate					
AC vs. WC paired t-test		p=0.001*	p=0.005*	p=0.001*	p=0.018
AC	speech vs. nonspeech	p=0.473		p=0.942	
WC	speech vs. nonspeech	p=0.006*		p=0.018	

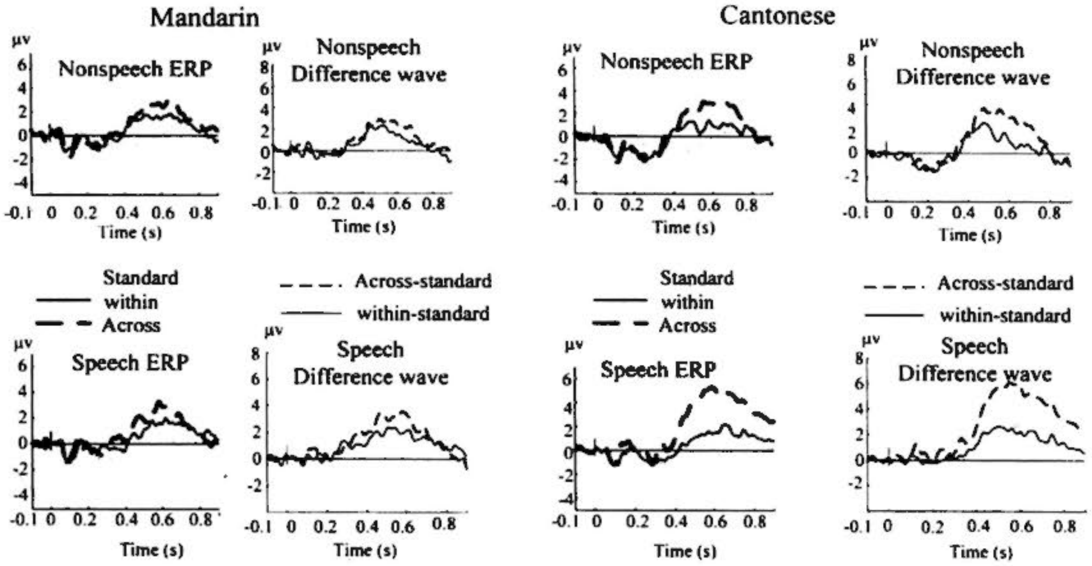


Figure 4.2.3 The EPR and difference waves from rising sets in Experiment II. The left two panels were data from Mandarin subjects, and the right two panels were data from Cantonese subjects. All the data were extracted on channel Poz.

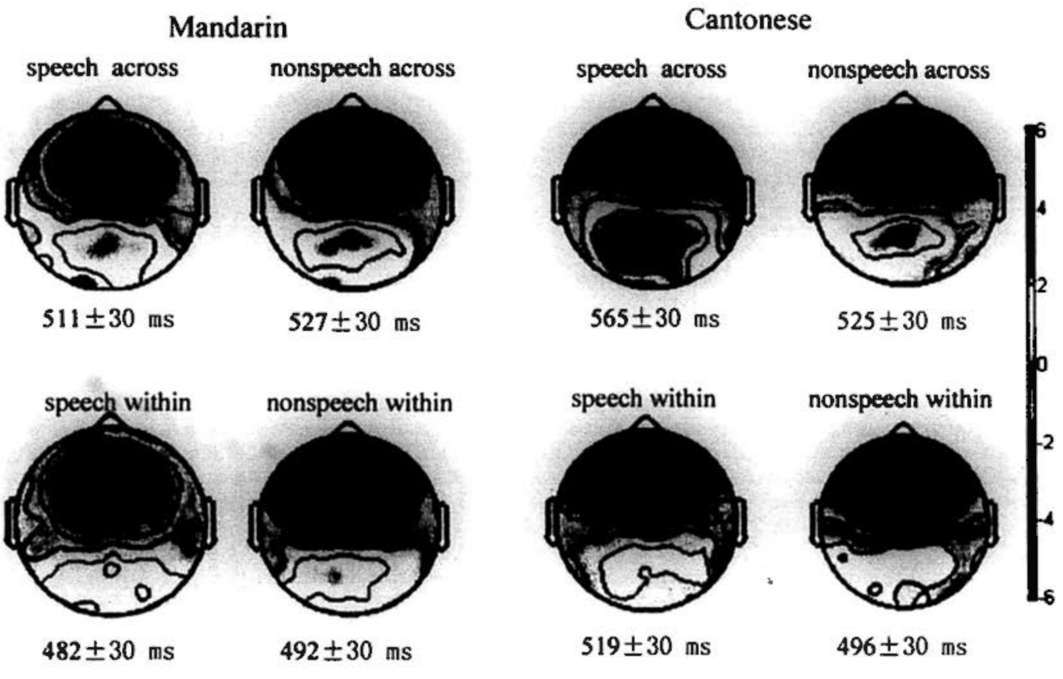


Figure 4.2.4 The topographic maps of difference waves from rising sets in Experiment II. The left two panels were data from Mandarin subjects, and the right two panels were data from Cantonese subjects. The topographic maps were centered at the peak of each condition.

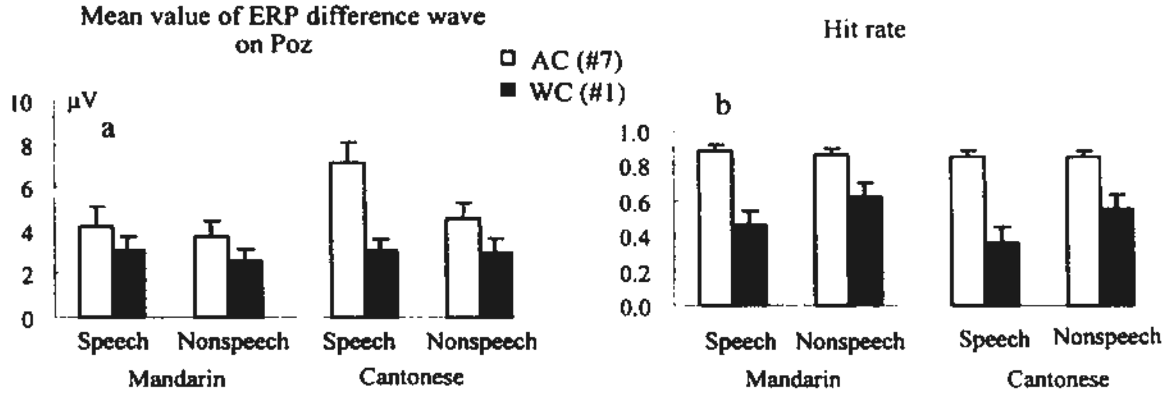


Figure 4.2.5 Summary of (a) ERP difference waves and (b) hit rate in rising set for four conditions (2 syllable types ×2 language groups).

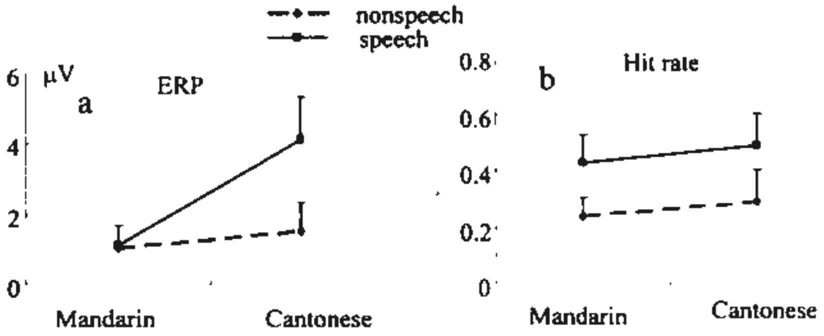


Figure 4.2.6 The value of CP (AC-WC) shown in (a) ERP and (b) hit rate data (see text for details). All data were from level set.

B. BEHAVIORAL RESULTS: HIT RATE.

The hit rate data was also presented in Table 4.2.1. The main test of three-way mix design showed a significant two-way interaction effect between CATEGORY and SYLLABLE, two significant main effect on SYLLABLE and CATEGORY respectively [SYLLABLE × CATEGORY: F(1,26)=15.318, p=0.001; SYLLABLE: F(1,26)=10.502, p=0.003; CATEGORY: F(1,26)=30.614, p<0.001]. There was no significant interaction or main effect on LANGUAGE. The results showed that CP was different in these two carrier types. Moreover, we recoded the data into a hit-rate-difference between AC and WC (AC-WC), which showed a CP. Then we conducted a paired t-test on this CP between speech and nonspeech conditions. The results showed that a significant effect on carrier syllable [t(27)=3.998,p<0.001], which indicated that CP was more prominent in speech condition than in nonspeech condition (see in Figure 4.2.6).

To further investigate what was CP pattern in speech and nonspeech continua, two types of *post hoc* analyses were conducted. The first type of *post hoc* test was two one-way

RMANOVA with CATEGORY as the within-subject factor on speech and nonspeech conditions respectively. The results showed significant main CATEGORY effects in both conditions [nonspeech: F(1,27)=17.36, $p<0.001$; speech: F(1,27)=38.885, $p<0.001$].

The second type of *post hoc* test was one-way RMANOVA with the within-subject factor of SYLLABLE. Two one-way RMANOVA were conducted on across-category and within-category deviants respectively. The results showed a significant main SYLLABLE effect in within-category deviant but not in across-category deviant [within-category: F(1,27)=17.354, $p<0.001$; across-category: F(1,27)=0.256, $p=0.617$].

In summary, the data from the hit rate showed that (1) CP was present in all conditions; (2) CP in speech condition was more prominent than that in nonspeech condition; (3) CP from two language groups was not significantly different.

C. ERP RESULTS IN LEFT AND RIGHT HEMISPHERES

(1) Mean value of difference wave

The mean value of difference wave from the left/right channels in different conditions was shown in Figure 4.2.7. The four-way mix design, with three within-subject factors (2 CATEGORY; 2 SYLLABLE; 2 HEMISPHERE) and one between-subject factors (2 LANGUAGE), showed a significant two-way effect about laterality [LANGUAGE × HEMISHPERE: F(1,26)=4.323, $p=0.048$)]. The results indicated that the lateralization pattern depended on the language experiences. Since the presence of significant CATEGORY and SYLLABLE main effects, the data was not grouped together. The results indicated that regardless the type of carrier syllable and the difference between AC and WC, the ERP from Cantonese subjects was more left lateralized than that from the Chinese subjects.

To further investigate how different language experiences affected the lateralization pattern, two types of *post hoc* analyses were carried out. First type of *post hoc* analysis was two three-way RMANOVA with three within-subject factors (2 SYLLALBE; 2 CATEGORY; 2HEMISHPERE) for Mandarin and Cantonese groups respectively. The result showed a significant two-way interaction effect and a significant main effect in Cantonese group [SYLLALBE ×HEMISHPERE: F(1,13)=5.369, $p=0.037$; SYLLABLE: F(1,13)=5.402, $p=0.037$]. No main or interaction effect was obtained in Mandarin group.

Furthermore, since the presence of the interaction effect in Cantonese group, the laterality of Cantonese data was further broken down into speech and nonspeech sets. The two-way RMAONVA with CATEGORY and HEMISPHERE as the within-subject factors on speech set showed a significant laterality effect [HEMISPHERE: F(1,13)=9.871, *p*=0.008]. No significant laterality effect was observed in nonspeech set. The results indicated that in the speech condition, the left hemisphere had higher ERP amplitude than the right hemisphere did; while no significant laterality effect was observed in nonspeech condition.

Another *post hoc* analysis was to investigate how the effect of language experience modulated the P300 in different hemispheres. Two three-way mix designs were conducted with two within-subject factors (2 SYLLALBE; 2 CATEGORY) and one between-subject factor (2 LANGUAGE) for left and right hemispheres respectively. The results showed a LANAGUAGE, a SYLLABLE and a CATEGORY main effect in left hemisphere [LANAGUAGE: F(1,26)=7.635, *p*=0.007; SYLLABLE: F(1,26)=5.558, *p*=0.026; CATEGORY: F(1,26)=8.264, *p*=0.008]. Only a main CATEGORY effect was observed in right hemisphere [CATEGORY: F(1,26)=4.787, *p*=0.038;], but no main or interaction effect on LANGUAGE. The results indicated that CP, the language effect, and the syllable effect were all presented in the left hemisphere, while only CP was presented in the right hemisphere.

In summary, the laterality analyses on rising set showed that (1) laterality effect was more prominent in Cantonese group, especially in speech set by showing more left lateralized of ERP for speech sounds (see Figure 4.2.7); (2) although both hemispheres showed CP, only left hemisphere showed a language and a syllable effect (see Figure 4.2.8).

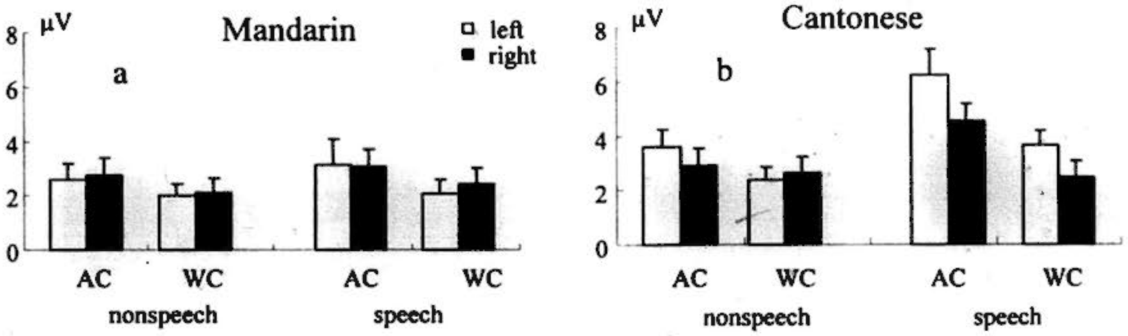


Figure 4.2.7 The value of difference waves on left and right channels in different conditions. a) Data for Mandarin subjects; b) data for Cantonese subjects. All data were for rising set. AC: across-category deviant; WC: within-category deviant.

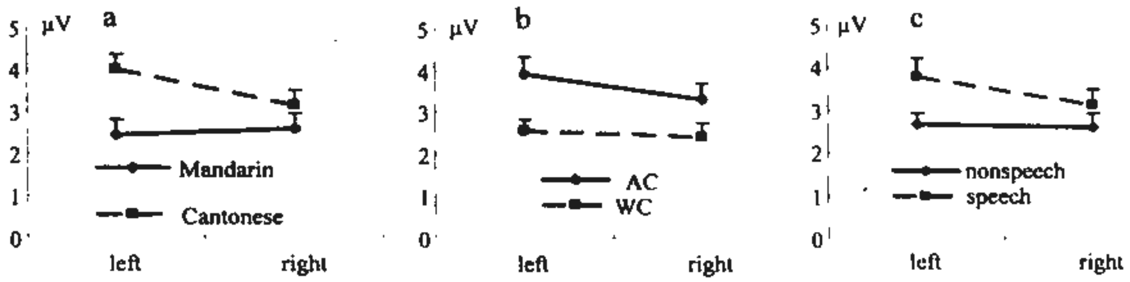


Figure 4.2.8 Summary of hemisphere lateralization pattern contributed by a)language experiences, b) types of deviants, and c)types of carrier syllables. All data were for rising set. AC: across-category deviant; WC: within-category deviant.

(2) Peak latency of difference wave

The main test on the peak latency for laterality on Po3 and Po4 was a five-way mix design with three within-subject factors (2 CATEGORY; 2 SYLLABLE; 2 HEMISPHERE) and two between-subject factors (2 LANGUAGE). There was no interaction or main effect. Therefore, no further analysis on peak latency for laterality of rising set was conducted.

4.2.3.2. Level set (#4,#7,and #10)

A. ERP RESULTS ON POZ

(1) Mean value of difference wave

ERP and difference waves on the Poz of level set were presented on Figure 4.2.9 and listed in Table 4.2.2. The topographic maps of difference waves were plotted on Figure 4.2.10. We observed a reverse pattern of expected on the definition of across-category and within-category. From the previous results (Wang, 1976) and the behavioral identification results, we expected the deviant #4 was an AC deviant to the standard #7, and the deviant #10 was a WC deviant to the standard #7. However, the peak magnitude of #4 was smaller than that of #10, which indicated a reverse the category distinction. We would discuss the reverse effect in the discussion section. To be consistent with the actual response, we still called the #4 as WC deviant and #10 as AC deviant.

The three-way design main test revealed a significant two-way interaction effect [CATEGORY×LANGUAGE: F(1,26)=4.77, p=0.038], two significant main effect [CATEGORY: F(1,26)=14.388, p=0.001; LANUGAGE: F(1,26)=12.88, p=0.001]. The main CATEGORY effect indicated the presence of CP (difference between AC and WC). The interaction effect between CATEGORY and LANGUAGE indicated that the degree of

CP was different in two languages. Since no SYLLABLE main or interaction effect, the speech data and nonspeech data were pooled together in the later analysis.

To further test what kind of CPs in different language groups, two *post hoc* one-way (2 CATEGORY) RMANOVA tests were carried out on Mandarin and Cantonese groups respectively, with the dependent variable of averaged value of speech and nonspeech. There was only a significant CATEGORY main effect in Cantonese group [Cantonese: F(1,13)=14.675, p=0.002]. There was no significant effect in Mandarin group. The results showed that CP was more prominent in Cantonese group, while there was no evidence showing the presence of CP in the Mandarin group.

Table 4.2.2 Mean value and latency of difference wave (deviant-minus-standard) and hit rate for deviants in level set (#4, #7, and #10) . *: $p<0.0125$ (significant level $p<0.0125$ after correction)

		Mandarin		Cantonese	
		Speech	Nonspeech	Speech	Nonspeech
Mean value (μV)	WC (#4)	2.41	2.55	3.30	4.69
	AC (#10)	2.73	3.67	7.00	6.33
Peak latency (ms)	WC (#4)	495.36	490.07	546.71	507.43
	AC (#10)	525.93	529.50	542.57	599.79
Hit rate	WC (#4)	0.63	0.60	0.44	0.65
	AC (#10)	0.81	0.93	0.91	0.95
paired t-test on mean value					
AC vs. WC paired t-test		p=0.692	p=0.107	p<0.001*	p=0.186
WC	speech vs. nonspeech	p=0.847		p=0.273	
AC	speech vs. nonspeech	p=0.327		p=0.470	
paired t-test on hit rate					
AC vs. WC paired t-test		p=0.092	p<0.001*	p<0.001*	p=0.002*
WC	speech vs. nonspeech	p=0.62		p=0.012*	
AC	speech vs. nonspeech	p<0.001*		p=0.072	

The second type of *post hoc* test was carried out to examine the how the language experience affects the two categories. Two one-way ANOVA, with LANGUAGE as the between-subject factor and the pooled value of speech and nonspeech, were carried out on deviant #4 (WC) and deviant #10 (AC) sets respectively. The significant main LANGUAGUE effect was only observed in deviant #10 by showing the higher peak value in Cantonese than in Mandarin subjects [F(1,26)=15.948, $p<0.001$; Cantonese:

mean=6.663μV, std.err=0.631μV; Mandarin: mean=3.201, std.err=0.631μV]. No significant interaction or main effect was observed in the analysis of deviant #4 (WC).

In summary, the results, presented in Figure 4.2.11a and Figure 4.2.12a, indicated that (1) CP in level set is only observed in Cantonese group, but not in Mandarin group; (2) Cantonese had a higher P300 response to deviant #10 (AC) than Mandarin did; and (3) there was no significant difference between speech and nonspeech.

(2) Peak latency of difference wave

The three-way mix design showed a significant CATEGORY main effect [F(1,26)=12.752, p=0.001]. There was no other significant interaction or main effect observed. The results showed that the subject responded to the WC deviant #4 (509.2±15 ms) faster than to the AC deviant #10 (550.0±14 ms).

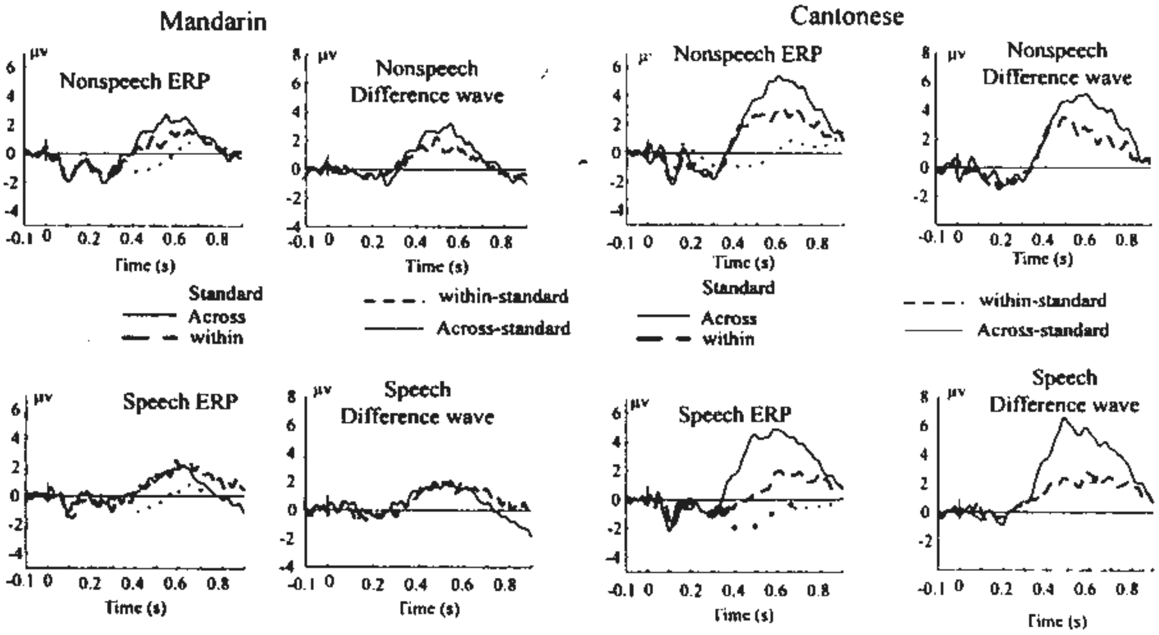


Figure 4.2.9 The EPR and difference waves from level set in Experiment II. The left two panels were data from Mandarin subjects, and the right two panels were data from Cantonese subjects. All the data were extracted on channel Poz.

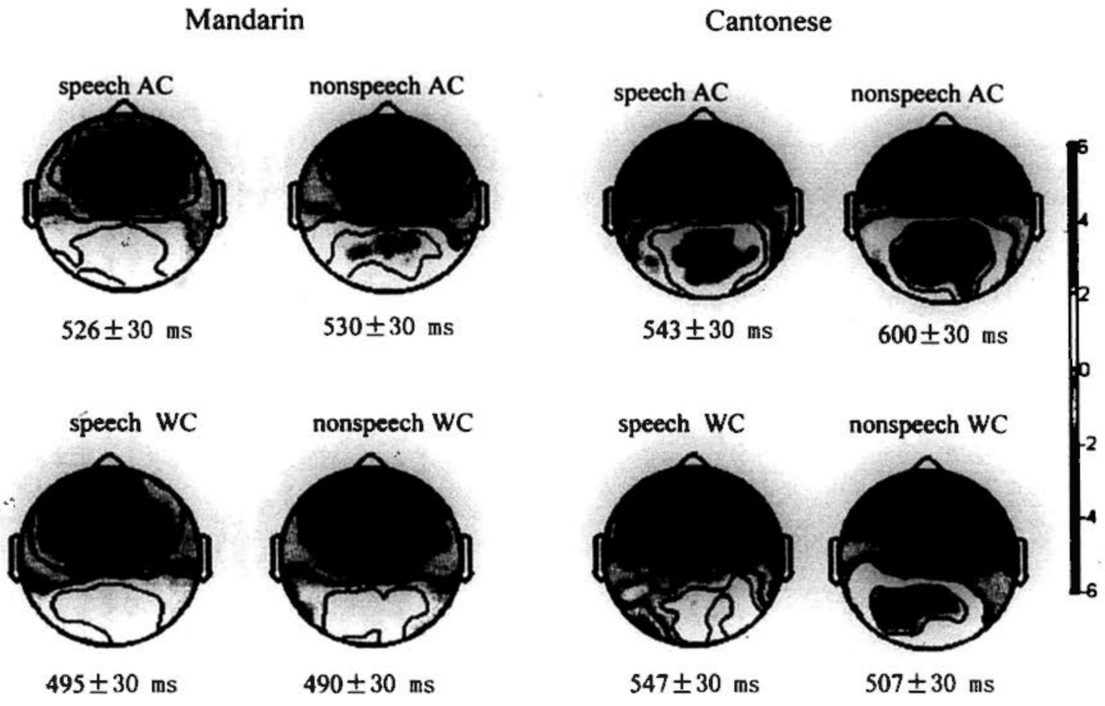


Figure 4.2.10 The topographic maps of difference waves from level set in Experiment II. The left two panels were data from Mandarin subjects, and the right two panels were data from Cantonese subjects. The topographic maps were centered at the peak of each condition.

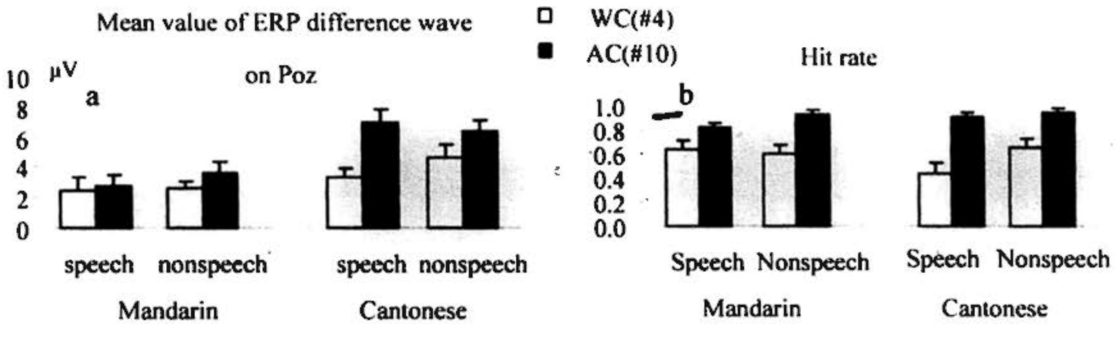


Figure 4.2.11 Summary of (a) ERP difference waves and (b) hit rate in level set for four conditions (2 syllable types ×2 language groups).

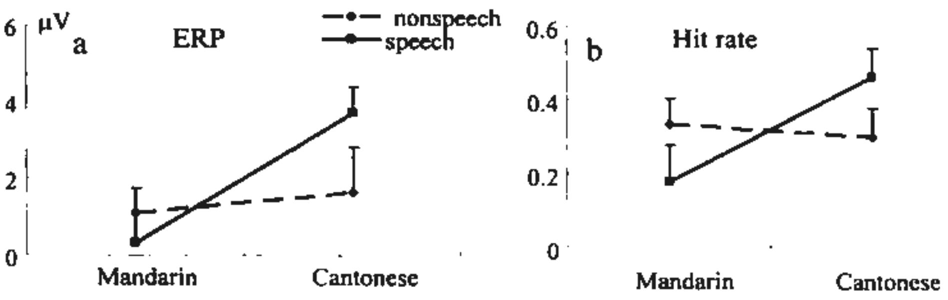


Figure 4.2.12 The value of CP (AC-WC) shown in (a) ERP and (b) hit rate data (see text for details). All data were from level set.

B. BEHAVIORAL RESULTS: HIT RATE

The main test was a three-way mix design showed a significant three-way interaction effect [F(1,26)=7.208, *p*=0.012]. The results indicated that both factors of language experience and types of carrier syllables contributed to CP. Therefore, some *post hoc* analyses were carried out.

The first type *post hoc* was to investigate how CP influenced by the types of carrier syllables in different language groups. Consequently, two RMAOVA with two within-subject factors (2 SYLLABLE; 2 CATEGORY) were carried out on Mandarin and Cantonese subjects respectively. The results showed no interaction effect in either language group. However, there was a significant CATEGORY main effect in both language subjects [Mandarin: F(1,13)=12.72, *p*=0.005; Cantonese: F(1,13)=33.613, *p*<0.001], and a significant SYLLABLE effect in Cantonese group [F(1,13)=16.122, *p*=0.001]. The results indicated that although CP was observed in both language groups, Cantonese speaker responded to speech and nonspeech differently.

The second type of *post hoc* analysis was to investigate how language experience influenced the distinguishing between speech and nonspeech in each category. Consequently, two two-way mix designs with one within-subject factor (2 SYLLABLE) and one between-category factor (2 LANGUAGE) were carried out on AC and WC deviants respectively. Both sets had a significant two-way (LANGUAGE×SYLLABLE) interaction effect [AC: F(1,26)=6.4, *p*=0.018; WC: F(1,26)=4.733, *p*=0.03] and the WC set also had a significant main SYLLABLE effect [F(1,26)=23.535, *p*<0.001]. Break down the data further and conduct the paired t-test between speech and nonspeech on each language group. There were four pairs of t-tests. The results showed that Mandarin group had a significant speech vs. nonspeech difference in AC set [t(13)=4.812, $p<0.001$], while Cantonese group had a significant speech vs. nonspeech difference in WC set [t(13)=2.9, $p=0.012$].

The last type of *post hoc* analysis was to investigate how language experience influenced the degree of CP under each type of carrier syllable. Consequently, two two-way mix designs, with one within-subject factor (2 CATEGORY) and one between-category factor (2 LANGUAGE), were carried out on the speech and nonspeech continuum respectively. Both sets had a significant main CATEGORY effect [nonspeech: F(1,26)=36.871, $p<0.001$; speech: F(1,26)=25.802, $p<0.001$]. The speech set also had a significant two-way (LANGUAGE × CATEGORY) interaction effects [F(1,26)=5.013, $p=0.034$]. Break down the speech data further and conduct the paired t-test between AC and WC on each language group. There were two pairs of t-tests. The results showed that Mandarin group did not have a significant CP but the Cantonese group had a significant CP in speech condition [t(13)=5.013, $p<0.001$].

In summary, the results indicated that 1) in speech condition, two language groups had different CPs by showing that Cantonese had a CP but Mandarin group did not; (2) in nonspeech condition, both language groups had CPs and there was no language difference; (3) Mandarin responded to within-category deviant differently between speech and nonspeech conditions, while Cantonese responded to across-category deviant differently between speech and nonspeech.

C. ERP RESULTS ON LEFT AND RIGHT HEMISPHERES

(1) Mean value of difference wave

The four-way mix design on Po3 and Po4, with three within-subject factors (2 CATEGORY; 2 SYLLABLE; 2 HEMISPHERE) and one between-subject factors (2 LANGUAGE), showed a significant three-way interaction effect on HEMISHPERE [(SYLLABLE × CATEGORY × HEMISHPERE: F(1,26)=7.389, $p=0.012$)]. The results indicated that across-category and within-category deviants had a different laterality pattern regarding the syllable type. From Figure 4.2.14, we observed that CP (difference between AC and WC) in speech and nonspeech conditions was better presented in the left and right hemisphere respectively. Recode the data into CP (AC-WC) in four conditions (speech+left, speech+right, nonspeech+left, and nonspeech+right) and conduct the four paired t-tests. The results confirmed that CP of speech in the left hemisphere was significantly higher than that in the right hemisphere [t(27)=2.782, *p*=0.01]. However, CP of nonspeech did not show significant lateralization on either hemisphere [t(27)=1.563, *p*=0.130].

To further test how the different hemispheres responded to CP in different carrier syllables, two *post hoc* two-way (2 SYLLALBE × 2 CATEGORY) RMANOVA tests were carried out on left and right hemisphere receptivity. There was a significant main CATEGORY effect in both hemispheres [left: F(1,27)=10.624, *p*=0.003; right: F(1,13)=5.752, *p*=0.024]. There was no other interaction or main effect in either hemisphere. The results indicated that CP was present in both hemispheres as shown in Figure 4.2.15b.

To further test how the laterality pattern of CP was present in speech and nonspeech conditions, two *post hoc* two-way (2 HEMISPHERE × 2 CATEGORY) RMANOVA tests were carried out on speech and nonspeech respectively. There was no significant main or interaction effect observed in nonspeech condition. However, there was a significant main CATEGORY effect and a significant interaction effect in speech condition [CATEGORY: F(1,27)=8.546, *p*=0.007; HEMISPHERE × CATEGORY: F(1,13)=7.74, *p*=0.01]. The results showed that in speech condition, deviant #10 (AC) had a left lateralization and deviant #4 (WC) had a right lateralization pattern (See Figure 4.2.14).

In summary, the laterality analyses on the level set showed that CP was presented in both hemispheres. However, speech and nonspeech had different preferable hemisphere to present CP. CP of speech was better presented in the left hemisphere, while, CP of nonspeech was more bilateral. In speech condition, AC was right lateralized and WC was left lateralized.

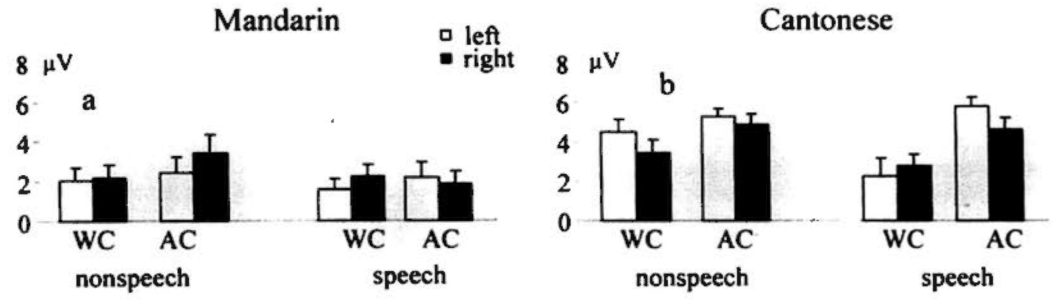


Figure 4.2.13 The value of difference waves on left and right channels in different conditions. a) Data for Mandarin subjects; b) data for Cantonese subjects. All data were for level set. AC: across-category deviant; WC: within-category deviant.

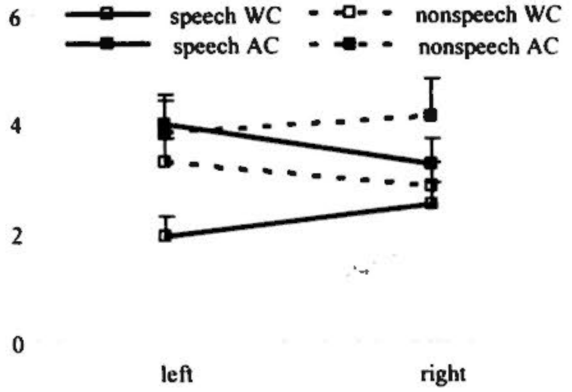


Figure 4.2.14 CP in speech and nonspeech conditions presented in two hemispheres (the level set)

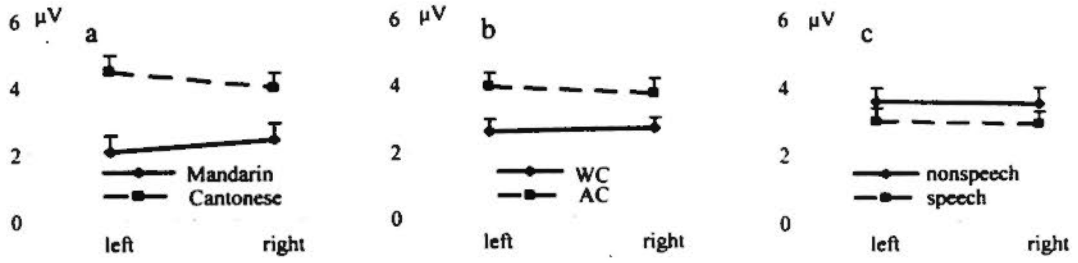


Figure 4.2.15 Summary of hemisphere lateralization pattern contributed by a)language experiences, b) types of deviants, and c)types of carrier syllables. All data were for rising set. AC: across-category deviant; WC: within-category deviant.

(2) Peak latency of difference wave

The main test on the peak latency for laterality on Po3 and Po4 was a five-way mix design with three within-subject factors (2 CATEGORY; 2 SYLLABLE; 2 HEMISPHERE) and two between-subject factors (2 LANGUAGE; 2 SEX). There was no any interaction or main effect on HEMISPHERE. Therefore, no further analysis on peak latency for laterality of level set was conducted.

4.2.3.3. Identification results

Proportion of the stimulus labeled as rising tone (tone 1) was analyzed. The data for each language group was presented in Figure 4.2.16. From the previous data (Wang, 1976), the stimulus #7 was the category boundary. The position of category boundary was confirmed by the averaged data shown in Figure 4.2.16, and showed that 50% mid point was located between stimulus #4 and #7. Since we only cared about the boundary, only the data from stimulus #4 and #7 was analyzed. The difference between stimulus #4 and #7 was defined as the boundary slope.

A three-way mix design was carried out with two within-subject factors (2 STIMULUS: #4 and #7; 2 SYLLABLE: speech and nonspeech) and one between-subject factors (2 LANGUAGE: Cantonese and Mandarin). The result showed no main or interaction effect relevant to LANGUAGE. Therefore, there was no significant identification performance difference between two language groups. The results showed a significant two-way interaction effect, two significant main effect [STIMULUS × SYLLABLE: F(1,26)=13.328, $p=0.001$; STIMULUS: F(1,26)=354.4, $p<0.001$; SYLLABLE: F(1,26)=25.005, $p<0.001$]. The two-way interaction effect indicated that the boundary slopes were different in speech and nonspeech condition.

Two types of *post hoc* analyses were carried out. The first type of analysis was a paired t-test on the boundary slope (difference between #4 and #7) between speech and nonspeech. The results showed that boundary slope in speech was significantly steeper than that in nonspeech condition [t(27)=3.719, $p=0.001$]. The second type of analysis was to investigate which stimulus was labeled differently in speech and nonspeech conditions. Therefore, two paired t-tests were carried out between speech and nonspeech on two stimuli. The results showed stimulus #7 was labeled significantly different in two conditions [t(27)=4.793, $p<0.001$], while labeling of stimulus #4 reached a marginal significant effect [t(27)=2.038, $p=0.051$].

In summary, the identification results confirmed the category boundary was between stimulus #4 and #7, which was consistent with previous results. Moreover, we found that the boundary was sharper in speech than that in nonspeech. However, we did not find a language difference on identification performance.

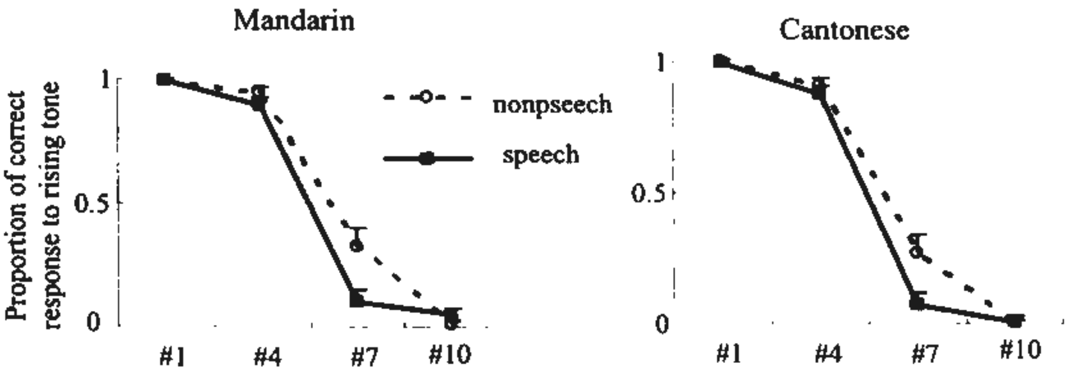


Figure 4.2.16 Identification results on stimulus #1,#4,#7, and #10 from Mandarin and Cantonese subjects. Error bar represented standard error.

4.2.4. Discussion

4.2.4.1. Category boundary

The *post hoc* labeling results showed that subjects quite reliably located the category boundary between stimulus #4 and #7, which met our expectation and was consistent with previous results (Wang, 1976). In other words, #4 was labeled as rising tone and #7 was labeled as Level Tone. There was no performance difference between Mandarin and Cantonese subjects.

However, when the stimuli were presented in the oddball paradigm, the category of stimulus #7 was changed according to its role in the oddball. In the rising set, when the stimulus #7 was presented as a deviant, comparing to the standard #4, the responses showed #7 had a different category to the #4. In other words, in the rising set oddball, deviant #7 was labeled as Level Tone. In the level set, when the stimulus #7 was presented as a standard, comparing to the deviant #4 and #10, the responses showed #7 had the same category to the #4. In other words, in the level set oddball, the standard #7 was labeled as Rising Tone. Both hit rate and the P300 value showed the similar pattern.

Such phenomenon of change of boundary depending on which stimulus was used for standard is not unusual, as stated by Rivera-Gaxiola et al.:

> *"However, as discussed earlier, in this study the ERP responses to the stimuli varied depending on both the stimulus used as a standard as well as the characteristics of the stimuli within the acoustic continuum and its linguistic relevance (closeness to the boundary, apparent flexibility on being 'dragged' into a different category and their representativeness...). This is again in agreement with both ERP and behavioural data from other groups, e.g. with those who have reported that boundaries are*

flexible and depend upon the stimulus with which one adapts the subjects as well as those who have worked with prototypicality of stimuli and who argue that depending on the goodness of a certain stimulus, the rest of the stimuli will be attracted to one or other category. Asymmetric responses (i.e. behavioural responses are different depending on been found by many groups, although they are rarely reported in the literature." P21 from (Rivera-Gaxiola, Csibra, Johnson, & Karmiloff-Smith, 2000)

In the oddball paradigm, the building up of a phonemic memory depends on the nature of the standard: if the standard is a prototype of a phoneme, the phonemic memory is built up stable and updating of an allophone deviant is easy; while, if the standard is an allophone, the phonemic memory is not stable and updating of a prototype deviant is difficult (Ikeda, Hayashi, Hashimoto, Otomo, & Kanno, 2001). When the updating procedure is easier, the amplitude of ERP is higher. In the present experiment, compared with stimulus #7, stimulus#4 was more like a prototypical rising tone. Therefore, the phonemic memory for stimulus #7 was not stable. Due to the selective adaptation effect in speech perception, the category boundary in the level set was changed. In a selective adaption paradigm (Eimas, et al., 1973), every stimulus from the testing continuum was followed by a repeated speech stimulus. After that, the locus of category boundary of the continuum was shifted to the repeated stimulus. This selective adaption experiment was reported to be evidence of the presence of auditory feature detectors in speech processing (Roberts & Summerfield, 1981).

In the present experiment, four stimuli #1, #4, #7, and #10 were used. Stimulus #1 and #10 have distinct features to be identified as rising and level tone respectively without confusion. The most ambiguous stimulus was #7, which had the least consistent category (See Figure 4.2.15). Linguistically, in a two alternative-force choice paradigm, #7 was labeled as level tones, even though it has the rising feature in the pitch contour. Therefore, #7 has the double association of categories: linguistically level tone and acoustically rising tone. This double association was confirmed by the results from Experiment III in this chapter and previous study (Wang, 1976). In both studies, stimulus #7 had a different category than the perfect level tone (stimulus #11) by non-tone language subjects, while it had a same category as the perfect level tone by native tone language subjects. These studies confirmed that the distinct acoustic feature of stimulus #7 is rising. However, with the experience of lexical tone, the linguistic category overrides the psychoacoustic category. When the stimulus #7 was repeatedly presented in the oddball paradigm as a standard, its distinct acoustic feature—rising contour—was picked up. Therefore, the stimulus #10, which has a much more level pitch contour, was categorized as across-category deviant to the standard #7.

The phenomenon that stimulus #10 was categorized as AC deviant in the level set reflected the presence of a psychoacoustic boundary. Another result obtained from the present experiment favored this hypothesis. In the rising set, Mandarin subjects had a higher hit rate for WC deviant (#1) in nonspeech condition than in speech condition. Since this category was consistent with linguistic category, the higher discrimination sensitivity for WC deviant in nonspeech condition reflected an auditory processing. However, in the level set, Mandarin subject had a higher hit rate for WC deviant (#4) in the speech condition than in the nonspeech condition. Since the category was consistent with psychoacoustic boundary and violated the linguistic boundary, the higher discrimination sensitivity on WC in speech condition indicated the effect from linguistic training. In the linguistic training, Stimulus #4 was an AC deviant and should have higher discrimination sensitivity. Therefore, in the speech condition, even though due to the experimental paradigm, which favored the psychoacoustic boundary, the linguistic influence still existed and affected the performance. The results indicated that both auditory and linguistic experiences affect the speech perception, although they have different weights, which is depended on the nature of a task.

Some longitudinal studies from infant to adults showed that the perception reorganization results in an increase of the discriminability for native phonemic contrasts but a decrease of the discriminability for the non native contrasts (Kuhl, 2004; Werker & Tees, 2002). This reorganization of perceptual space is very important for humans to communicate, because it causes within-category differences to become less discriminated, thereby preventing non relevant information from reaching the mental lexicon. However, such reorganization does not mean that a complete or permanent loss of the discriminability for non-native phonemic contrasts. Several lines of evidence showed that the discriminability of within-category contrast is reserved and can be exhibited under certain conditions. First line of such evidence was from the behavioral studies by manipulation of the auditory memory (Reviewed in Section 2.1.3). For example, more interference to the auditory memory will increase the across-category discrimination but reduce the within-category discrimination (Fujisaki & Kawashima, 1969; Lane, 1965; Pisoni, 1973). Therefore, a less interference to the auditory memory will leave the auditory memory intact, and thus better within-category discrimination will be observed. Second line of such evidence was from the neurophysiologic responses to native and non-native phonemic contrast. Although non-native phonemic contrast elicited either a smaller or a different brain response than the native phonemic contrast did, the presence of the brain responses to the non-native phonemic contrast indicated that such ability was just decreased but not diminished (Kasai, et al., 2002; Sittiprapaporn, Tervaniemi, Chindaduangratn, & Kotchabhakdi, 2005; Tampas, et al., 2005).

Therefore, we concluded that there are two types of category boundaries were observed in the experiment: one was the linguistic boundary which was located in the middle of the continuum and was observed from both language listeners; and the other was the psychoacoustic boundary which was located at the level end of the continuum and was especially observed from Mandarin listeners. It is hard to label the class (linguistic or psychoacoustic) for the category boundary at the level end of the continuum from Cantonese listeners. In the Cantonese tone system, there is a contrast between slightly rising (tone 5: low rising) and level (tone 3: mid level), which corresponds to the category boundary at the level end. Through the comparison of the performance between speech and nonspeech, which was described in the section 4.2.4.2, we inclined to call it a linguistic boundary. Nonetheless, we could still conclude from the results that both linguistic and psychoacoustic boundary is influenced by linguistic experience and general auditory processing.

4.2.4.2. Difference between speech and nonspeech

The difference between speech and nonspeech was observed in three types of data: behavioral identification data, P300 data, behavioral hit rate data. However, the pattern from such difference in three types of data was different. (1) The results from behavioral identification task showed that category boundary was sharper in speech condition than that in nonspeech condition, which indicated a greater CP in speech condition than in nonspeech condition (See Figure 4.2.15). (2) There was no difference between Cantonese and Mandarin listeners. In contradict to the identification results, the brain responses from the oddball paradigm showed that different language groups had different abilities to distinguish speech from nonspeech. Greater CP reflected by P300 in speech condition than that in nonspeech condition was only observed in Cantonese listeners but not in Mandarin listeners, especially in the rising sets (See Figure 4.2.6a). The P300 from the level set did not change the general pattern but the difference between speech and nonspeech did not reach a statistically significant level from both language groups. (3)Although the behavioral responses were recorded simultaneously during the recording, the hit rate data did not show a same pattern as P300 did. (a) In the rising set, hit rate data from Cantonese and Mandarin listeners showed a greater CP in speech than that in nonspeech conditions (See Figure 4.2.6b). (b) In the level set, hit rate data from Cantonese and Mandarin listeners had a reverse pattern regarding the difference of CP between speech and nonspeech. Mandarin listeners had a higher CP in nonspeech than that in speech condition; while, Cantonese listeners had a higher CP in speech than that in nonspeech condition (See Figure 4.2.12b).

The behavioral data consistently showed that the linguistic CP, which presented in rising set by listeners from both languages and in level set from Cantonese listeners, was significantly greater in speech condition than that in nonspeech condition. The results suggested that the linguistic CP was selectively exhibited in speech context. A similar but weaker CP of the same position in the continuum was also observed in nonspeech condition. We argued that it was a generalization from the experience in speech condition, which was benefited from sharing the same critical feature (the pitch contours) as speech. On the other hand, since the psychoacoustic boundary is irrelevant to linguistic training, CP for psychoacoustic boundary should be greater in nonspeech condition than that in speech condition. Indeed, such pattern was also observed, which is exhibited by CP on the boundary at the level end of the continuum from Mandarin listeners. If a contrast is absent in a language environment, through the procedure of perceptual reorganization, discriminability to this contrast will be gradually reduced. Many experiments have shown that such declination is more prominent in speech context. For example, Xu et al. (2006) investigated CP on the continuum spanning from level to rising by English listeners, whose speech system does not contain this phonemic contrast, in both speech and nonspeech conditions. The results showed that CP from English listeners was more prominent in nonspeech condition than that in speech condition. More interesting, 6 months old English infants are able to discriminate this contrast, while when the same infants grow up into 9 months old, the discriminability to the same contrast was declined but only under the speech condition (Mattock & Burnham, 2006).

The behavioral data was consistent and met our expectation by showing that linguistic experiences selectively affected the speech sounds more than the nonspeech sounds. The P300 data from Cantonese listeners was consistent with the behavioral data. However, the P300 data from Mandarin listeners showed no difference between speech and nonspeech. The difference between two groups of listeners would be discussed this in the section 4.2.4.3 and the difference between P300 and behavioral data would be discussed in section 4.2.4.4. Nonetheless, the data from the present study suggested that the linguistic environment enhances the discriminability for the native phonemic contrast but reduces the discriminability for non native phonemic contrast.

4.2.4.3. Influence of language experiences on CP

The influence of language experience on CP was exhibited in both P300 data and behavioral hit rate data. (1) In the rising set, the P300 data showed that CP was more prominent in speech condition than that in nonspeech condition, and this discrepancy was only observed from Cantonese listeners but not from Mandarin listeners (See Figure 4.2.6a). Moreover, the difference between Mandarin and Cantonese was mainly contributed by the higher P300 to the AC deviant in speech condition. (2) In the level set, Cantonese and Mandarin listeners showed a different pattern between speech and nonspeech conditions on CP. Mandarin group had a relatively greater CP in nonspeech than that in speech condition; while, Cantonese group had a relatively greater CP in speech condition than that in nonspeech condition. Both hit rate data and P300 data had the same trend but only hit rate data reached the statistically significant level.

The selectively higher P300 to AC deviant in speech condition by Cantonese listeners than by Mandarin listeners in the rising set may be due to different tone systems. Although Mandarin and Cantonese are both tone languages, they have different tone inventories. Mandarin has four tones while Cantonese has six tones. The tone system of Cantonese is more compact than that of Mandarin (Peng, 2006). Therefore, such system difference may contribute to the neurophysiological difference we have observed. The tone inventory affects the tone perception was also reported in Gandour's work, which investigated tone perceptions from different language listeners (Gandour, 1983). In his study, different tones were organized into "height" and "direction" dimensions. Cantonese listeners can be separated from Mandarin listeners through their different patterns to put relative weights on these two dimensions. Moreover, in this study, the confusion matrix also suggested that Cantonese listeners made more errors between level tone and falling tone, while Mandarin listeners made more errors between rising tone and level tones. Gandour (1983) suggested that it was due to different tone sandhi rules in these two tone systems. Therefore, the data in Gandour (1983) suggested that Cantonese listeners were more sensitive to the difference between the level and the rising tones than Mandarin listeners did, which was consistent with our data. P300 is sensitive to the difference of phonemic inventories between two dialects in a same language has been reported (Conrey, Potts, & Niedzielski, 2005). However, the contrast they studied in the experiment was present in one dialect but merged in the other. In another cross-language study, which used passive listening paradigm, also unexpectedly reported a P300 like component, which they called late positive deflection since it was elicited without attention, also sensitive to the difference between native and non-native contrasts (Rivera-Gaxiola, et al., 2000). In summary, the difference in tone inventories and tone sandhi rules resulted in the observation in the present study: Cantonese had a higher CP in AC deviant in speech condition than Mandarin did. It's the first time to show that different phonemic inventories affect the perception of the same contrast. We would discuss the difference between ERP and behavioral data in the section 4.2.4.

The hypothesis of difference on tone systems could also explain the second aspect of the language difference: Mandarin and Cantonese had a reverse pattern on CP difference between speech and nonspeech. Cantonese tone inventory includes a linguistic boundary between the level (tone 3: mid level tone) vs. the slightly rising (tone 5: low rising). Moreover, many tones in the Cantonese system share the same dynamic pitch pattern but only differ in pitch height. For example, in Cantonese, there are three level tones, as well as two rising tones, which differed in the pitch height but not in the dynamic pitch pattern. Therefore, Cantonese listeners are more relied on adjacent context to make the decision (See experiments in Chapter 3). In our paradigm on the level set, the boundary was located between the level vs. the slightly rising, which corresponded to the Cantonese linguistic boundary regarding on the pitch pattern. Although the pitch height was a bit higher than natural speech, since Cantonese listeners are very sensitive to context information, the absolute pitch height seems less important for them in the present paradigm. Therefore, CP observed in level set was consistent with Cantonese's linguistic boundary. Linguistic boundary is formed or enhanced in language environment and thus CP is greater in speech condition than that in nonspeech condition. However, Mandarin tone system does not have the contrast between slightly rising and level. Consequently, the boundary in the level set is a psychoacoustic boundary for Mandarin subjects. Ideally, a psychoacoustic boundary is irrelevant to language experience, therefore the boundary in nonspeech condition will be kept but in speech condition will be lost if this contrast is not present in the listeners' language environment. Indeed, CP of behavioral data reached a statistically significant level only in nonspeech condition by Mandarin listeners. Although the ERP data showed a similar pattern, CP obtained from ERP data did not reach a statistically significant level in either speech or nonspeech condition by Mandarin listeners. We would discuss the difference between ERP and behavioral data in section 4.2.4.4.

However, the difference of the tone systems cannot explain all the observed results. On one hand, the complexity of the Cantonese tone inventory and the tone sandhi rule leads to a greater discriminability to the tone contrast between the level and the rising in speech condition by Cantonese listeners than by Mandarin listeners. On the other hand, it cannot explain why Mandarin listeners showed a similar CP in both speech and nonspeech (ERP data), while Cantonese listeners showed a greater CP in speech than that in nonspeech condition. Therefore, other mechanisms may be responsible for the results we observed.

4.2.4.4. Brain activities and overt behavioral responses

CP was present in both P300 and behavioral hit rate data indicated that the across-category deviant was easier to be detected from the standard than the within-category one. In a general trend, brain activities were consistent with behavioral responses. However, in the rising set, the language difference was only observed in P300 data but not in hit rate data; while in the level set, the language difference was more prominent in hit rate data than that in ERP data. The behavioral identification data was consistent with hit rate data in the rising set. At a first glance, the difference between ERP and behavioral data seemed to suggest that ERP had different sensitivities to a language difference in the two sets: more sensitive in the rising set and less sensitive in the level set. However, when we examined the data in details and considered the mechanisms underlying the language effects in different sets (discussed in section 4.2.3), we found that one mechanism can explain the difference between P300 and behavioral data in both sets. The difference between P300 and behavioral data in both sets suggested the same thing: Compared with behavioral data, P300 was more sensitive to long-term linguistic training.

Previous studies have suggested that P300 responses index the completion of stimulus evaluation and the size of P300 covaries with the certainty when the same amount of attention is given (Picton, 1992). The increased memory load, which is associated with a hard task, will deduce the amplitude of P300. In other words, a higher amplitude P300 is associated with more confidence of the stimulus (Polich, 2003a, 2007). Moreover, the P300 is elicited by successful updating of memory. Therefore, a conceptually more salient deviant from an established phonemic memory will elicit a larger amplitude P300 in the present experiment.

In the rising set, the language effect was exhibited by the fact that different language listeners had different degrees of CP regardless of the speech context. Both P300 and behavioral data exhibited such effect. More importantly, P300 data also showed that Cantonese listeners had a greater CP in speech condition than that in nonspeech condition, but Mandarin listeners did not have such difference between two speech contexts. Moreover, such difference between two language listeners was primarily due to a higher P300 to AC deviant in speech condition by Cantonese listeners. Such language effect relevant to speech context did not exhibit in behavioral hit rate data, which was simultaneously recorded with EEG recording. As discussed in section 4.2.3, the language effect relevant to speech context was due to the difference of two tone systems, which includes tone inventories and tone sandhi rules. Since P300 is associated with more confidence of the stimulus, the selective higher P300 to AC deviant in speech condition by Cantonese suggested that Cantonese listeners were more confident with this stimulus. In other words, the results indicated that Cantonese listeners are particularly good at building up phonemic memory for speech tones, and so they are easier to update the memory and increase the arousal when they hear the deviant from the other phonemic categories. Moreover, the difference between speech and nonspeech from Cantonese data suggested that the phonemic memory for lexical tones in Cantonese listeners relies on the segmental information or familiarity of the sounds. On the other hands, the P300 for Mandarin listeners was similar in speech and nonspeech conditions, which suggested that the building up of phonemic memory for lexical tones by Mandarin listeners rely on pitch contours. Therefore, speech and nonspeech are processed similarly in the phonological level. The hit rate data of the rising set did not show the similar interaction effect between language and syllable, because the cognitive stage reflected by the overt responses is different from that by P300. During the overt responses, the subjects can rely on previous presented stimuli to do the judgment. Therefore, Mandarin listeners had a higher within-category hit rate than Cantonese did, especially in the nonspeech condition (See Figure 4.2.5b).

In the level set, CP category from Mandarin listeners was a psychoacoustic boundary as discussed in section 4.2.4.1 and 4.2.4.2. Therefore, the degree of CP was higher in nonspeech condition than that in speech condition, which was observed in the hit rate data. In P300 data, since the phonemic memory was not well established, therefore, the P300 to both types of deviants in both speech context conditions was very low from the Mandarin listeners. However, CP category from Cantonese listeners met with their linguistic boundary; therefore, similar results were obtained as those observed in the rising set.

In summary, the difference between P300 and hit rate data further confirmed that ERP in the study of phonological study provides more information about the speech processing in different stages. The P300 data suggested that the establishing of phonemic memory is affected by linguistic environment outside the single phonemic training. Some other factors including the size of phonemic inventory, tone sandhi rules contribute to the phonemic awareness.

4.2.4.5. Hemispheric lateralization

A number of studies have concluded that P300 can be used as a marker for hemispheric asymmetry (Alexander, et al., 1996; Alexander, et al., 1995; Polich & Hoffman, 1998). Asymmetry of P300 amplitude favors the RH for simple visual or auditory discrimination (Alexander, et al., 1996; Alexander, et al., 1995), but favors the LH when verbal stimuli were used (Bentin & Feinsod, 1983; Goodin, Waltz, & Aminoff, 1985; Itzchak, Babkoff, & Faust, 2007). Asymmetric P300s between left and right hemispheres were also observed in the present experiment. In the rising set, effects of language experiences (Mandarin vs. Cantonese) and carrier syllables (speech vs. nonspeech) were more expressed in the left hemisphere than in the right hemisphere, but there was no evidence for asymmetric hemispheric lateralization pattern for the effect of category (AC vs. WC). In the level set, effect of category in speech condition was more expressed in the left hemisphere than in the right hemisphere. There was no evidence for asymmetric hemispheric lateralization pattern for the effect of category in the nonspeech condition, but the marginal mean value showed a general trend that the effect of category in nonspeech condition was more expressed in the right hemisphere.

In the early studies, since Pierre Paul Broca (1824-1880) (Broca, 1861) and Wernicke (1794-1969) (Wernicke, 1874,1969) reported two cases that patients with different brain areas damaged in the left hemisphere lost the speech functions of production and comprehension respectively but left other functions intact, it is generally accepted that the speech function is left hemispheric lateralized. However, new evidence from brain imaging studies, behavioral dichotic listening and patients with disorder language functions suggested that the early notion about language function in the brain is too rough. Depending on different processing demands and different acoustic properties, the lateralization patterns of the speech sounds are different. For example, Friederici suggested that syntactic, semantic information is primarily processed in the left hemisphere, whereas sentence level prosody is processed in the right hemisphere (Friederici, 2009; Friederici & Alter, 2004). On the other hand, some researchers suggested that the acoustic properties of a stimulus dominate the pattern of hemispheric lateralization. They argued that auditory cortices in the two hemispheres are relatively specialized, such that temporal resolution is better in left auditory cortical areas and spectral resolution is better in right auditory cortical areas (Tervaniemi & Hugdahl, 2003; Zatorre, Belin, & Penhune, 2002). Since speech is highly dependent on rapidly changing broadband sounds, it is left lateralized. However, they also admit that some top-down information such as the familiarity of speech sound also modifies the degree in which the left vs. right auditory areas contribute to sound encoding (Thompson, Schellenberg, & Husain, 2003).

In the present experiment, the stimuli differed in type of category (AC or WC) and type of carrier syllable (speech or nonspeech), but did not differ in language function. They even did not differ too much on the acoustic properties, since both types of carrier syllables had similar complexity of spectrum information, intensity profiles and pitch contours. Therefore, only the difference of familiarity between speech and nonspeech can explain why linguistic tones were more left lateralized in the present study. That the processing of familiar sounds was more left lateralized was also reported from a passive oddball study. Evoke of MMNm (magnetic equivalent to the MMN-wave of the electroencephalogram) to very frequent word deviant significantly earlier in the left hemisphere in comparison with the right side. In conclusion, processing speed may be an important aspect of the hemispheric specialization of language (Ackermann, Lutzenberger, & Hertrich, 1999).

The same reason can also be used to explain why Cantonese listeners activate left hemisphere more than Mandarin listeners did. From the statistical results, we observed that the difference of hemispheric lateralization pattern regarding on the language experience was majorly contributed by the performance in speech condition by Cantonese listeners. In other words, when Cantonese listeners heard speech sounds, the left hemisphere was more active than the right hemisphere was; whereas, there was no obvious asymmetric hemisphere activation to other conditions (Mandarin+speech, Mandarin+nonspeech, Cantonese+nonspeech). Recall the results showing the language difference, Cantonese listeners particular had a significant higher P300 to AC deviant in speech condition. From the previous studies, higher amplitude of P300 is associated with a more familiarity of the stimulus. Therefore, we can conclude that the familiarity of speech sounds drove Cantonese listeners activated the left hemisphere more in the speech condition.

Hemispheric lateralization pattern on AC and WC stimuli has been well studied in the domain of color perception and visual perception recently (Drivonikou, et al., 2007; Franklin, Drivonikou, Clifford, et al., 2008; Gilbert, 2007; Regier & Kay, 2009; Siok, et al., 2009). In these studies, subjects were asked to point out whether an odd color patch, which was presented against a background with the other color, was at the right or the left side as quickly as possible. The odd color patch either had the same color name or different color names with the background color. The results showed that when the odd color patch was presented to the right visual field (see through from the left hemisphere), the patches with a different color name than the background color were identified faster than those with the same color name, even though, the odd color always had the same chromatic separations to the background color. When the odd color patch was presented to the left visual field (see through from the right hemisphere), there was not a clear pattern. Although there were many versions of the paradigm, they followed the same logic. Therefore, the data implied that categorical perception of color is more expressed in the right visual field (the left hemisphere). More surprisingly, some follow-up studies compared the performance from different ages of subjects including infants, children who had acquired the color terms and who did not successfully acquire the color terms, and adults (Franklin, Drivonikou, Bevis, et al., 2008; Franklin, Drivonikou, Clifford, et al., 2008). The results showed that CP of color was more expressed in the left hemisphere in the population who had acquired the color terms (adults, children who had acquired the color terms), while CP of color was more expressed in the right hemisphere in the population who did not successfully acquire the color terms (infants, children who did not acquire the color terms). These results indicated that linguistic CP of color is more expressed in the left hemisphere. Moreover, the left lateralization pattern is probably due to the facilitation of language labeling (Siok, et al., 2009). The hemispheric lateralization pattern is used to support the view that the linguistic labeling contribute to the formation of CP in color.

In the domain of tone perception, there were few reports about the interaction of CP and hemispheric lateralization according to our knowledge, although two conference abstracts from the same group reported that the AC deviant of lexical tone was more left hemispheric lateralized (Xi, Zhang, Xu, & Shu, 2009; G. Xu, Xi, Zhang, & Shu, 2009). Moreover, there is no conclusive conclusion on the lateralization pattern of tones. On one hand, lexical tones are slow changing signals, so it should be right hemispheric lateralized (L. Liu, et al., 2006; Zatorre, 1988; Zatorre, et al., 1992). On the other hand, lexical tones carry linguistic meanings and familiar by native listeners, so it should be left hemispheric lateralized (Gandour, et al., 2003; Gandour, et al., 1998; Tervaniemi & Hugdahl, 2003). Recently, there is a new opinion about the lateralization pattern of lexical tones with the findings from the ERP study, which tries to reconcile the two aspects together: lexical tones have different lateralization patterns depending on the task and the cognitive stages and both the acoustic and linguistic aspects contribute to the lateralization pattern of tones (Shuai, 2009; Zatorre & Gandour, 2008).

Therefore, due to the complex and dynamic lateralization pattern of lexical tones, it is hard to use lateralization patterns to support or against the linguistic function of CP in lexical tones. However, since lexical tones are defined within the linguistic domain, we don't need to use that to "confirm" or "support" the linguistic aspect of CP in lexical tones. Rather, we can use the lateralization pattern from CP to provide more information about the lateralization pattern of lexical tones. In the experiment of CP, discrimination of tones across a category is facilitated by the linguistic labeling, in which case the linguistic function is more prominent and the left hemispheric lateralization is predicted. On the other hand, discrimination of tones from the same category relies on acoustic features, in which case acoustic features are dominant and the lateralization pattern is depended on the relative features of the two stimuli. A relative fast changing sound is more left lateralized than slow changing one (Robertson & Ivry, 2000).

In the present experiment, we found that CP of speech tones was more expressed in the left hemisphere in the level set. In the same level set, the lateralization pattern of CP was more right hemispheric lateralized, although it did not reach a significant level. The results from the level set were quite consistent with those from the color CP's studies, which showed that the left hemisphere is more involved in the language processing than the right one.

Although in the rising set, there is no evidence to support the interaction effect between category type and lateralization pattern, the marginal mean value showed a similar trend. In other words, in the rising set, brain responses to CP for both speech tones and nonspeech pitches were more bilaterally activated, but did not change the general trend observed from the level set. The lack of significant left hemispheric lateralization pattern for CP in speech condition of this set might be contributed by a combination effect from linguistic aspect and acoustic properties. If only a linguistic function was considered, CP was expected to be more expressed in the left hemisphere. In other words, the discrepancy between AC and WC was larger in the left hemisphere than in the right hemisphere, as shown in the level set (see Figure 4.2.14). However, we also observed that the WC deviant in the rising set had a steeper slope and changed faster than AC deviant did, which should be left hemispheric lateralized and thus reduced the discrepancy on the left hemisphere. Therefore, both linguistic function and acoustic properties contributed to the observed results.

4.2.5. Summary

In this experiment, we conducted an oddball attentive P300 experiment to investigate CP on the level vs. rising tone continuum with the subject's attention. Two types of carrier syllables (speech and nonspeech) were used. Subjects from two different dialect groups (Mandarin and Cantonese) participated in the study. The brain activities and the overt subject responses were recorded simultaneously in the experiment and a *post hoc* behavioral labeling task was also carried out to check the category boundary in the continuum.

The results showed that CP was presented in both P300 and overt behavioral responses. However, there were two types of CPs (discussed in 4.2.4.1): one was consistent with linguistic category (the rising set) and the other was consistent with psychoacoustic category (the level set in Mandarin data). We also observed that experiences from different tone systems (including the tone inventories and tone sandhi rules) influenced the performance on lexical tones, even though both systems had the testing tones (discussed in 4.2.4.2). Cantonese subjects had greater CP than Mandarin subjects did. More importantly, Cantonese subjects responded to speech and nonspeech differently, while there was no obvious difference was observed from Mandarin subjects. At last, the data showed that hemisphere laterality was influenced by language backgrounds, types of carrier syllables, and types of deviants (AC vs. WC). All influences were more expressed by the left hemisphere (discussed in 4.2.4.3), and the hemispheric lateralization pattern was affected by both linguistic function and acoustic properties of the stimuli.

4.3. Experiment III: Behavioral Test

4.3.1. Method

4.3.1.1. Subjects

Fifty-nine right-handed subjects, with normal hearing and no reported history of neurological illness, participated in the experiment with payment. Nineteen of them were native Cantonese subjects [9 female and 10 male, mean age = 21.6 ± 1.6; Laterality Quotient = 71.3 ± 22.9 (Oldfield, 1971)]; 20 of them were native Mandarin subjects [10 female and 10 male; age = 23.7 ± 2.3; Laterality Quotient = 79.6 ± 22.3]; and 20 of them were native German subjects [10 female and 10 male; age = 28.5 ± 4.6; Laterality Quotient = 87.2 ± 11.5]. Informed written consent was obtained from each subject. Ethics approval was obtained from Survey and Behavioral Research Ethics Committee of The Chinese University of Hong Kong.

4.3.1.2. Stimuli

The stimuli were the exactly the same as used in Experiment I, which included 11 real speech stimuli and 11 pure tone stimuli with the same pitch contours. The stimuli were shown in Figure 4.1.1.

4.3.1.3. Identification task

The 11 stimuli in speech and nonspeech conditions were repeated twice and were randomly presented to the subjects in a block with 4s ISI. There were 5 such blocks for each condition. Therefore, each stimulus had 10 repetitions. Presentations of speech and nonspeech were counterbalanced among the subjects. The subjects were asked to identify the category of the heard sound by pressing key "1" for category "1" and pressing key "2" for category "2". Demo of the sound categories for two end points (#1 and #11) was shown to the subjects before the task. To make the procedure comparable in speech and nonspeech conditions, the categories were demoed as "sound 1" and "sound 2" without linguistic labeling in speech set. At the end of each block, the score for correct responses to two endpoints was shown on the screen to attract the subject's attention. If there was any block with a score lower than 75%, a supplement block would be presented.

4.3.1.4. Discrimination task

Discrimination task was only conducted on the speech continuum by the same three groups of subjects. Subjects discriminated between a pair of stimuli in a behavioral same/different test. The stimulus pairs included both two-step and zero-step pairs (e.g. 6-6, 8-8, 6-8 and 8-6). A block contained 29 such pairs and the stimuli in a block were presented to the subjects in a random order. There were 7 repetitions for each pair. The ISI was 500 ms between two stimuli in a pair (trial), and ITI was 2500 ms between adjacent trials. One extra practice block (data was excluded from the analysis) was conducted to familiarize subjects with the procedure.

4.3.2. Data Analysis

4.3.2.1. Identification data

As introduced in the chapter 2, there are several methods to analyze the identification data. In this experiment, predicted discrimination and the probit approximation were used.

A. PREDICTED DISCRIMINATION.

Follow the equation (2.1), which was rewritten here, to calculate the predicted discrimination.

$$P_{pred.disc_{(i,j)}} = (1 + (P_{A_i} - P_{A_j})^2)/2 = (1 + (P_{B_i} - P_{B_j})^2)/2 \qquad (2.1)$$

According to the crossover points in identification curves, the predicted discrimination values were grouped into across- and within- category (AC and WC) results. Then the comparison was conducted on these across- and within- category pairs.

B. PROBIT APPROXIMATION.

Follow the equation (2.10) –(2.13), which were rewritten here.

$$P_{probit} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{y-5} e^{\frac{-u^2}{2}} du$$

(2.10)

$$y = \frac{1}{\sigma} x - \frac{\mu}{\sigma} + 5 = b_0 + b_1 x$$

(2.11)

Position of CP boundary: $x_{cb} = \frac{5 - b_0}{b_1}$

(2.12)

Boundary width or sharpness: $w = x_{p=0.75} - x_{p=0.25} = \frac{1.35}{b_1}$ (2.13)

According to the crossover points in the identification curves (shown in Figure 4.3.1), the predicted discrimination values were grouped into across- and within- category results. Then the comparison was on these across- and within- category pairs. Two two-way mix designs were carried out on the position and width of the boundary with one within-subject factor (2 SYLLABLE: speech and nonspeech) and one between-subject factor (3 LANGUAGE: Mandarin, Cantonese and Germen). The statistical analyses were to investigate whether the language experience and the syllable types affected the performance.

4.3.2.2. Discrimination data

The score of discriminations for each pair was calculated following descriptions in (Francis, et al., 2003). Obtained discrimination scores were calculated by equation (2.14), which was described in chapter 2. The equation (2.14) was repeated here.

$$D_{obt} = P('S'/S) \bullet P(S) + P('D'/D) \bullet P(D) \quad (2.14)$$

4.3.3. Results

4.3.3.1. Identification curves.

A. IDENTIFICATION CURVE AND PROBIT APPROXIMATION

Identification responses averaged across listeners according to conditions and language groups were presented in Figure 4.3.1. The figures showed that all the listeners exhibited crossovers in their identification function. Two categories were obtained along each continuum. The probit approximation fitted the curve quite nicely with small variance. The boundary position and the boundary width, as well as the approximation variance, were shown in Table 4.3.1.

Boundary position. There was no significant interaction or main effect on the boundary position. All the boundary positions were near stimulus #5. Therefore, the position of boundary was not analyzed further.

Boundary width. However, there was a significant two-way interaction effect between SYLLABLE and LANGUAGE [F(2,56)=3.634, p=0.033] and a main LANGUGE effect [F(1,56)=8.042, p=0.007] on boundary width. From the average curve shown in Figure 4.3.2, we could observe that Mandarin and Cantonese subjects had a larger boundary width in nonspeech condition than that in speech condition, while German speaker had the reverse pattern. This observation was confirmed by three pairwise t-tests on the boundary width difference between speech and nonspeech between every two languages. The results showed that there was a significant language effect on the difference of boundary width between Mandarin and German subjects ($p<0.05$), but no difference between Mandarin and Cantonese (p=1) or between Cantonese and German (p=0.186) (all were after Bonferroni correction).

However, the difference in speech and nonspeech were not significant by three *post hoc* paired t-tests between speech and nonspeech after Bonferroni correction. Another *post hoc* analysis was carried out on boundary width to investigate the effect of language groups on each syllable type. The results showed that language experience made a difference on speech condition [F(2,56)=10.401, $p<0.001$], but not in nonspeech condition [F(2,56)=0.483, p=0.619]. In the speech condition, pairwise t-test (with Bonferroni correction) showed that Mandarin and Cantonese groups had significantly narrower boundary than German groups did [Mandarin vs. German: $p<0.001$; Cantonese vs. German: $p=0.004$], but Mandarin and Cantonese did not significantly differ from each other.

In summary, the identification curve and probit approximation showed that two tone language groups (Mandarin and Cantonese) had a shaper boundary in speech condition than non tone language group (German) did. On the other hand, there was no significant language effect on nonspeech condition. Moreover, tone language subjects had the similar trend that boundary was sharper in speech condition than that in nonspeech condition, while German subjects had the reverse pattern. The results indicated that the language experience was more easily exhibited by the speech sounds even though both sounds shared the same critical phonetic feature of pitch.

Table 4.3.1 Boundary position and boundary width on speech and nonspeech continua from three language groups calculated by probit approximation.

	Mandarin		Cantonese		German	
	speech	nonspeech	speech	nonspeech	speech	nonspeech
position	5.47	5.30	5.14	5.00	5.59	5.38
width	1.14	1.83	1.43	1.85	2.51	2.15
approximation variance	0.006	0.006	0.006	0.011	0.009	0.010

Table 4.3.2 Grouping information, grouped mean value, and statistical analysis of predicted discrimination scores for continuum spanning from level to rising tones in six conditions (2 SYLLABLE ×3 LANGUAGE). * : $p<0.004$; (significant level $p<0.004$ after correction)

Categories (pairs)		Mandarin Speech	Mandarin Nonspeech	Cantonese Speech	Cantonese Nonspeech	German Speech	German Nonspeech
mean value	AC (4-6, 5-7)	0.71	0.64	0.65	0.62	0.56	0.62
	WC (others)	0.53	0.53	0.53	0.54	0.54	0.53
Paired t-tests							
AC vs. WC		$p<0.001$*	$p=0.001$*	$p<0.001$*	$p=0.003$*	$p=0.245$	$p<0.001$*
AC	speech vs. nonspeech	$p=0.018$		$p=0.156$		$p=0.004$*	
WC	speech vs. nonspeech	$p=0.563$		$p=0.230$		$p=0.051$	

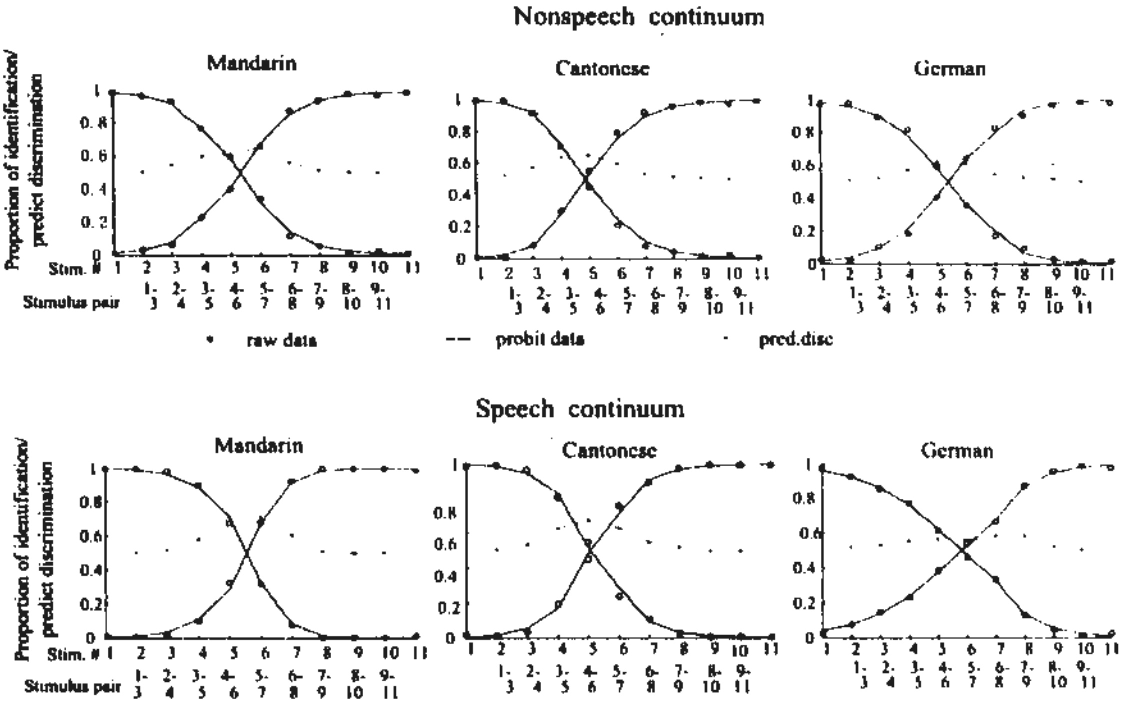


Figure 4.3.1 Results of identification of level vs. rising tone continuum (open circles) and their corresponding probit approximation (solid line), with the predicted discrimination results (dotted line) superimposed over the identification curves. Upper panel represented results from the nonspeech continuum and lower panel represented results from the speech continuum. From left to right, three panels represented the performance of Mandarin, Cantonese and German subjects respectively.

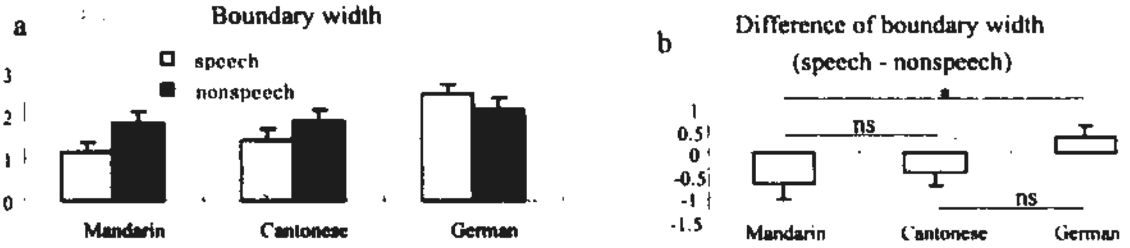


*Figure 4.3.2 a)Boundary width of continuum spanning from level to rising tones in six conditions [2 SYLLABLE (speech and nonspeech) ×3 LANGUAGE (Mandarin, Cantonese, and German)]. A smaller number of boundary width indicted a sharper category boundary. b) Difference of boundary width (speech–nonspeech) on three language groups. *: $p<0.05$, ns: not significant. Error bar represented standard error.*

B. PREDICTED DISCRIMINATION.

The predicted discrimination score was grouped into AC and WC according to the cross-over (or position of boundary). A three-way mix design was carried out with two within-subject factors [2 CATEGORY: AC and WC; 2 SYLLABLE: speech and nonspeech] and one between-subject factor [3 LANGUAGE: Mandarin, Cantonese and

German]. There was a significant three-way interaction effect [F(2,56)=7.896, p=0.001], two significant two-way interaction effect [CATEGORY × LANGUAGE : F(2,56)=6.179, p=0.004; SYLLABLE × LANGUAGE: F(2,56)=7.803, p=0.001], and one main effect [CATEGORY: F(2,56)=90.6, p<0.001]. This indicated that both types of carrier syllables (speech and nonspeech) and language experience contributed to the difference of the degree of CP.

To investigate how the language affected the degree of CP in different carrier syllables, two two-way mix designs were carried out on speech and nonspeech continua respectively. The two-way mix design used CATEGORY as the within-subject factor and LANGUAGE as the between-subject factor. In the nonspeech condition, only the CATEGORY main effect was significant [F(1,56)=44.647, p<0.001], which showed that language effect did not exhibit in the nonspeech condition. However, in the speech condition, two main effects (CATEGORY and LANGUAGE) were significant, and more importantly, the two-way interaction effect was also significant [F(2,56)=14.523, p<0.001]. This result showed that language only significantly influence CP in speech condition. Three paired t-tests between AC and WC in speech condition were carried out for three language groups respectively. The results showed that in the speech condition, two tone language groups had significant CPs, but non-tone language group did not (Shown in Table 4.3.2 and Figure 4.3.3). Two one-way ANOVA on AC and WC in speech condition showed that the group difference was presented in AC but not in WC [AC: F(2,56)=18.582, p<0.001; WC: F(2,56)=1.646, p=0.202]. *Post hoc* pairwise t-test revealed that two tone language groups significantly differed from the non-tone language group on AC pair in speech condition.

Another type of analysis was to test how types of carrier syllables affected CP in each language group. In other words, the test was to test the difference of CP between speech and nonspeech in each language group. Three two-way RMANOVA with two within-subject factors (2 CATEGORY; 2 SYLLABLE) were done on three language groups respectively. The significant two-way interaction effect was observed in both Mandarin and German groups [Mandarin: F(1,19)=5.671, p=0.028; German: F(1,19)=10.446, p=0.004]. Paired t-tests between speech and nonspeech on AC and WC in each language group showed that speech and nonspeech differed in AC [Mandarin: t(19)=2.582, p=0.018; German: t(19)=-3.266, p=0.004]. However, the direction of the difference was different in Mandarin and German groups. Mandarin group had a higher AC value in speech condition than that in nonspeech, but the German had a reverse direction by showing a higher AC value in nonspeech than that in speech condition.

The last type of *post hoc* analysis was to test how the difference of speech and nonspeech was exhibited in each category pair and how it was modulated by the language experience. Two two-way mix designs were carried out on AC and WC respectively, with the within-subject factor of SYLLABLE and the between-subject factor of LANGUAGE. The WC showed a marginal significant two-way interaction effect [F(2,56)=3.242, p=0.047], and AC showed a significant two-way interaction effect [F(2,56)=8.244, p=0.001]. The results showed that the degree of speech and nonspeech difference depended on language experience and the dependence was more exhibited in AC pairs.

In summary, the results from the predicted discrimination scores showed that both tone language groups had greater CP in speech than that in nonspeech, but non-tone language group reversed the pattern. Moreover, we also observed that different language groups had significantly different degrees of CP in speech condition. However, such difference was not significantly presented in nonspeech condition. The language effect was most exhibited in AC pairs in speech condition. The results from the predicted discrimination were consistent with that obtained from the probit approximation.

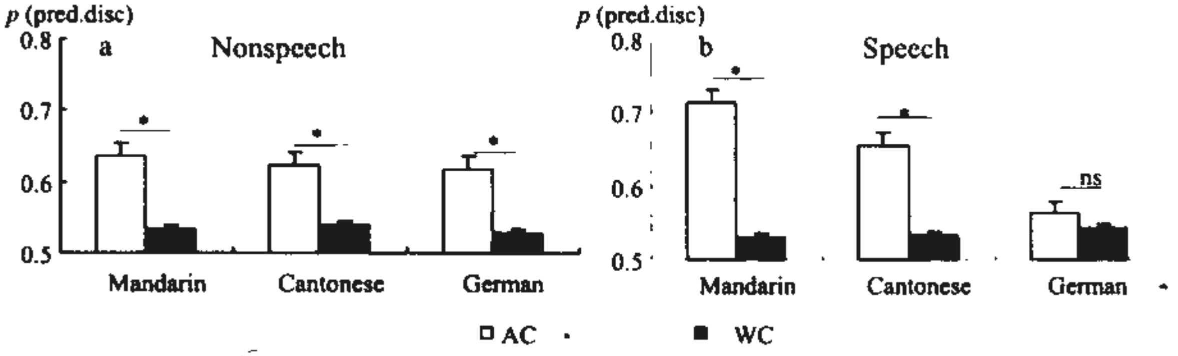


*Figure 4.3.3 Grouped value of predicted discrimination of level to rising continuum according to AC (Across-category) and WC(within-category). a) In the nonspeech continuum; b) in the speech continuum. *: p<0.004; ns. Non significan. Error bar represented standard error.*

4.3.3.2. Discrimination results

The discrimination was conducted in speech condition only. The curve of discrimination score for each language was presented in Figure 4.3.4. From the average curve, there was one peak (in the middle of the curve) from the Mandarin group; two peaks (one in the middle and the other at the level end of the curve) from the Cantonese group; and one peak (at the level end of the curve) from the German subjects. One-way ANOVA with LANGUAGE as a between-subject factor on each of the nine discrimination pairs showed that some pairs were different in different language groups [$p<0.05$]. Those pairs included WC pairs (1-3, 2-4), AC pairs (5-7, 6-8) and the level end pair (9-11).

Similar to the description in chapter 3, the discrimination data was grouped into AC and WC. Moreover, the level end pair was also singled out. Therefore, there were three categories (WC, AC and Level end) in the analysis. A two-way mix design was carried out with CATEGORY (3 levels) as within-subject factor and LANGUAGE (3 levels as the between-subject factor. There was a significant two-way interaction effect [F(2,54)=14.906, $p<0.001$].

To further investigate how language effect was presented in each category, three one-way ANOVA were carried out. The results showed that all three categories had significant language effects [WC: F(2,55)=8.579, $p=0.001$; AC: F(2,55)=5.99, $p=0.004$; level end: F(2,55)=9.676, $p<0.001$]. The *post hoc* pairwise t-tests on each category showed that language experience had different effects on different pairs. (1) German group had higher within-category discrimination than Mandarin and Cantonese groups did, while there was no significant difference between Mandarin and Cantonese groups. (2) Mandarin group had higher between-category discrimination in the middle of the continuum than German group did, while Cantonese group's performance was in between. (3) German subjects had higher level end discrimination than Mandarin group, while Cantonese group's performance was in between.

Another type of *post hoc* analysis was to investigate what was the performance of different categories in each language group. Three one-way RMANOVA with within-subject factor of CATEGORY were carried out on each language group. The results showed that all the language groups had main CATEGORY effects. However, the *post hoc* pairwise t-tests showed that different language groups had different types of category effect. In Mandarin group, the AC had the significantly highest score, while the scores of WC and the level end were smaller and did not differ from each other. In Cantonese group, the WC had the significantly lowest score, while there were no differences between AC and the level end point. At last, in the German group, the level end point had the significantly highest score, while no difference between AC and WC was obtained. The mean values and the statistical results were presented in Table 4.3.3.

In summary, the obtained discrimination score showed that:

(1) Mandarin and Cantonese groups had better discrimination on the category boundary, which was located in the middle of the continuum.

(2) The within-category discrimination in tone language groups (Mandarin and Cantonese) was lower than that in the non-tone language group (German).

(3) Cantonese group also had a peak on the level end, which was presumably due to the two rising tones and three level tones in Cantonese tone system. The level end pair may be corresponded to the category boundary between low rising tone (tone 5) and mid level tone (tone 3).

(4) Finally, the German group showed a peak on the level end pair, which may reflect the psychoacoustic boundary between the level and rising pitches as that reported previously in English group (Wang, 1976).

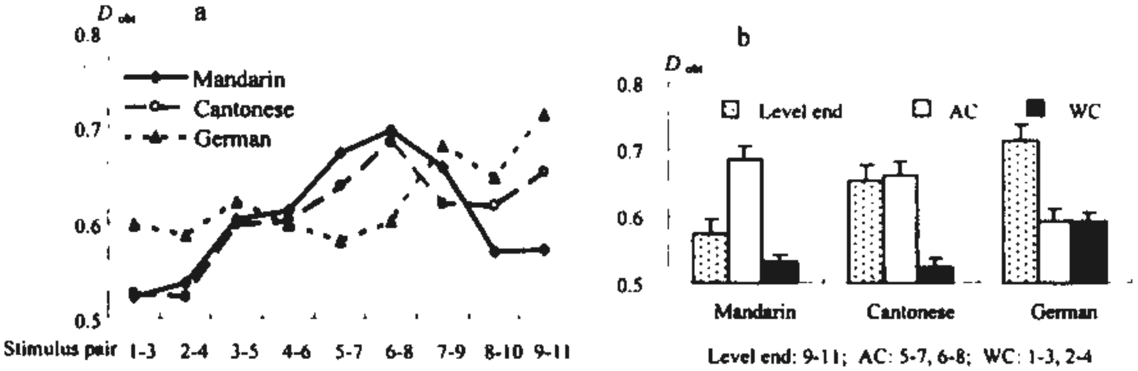


Figure 4.3.4 Obtained discrimination data of the speech continuum spanning from level to rising tones by Mandarin, Cantonese and German subjects. a) Discrimination curves along the speech continuum; b) grouped values of within-category pairs (WC), across-category pairs (AC), and level end pair.

Table 4.3.3 Grouping information, grouped mean value, and statistical analysis of obtained discrimination scores for continuum spanning from level to rising tones in three language groups. * : $p<0.008$; (significant level $p<0.008$ after correction)

Categories (pairs)		Mandarin	Cantonese	German
mean value	level end (9-11)	0.57	0.65	0.71
	AC (5-7, 6-8)	0.68	0.66	0.59
	WC (1-3, 2-4)	0.53	0.52	0.59
pairwise t-test (Bonferroni correction)				
		Mandarin	Cantonese	German
AC vs. WC		$p<0.001$*	$p<0.001$*	$p=1.0$
AC vs. level end		$p=0.002$*	$p=1.0$	$p<0.001$*
WC vs. level end		$p=0.052$	$p<0.001$*	$p<0.001$*
		Mandarin vs. Cantonese	Mandarin vs. German	Cantonese vs. German
level end		$p=0.059$	$p<0.001$*	$p=0.201$
AC		$p=1.0$	$p=0.005$*	$p=0.056$
WC		$p=1.0$	$p=0.003$*	$p=0.002$*

4.3.4. Discussion

4.3.4.1. *Effect of language experience on category boundary*

The identification results, using probit approximation and predicted discrimination, showed that two tone language groups (Mandarin and Cantonese) had a shaper boundary in speech condition than the non-tone language group (German) did. This sharper boundary in tone language resulted in the better across-category discrimination than within-category discrimination (See Figure 4.3.3 and 4.3.4). On the other hand, there was no significant language effect on nonspeech condition. The results indicated that the language experience was more easily exhibited by the speech sounds even though both sounds shared the same critical phonetic feature of pitch. The results were not surprising and consistent with previous results on the effect of language experience (Hallè, et al., 2004; Wang, 1976; Y. S. Xu, Gandour, & Francis, 2006).

4.3.4.2. *Difference between speech and nonspeech*

In addition to CP, we also observed the difference between speech and nonspeech. Tone language subjects had the similar trend that category boundary was sharper in speech condition than that in nonspeech condition, while the non-tone language subjects (German) had the reverse pattern. Similar trend has been reported in another study (Y. S. Xu, Gandour, & Francis, 2006), which compared the performance between Mandarin and English subjects on the similar level to rising continuum. According to the multistore model proposed by Xu et al., the categorical performance by tone language subjects is contributed by the long-term categorical storage. The long-term categorical storage has a template in the memory, and does not need the online decoding. Therefore, no significant difference on the stimulus complexity will be observed in the tone language subjects. On the other hand, CP observed in non-tone language subjects is due to the short-term categorical memory which relies on online decoding. The more complex sound (speech) makes the perception of pitch more difficult. Therefore, the multistore model and the difference of complexity between speech and nonspeech predict that CP in nonspeech is greater than in speech condition by non-tone subjects. This type of prediction is consistent with our obtained results. Indeed, a cross-sectional and longitudinal study (Mattock & Burnham, 2006) compared the performance of tone perception by 6-9 months infants from Mandarin and English population. The results showed that Mandarin infants performed equally well at 6 and 9 months for both speech and nonspeech tone discrimination; while, English infants showed a declination between 6 and 9 months for speech tone discrimination, whereas their nonspeech tone discrimination remained constant. These results supported that the reorganization of tone perception is a function of the native language environment, and the reorganization is linguistically based.

However, we also observed a trend that higher CP in speech condition than in nonspeech condition by tone language listeners, especially by Mandarin listeners. Moreover, although there was no significant effect observed for the boundary position, there was a general trend that position of speech was more biased to level end than that of nonspeech. Both trends cannot be explained by the Xu et al.'s "complexity theory". We used pure tone in nonspeech condition, which is even less complex than complex tone used in Xu et al's study. We hypothesized two reasons behind our observation, which were not considered by Xu et al. One was the difference of experience. Tone language listeners had more experiences of speech sound than nonspeech sound, so the sensitivity to speech was higher than to the nonspeech sounds. However, such "experience hypothesis" can not explain Xu et al.'s results. Therefore, there may be another mechanism to explain the discrepancy. This mechanism still relates to the acoustic information of stimuli. In the experiment, we use pure tone and all stimuli have level portions at the beginning which may the stimuli more nature. The richness of harmonic information facilitates the pitch discrimination has been reported neurally (Tervaniemi, et al., 2000). Also, a neuromagnetic study showed that pitch, but not frequency, was tonotopically coded in the auditory cortex (Pantev, Hoke, Lütkenh ner, & Lehnertz, 1989). Therefore, the higher frequency harmonic information of a complex tone which miss the fundamental frequency (missing F0) provides secondary and additional information to identify the pitch in the similar brain region as the pure tone with the same F0 does (Bendor & Wang, 2005). We proposed that the presence of higher frequency harmonic information enhances pitch perception and resulted in the observed trend in our experiment.

The interaction effect between the factors of language experience and syllable types indicated that CP observed in tone language subjects did not result simply from the presence of an innate sensitivity of pitch direction. Otherwise German subjects would be expected to show the same CPs in the speech condition as the tone language subjects did in the nonspeech condition. Therefore, tone language subjects' nonspeech performance must derive at least in part from their experience with listening to lexical tone pitch patterns. In other words, the experience of lexical tones was generalized into the perception of nonspeech pitches.

4.3.4.3. Linguistic boundary and psychological boundary in discrimination curve

In the obtained discrimination curves, we observed one peak in the middle of the speech continuum from Mandarin and Cantonese groups. The peak in the discrimination curve indicated a better discrimination on the category boundary. This category boundary corresponds to the linguistic boundary between the level and the rising tones. Such linguistic boundary is not present in non-tone language subjects. A better within-category discrimination but poorer across-category discrimination in non-tone language subjects than in tone language subjects suggested the reorganization of pitch perception through the linguistic training.

Longitudinal studies showed that infants can discriminate non-native speech contrasts without relevant experience, and that there is a decline in this ability during ontogeny (Mattock, Molnar, Polka, & Burnham, 2008; Werker & Tees, 2002). Such declination in foreign-language consonant perception and increment in native-language consonant perception occurs by around 11 months(Kuhl, 2004), and the reorganization of linguistic tones was even earlier at around 9 months (Mattock, et al., 2008). The progressive development of categorical perception from babyhood to adolescence has probably been influenced by spoken communication. The increase in CP causes within-category differences to become less discriminated, thereby preventing non relevant information from reaching the mental lexicon. This should facilitate word recognition, especially under difficult listening conditions. CP, which develops as the infant ages, enhances communication. In addition to the behavioral evidence, neural correlates of this dynamic language development with ERP measurement further discovered pointed out that individual developmental differences on CP might have an impact on language development (Rivera-Gaxiola, Silva-Pereyra, & Kuhl, 2005).

In addition to a discrimination peak in the middle of the continuum, which was corresponding to the linguistic boundary, Cantonese group also had a peak on the level end. The peak on the level end in Cantonese speaker was presumably due to the presence of two rising tones and three level tones in the Cantonese tone system. The level end pair might correspond to the category boundary between low rising tone (tone 5) and mid level tone (tone 3).

Different tone inventories will affect the tone perception has been reported in previous studies. Gandour (1983) investigated the perceptual dimensions of tones and the effect of linguistic experience on listeners' perception of tones with five language groups (Cantonese, Mandarin, Taiwanese, Thai, and English). It was not surprising that tone language vs. non-tone language listeners can be classified on the basis of their patterns of dimension weights. More interesting, tone language listeners from different tone systems can also be classified (for example, Thai vs. Chinese listener; Cantonese vs. Mandarin and Taiwanese listener) through this "hierarchical clustering". Such linguistic experience affecting the perception was also exhibited by the experimental results that tone listeners were easier to discriminate native tones. For example, Cantonese listeners were better than both Mandarin and English listeners at discriminating Cantonese tones, and Mandarin listeners did better than both Cantonese and English subjects at discriminating Mandarin tones(Y. S. Lee, et al., 1996).

The second peak on the level end observed from Cantonese listeners may also be a psychoacoustic boundary as that from German listeners. That peak on the level end from

German group reflects the psychoacoustic boundary between the perfect level and slightly rising pitches as that reported previously in English group (Wang, 1976) and French group (Hallè, et al., 2004). In Cantonese, the rich tone inventory helps to keep the psychoacoustic boundary without declination. Whether the second peak from Cantonese listeners' discrimination reflected a psychoacoustic boundary or a linguistic boundary remained to be explored in feature. Nonetheless, both hypotheses do not deny the influence of the richness of inventory in the Cantonese tone system.

4.3.4.4. Comparison between identification and discrimination results

CP shown in identification task and discrimination tasks were quite consistent. The location of discrimination peak had the trend to be consistent with the identification boundary. However, we also observed the boundary shift in the discrimination curve compared with the identification results. In identification results, the category boundary was near the stimulus #5 in all conditions, which should be corresponded to discrimination on 4-6 and 5-7 pairs. In contrast, in the obtained discrimination results, the discrimination peak was located on 6-8 pair, although a smaller peak was also located on 5-7 pair. Therefore, the location of discrimination peak shifted one pair towards the level part.

The peak shift was not an odd phenomenon. In an earlier experiment (Hallè, et al., 2004), similar phenomenon was also reported. In that study, a supplementary test on rating the naturalness of each stimulus was conducted. A lower rate would be given to a more ambiguous stimulus. The "shifted" discrimination peak was consistent with the change from the most ambiguous stimulus to confident stimulus.

The discrimination shift towards to level portion may be partly due to the artifact introduced from the paradigm in our experiment. We conducted the identification task first and then the discrimination task. The identification task was conducted on both speech and nonspeech continua. Subjects may be fatigue after the identification task. Therefore, the phonetic storage (information) was temporally overridden by auditory storage and thus the boundary was more biased to the psychoacoustic boundary, the level part. We hope future studies would solve this puzzle.

4.3.5. Summary

The behavioral Experiment Investigated the effect of language experience and the carrier syllables on CP along the continuum spanning from level to rising tones. Identification results were obtained from three language subjects (Mandarin, Cantonese, and German) in two types of carrier syllables (speech and nonspeech). The discrimination results were obtained from the same three language subjects in speech condition only. It's generally known that other suprasegmental features, especially intensity profiles, do highly correlate with tone perception (Abramson, 1972; Howie, 1976; S. Liu & Samuel, 2004; Whalen & Xu, 1992), but in this study, we just focused on the primary cue, fundamental frequency (the physical correlate of pitch), to lexical tone perception, and fixed other features as constant.

We observed CPs from all conditions, although the degrees of CP and the nature of CP were different regarding the language experience and the type of carrier syllable. Tone language subjects had a higher degree of CP than non-tone language subjects did, which was expressed by sharper category boundary and greater across-category vs. within-category discrimination.

The influence of language experience was not only expressed by degree of CP but also by the difference between speech and nonspeech. In tone language, the degree of CP was greater in speech condition than that in nonspeech condition; while, in the non-tone language, the pattern was reverse. Such difference between speech and nonspeech was modulated by linguistic experience inferred that linguistic experience is more specific to speech sounds although it can generalize to nonspeech perception. Such difference, also suggested that the complexity of spectral information in speech sounds makes the subject more difficult to track the pitch contours.

In addition to the difference of linguistic background between tone and non-tone language, we also observed the difference between two tone languages. Probably due to the difference of the tone learning system, Mandarin subjects have better labeling performance on speech than Cantonese subjects do. Moreover, due to different tone systems, Cantonese subjects have an additional category boundary on the level end of discrimination curves, which probably corresponds to discrimination between low rising and mid level tones.

In this experiment, we observed not only a linguistic category, but also a psychoacoustic boundary. The psychoacoustic boundary was located between perfect level vs. slightly rising tones. The psychoacoustic boundary was observed from German subjects. The presence of a psychoacoustic boundary indicated CP is not a linguistic specific phenomenon; rather, it is also existed in general acoustic processing. The presence of psychoacoustic boundary in German subject but lack in Mandarin subjects indicated that linguistic training changes the pitch perception.

4.4. Summary of chapter 4

Three experiments were carried out to investigate CP along the continuum spanning from the level to rising tones. Two aspects were studied in this chapter through the experiments: (1) the temporal dynamics of CP, and (2) influences of linguistic background on CP. Although experiments in chapter 3 have investigated the second point (influence of linguistic background on CP) very well, it did not examine whether speech sounds that are native in distinct languages are perceived differently by first language listeners of those languages. In this chapter, we specifically compared Mandarin with Cantonese on CP of the level and rising tones, present in the tonal inventories of both languages and investigated the extent to which native Mandarin and Cantonese listeners perceive these tones differently.

Pervious studies and experiments in chapter 3 have shown lexical tones were perceived categorically, and there was a perception difference between speech and nonspeech. However, all the data were from the behavioral data. Experiment I used passive oddball listening paradigm to investigate whether CP of lexical tones by native Mandarin listeners was present in the preattentive stage, as well as whether the difference of speech and nonspeech was present in the preattentive stage. The results showed that although listeners could detect the minor pitch difference between deviants and the standard in all conditions (reflected by MMN), only across-category deviant in the speech condition elicited an earlier MMN than others. Even though nonspeech pitches were not perceived categorically in the preattentive stage, they were in the late cognitive stage with full attention and in the classical testing paradigm. This was verified by *post hoc* behavioral identification and discrimination tasks. The observation of CP in the preattentive stage suggested the articulation planning is not a prerequisite for CP, which is not consistent with the motor theory, and thus again favored the multistage processing model for speech perception.

Experiment II used a similar oddball paradigm as that used in Experiment I but subjects were asked to pay attention to the stimuli and respond, whenever they heard a deviant. Both Mandarin and Cantonese subjects participated in the experiment and both speech and nonspeech (complex tones) stimuli were used. Although the level and rising tones are present in both Mandarin and Cantonese tone systems, these two tone systems have different inventories, different tone sandhi rules, and different teaching and learning systems as discussed in section 4.2.4.3. Therefore, although CP was observed in all conditions, there were some differences between these two groups of listeners. Cantonese listeners relied more on the carrier syllables to build up the phonological (toneme) memory than Mandarin listeners did, and Cantonese listeners had better discrimination on the continuum than Mandarin listeners did. Moreover, the results showed that brain activities were more informative than the behavioral responses which reflected the establishment of phonological memory, as detailed in section 4.2.4.4.

Table 4.4.1 CP effect (Across-category > Within-category) of level vs. rising tones in different cognitive stages for Mandarin listeners.

	Preattentive stage (MMN)	Attentive stage (P300)	Behavioral response (Hit Rate)
Speech	CP	CP	CP
nonspeech	No	CP	CP
Difference between speech and nonspeech	Yes	No	Yes

Experiment I and II were designed to investigate different stages of CP in the same oddball paradigm. Although in all stages, Mandarin listeners exhibited CP on the continuum from the level to rising tones, CP performance was different in different stages (See in Table 4.4.1). In the preattentive stage, CP was only observed in speech conditions and happened very early after the stimulus onset. In the attentive stage, CP was observed in both speech and nonspeech conditions and there was no evidence to show the difference of CP between speech and nonspeech conditions. At the last stage from the behavioral performance in the same paradigm, CP was again observed in both and nonspeech conditions; however, the degree of CP was greater in speech condition than that in nonspeech condition. This temporal dynamic process of CP from Mandarin listeners suggested that different mechanisms underlie CP in different cognitive stages.

Experiment III was a cross-language study to investigate CP on the level and rising tones by behavioral identification and discrimination tasks. The factor of language background affected many aspects of the performance. (1) The category boundaries were sharper for both groups of tone language listeners (Mandarin and Cantonese) than for the non-tone language (German) listeners. (2) Moreover, Cantonese listeners had two category boundaries (one in the middle, and the other at the level end) on the discrimination curves similar to that observed in Experiment II, which presumably corresponded to the boundaries between Cantonese high level and high rising tones and between Cantonese mid level and low rising tones. German listeners had one psychoacoustic category boundary at the level end from the discrimination curve. (3) Both groups of tone language listeners had a sharper category boundary in speech condition than in nonspeech condition, while non-tone language listeners had a sharper category boundary in nonspeech condition than in speech condition (See section 4.3.3).

The results from the three experiments reported in this chapter showed that the linguistic experience influenced CP in many aspects including those irrelevant to the articulation gesture or their neural commands. Moreover, both linguistic and psychoacoustic category boundaries (the detailed presentation of these two category boundaries was shown in Table 4.4.2) were observed in the experiments which suggested that CP was influenced by both linguistic experience and general auditory processing. Finally, the temporal dynamic processing of CP in different cognitive stages was revealed, which indicated different mechanisms underlie different stages of CP. All in all, the experiments in chapter 4 favored the multistage model in tone perception more than the motor theory model and obtained a similar conclusion as that from chapter 3.

Table 4.4.2 Presentation of psychoacoustic and linguistic category boundaries in the continuum spanning from the level to the rising tone

	Linguistic category	Psychoacoustic category
location (Language group)	Middle (Mandarin, Cantonese) Level end (Cantonese)	level end (Mandarin, German)
CP (AC>WC)	Speech > Nonspeech	speech < Nonspeech

CHAPTER 5 GENERAL DISCUSSIONS AND MAJOR CONTRIBUTIONS

Two sets of experiments reported in the present thesis investigated the categorical perception of lexical tones on two continua, spanning among Cantonese three level tones and spanning from the level to the rising tones in both Cantonese and Mandarin tone systems, respectively. Two major points were investigated through these experiments: the factors influencing *Categorical perception* (CP) of lexical tones, and the temporal dynamics of CP of lexical tones.

5.1. General discussions

5.1.1. Factors influencing CP

CP is extremely important, because we can efficiently and effectively organize the world through it. In contradict to CP, there is a concept of continuous perception. It has been generally accepted that the consonant perception is good evidence for CP, while vowel perception is more like continuous perception (Repp, 1984). As for tones, there is no strict dichotomy between CP and continuous perception. Many factors have been studied in previous studies to investigate how they influence the degrees of CP of lexical tones. However, those factors are reported by different groups of researchers and in different studies. Those reported factors covary with some other factors irrelevant to our research interests (e.g., different instructions given to the subjects, different presentation paradigms, and different stimuli synthetic methods used in different studies). Therefore, it is hard to obtain a complete and clear picture on the influence from factors of our interests. The present thesis systematically investigated different types of factors influencing CP of lexical tones. They were:

(1) the intrinsic acoustic properties of tones (contrast between level tones and contrast between level and contour tones), between which the difference emphasizes the general auditory processing;

(2) different positions of target syllables [isolation (without a contextual sentence), LC (Left Context, at the end of a contextual sentence), and RC (Right Context, at the beginning of a contextual sentence)], between which the difference emphasizes the general auditory processing;

(3) different language backgrounds (native tone language listeners from different tone systems, non native tone language listeners with experience of other tone systems, and non-tone language listeners), between which the difference emphasizes the language specific processing;

(4) different types of carrier syllables (real word speech, non-word speech, and nonspeech), between which the difference emphasized the language specific processing.

5.1.1.1. Intrinsic acoustic properties

Level tones, which are present in the Cantonese tone system, were perceived by native listeners less categorically when they were presented in isolation than in a contextual sentence (Refer to Experiment I in chapter 3). When level tones were presented in isolation, the identification curves were messy with large variations along the curves. However, similarly presented in isolation, continuum spanning from the level to the rising tones was perceived categorically by native listeners (Refer to three experiments in chapter 4). The observation that level tones were perceived non-categorically while contour tones were perceived categorically in the isolation form is consistent with earlier findings [level tones: (Francis, et al., 2003; Francis, Ciocca, Wong, Leung, & Chu, 2006), and contours tones: (Francis, et al., 2003; Wang, 1976; Y. S. Xu, Gandour, & Francis, 2006)].

Speech perception relies on both intrinsic and extrinsic references (Y. S. Xu, Gandour, & Francis, 2006). The intrinsic references are computed locally, which refer to the relative relationship between two or more acoustic features within a syllable (e.g., temporal order of acoustic cues for voicing relative to the release of stop consonants), or relative relationship of one acoustic feature across time in a syllable (e.g., direction of pitch movement or formant transition). Categorization of intrinsic features is less demanding on working memory, because computation can be carried out within each stimulus. For example, judgments of pitch direction (the contour tones herein) rely on the intrinsic reference. In contrast, steady-state features such as the pitch contrast of level tones lack intrinsic references. Categorical encoding of these acoustic features is dependent on extrinsic references. Extrinsic references are computed globally, which are based on either a normalized acoustic level derived from other stimuli in the context, or the best matched exemplar in memory. This explains why steady-sate level tones were not perceived categorically when they were presented in isolation.

According to the intrinsic acoustic properties of level tones and contour tones, the perception of level tones relies on extrinsic references, while the perception of contour tones relies on intrinsic references. Since the computation of intrinsic or extrinsic references is not restricted to speech sounds but rather a general auditory processing, the observation of different degrees of CP on level tones and contour tones supports the function of general auditory processing in CP.

5.1.1.2. Position of target syllables relative to context

As discussed in the section 5.1.1.1, perception of steady-state level tones relies on extrinsic reference. Indeed, improvements on both identification consistency and the degree of CP were obtained after the targets (level tones) were presented in a contextual sentence. We also manipulated the positions of the contextual sentence to be before the target (LC, left context) or to be after the target (RC, right context). The results showed that the position of the contextual sentence affected both the degree of CP and the discrimination score, in that both were higher for the across-category pair when the target level tones were presented in RC than in LC sentence. This *asymmetric effect of context position* (AECP) on speech perception has been reported in earlier studies (Francis, et al., 2006; Mann & Repp, 1981; Ohde & Sharf, 1977; van Son & Pols, 1999).

Several mechanisms may be responsible for this observed AECP. First possible reason is the influence from coarticulation. Previous studies have shown that the coarticulation affects the acoustic properties on the transition between the context and the target (Magen, 1997). Moreover, in lexical tones, carry-over effect (LC herein) is greater than the anticipatory effect (RC herein) (Y. W. Wong, 2006; Y. Xu, 1994, 1997). The coarticulation effect results in a smaller pitch distance between the context and the target in LC than in RC sentence. Therefore, a stronger coarticulation effect leads to a weaker discriminability and a smaller CP effect in LC than in RC condition. Such prediction was consistent with the findings from four experiments in chapter 3, especially Experiment III. The results from Experiment III showed that when the transition was more natural (coarticulation effect was more obvious, and the distance between the target and the context was smaller, especially in LC), the AECP was more prominent. When the target syllables were nonspeech, a similar AECP effect on the discrimination score was also observed from Experiment III in the present thesis. Moreover, a series of experiments showed that nonspeech with similar acoustic features also has a similar effect on perception of the target (Holt, 2005, 2006). The studies from nonspeech indicated that speech categorization is sensitive to statistical distributions of acoustic information, even if the distributions are composed of nonlinguistic elements.

The observation that a more natural transition resulted in a more obvious AECP from Experiment III of chapter 3 contradicts the predictions of the motor theory. The motor theory proposed that invariance of articulatory gesture is the main reason for CP, therefore, natural coarticulation does not affect speech perception. In other words, motor theory predicts that no AECP will be observed when the transition is as natural as it was produced and that different degrees of distortion in LC and RC from the natural sounds will induce an AECP. In Experiment III of chapter 3, two sets of continua were compared: INT continuum had a more natural transition by interpolating pitch contours from the natural speech sounds, and SSC continuum had a "distorted" transition by simply shifting the context (See section 3.3 for details). The result showed that AECP was more prominent for INT continuum than for SSC, which was contradicted to the prediction from the motor theory. Consequently, the motor theory cannot explain the observed results.

If the asymmetric coarticulation effect is the only reason to the observed AECP, a symmetric CP and discrimination will be predicted if the transitions are the same in the LC and RC conditions. Experiment II of chapter 3 was designed to test this hypothesis. In this experiment, stimuli were synthetic speech sounds. Therefore, same transition was present in both LC and RC conditions. Nonetheless, similar AECP on CP was observed. This result was consistent with a previous study on the perceptual order and influence of the vocalic sound on the fricative consonant (Mann & Soli, 1991). Their study showed that perception of the fricative was affected by a following vocalic sound (FV) more than by a preceding one (VF). Moreover, this asymmetric effect was still present when the stimuli were presented in *reverse*, where the acoustic properties were asymmetric in a reverse direction if any. Therefore, the asymmetric acoustic properties due to coarticulation cannot explain the observations either from our Experiment II or from the experiment by Mann and Soli (1991). In addition to the asymmetric acoustic properties of transition due to coarticulation effect, other mechanisms should be responsible for the AECP too.

The second possible reason for AECP was the timing of reference in relation to the target. In both LC and RC conditions, tonal categories of the context are well identified by the semantic meaning, even though actual pitch contours of the context vary along the continuum. However, the identification of tonal categories of the target totally relies on the context because pitch contours of the target are kept unchanged. In LC condition, the reference is established before the target is presented, therefore, auditory memory does not decay when subjects make a judgment on the target. However, in RC condition, the reference is presented after the target has been heard, therefore, auditory memory may decay when the reference is identified. Decay of auditory memory enhances the CP has been evident in several studies discussed in section 2.1.3. Since building up of phonemic memory relies on linguistic experience, a stronger perceptual order effect will be predicted with more experience. This prediction is consistent with a cross-sectional study on populations with different ages (Nittrouer, Miller, Crowther, & Manhart, 2000). The study showed that such asymmetric effect regarding on the context order was developed with experience: no AECP was observed by 5-year-old children, a weaker AECP was observed by 7-year-old children, and a stronger AECP was observed by adults.

Both possible reasons (coarticulation and timing of reference) to explain the observed AECP reflect the role of general auditory processing even though it is affected by the linguistic experience. Therefore the obtained AECP favors the multistage model of speech perception, which includes both general processing and linguistic experience.

5.1.1.3. Language backgrounds

Cross-language comparisons on CP of lexical tones were conducted. In addition to contrasting the responses from the tone and non-tone language listeners, we also compared responses from two groups of tone language listeners with different tone systems. Two sets of tones were tested: one set of tones are only present in one tone system, the other set of tones are present in both tone systems. The aims of these comparisons were to test (1) whether listeners without a tone experience perceived tone categorically; (2) whether experience from other tone systems affected the perception of non-native tones; (3) whether experience from different tone systems affected the perception of the same tones present in both systems.

Results showed that tone language listeners perceived native tone contrasts categorically. Tone language listeners also perceived non-native tone contrasts categorically, but the degree of CP was less than that from native listeners and the position of the category boundary was affected by their tone inventories. For non-tone language listeners, they did not perceive level tones categorically; while they had a psychoacoustic boundary on the continuum spanning from the level to the rising tone. This difference between tone and non-tone language groups demonstrated that listeners' native language experience with tone categories affected the perception non-native suprasegmental categories. This finding is also consistent with a study on training Mandarin and English subjects to learn Cantonese tones and showing the perception and acquisition of Cantonese tones is affected by the learners' linguistic experience (Francis, Valter, & Lian, 2004).

It is also interesting to find that the AECP of discrimination score (higher discrimination for RC than for LC condition) and the CP effect (better discrimination for across-category than for within-category pairs and sharp category boundary) were more prominent in speech condition than in nonspeech condition for tone language listeners, but were more prominent in nonspeech condition than in speech condition for non-tone language listeners. The tone and non-tone language listeners had different patterns on the difference of performance between speech and nonspeech indicates that AECP and CP effect can be stemmed from the general auditory processing and be shaped by linguistic experience. Some longitudinal and cross-sectional studies on speech perception showed that infants can discriminate against non native phonemic contrast in early infancy, but such ability declines as a function of age and linguistic experience (Kuhl, et al., 2006; Werker & Tees, 1992, 2002). However, this reorganization of perceptual space is limited to the speech condition. For example, tone contrast between the level vs. the rising is not present in English system, but 6 months old English infants are able to discriminate this contrast. However, when the same infants grow up into 9 months old, they fail to discriminate against the same contrast but only under the speech condition (Mattock & Burnham, 2006). Another study is consistent with the developmental change by showing that adult English listeners have a higher discrimination on the contrast in nonspeech condition than in speech condition (Y. S. Xu, Gandour, & Francis, 2006).

Different tone systems affect the perception not only on native tone contrast but also on native tone contrast. Level and rising tones are present in both Mandarin and Cantonese tone systems. However, our results showed that due to different tone inventories, tone sandhi[8] rules (for detailed discussion, please refer to section 4.2.4.3), Cantonese listeners are better to discriminate the level and rising tone contrast than Mandarin listeners did, especially in speech condition. Moreover, Cantonese listener had a second discrimination peak on the level end for the continuum spanning from the level to the rising tone. The second discrimination peak presumably corresponds to discrimination for the contrast between Cantonese mid-level tone and low-rising tone.

5.1.1.4. Carrier syllables

Three types of carrier syllables (real word, non-word, and nonspeech) were used in chapter 3 to study the CP of Cantonese level tones; two types of carrier syllables (speech and nonspeech) were used in chapter 4 to study the CP of level vs. rising tones. For native tone listeners, there was no difference on AECP or CP effect between real word and non-word but higher CP effect for speech stimuli than for nonspeech stimuli. For non-tone listeners, a reverse pattern was observed: higher AECP and CP effect for nonspeech than for speech stimuli. Summary of the pattern were shown in table 5.1.1.

Table 5.1.1 Summary of the effect of carrier syllables on discrimination performance

Subjects	Pattern	Effect on carrier syllables
Native tone language listener	Discrimination RC>LC, AC>WC	real word = non word>nonspeech (or, Speech > nonspeech)
Non-tone language listener	Discrimination RC>LC, AC>WC	nonspeech > speech

---

[8] Gandour's (1983) nonstandard use term.

No significant difference between real word and non-word, but a significant difference between speech and nonspeech for native listeners, suggests that the CP processing is a phonological (or tonological) processing regardless whether the speech had semantic meanings. We used complex tones as the carrier for nonspeech; therefore, the spectrum complexity between speech and nonspeech is comparable. Nonetheless, we observed a significant difference between speech and nonspeech. Moreover, the difference was modulated by linguistic experience. Therefore, the difference between speech and nonspeech indicated that linguistic experience reflected by behavioral responses was more confined to speech stimuli but may be generalized to nonspeech stimuli.

The observation that language experience modulated speech vs. nonspeech difference may be partly explained by the multistorage model proposed by Y.S. Xu et al. (2006). Y.S. Xu et al. proposed that complex tones have simpler spectra than real speech. Therefore, pitch contours are easier to be extracted on-line from the complex tones than spectrally more complex speech stimuli, and thus results in higher CP in nonspeech than in speech by non-tone listeners. This prediction from the multistorage model is consistent with our results. However, their model also suggested that CP of lexical tones primarily relies on pitch contours. For native listeners, the long-term memory for phonological information of pitch contours is built up and therefore no significant difference between speech and nonspeech, which is consistent with their results but not with ours. Our results showed that CP in speech is higher than in nonspeech condition.

We proposed that the building up of phonological information of lexical tones not only relies on the pitch contours themselves, but also on the segmental information. Speech sounds are more familiar to native listeners, and the linguistic experience helps to build up a long-term memory. Therefore, no explicit on-line analysis on pitch contours is required. This automatic arousal of phonological information about lexical tones results in native listeners perceiving speech sounds more categorically than nonspeech sounds. No significant difference between real word and non-word[h] suggested that the phonemic information contributed to tonal identification to some extent, even though the degree of integration may also be affected by different tone systems and different degrees of phonological awareness.

---

[h] The non-word stimuli used in the experiment may be also meaningful with tone 1 and tone 3 to some of the native listeners. However, according to the *post hoc* self reports by the listeners, none of them reported they recognize the sounds.

This hypothesis is consistent with some previous behavioral studies which suggested that perception of tone and segmental information influence each other (L. Lee & Nusbaum, 1993; Tong, Francis, & Gandour, 2008). Brain-imaging data from perception study also suggests that distinct brain regions are involved when pitch contours are superimposed on linguistic (pseudoword) or nonlinguistic (hum) contexts (Gandour, et al., 2002). The brain-imaging finding suggests the representation of tones in the brain depends on speech context even when none of the carrier bases carry semantic meaning.

5.1.2. Temporal dynamics of CP

In addition to different factors influencing CP, we also investigated the temporal dynamics of CP on different cognitive stages using ERP measure. Three stages — preattentive stage, attentive stage, and overt response stage — were studied in a similar odd-ball paradigm (Shown in Table 4.4.1). Stimuli used in this set of experiments were speech and nonspeech continua spanning from the level to the rising tone. Difference between nonspeech and speech reflects the difference between general auditory processing and language specific processing.

In preattentive stage, even though the subjects did not pay attention to the stimuli, they could detect all the deviants from the standard by eliciting significant MMNs. However, the early MMN was only elicited from the across-category speech deviants, while the late MMN was elicited from the other deviants. Moreover, the CP effect was present only in speech, with a higher MMN eliciting from across-category than from within-category. In the discussion (in section 4.1.4), we attribute the observed CP in speech condition to the stored phonemic memory on tones, which reflects the language processing. The fact that a late MMN was elicited from two types of deviants in the nonspeech condition as well as from within-category deviants in the speech condition indicated that the auditory processing was also present in the preattentive stage.

The observation of CP of lexical tones in preattentive stage suggests the automatic phonemic awareness without subjects' consciousness. This early automatic detection of phonemic change for speech stimuli can be considered as a reorganization of neural network through training. A new study showed that the interconnection between the trained area and the attention related neural network was increased even in the resting state (without any tasks), while the interconnection between the untrained areas and the attention related neural network was decreased. This result suggested the top-down reorganization of neural network through learning (Lewis, Baldassarre, Committeri, Romani, & Corbetta, 2009; Welberg, 2009). The result is quite consistent with the finding that MMN neural generators normally involve the frontal lobe (Näätänen, et al., 2007). The neural generator from the frontal lobe corresponds to the involuntary shift of attention to a memory change when a deviant is heard. Therefore, in the case of speech perception, the training enhances the neural connection between the auditory cortex and the attention related neural network, and thus increases the neural activities (MMN herein) of these areas when the trained sounds are heard.

This preattentive phonemic awareness is an indicator of language processing. It was observed from many other types of phonemic contrasts [i.e., consonant-vowel syllables (Kazanina, et al., 2006), vowels (Jacobsen, Schröger, & Alter, 2004; Näätänen, et al., 1997), and phoneme combinations (Bonte, Poelmans, & Blomert, 2007)]. Moreover, the successful preattentive phonemic awareness predicts a good mastery of phonemic contrasts in the second language learning (Díaz, Baus, Escera, Costa, & Sebastián-Gallés, 2008), while the unsuccessful preattentive phonemic awareness is associated with dysfunction in language processing (Bonte & Blomert, 2004; Kasai, et al., 2002).

Similar to the preattentive stage, in attentive stage, when the subjects paid attention to the stimuli, all deviants were detected from the standard; moreover, CP was present in both speech and nonspeech conditions, with across-category deviants eliciting a larger P300 than within-category deviants. Both linguistic and psychoacoustic boundaries were obtained from the present experiment (Refer to table 4.4.2 for detailed presentation of these two category boundaries). The presence of a linguistic category boundary indicates the presence of language processing in this attentive stage. The language processing is also exhibited by the presence of different performances from Cantonese and Mandarin listeners, which indicates influences from different tone systems. On the other hands, the presence of a psychoacoustic category boundary indicates the presence of general auditory processing in this attentive stage (See details in section 4.2.3 and section 4.2.4). The presence of general auditory processing is also exhibited by the fact that within-category deviants in both speech and nonspeech conditions elicited a P300.

The P300 component reflects the ease with which subjects update mental schema in response to deviant stimuli (Polich, 2007). Therefore, the observed CP effect, with across-category deviants eliciting a higher P300 than within-category deviants, indicates that across-category deviants are more salient to the mental schema (standard) than within-category deviants. Since these two types of deviants have the same acoustic differences as the standard, the larger difference of across-category deviants is coming from the different categories between this type of deviants and the standards. The observed linguistic CP in both speech and nonspeech conditions indicates that subjects have intuitive or explicit phonological (or tonological) knowledge to extract pitch contours from the speech stimuli and this ability is also generalized to nonspeech condition. Some researchers suggested that exposure to a particular language produces a neural commitment to the acoustic properties of that language (Zhang, Kuhl, Imada, Kotani, & Tohkura, 2005). In the present experiment, a subtle difference of the perception performance was obtained by listeners from two different tone systems: no difference of CP from speech and nonspeech conditions by Mandarin listeners; a higher degree of CP from the speech condition than from the nonspeech condition by Cantonese listeners. This subtle difference further indicated that the degree of neural commitment to the acoustic properties depends on linguistic experiences. Mandarin listeners have a good and explicit phonological knowledge to extract pitch contours from both speech and nonspeech conditions, and thus they have a high degree of neural commitment to the acoustic properties (pitch contours) (Details were discussed in section 4.2.4). However, Cantonese listeners only have intuitive phonological knowledge about tones, which is attributed to lack of the phonemic training during early school. Consequently, the degree of neural commitment to pitch contours is lower by Cantonese listeners.

Finally, in the decision stage, when the subjects were asked to do the overt responses in the oddball paradigm, CP was present in both speech and nonspeech conditions as in the attentive stage, with a higher hit rate from across-category deviants than from within-category deviants. Moreover, similar to the P300 data, both linguistic and psychoacoustic category boundaries were obtained from the hit rate data (Refer to table 4.4.2 for detailed presentation of these two category boundaries). The presence of a linguistic category boundary indicates the presence of language processing in this overt response stage. On the linguistic category boundary, speech and nonspeech condition differed in within-category deviants with a higher hit rate in the nonspeech condition than in the speech condition. This difference between speech and nonspeech conditions indicates that even though the phonological knowledge and linguistic experience reduce the sensitivity to the within-category (acoustic) difference, the discriminability for the acoustic difference was preserved, especially in the nonspeech condition. On the other hands, the presence of a psychoacoustic boundary indicates the presence of general auditory processing.

Different patterns of speech perception were observed in three stages from the present thesis. Firstly, the within-category deviants of linguistic category in the nonspeech condition elicited a similar size of MMN as across-category deviants did in the preattentive stage, but elicited a smaller size of P300 in the attentive stage. Such loss of sensitivity for non native phonemic contrasts attentively or behaviorally was also observed in other studies. For example, a study examined ERP from subjects during a passive oddball task when they are presented with different types of syllabic contrasts: native phonemic contrast, non-native phonemic contrast, and within-category contrasts. The analyses showed that subjects pre-attentively perceived the differences in all experimental conditions, but the difference in the non-native and within-category conditions did not perceive behaviorally. These results support the notion that there is no permanent loss of the initial perceptual abilities that humans have as infants, but that there is an important neural reorganization which allows the system to overcome the differences detected and only be aware of contrasts that are relevant in the language which will become the subjects native tongue (Rivera-Gaxiola, et al., 2000). In their experiment, different types of contrasts do not share the same contrasts; therefore, nonnative contrasts are not detectable behaviorally due to the neural reorganization. Our results and Rivera-Gaxiola et al.'s results are consistent with another study, which reported that only explicit knowledge of deviants can be detected in the attentive stage but all types of difference can be automatically detected preattentively (van Zuijen, et al., 2006). Secondly, the degree of linguistic CP reflected by P300 was similar in speech and nonspeech conditions by Mandarin listeners, but was more prominent in speech condition than in nonspeech condition reflected by the hit rate data. This result indicates different strategies used in different stages, which was discussed in section 4.2.4.4.

Table 5.1.2 Presentation of general auditory processing and language specific processing in three stages tested in the oddball paradigm.

Stage	**Measure**	**General auditory** processing	**Language specific** processing
preattentive stage	MMN	1. Late MMN from the WC deviant in the speech condition; 2. Late MMN from both AC and WC deviants in the nonspeech condition	1. Early MMN from the AC deviant in the speech condition; 2. CP in the speech condition;
Attentive stage	P300	1. P300 from all types of deviants in both speech and nonspeech conditions; 2. presence of a psychoacoustic category	1. Presence of a linguistic category; 2. Difference between two language groups (Cantonese and Mandarin)
Behavioral stage	Hit Rate	1. Detection of all types of deviants; 2. Presence of a psychoacoustic boundary	1. Presence of a linguistic category; 2. Difference between speech and nonspeech conditions

The same oddball paradigms were used in different stages to investigate CP of tones. Both the general auditory processing and linguistic specific processing are exhibited by different types of presentation, which are summarized in table 5.1.2. These different patterns of two types of processing indicated that different mechanisms underlying the CP in different stages. That CP was observed in the preattentive stage suggested that no articulatory planning or corresponding neural command is needed for CP. This result is inconsistent with the motor theory, which occurs with subjects' attention. The observed CP of speech stimuli in preattentive stage may be due to long-term phonemic storage and thus the multistage model is more reasonable. In the attentive stage, the phonological arousal is the major reason for the observed linguistic CP. Finally, the fact that a better discrimination on the within-category difference in the nonspeech condition from the overt responses for linguistic category suggested that although the discriminability for a nonnative tone contrast is reduced through linguistic experiences, the ability is not lost permanently. The facts that different mechanisms are responsible for CP in different stages further confirm that CP processing is a multistage processing, which involves both processing from general auditory processing and language processing in different stages. Therefore, the single mechanism proposed in the motor theory is not sufficient to explain CP.

5.1.3. Multistage model for tone perception

The results in the present thesis suggest that speech perception is a combination process of auditory processing and language specific processing. The contribution of auditory processing on CP was mostly exhibited by the presence of psychoacoustic boundaries, influence from intrinsic acoustic properties, and general better discrimination in RC than in LC context. The contribution of language specific processing was mostly exhibited by the presence of phonological boundaries which are consistent with native tone systems and different degrees of context effect which are influenced by linguistic experience.

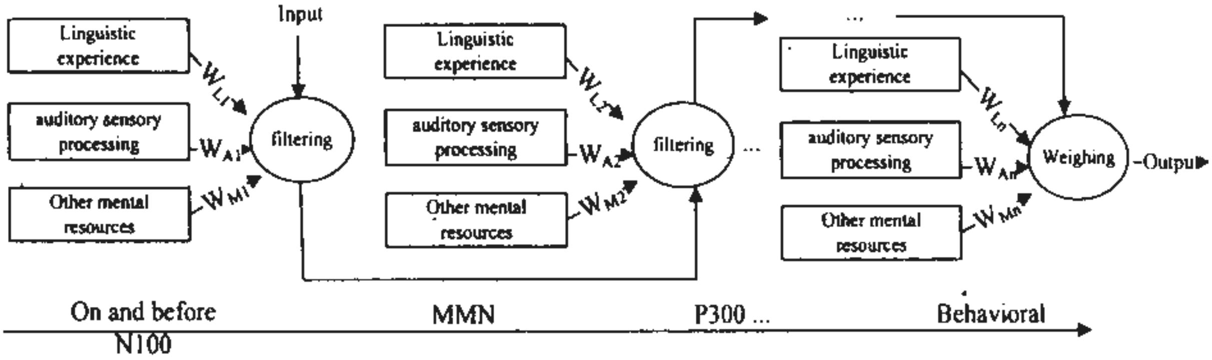


Figure 5.1.1 Conceptual diagram of multistage processing model

The combination of two types of processing is not only exhibited in several ways on the final cognitive stage (behavioral responses), these two types of processing, as well as some other mental resources such as attention arousal, also have different weights in various cognitive stages. Therefore, the output of each stage is based on the weighting from different types of sources (shown in Figure 5.1.1). In the model, the representation at different stages is determined by the weighting from different types of resources. For example, in the preattentive stage, awareness of tone categories was evoked automatically without attention. However, such process relies on some spectral information, which is necessary to separate speech from nonspeech. In other words, the attention arousal from the pathway of the mental resource is lest weighted in the preattentive stage. However, when the attention arousal plays an important role in the processing with a higher weight, the phonological arousal becomes clearer such as in the attentive stage (P300) or behavioral stage. In the attentive stage, building up of the mental schema (the phonological arousal) of lexical tones relies on the representation features — the pitch contours. At the final stage, with the overt response, since listeners can also rely on immediate adjacent stimuli to make part of the decision, the within-category difference is also aware, especially in nonspeech condition.

All in all, no matter different types of exhibition or different weights of exhibition, all the evidence suggested that the articulatory gestures or their corresponding neural command is not the only reason for CP. Although there is some indirect evidence to show that motor system is recruited for perceiving speech [See review (Galantucci, et al., 2006)], it doesn't provide the direct evidence that the motor system is necessary for perception. In a review, Scott et al. suggested that the function of the motor system in the perception is modulation but not central (Scott, et al., 2009). Speech perception is a complex process, which has multiple stages and involves many resources.

Brain imaging studies showed that different brain regions are involved when subjects perceive speech sounds. Moreover different brain regions have different timing to be involved in speech perception The importance of temporal-parietal regions on speech perception has been widely accepted, since Wernicke reported the case study (Wernicke, 1874,1969). The regions are also important for categorical perception (Ravizza, 2003). In addition to the temporal region, the prefrontal cortex and basal ganglia are also reported to be important for category learning (Ashby & Ell, 2001). Recently, motor areas are also found to play some important roles in speech perception, but the involvement of the motor areas in speech perception is later than the involvement of the temporal regions (Pulvermüller & Shtyrov, 2006; Pulvermüller, et al., 2005).

In conclusion, we proposed a multistage model, which integrated the virtues from multistore model and multilevel model, to explain the CP data observed in the present study and early studies. The multistage model emphasized that both linguistic specific processing and general auditory processing have different weights in different conditions and in different stages along the time axis. Many kinds of neural resources act together when the subjects perceive a speech sounds. This activated neural network is shaped by linguistic experience and affected by the level attentional arousal. Moreover, many factors, such as stimulus complexity, familiarity, and context etc. all contribute to finely attune the activation of the neural network. Although Xu et al. (2006) proposed a multistage model, which included different types of memory systems, to explain CP. However, no temporal information is available for the function of each memory system in their model. Phillip (2003) suggested a serial or near-parallel multilevel model to explain that the speech perception is from the general auditory processing to phonetic processing and to phonemic processing. However, this model overlooks the linguistic effect during the early responses and ignores the complexity of the conditions where speech is perceived. Our proposed multistage model introduces the concept of different weights for two types of processing and thus encourages reconsidering the model of speech perception.

5.2. Major contributions of this work

In the present thesis, different factors influencing CP of lexical tones and temporal dynamics of CP were systematically investigated by behavioral and ERP experiments. Two types of comparisons were carried out: syllable carriers (speech and nonspeech sounds) and linguistic backgrounds of the listeners (from native tone language system, non-native tone language system and non-tone language systems). The results suggest that linguistic experience and general auditory processing both contribute to CP.

We reported that experience of one tone system influences the perception on non-native tone contrasts. More importantly, for the first time in the literature, we reported that even though a tone contrast (i.e., level vs. rising) is present in both tone systems, the same contrast is perceived differently by the two groups of subjects by virtue of their different language experience.

The temporal dynamics of CP of lexical tones from preattentive stage through the attentive stage to the overt response stage was also revealed in the present thesis. This complete procedure of CP under the same oddball paradigm is new to the literature. Different patterns of CP were observed in different stages. This suggests that the general auditory processing and language processing have different weights in different stages.

5.3. Future work

Although ERP technology is good for investigating the process of CP along the time domain, it is not an appropriate method to study where the CP happens in the brain due to its poor spatial resolution. Processing of native phonemic contrast activates the left hemisphere more than the right hemisphere even in preattentive stage has been reported (Näätänen, et al., 1997). Impairment of language function is associated with increased involvement of the right hemisphere. For example, aphasia patients showed increased right hemisphere activities to the consonant-vowel syllables in the preattentive stage (Becker & Reinvang, 2007).

However, the ERP study in the present thesis did not show a clear asymmetric activation of two hemispheres comparing across-category and within-category stimuli even in the attentive stage. It seems that the brain regions involved in the tone processing are not clear yet, although Zatorre and Gandour proposed that the activation of hemispheres for tone processing is different in different stages (Zatorre & Gandour, 2008). A PhD thesis reported the experimental results through ERP studies, which is consistent with their proposal (Shuai, 2009). However, due to the low spatial resolution of ERP techniques, her results did not show the source localization of the neural generators. We hope using brain imaging method like fMRI, which has a better spatial resolution than ERP, will identify which brain regions respond to tone CP better.

On the other hands, the different performance on the same native tone contrast from subjects of different tone systems was only investigated in the attentive stage in the present study. We would like to investigate an earlier stage such as those measured from MMN and FFR (*frequency following response*) to investigate how early such a subtle difference in experience will affect the brain activities.

REFERENCES

Abramson, A. S. (1972). Tonal experiments with whispered Thai. In A. Valdman (Ed.), *Papers in linguistics and phonetics to the memory of pierre delattre*. Mouton: The Hague.

Abramson, A. S. (1979). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Ohman (Eds.), *Frontiers of speech communication research*. New York: Academic Press.

Ackermann, H., Lutzenberger, W., & Hertrich, I. (1999). Hemispheric lateralization of the neural encoding of temporal speech features: A whole-head magnetencephalography study. *Cognitive Brain Research, 7*(4), 511-518.

Akatsuka, K., Wasaka, T., Nakata, H., Inui, K., Hoshiyama, M., & Kakigi, R. (2005). Mismatch responses related to temporal discrimination of somatosensory stimulation. *Clinical Neurophysiology, 116*(8), 1930-1937.

Alain, C., Woods, D. L., & Knight, R. T. (1998). A distributed cortical network for auditory sensory memory in humans. *Brain Research, 812*(1-2), 23-37.

Alexander, J. E., Bauer, L. O., Kuperman, S., Morzorati, S., O'Connor, S. J., Rohrbaugh, J., et al. (1996). Hemispheric differences for P300 amplitude from an auditory oddball task. *International Journal of Psychophysiology, 21*(2-3), 189-196.

Alexander, J. E., Porjesz, B., Bauer, L. O., Kuperman, S., Morzorati, S., O'Connor, S. J., et al. (1995). P300 hemispheric amplitude asymmetries from a visual oddball task. *Psychophysiology, 32*(5), 467-475.

Alho, K. (1995). Cerebral generators of mismatch negativity (MMN) and its magnetic counterpart (MMNm) elicited by sound changes. *Ear and Hearing, 16*(1), 38-51.

Alho, K., Winkler, I., Escera, C., Huotilainen, M., Virtanen, J., Jaaskelainen, I. P., et al. (1998). Processing of novel sounds and frequency changes in the human auditory cortex: Magnetoencephalographic recordings. *Psychophysiology, 35*(2), 211-224.

Arikan, M. K., Devrimb, M., Oran, Ö., Inan, S., Elhih, M., & Demiralp, T. (1999). Music effects on event-related potentials of humans on the basis of cultural environment. *Neuroscience Letters, 268*(1), 21-24.

Ashby, F. G., & Ell, S. W. (2001). The neurobiology of human category learning. *Trends in Cognitive Sciences, 5*(5), 204-210.

Atienza, M., Cantero, J. L., & Dominguez-Marin, E. (2002). Mismatch negativity (MMN): An objective measure of sensory memory and long-lasting memories during sleep. *International Journal of Psychophysiology, 46*(3), 215-225.

Baddeley, A., & Wilson, B. A. (1993). A developmental deficit in short-term phonological memory:Implications for language and reading. *Memory, 1*(1), 65-78.

Bastian, J., Eimas, P. D., & Liberman, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *The Journal of the Acoustical Society of America, 33*(6), 842.

Baudena, P., Halgren, E., Heit, G., & Clarke, J. M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. *Electroencephalography and Clinical Neurophysiology, 94*(4), 251-264.

Becker, F., & Reinvang, I. (2007). Mismatch negativity elicited by tones and speech sounds: Changed topographical distribution in aphasia. *Brain and Language, 100*(1), 69-78.

Belin, P. (2006). Voice processing in human and non-human primates. *Philosophical Transactions of the Royal Society B-Biological Sciences, 361*(1476), 2091-2107.

Bendor, D., & Wang, X. Q. (2005). The neuronal representation of pitch in primate auditory cortex. *Nature, 436*(7054), 1161-1165.

Bent, T., Bradlow, A. R., & Wright, B. A. (2006). The influence of linguistic experience on the cognitive processing of pitch in speech and nonspeech sounds. *Journal of Experimental Psychology: Human Perception and Performance, 32*(1), 97-103.

Bentin, S., & Feinsod, M. (1983). Hemispheric asymmetry for word perception: Behavioral and ERP evidence. *Psychophysiology, 20*(5), 489-497.

Best, C., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *perception & Psychophysics, 29*(3), 191-211.

Birbaumer, N., & Cohen, L. G. (2007). Brain–computer interfaces: Communication and restoration of movement in paralysis. *Journal of Physiology, 579*(3), 621-636.

Blau, V., Maurer, U., Tottenham, N., & McCandliss, B. (2007). The face-specific n170 component is modulated by emotional facial expression. *Behavioral and Brain Functions, 3*(1), 7.

Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer (Version 4.6.33): Institute of Phonetic Sciences. .

Bonte, M. L., & Blomert, L. (2004). Developmental dyslexia: ERP correlates of anomalous phonological processing during spoken word recognition. *Cognitive Brain Research, 21*(3), 360-376.

Bonte, M. L., Poelmans, H., & Blomert, L. (2007). Deviant neurophysiological responses to phonological regularities in speech in dyslexic children. *Neuropsychologia, 45*(7), 1427-1437.

Brace, N., Kemp, R., & Snelgar, R. (2003). *Spss for psychologists: A guide to data analysis using spss for windows* (2d ed.). Great Britain: Ashford colour press.

Breier, J. I., Fletcher, J. M., Denton, C., & Gray, L. C. (2004). Categorical perception of speech stimuli in children at risk for reading difficulty. *Journal of Experimental Child Psychology, 88*(2), 152-170.

Breier, J. I., Gray, L., Fletcher, J. M., Diehl, R. L., Klaas, P., Foorman, B. R., et al. (2001). Perception of voice and tone onset time continua in children with dyslexia with and without attention deficit/hyperactivity disorder. *Journal of Experimental Child Psychology, 80*, 245-270.

Broca, P. (1861). Remarques sur le siège de la faculté du langage articulé; suivies d'une observation d'aphémie (perte de la parole), translated as 'remarks on the seat of the faculty of articulated language, following an observation of aphemia (loss of speech)' by C.D. Green. Retrieved Dec., 2009, from http://psychclassics. yorku.ca/Broca/aphemie-e.htm

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., & Vanlancker, D. (1994). Influence of language structure on brain-behavior development. *Brain and Language, 46*(4), 607-619.

Burton, M. W., & Small, S. L. (2006). Functional neuroanatomy of segmenting speech and nonspeech. *Cortex, 42*(4), 644-651.

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Categorical perception of happiness and fear facial expressions: An ERP study. *Journal of Cognitive Neuroscience, 14*(2), 210-227.

Carden, G., Levitt, A., Jusczyk, P. W., & Walley, A. (1981). Evidence for phonetic processing of cues to place of articulation - perceived manner affects perceived place. *Perception & Psychophysics, 29*(1), 26-36.

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 & Neuroscience, 25*(3/4), 195-210.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007a). Experience-dependent neural plasticity is sensitive to shape of pitch contours. *Neuroreport, 18*(18), 1963.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2007b). Mismatch negativity to pitch contours is influenced by language experience. *Brain Research, 1128*(1), 148-156.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Relative influence of musical and linguistic experience on early cortical processing of pitch contours. *Brain and Language, 108*(1), 1-9.

Cheour, M., Ceponiene, R., Lehtokoski, A., Luuk, A., Allik, J., Alho, K., et al. (1998). Development of language-specific phoneme representations in the infant brain. *Nature Neuroscience, 1*(5), 351.

Cheung, H., Chung, K. K. H., Wong, S. W. L., McBride-Chang, C., Penney, T. B., & Ho, C. S. H. (2009). Perception of tone and aspiration contrasts in Chinese children with dyslexia. *Journal of Child Psychology and Psychiatry, 50*(6), 726-733.

Chong, T. T. J., Cunnington, R., Williams, M. A., Kanwisher, N., & Mattingley, J. B. (2008). fMRI adaptation reveals mirror neurons in human inferior parietal cortex. *Current Biology, 18*(20), 1576-1580.

Chope, M., Metzlutz, M. N., Wioland, N., Rumbach, L., & Kurtz, D. (1994). Event-related potentials and language processing. *Neurophysiologie Clinique-Clinical Neurophysiology, 24*(4), 275-300.

Cienfuegos, A., March, L., Shelley, A.-M., & Javitt, D. C. (1999). Impaired categorical perception of synthetic speech sounds in schizophrenia. *Biological Psychiatry, 45*(1), 82-88.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002). Mismatch negativity evoked by the McGurk-MacDonald effect: A phonetic representation within short-term memory. *Clinical Neurophysiology, 113*(4), 495-506.

Conrey, B., Potts, G. F., & Niedzielski, N. A. (2005). Effects of dialect on merger perception: ERP and behavioral correlates. *Brain and Language, 95*(3), 435-449.

Cooper, W. E. (1979). *Speech perception and production: Studies in selective adaptation*. Norwood, NJ: Ablex Pub.

Crystal, D. (2003). *A dictionary of linguistics and phonetics: 5th edition*: Blackwell.

D'Ausilio, A., Pulvermüller, F., Salmas, P., Bufalari, I., Begliomini, C., & Fadiga, L. (2009). The motor somatotopy of speech perception. *Current Biology, 19*(5), 381-385.

Dalebout, S., & Stack, J. (1999). Mismatch negativity to acoustic differences not differentiated behaviorally. *Journal of the American Academy of Audiology, 10*(7), 388-399.

Dehaene-Lambertz, G., Pallier, C., Serniclaes, W., Sprenger-Charolles, L., Jobert, A., & Dehaene, S. (2005). Neural correlates of switching from auditory to speech perception. *NeuroImage, 24*(1), 21-33.

Delorme, A., & Makeig., S. (2004). Eeglab: An open source toolbox for analysis of single-trial EEG dynamics. *Journal of Neuroscience Methods 134*, 9-21.

Díaz, B., Baus, C., Escera, C., Costa, A., & Sebastián-Gallés, N. (2008). Brain potentials to native phoneme discrimination reveal the origin of individual differences in learning the sounds of a second language. *Proceedings of the National Academy of Sciences*, -.

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

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research, 17*(3), 637-650.

Dinstein, I., Gardner, J. L., Jazayeri, M., & Heeger, D. J. (2008). Executed and observed movements have different distributed representations in human aips. *Journal of Neuroscience, 28*(44), 11231-11239.

Donchin, E. (1981). Surprise!...Surprise? *Psychophysiology, 18*(5), 493-513.

Donchin, E. (2006). fMRI:Not the only way to look at the human brain. *Observer, 19*(11).

Donchin, E., Tueting, P., Ritter,.W., Kutas, M., & Heffley, E. (1975). Independence of CNV and P300 components of human averaged evoked-potential. *Electroencephalography and Clinical Neurophysiology, 38*(5), 449-461.

Dooling, R. J., Okanoya, K., & Brown, S. D. (1989). Speech perception by budgerigars (melopsittacus undulatus): The voiced-voiceless distinction. *Perception & Psychophysics, 46*(1), 65-71.

Drivonikou, G. V., Kay, P., Regier, T., Ivry, R. B., Gilbert, A. L., Franklin, A., et al. (2007). Further evidence that whorfian effects are stronger in the right visual field than the left. *Proceedings of the National Academy of Sciences of the United States of America, 104*(3), 1097-1102.

Ehret, G. (1987). Categorical perception of sound signals: Facts and hypotheses from animal studies. . In S. Harnad (Ed.), *Categorical perception: The groundwork of cognition*. New York, NY, US: Cambridge University Press.

Ehret, G. (1992). Categorical perception of mouse-pup ultrasounds in the temporal domain. *Animal Behaviour, 43*(3), 409-416.

Eimas, P. D., Cooper, W. E., & Corbit, J. D. (1973). Some properties of linguistic feature detectors. *Perception & Psychophysics, 13*(2), 247-252.

Eimas, P. D., Siquelan.Er, Jusczyk, P., & Vigorito, J. (1971). Speech perception in infants. *Science, 171*(3968), 303-306.

Electrical Geodesics, I. (2006). Net station viewer and waveform tools tutorial s-man-200-tvwr-001.

Eng, J. (2003). Sample size estimation: How many individuals should be studied? *Radiology 227*, 309-313.

Escera, C., Alho, K., Schröger, E., & Winkler, I. (2000). Involuntary attention and distractibility as evaluated with event-related brain potentials. *Audiology & Neuro - Otology, 5*(3-4), 151.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15*(2), 399-402.

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology, 73*(6), 2608-2611.

Farwell, L. A., & Donchin, E. (1988). Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials. *Electroenceph.Clin. Neurophysiol, 70*, 510-523.

Finney, D. J. (1971). *Probit analysis* (Third ed.). New York: Cambridge Univ. Press Cambridge.

Fischer, C., Morlet, D., & Giard, M. H. (2000). Mismatch negativity and n100 in comatose patients. *Audiology and Neurotology, 5*(3-4), 192-197.

Fitch, H., Halwes, T., Erickson, D., & Liberman, A. (1980). Perceptual equivalent of two acoustic cues for stop-consonant manner. *perception & Psychophysics, 27*(4), 343-350.

Fowler, C. A., & Rosenblum, L. D. (1990). Duplex perception: A comparison of monosyllables and slamming doors. *Journal of experimental Psychology: Human Perception and Performance, 16*(4), 742-754.

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics, 65*(7), 1029-1044.

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America, 119*(3), 1712-1726.

Francis, A. L., Valter, C., & Lian, M. (2004). Effects of native language experience on perceptual learning of Cantonese lexical tones. *The Journal of the Acoustical Society of America, 115*(5), 2544.

Franklin, A., Drivonikou, G. V., Bevis, L., Davies, I. R. L., Kay, P., & Regier, T. (2008). Categorical perception of color is lateralized to the right hemisphere in infants, but to the left hemisphere in adults. *Proceedings of the National Academy of Sciences, 105*(9), 3221-3225.

Franklin, A., Drivonikou, G. V., Clifford, A., Kay, P., Regier, T., & Davies, I. R. L. (2008). Lateralization of categorical perception of color changes with color term acquisition. *Proceedings of the National Academy of Sciences. 105*(47), 18221-18225.

Frenck-Mestre, C., Meunier, C., Espesser, R., Daffner, K., & Holcomb, P. (2005). Perceiving nonnative vowels: The effect of context on perception as evidenced by event-related brain potentials. *J Speech Lang Hear Res, 48*(6), 1496-1510.

Friederici, A. D. (2009). Pathways to language: Fiber tracts in the human brain. *Trends in Cognitive Sciences, In Press, Corrected Proof.*

Friederici, A. D., & Alter, K. (2004). Lateralization of auditory language functions: A dynamic dual pathway model. *Brain and Language, 89*(2), 267-276.

Frisch, S. A., Pierrehumbert, J. B., & Broe, M. B. (2004). Similarity avoidance and the OCP. *Natural Language & Linguistic Theory, 22*(1), 179-228.

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

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). Identification and discrimination of synthetic vowels. *Language and Speech, 5*, 171-189.

Fujisaki, H., & Kawashima, T. (1969). *On the modes and mechanisms of speech perception.* Paper presented at the Annual Report of the Engineering Research Institute.

Fujisaki, H., & Kawashima, T. (1971). *A model of the mechanisms for speech perception: Quantitative analyses of categorical effects in discrimination.* Paper presented at the Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo.

Galantucci, B., Fowler, C. A., & Turvey, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review, 13*(3), 361-377.

Gandour, J. T. (1983). Tone perception in far eastern-languages. *Journal of Phonetics, 11*(2), 149-175.

Gandour, J. T., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., et al. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language, 84*(3), 318-336.

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

Gandour, J. T., Wong, D., Lowe, M., Dzemidzic, M., Satthamnuwong, N., Tong, Y. X., et al. (2002). A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *Journal of Cognitive Neuroscience, 14*(7), 1076-1087.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology, 120*(3), 453-463.

Gilbert, A. L. (2007). *Lateralized effects of language on perception.* University of California, Berkeley.

Gilbert, A. L., Regier, T., Kay, P., & Ivry, R. B. (2006). Whorf hypothesis is supported in the right visual field but not the left. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 489-494.

Goodin, D. S., Waltz, D. A., & Aminoff, M. J. (1985). Task-dependent hemisphere asymmetries of the visual evoked potential. *Neurology, 35*(3), 378-.

Hagoort, P., & Levelt, W. J. M. (2009). The speaking brain. *Science, 326*(5951), 372-373.

Halgren, E., Baudena, P., Clarke, J. M., Heit, G., Marinkovic, K., Devaux, B., et al. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. II. Medial, lateral and posterior temporal lobe. *Electroencephalography and Clinical Neurophysiology, 94*(4), 229-250.

Hallè, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Harnad, S. (Ed.). (1987). *Categorical perception: The groundwork of cognition*.

He, B., Lian, J., Spencer, K. M., Dien, J., & Donchin, E. (2001). A cortical potential imaging analysis of the P300 and novelty p3 components. *Human Brain Mapping, 12*(2), 120-130.

Hickok, G., Holt, L. L., & Lotto, A. J. (2009). Response to wilson: What does motor cortex contribute to speech perception? *Trends in Cognitive Sciences, 13*(8), 330-331.

Hillyard, S. A., & Kutas, M. (1983). Electrophysiology of cognitive processing. *Annual Review of Psychology, 34*, 33.

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science, 16*(4), 305-312.

Holt, L. L. (2006). Speech categorization in context: Joint effects of nonspeech and speech precursors. *The Journal of the Acoustical Society of America, 119*(6), 4016-4026.

Holt, L. L., Lotto, A. J., & Diehl, R. L. (2004). Auditory discontinuities interact with categorization: Implications for speech perception. *The Journal of the Acoustical Society of America, 116*(3), 1763-1773.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America, 109*(2), 764-774.

Horev, N., Most, T., & Pratt, H. (2007). Categorical perception of speech (VOT) and analogous non-speech (FOT) signals: Behavioral and electrophysiological correlates. *Ear and Hearing, 28*(1), 111-128.

Hot, P., Saito, Y., Mandai, O., Kobayashi, T., & Sequeira, H. (2006). An ERP investigation of emotional processing in european and Japanese individuals. *Brain Research, 1122*(1), 171-178.

Howie, J. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, U.K.:: Cambridge Univ Press.

Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (1999). Cortical mechanisms of human imitation. *Science, 286*(5449), 2526-2528.

Ikeda, K., Hayashi, A., Hashimoto, S., Otomo, K., & Kanno, A. (2001). Asymmetrical mismatch negativity in humans as determined by phonetic but not physical difference. *Neuroscience Letters, 321*(3), 133-136.

Ila, A. B., & Polich, J. (1999). P300 and response time from a manual stroop task. *Clinical Neurophysiology, 110*(2), 367-373.

Itzchak, E. B., Babkoff, H., & Faust, M. (2007). Event related potentials (ERP) and behavioral measurements to verbal stimulation of visual fields. *Cortex, 43*(4), 511-523.

Jacobsen, T., Schröger, E., & Alter, K. (2004). Pre-attentive perception of vowel phonemes from variable speech stimuli. *Psychophysiology, 41*(4), 654-659.

Johnson, K., & Ralston, J. V. (1994). Automaticity in speech-perception - some speech/nonspeech comparisons. *Phonetica, 51*(4), 195-209.

Jusczyk, P. W., Rosner, B. S., Cutting, J. E., Foard, C. F., & Smith, L. B. (1977). Categorical perception of non-speech sounds by 2-month-old infants. *Perception & Psychophysics, 21*(1), 50-54.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., et al. (2001). Brain lateralization for mismatch response to across- and within-category change of vowels. *Neuroreport, 12*(11), 2467-2471.

Kasai, K., Yamada, H., Kamio, S., Nakagome, K., Iwanami, A., Fukuda, M., et al. (2002). Neuromagnetic correlates of impaired automatic categorical perception of speech sounds in schizophrenia. *Schizophrenia Research, 59*(2-3), 159-172.

Kazanina, N., Phillips, C., & Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences of the United States of America, 103*(30), 11381-11386.

Keating, P. A. (2006). UCLA phonetics lab statistics, 2009, from http://www.linguistics.ucla.edu/faciliti/facilities/statistics/statistics.html

Keating, P. A., Mikoś, M. J., & Ganong III, W. F. (1981). A cross-language study of range of voice onset time in the perception of initial stop voicing. *The Journal of the Acoustical Society of America, 70*(5), 1261-1271.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, french, and English. *Journal of Phonetics, 25*(2), 143-168.

Klein, D., Zatorre, R. J., Milner, B., & Zhao, V. (2001). A cross-linguistic PET study of tone perception in mandarin Chinese and English speakers. *Neuroimage, 13*(4), 646-653.

Kluender, K. R., Diehl, R. L., & Killeen, P. R. (1987). Japanese-quail can learn phonetic categories. *Science, 237*(4819), 1195-1197.

Kluender, K. R., & Lotto, A. J. (1994). Effects of first formant onset frequency on [-voice] judgments result from auditory processes not specific to humans. *The Journal of the Acoustical Society of America, 95*(2), 1044-1052.

Kohler, E., Keysers, C., Umilta, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science, 297*(5582), 846-848.

Korpilahti, P., Krause, C. M., Holopainen, I., & Lang, A. H. (2001). Early and late mismatch negativity elicited by words and speech-like stimuli in children. *Brain and Language, 76*(3), 332-339.

Krishnan, A., Gandour, J. T., Bidelman, G. M., & Swaminathan, J. (2009). Experience-dependent neural representation of dynamic pitch in the brainstem. *Neuroreport, 20*(4), 408-413.

Krishnan, A., Xu, Y. S., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research, 25*(1), 161-168.

Kuhl, P. K. (1986). Theoretical contributions of tests on animals to the special-mechanisms debate in speech. *Experimental Biology, 45*(3), 233-265.

Kuhl, P. K. (2004). Early language acquisition: Cracking the speech code. *Nature Reviews Neuroscience, 5*(11), 831-843.

Kuhl, P. K., & Miller, J. D. (1975). Speech perception by the chinchilla - voiced-voiceless distinction in alveolar plosive consonants. *Science, 190*(4209), 69-72.

Kuhl, P. K., & Padden, D. M. (1983). Enhanced discriminability at the phonetic boundaries for the place feature in macaques. *The Journal of the Acoustical Society of America, 73*(3), 1003-1010.

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science, 9*(2), F13-F21.

Kujala, T., & Näätänen, R. (2001). The mismatch negativity in evaluating central auditory dysfunction in dyslexia. *Neuroscience and Biobehavioral Reviews, 25*(6), 535-543.

Kujala, T., Tervaniemi, M., & Schröger, E. (2007). The mismatch negativity in cognitive and clinical neuroscience: Theoretical and methodological considerations. *Biological Psychology, 74*(1), 1-19.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences - brain potentials reflect semantic incongruity. *Science, 207*(4427), 203-205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*(5947), 161-163.

Kutas, M., McCarthy, G., & Donchin, E. (1977). Augmenting mental chronometry: The P300 as a measure of stimulus evaluation time. *Science, 197*(4305), 792-795.

Kutas, M., & Van Petten, C. K. (1994). Psycholinguistics electrified: Event-related brain potential investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). New-York: Academic Press,.

Lane, H. (1965). The motor theory of speech perception: A critical review. *Psychological Review, 72*(4), 275-309.

Larkey, L., Wald, J., & Strange, W. (1978). Perception of synthetic nasal consonants in initial and final syllable position. *perception & Psychophysics, 23*(4), 299-312.

Lasky, R. E., Syrdallasky, A., & Klein, R. E. (1975). VOT discrimination by 4 to 6 1/2 month-old infants from spanish environments. *Journal of Experimental Child Psychology, 20*(2), 215-225.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (de)constructing the n400. *Nat Rev Neurosci, 9*(12), 920-933.

Lee, L., & Nusbaum, H. C. (1993). Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics, 53*, 157 - 165.

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

Lemmetty, S. (1999). *Review of speech synthesis technology*. Helsinki University of Technology.

Lewis, C. M., Baldassarre, A., Committeri, G., Romani, G. L., & Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences, 106*(41), 17558-17563.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461.

Liberman, A. M., Harris, K. S., Eimas, P. D., Lisker, L., & Bastian, J. (1961). An effect of learning on speech perception: The discrimination of durations of silence with and without phonemic significance. *Language and Speech, 4*, 175-195.

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*(5), 358-368.

Liberman, A. M., Harris, K. S., Kinney, J. A., & Lane, H. (1961). The discrimination of relative onset time of the components of certain speech and nonspeech patterns. *Journal of Experimental Psychology, 61*, 379-388.

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

Liberman, A. M., & Mattingly, I. G. (1989). A specialization for speech-perception. *Science, 243*(4890), 489-494.

Lin, T., & Wang, W. S. Y. (1984). An experiment in tone perception (trans.) (聲調感知問題). *Journal of Chinese Linguistics (中國語言學報), 2*, 59-69.

Linden, D. E. J. (2005). The P300: Where in the brain is it produced and what does it tell us? *Neuroscientist, 11*(6), 563-576.

Linden, D. E. J., Prvulovic, D., Formisano, E., Vollinger, M., Zanella, F. E., Goebel, R., et al. (1999). The functional neuroanatomy of target detection: An fMRI study of visual and auditory oddball tasks. *Cereb. Cortex, 9*(8), 815-823.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustic measurements. *Word, 20*, 384-422.

Liu, L., Peng, D., Ding, G., Jin, Z., Zhang, L., Li, K., et al. (2006). Dissociation in the neural basis underlying Chinese tone and vowel production. *NeuroImage, 29*(2), 515-523.

Liu, S., & Samuel, A. G. (2004). Perception of Mandarin lexical tones when f0 information is neutralized. *Language and Speech, 47*(2), 109-138.

Lotto, A. J., Hickok, G. S., & Holt, L. L. (2009). Reflections on mirror neurons and speech perception. *Trends in Cognitive Sciences, 13*(3), 110-114.

Lü, Z.-L., Williamson, S. J., & Kaufman, L. (1992). Human auditory primary and association cortex have differing lifetimes for activation traces. *Brain Research, 572*(1-2), 236-241.

Luck, S. J. (2005). *An introduction to the event-related potential technique*. Cambridge, Massachusetts

London, England: The MIT press.

Luo, H., Husain, F. T., Horwitz, B., & Poeppel, D. (2005). Discrimination and categorization of speech and non-speech sounds in an MEG delayed-match-to-sample study. *Neuroimage, 28*(1), 59-71.

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

Maess, B., Jacobsen, T., Schröger, E., & Friederici, A. D. (2007). Localizing pre-attentive auditory memory-based comparison: Magnetic mismatch negativity to pitch change. *Neuroimage, 37*(2), 561-571.

Magen, H. S. (1997). The extent of vowel-to-vowel coarticulation in English. *Journal of Phonetics, 25*(2), 187-205.

Magliero, A., Bashore, T. R., Coles, M. G. H., & Donchin, E. (1984). On the dependence of P300 latency on stimulus evaluation processes. *Psychophysiology, 21*(2), 171-186.

Maiste, A. C., Wiens, A. S., Hunt, M. J., Scherg, M., & Picton, T. W. (1995). Event-related potentials and the categorical perception of speech sounds. *Ear and Hearing, 16*(1), 68-90.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *The Journal of the Acoustical Society of America, 69*(2), 548-558.

Mann, V. A., & Soli, S. D. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics, 49*(5), 399-411.

Mattock, K., & Burnham, D. (2006). Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy, 10*(3), 241-265.

Mattock, K., Molnar, M., Polka, L., & Burnham, D. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition, 106*(3), 1367-1381.

Max, L., & Onghena, P. (1999). Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research, 42*(2), 261-270.

May, B., Moody, D. B., & Stebbins, W. C. (1989). Categorical perception of conspecific communication sounds by Japanese macaques, macaca fuscata. *The Journal of the Acoustical Society of America, 85*(2), 837-847.

McCarthy, G., & Donchin, E. (1981). A metric for thought: A comparison of P300 latency and reaction time. *Science, 211*(4477), 77-80.

McCarthy, G., Luby, M., Gore, J., & Goldman-Rakic, P. (1997). Infrequent events transiently activate human prefrontal and parietal cortex as measured by functional mri. *J Neurophysiol, 77*(3), 1630-1634.

McCarthy, G., Wood, C., Williamson, P., & Spencer, D. (1989). Task-dependent field potentials in human hippocampal formation. *J. Neurosci., 9*(12), 4253-4268.

McGarry-Roberts, P. A., Stelmack, R. M., & Campbell, K. B. (1992). Intelligence, reaction time, and event-related potentials. *Intelligence, 16*(3), 289-313.

Miller, J. L., & Eimas, P. D. (1977). Studies on the perception of place and manner of articulation: A comparison of the labial-alveolar and nasal-stop distinctions. *The Journal of the Acoustical Society of America, 61*(3), 835-845.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *perception & Psychophysics, 25*(6), 457-465.

Mirman, D., Holt, L. L., & McClelland, J. L. (2004). Categorization and discrimination of nonspeech sounds: Differences between steady-state and rapidly-changing acoustic cues. *The Journal of the Acoustical Society of America, 116*(2), 1198-1207.

Molfese, D. L., Laughlin, N. K., Morse, P. A., Linnville, S. E., Wetzel, W. F., & Erwin, R. J. (1986). Neuroelectrical correlates of categorical perception for place of articulation in normal and lead-treated rhesus monkeys. *Journal of Clinical and Experimental Neuropsychology, 8*(6), 680 - 696.

Morse, P., & Snowdon, C. (1975). An investigation of categorical speech discrimination by rhesus monkeys. *Perception and Psychophysics, 17*(1), 9-16.

Näätänen, R. (2000). Mismatch negativity (MMN): An objective measure of central auditory processing deficits. *International Journal of Psychophysiology, 35*(1), 27-27.

Näätänen, R. (2002). Special didactic 1: Mismatch negativity (MMN): Clinical applications. *International Journal of Psychophysiology, 45*(1-2), 9-9.

Näätänen, R. (2003). Mismatch negativity: Clinical research and possible applications. *International Journal of Psychophysiology, 48*(2), 179-188.

Näätänen, R., & Alho, K. (1997). Mismatch negativity - the measure for central sound representation accuracy. *Audiology & Neuro - Otology, 2*(5), 341-353.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): A review of the evidence. *Psychophysiology, 42*(1), 25-32.

Näätänen, R., & Kähkönen, S. (2008). Central auditory dysfunction in schizophrenia as revealed by the mismatch negativity (MMN) and its magnetic equivalent MMNm: A review. *The International Journal of Neuropsychopharmacology, Forthcoming*(1), 1-11.

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*(6615), 432-434.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: A review. *Clinical Neurophysiology, 118*(12), 2544-2590.

Näätänen, R., & Picton, T. W. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology, 24*(4), 375-425.

Nan, Y., Knösche, T. R., & Friederici, A. D. (2006). The perception of musical phrase structure: A cross-cultural ERP study. *Brain Research, 1094*(1), 179-191.

Nan, Y., Knösche, T. R., & Friederici, A. D. (2009). Non-musicians' perception of phrase boundaries in music: A cross-cultural ERP study. *Biological Psychology, 82*(1), 70-81.

Nelson, D., & Marler, P. (1989). Categorical perception of a natural stimulus continuum: Birdsong. *Science, 244*(4907), 976-978.

Nenonen, S., Shestakova, A., Huotilainen, M., & Näätänen, R. (2005). Speech-sound duration processing in a second language is specific to phonetic categories. *Brain and Language, 92*(1), 26-32.

Nittrouer, S., Miller, M., Crowther, C., & Manhart, M. (2000). The effect of segmental order on fricative labeling by children and adults. *perception & Psychophysics, 62*(2), 266-284.

Ohde, R. N., & Sharf, D. J. (1977). Order effect of acoustic segments of vc and cv syllables on stop and vowel identification. *Journal of Speech and Hearing Research, 20*(3), 543-554.

Ohlemiller, K. K., Jones, L. B., Heidbreder, A. F., Clark, W. W., & Miller, J. D. (1999). Voicing judgements by chinchillas trained with a reward paradigm. *Behavioural Brain Research, 100*(1-2), 185-195.

Oldfield, R. C. (1971). Assessment and analysis of handedness - edinburgh inventory. *Neuropsychologia, 9*(1), 97-113.

Pantev, C., Hoke, M., Lütkenh ner, B., & Lehnertz, K. (1989). Tonotopic organization of the auditory cortex: Pitch versus frequency representation. *Science, 246*(4929), 486-488.

Pause, B. M., & Krauel, K. (2000). Chemosensory event-related potentials (cserp) as a key to the psychology of odors. *International Journal of Psychophysiology, 36*(2), 105-122.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: A review. *Biological Psychology, 63*(3), 199-236.

Pegg, J. E., & Werker, J. F. (1997). Adult and infant perception of two English phones. *Journal of the Acoustical Society of America, 102*(6), 3742-3753.

Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*(1), 176-180.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics, 34*(1), 134-154.

Phillips, C. (2001). Levels of representation in the electrophysiology of speech perception. *Cognitive Science, 25*(5), 711-731.

Phillips, C., Pellathy, T., Marantz, A., Yellin, E., Wexler, K., Poeppel, D., et al. (2000). Auditory cortex accesses phonological categories: An MEG mismatch study. *Journal of Cognitive Neuroscience, 12*(6), 1038-1055.

Picton, T. W. (1992). The P300 wave of the human event-related potential. *Journal of Clinical Neurophysiology, 9*(4), 456-479.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253-260.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *The Journal of the Acoustical Society of America, 61*(5), 1352-1361.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H. A., Roberts, T. P. L., & Marantz, A. (1997). Processing of vowels in supratemporal auditory cortex. *Neuroscience Letters, 221*(2-3), 145-148.

Polich, J. (2003a). *Detection of change : Event-related potential and fMRI findings*. Boston: Kluwer Academic Publishers.

Polich, J. (2003b). Theoretical overview of p3a and p3b. In J. Polich (Ed.), *Detection of change: Event-related potential and fMRI findings* (pp. 83-98): Springer.

Polich, J. (2004). Clinical application of the P300 event-related brain potential. *Physical Medicine and Rehabilitation Clinics of North America, 15*(1), 133-161.

Polich, J. (2007). Updating P300: An integrative theory of p3a and p3b. *Clinical Neurophysiology, 118*(10), 2128-2148.

Polich, J., & Hoffman, L. D. (1998). P300 and handedness: On the possible contribution of corpus callosal size to ERPs. *Psychophysiology, 35*(05), 497-507.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: An integrative review. *Biological Psychology, 41*(2), 103-146.

Prather, J. F., Nowicki, S., Anderson, R. C., Peters, S., & Mooney, R. (2009). Neural correlates of categorical perception in learned vocal communication. *Nature Neuroscience, 12*(2), 221-228.

Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., & Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds. *Proceedings of the National Academy of Sciences, 103*(20), 7865-7870.

Pulvermüller, F., & Shtyrov, Y. (2006). Language outside the focus of attention: The mismatch negativity. As a tool for studying higher cognitive processes. *Progress in Neurobiology, 79*(1), 49-71.

Pulvermüller, F., Shtyrov, Y., Hasting, A. S., & Carlyon, R. P. (2008). Syntax as a reflex: Neurophysiological evidence for early automaticity of grammatical processing. *Brain and Language, 104*(3), 244-253.

Pulvermüller, F., Shtyrov, Y., & Hauk, O. (2009). Understanding in an instant: Neurophysiological evidence for mechanistic language circuits in the brain. *Brain and Language, 110*(2), 81-94.

Pulvermüller, F., Shtyrov, Y., & Ilmoniemi, R. (2005). Brain signatures of meaning access in action word recognition. *Journal of Cognitive Neuroscience, 17*(6), 884-892.

Pulvermüller, F., Shtyrov, Y., Ilmoniemi, R. J., & Marslen-Wilson, W. D. (2006). Tracking speech comprehension in space and time. *Neuroimage, 31*(3), 1297-1305.

Ravizza, S. M. (2003). Dissociating the performance of cortical and subcortical patients on phonemic tasks. *Brain and Cognition, 53*(2), 301-310.

Ravizza, S. M., Henri, C., & Claire, L. (2005). Neural regions associated with categorical speech perception and production. In H. Cohen & C. Lefebvre (Eds.), *Handbook of categorization in cognitive science* (pp. 601-615). Oxford: Elsevier Science Ltd.

Regier, T., & Kay, P. (2009). Language, thought, and color: Whorf was half right. *Trends in Cognitive Sciences, 13*(10), 439-446.

Repp, B. H. (1984). Categorical perception: Issues, methods, findings. *Speech and Language: Advances in basic research and practice, 10*, 243-335.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance, 4*(4), 621-637.

Rivera-Gaxiola, M., Csibra, G., Johnson, M. H., & Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research, 111*(1-2), 13-23.

Rivera-Gaxiola, M., Silva-Pereyra, J., & Kuhl, P. K. (2005). Brain potentials to native and non-native speech contrasts in 7-and 11-month-old american infants. *Developmental Science, 8*(2), 162-172.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*(1), 169-192.

Roberts, M., & Summerfield, Q. (1981). Audiovisual presentation demonstrates that selective adaptation in speech perception is purely auditory. *perception & Psychophysics, 30*(4), 309-314.

Robertson, L. C., & Ivry, R. (2000). Hemispheric asymmetries: Attention to visual and auditory primitives. *Current Directions in Psychological Science, 9*(2), 59-63.

Sahin, N. T., Pinker, S., Cash, S. S., Schomer, D., & Halgren, E. (2009). Sequential processing of lexical, grammatical, and phonological information within broca's area. *Science, 326*(5951), 445-449.

Scerif, G., Worden, M. S., Davidson, M., Seiger, L., & Casey, B. J. (2006). Context modulates early stimulus processing when resolving stimulus-response conflict. *Journal of Cognitive Neuroscience, 18*(5), 781-792.

Schirmer, A., Simpson, E., & Escoffier, N. (2007). Listen up! Processing of intensity change differs for vocal and nonvocal sounds. *Brain Research, 1176*, 103-112.

Schulte-Körne, G., Deimel, W., Bartling, J., & Remschmidt, H. (2001). Speech perception deficit in dyslexic adults as measured by mismatch negativity (MMN). *International Journal of Psychophysiology, 40*(1), 77-87.

Scott, S. K., McGettigan, C., & Eisner, F. (2009). A little more conversation, a little less action — candidate roles for the motor cortex in speech perception. *Nature Review Neuroscience, 10*(4), 295-302.

Sellers, E. W., Kübler, A., & Donchin, E. (2006). Brain–computer interface research at the university of south florida cognitive psychophysiology laboratory: The P300 speller. *IEEE TRANSACTIONS ON NEURAL SYSTEMS AND REHABILITATION ENGINEERING*

Serniclaes, W., Sprenger-Charolles, L., Carre, R., & Demonet, J.-F. (2001). Perceptual discrimination of speech sounds in developmental dyslexia. *Journal of Speech Language Hearing Research, 44*(2), 384-399.

Serniclaes, W., Ventura, P., Morais, J., & Kolinsky, R. (2005). Categorical perception of speech sounds in illiterate adults. *Cognition, 98*(2), B35-B44.

Sharma, A., & Dorman, M. F. (1999). Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *Journal of the Acoustical Society of America, 106*(2), 1078-1083.

Shtyrov, Y., Pulvermüller, F., Naatanen, R., & Ilmoniemi, R. J. (2003). Grammar processing outside the focus of attention: An MEG study. *Journal of Cognitive Neuroscience, 15*(8), 1195-1206.

Shuai, L. (2009). *ERP studies of tone lateralization*. The Chinese University of Hong Kong, Hong Kong.

Simos, P. G., Diehl, R. L., Breier, J. I., Molis, M. R., Zouridakis, G., & Papanicolaou, A. C. (1998). MEG correlates of categorical perception of a voice onset time continuum in humans. *Cognitive Brain Research, 7*(2), 215-219.

Siok, W. T., Kay, P., Wang, W. S. Y., Chan, A. H. D., Chen, L., Luk, K.-K., et al. (2009). Language regions of brain are operative in color perception. *Proceedings of National Academy of Sciences of the United States of America, in press.*

Sittiprapaporn, W., Tervaniemi, M., Chindaduangratn, C., & Kotchabhakdi, N. (2005). Preattentive discrimination of across-category and within-category change in consonant-vowel syllable. *Neuroreport, 16*(13), 1513-1518.

Smith, M., Halgren, E., Sokolik, M., Baudena, P., Musolino, A., Liegeois-Chauvel, C., et al. (1990). The intracranial topography of the p3 event-related potential elicited during auditory oddball. *Electroencephalography and Clinical Neurophysiology, 76*(3), 235.

Snowdon, C. T. (1979). Response of nonhuman animals to speech and to species-specific sounds. *Brain, Behavior and Evolution, 16*(5-6), 409-429.

Spencer, K. M., Dien, J., & Donchin, E. (1999). A componential analysis of the ERP elicited by novel events using a dense electrode array. *Psychophysiology, 36*(03), 409-414.

Steinschneider, M., Volkov, I. O., Fishman, Y. I., Oya, H., Arezzo, J. C., & Howard, M. A. (2005). Intracortical responses in human and monkey primary auditory cortex support a temporal processing mechanism for encoding of the voice onset time phonetic parameter. *Cerebral Cortex, 15*(2), 170-186.

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to lane's critical review. *Psychological Review, 77*, 243-249.

Sussman, E., Kujala, T., Halmetoja, J., Lyytinen, H., Alku, P., & Näätänen, R. (2004). Automatic and controlled processing of acoustic and phonetic contrasts. *Hearing Research, 190*(1-2), 128-140.

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science, 150*(3700), 1187-&.

Szűcs, D., Soltész, F., & White, S. (2009). Motor conflict in stroop tasks: Direct evidence from single-trial electro-myography and electro-encephalography. *Neuroimage, 47*(4), 1960-1973.

Tampas, J. W., Harkrider, A. W., & Hedrick, M. S. (2005). Neurophysiological indices of speech and nonspeech stimulus processing. *Journal of Speech, Language and Hearing Research, 48*(5), 1147-1164.

Tervaniemi, M., & Hugdahl, K. (2003). Lateralization of auditory-cortex functions. *Brain Research Reviews, 43*(3), 231-246.

Tervaniemi, M., Ilvonen, T., Sinkkonen, J., Kujala, A., Alho, K., Huotilainen, M., et al. (2000). Harmonic partials facilitate pitch discrimination in humans: Electrophysiological and behavioral evidence. *Neuroscience Letters, 279*(1), 29-32.

Tervaniemi, M., Saarinen, J., Paavilainen, P., Danilova, N., & Naatanen, R. (1994). Temporal integration of auditory information in sensory memory as reflected by the mismatch negativity. *Biological Psychology, 38*(2-3), 157-167.

Thompson, W. F., Schellenberg, E. G., & Husain, G. (2003). Perceiving prosody in speech. Effects of music lessons. *Annals of the New York Academy of Sciences, 999*(1), 530-532.

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

Tremblay, K., Kraus, N., McGee, T., Ponton, C., Otis, & Brian (2001). Central auditory plasticity: Changes in the N1-p2 complex after speech-sound training. *Ear and Hearing, 22*(2), 79-90.

Tucker, D. M. (1993). Spatial sampling of head electrical fields: The geodesic sensor net. *Electroencephalography and Clinical Neurophysiology, 87*(3), 154-163.

van Son, R. J. J. H., & Pols, L. C. W. (1999). Perisegmental speech improves consonant and vowel identification. *Speech Communication, 29*(1), 1-22.

van Zuijen, T. L., Simoens, V. L., Paavilainen, P., Naatanen, R., & Tervaniemi, M. (2006). Implicit, intuitive, and explicit knowledge of abstract regularities in a sound sequence: An event-related brain potential study. *Journal of Cognitive Neuroscience, 18*(8), 1292-1303.

Verleger, R. (1997). On the utility of p3 latency as an index of mental chronometry. *Psychophysiology, 34*(2), 131-156.

Walhovd, K. B. C. A., & Fjell, A. M. (2001). Two- and three-stimuli auditory oddball ERP tasks and neuropsychological measures in aging. *Neuroreport, 12*(14), 3149-3153.

Wang, W. S. Y. (1967). Phonological features of tone. *International Journal of American Linguistics, 33*(2), 93-105.

Wang, W. S. Y. (1973). Chinese language. *Scientific American, 228*(2), 50-60.

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

Wang, W. S. Y., & Peng, G. (2006). *Language, speech and technology (trans.) (语言、语音与技术)*. Shanghai (上海）: Shanghai educational publishing house （上海教育出版社）.

Watkins, K. E., & Paus, T. (2004). Modulation of motor excitability during speech perception: The role of broca's area. *Journal of Cognitive Neuroscience, 16*(6), 978-987.

Watkins, K. E., Strafella, A. P., & Paus, T. (2003). Seeing and hearing speech excites the motor system involved in speech production. *Neuropsychologia, 41*(8), 989-994.

Welberg, L. (2008). Mirror neurons: Towards a clearer image. *Nature Review Neuroscience, 9*(12), 888-889.

Welberg, L. (2009). Neuroimaging: Learning changes the resting brain. *Nature Review Neuroscience, 10*(11), 766-767.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44.

Werker, J. F., & Tees, R. C. (1992). The organization and reorganization of human speech-perception. *Annual Review of Neuroscience, 15*, 377-402.

Werker, J. F., & Tees, R. C. (2002). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior & Development, 25*(1), 121-133.

Wernicke, C. (1874,1969). The symptom complex of aphasia: A psychological study on an anatomical basis. In R. S. C. a. M. W. Wartofsky (Ed.), *Boston studies in the philosophy of science* (pp. 34-97). Dordrecht: D. Reidel Publishing Company.

Whalen, D. H., & Xu, Y. (1992). Information for Mandarin tones in the amplitude contour and in brief segments. *Phonetica, 49*(1), 25-47.

Wong, P. C. M., Skoe, E., Russo, N. M., Dees, T., & Kraus, N. (2007). Musical experience shapes human brainstem encoding of linguistic pitch patterns. *Nature Neuroscience, 10*(4), 420-422.

Wong, Y. W. (2006). *Contextual tonal variations and pitch targets in Cantonese.* Paper presented at the Proceedings of 3rd International Conference on Speech Prosody, Dresden, Germany.

Wyttenbach, R. A., May, M. L., & Hoy, R. R. (1996). Categorical perception of sound frequency by crickets. *Science, 273*(5281), 1542-1544.

Xi, J., Zhang, L., Xu, G., & Shu, H. (2009). *Brain response to categorical perception of lexical tone in Mandarin - an ERP study.* Paper presented at the The 13th International Conference on the Processing of East Asian Languages.

Xu, G., Xi, J., Zhang, L., & Shu, H. (2009). *Neural correlates of categorical perception of lexical tone in Mandarin - a fMRI study.* Paper presented at the The 13th International Conference on the Processing of East Asian Languages.

Xu, Y. (1994). Production and perception of coarticulated tones. *The Journal of the Acoustical Society of America, 95*(4), 2240-2253.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of Phonetics, 25*(1), 61-83.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication, 33*(4), 319-337.

Xu, Y. S., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y. X., et al. (2006). Activation of the left planum temporale in pitch processing is shaped by language experience. *Human Brain Mapping, 27*(2), 173-183.

Xu, Y. S., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063-1074.

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 55-89.

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

Zatorre, R. J. (1988). Pitch perception of complex tones and human temporal-lobe function. *The Journal of the Acoustical Society of America, 84*(2), 566-572.

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences, 6*(1), 37-46.

Zatorre, R. J., Evans, A. C., Meyer, E., & Gjedde, A. (1992). Lateralization of phonetic and pitch discrimination in speech processing. *Science, 256*(5058), 846-849.

Zatorre, R. J., & Gandour, J. T. (2008). Neural specializations for speech and pitch: Moving beyond the dichotomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 363*(1493), 1087-1104.

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *Neuroimage, 26*(3), 703-720.

Zhu, W.-N., Zhang, J.-J., Liu, H.-W., Ding, X.-J., Ma, Y.-Y., & Zhou, C.-L. (2008). Differential cognitive responses to guqin music and piano music in Chinese subjects: An event-related potential study. *Neuroscience Bulletin, 24*(1), 21-28.

Zion-Golumbic, E., Deouell, L. Y., Whalen, D. H., & Bentin, S. (2007). Representation of harmonic frequencies in auditory memory: A mismatch negativity study. *Psychophysiology, 44*(5), 671-679.

APPENDIX 1 ABBREVIATIONS

Abbreviation	Full name
AC	*across-category*
ANOVA	*analysis of variance*
APEC	*asymmetric position effect of context*
CP	*categorical perception*
CPT	*complex tone continuum*
EEG	*electroencephalography*
ERP	*event-related-potential*
F0	*fundamental frequency*
F1	*first formant*
F2	*second formant*
fMRI	*functional magnetic resonance imaging*
INT	*interpolation continuum*
ISI	*inter-stimulus-interval*
ITI	*inter-trial-interval*
LC	*left context sentence*
MEG	*magnetoencephalography*
MEP	*motor evoked potential*
MMN	*mismatch negativity*
NWS	*no word continuum*
PET	*positron emission tomography*
POA	*place of articulatory*
RC	*right context sentence*
RM	*repeated measures*
RMANOVA	*repeated measures analysis of variance*
SOA	*stimulus onset asynchrony*
SSC	*simply shifted continuum*
TMS	*transcranial magnetic stimulation*
TOT	*tone-onset-time*
TS	*target syllable*
TWI	*time window integration*
VOT	*voice-onset-time*
WC	*within-category*

APPENDIX 2 SAMPLE SHEETS OF QUESTIONNAIRE (I) : CHAPTER 3 EXPERIMENT I

Name:________________ Sex:________________

廣東方言聲調感知測試

本測試的目的是考察廣東方言的第一聲，第三聲和第六聲三個調類在感知上的界限。所用的語料是以真人錄音爲基礎，部分特徵通過機器調節合成.

本測試包括兩部分,一部分是單字的測試,另一部份是短句測試. 要求受試者回答所聽到每個音節是哪個字,對於難以判斷的音節,請選擇你認爲較接近的字,答案沒有正誤之分,請根據您的實際感知作出選擇, 以下所有錄音請只聽一次.多謝您的合作!

第一部份 單字測試

以下六組聽辯語料,每組包括 5 個音節,每個音節播放兩次, 之間有短暫停頓, 然後播放下一個音節,不同音節之間有較長停頓. 其中每一個音節可能是”分”,”瞓”或者”份”中的一個,請在您選擇的答案上畫圈(三者必選其一).

例子,請聽錄音,該錄音含有三個音節(正式測試將含有 5 個音節),每個音節播放兩次,之間有短暫停頓, 下一個音節播放之前有較長停頓.黨聽到每個音節,就在對應答案上畫圈,如:

1. 瞓 份 [✓ 分]　2. [✓ 瞓] 份 分　3. 瞓 [✓ 份] 分

以下正式開始:

第一組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第二組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第三組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第四組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第五組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第六組:
1. 瞓 份 分　2. 瞓 份 分　3. 瞓 份 分　4. 瞓 份 分　5. 瞓 份 分

第二部分 短句測試

本部分包含兩套語料,第一套語料包含測試句子“呢個字係*”,其中*是選項中的一個字,受試者根據聽到的感覺選擇,兩個必選一個. 第二套語料測試句子是”*係乜意思”,同樣*也是選項中的一個字,受試者根據聽到的感覺選擇. 每組測試語料中包含 5 個這樣的句子,每個句子播放兩次,之間有 0.5 秒停頓, 不同句子之間間隔 4 秒, 每次的*可能代表不同的字.

第一套: 語料內容是"呢個字係*" ,請根據您的聽覺,在下類選項中選擇最接近您聽到的*

例子:本例中包含3個句子,每句內容都是"呢個字係*" ,請根據感覺選擇每次聽到的" *" , 請聽錄音
答案:1. 瞓 份 [✓ 分] 2. [✓ 瞓] 份 分 3. 瞓 [✓ 份] 分

以下正式開始:

第1組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第2組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第3組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第4組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第5組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第二套: 語料內容是" *係乜意思" ,請根據您的聽覺,在下類選項中選擇最接近您聽到的*.

例子:本例中包含3個句子,每句內容都是" *係乜意思" ,請根據感覺選擇每次聽到的" *" , 請聽錄音
答案: 1. 瞓 份 [✓ 分] 2. 瞓 [✓ 份] 分 3. [✓ 瞓] 份 分

以下正式開始:

第1組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第2組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第3組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第4組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

第5組:
1. 瞓 份 分 2. 瞓 份 分 3. 瞓 份 分 4. 瞓 份 分 5. 瞓 份 分

APPENDIX 2 SAMPLE SHEETS OF QUESTIONNAIRE (II) : CHAPTER 3 EXPERIMENT II

廣東方言聲調感知測試

本測試是考察粵語的第一聲，第三聲和第六聲三個調類在感知上的界限. 所用的語料是機器合成的。請直接雙擊喇叭符號 聽取錄音(因爲文件較大,可能需要等待片刻),下所有錄音請只聽一次. 多謝您的合作!

這部份您將聽到句子"頭先你聽*" ,其中*是三個選項中的一個字, 可能是" 詩歌"的"詩"; 或者是"嗜好"的"嗜"; 或者是"大事"的"事".請您根據聽到的感覺選擇, 三個必選一個. 在每組測試語料中,包含 13 個句子, 每次的*可能代表不同的字.
例子:假設您聽到 5 個句子, 根據感覺選擇每句內容中"*" 並在對應答案上畫圈如:

1. [✓ 詩]	2. 詩	3. 詩	4. 詩	5. 詩
嗜	嗜	[✓ 嗜]	嗜	[✓ 嗜]
事	[✓ 事]	事	[✓ 事]	事

注:答案也可以寫在另外一張獨立的答題紙上.

以下正式開始:
第一組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第二組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第三組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第四組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩

嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第五組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第六組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第七組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第八組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第九組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

這部份您將聽到句子" *聽得好好" ,同樣*也是選項中的一個字,受試者根據聽到的感覺選擇, 每組測試語料中,包含 13 個這樣的句子,句子之間間隔 1.5 秒, 每次的*可能代表不同的字.

以下正式開始:
第一組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第二組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第三組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第四組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第五組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第六組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第七組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第八組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

第九組:

1. 詩	2. 詩	3. 詩	4. 詩	5. 詩	6. 詩	7. 詩	8. 詩	9. 詩	10. 詩	11. 詩	12. 詩	13. 詩
嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜	嗜
事	事	事	事	事	事	事	事	事	事	事	事	事

APPENDIX 3 STATISTICAL ANALYSIS ON CHAPTER 3 EXPERIMENT III

Appendix Table 3.1. Statistical analysis on predicted discrimination scores on simple shifted (SSC) and interpolation (INT) continua

		F(1,7)	p
Three-way RMANOVA	CATEGORY (AC vs. WC)	22.252	0.002*
	CONTINUUM (SSC vs. INT)	1.827	0.219
	POSITION (LC vs. RC)	11.915	0.011*
	CATEGORY × CONTINUUM	0.875	0.381
	CATEGORY × POSITION	7.305	0.031*
	POSITION × CONTINUUM	3.157	0.119
	CATEGORY × CONTINUUM × POSITION	2.098	0.191
*: significant level $p<0.05$			
two-way RMANOVA on INT	CATEGORY (AC vs. WC)	16.805	0.005*
	POSITION (LC vs. RC)	17.955	0.004*
	CATEGORY × POSITION	26.773	0.001*
two-way RMANOVA on SSC	CATEGORY (AC vs. WC)	24.231	0.002*
	POSITION (LC vs. RC)	1.901	0.210
	CATEGORY × POSITION	0.597	0.465
*: significant level $p<0.05$			
		t(7)	p
paired t-test	SSC-LC-AC vs. SSC-RC-AC	-1.016	0.344
	INT-LC-AC vs. INT-RC-AC	-5.105	0.001*
	SSC-LC-WC vs. SSC-RC-WC	0.269	0.796
	INT-LC-WC vs. INT-RC-WC	2.966	0.021
	SSC-LC-AC vs. SSC-LC-WC	2.551	0.038
	SSC-RC-AC vs. SSC-RC-WC	5.050	0.001*
	INT-LC-AC vs. INT-LC-WC	2.455	0.044
	INT-RC-AC vs. INT-RC-WC	4.784	0.002*
*: significant level $p<0.0125$			

Appendix Table 3.2. Statistical analysis on obtained discrimination scores on simple shifted (SSC) and interpolation (INT) continua

		F(1,7)	p
Three-way RMANOVA	CATEGORY (AC vs. WC)	64.009	0.000*
	CONTINUUM (SSC vs. INT)	0.020	0.891
	POSITION (LC vs. RC)	30.389	0.001*
	CATEGORY × CONTINUUM	4.933	0.062
	CATEGORY × POSITION	16.699	0.005*
	POSITION × CONTINUUM	0.055	0.821
	CATEGORY × CONTINUUM × POSITION	1.241	0.302
	significant level $p<0.05$		
INT two-way RMANOVA	CATEGORY (AC vs. WC)	89.558	0.000*
	POSITION (LC vs. RC)	39.899	0.000*
	CATEGORY × POSITION	16.701	0.005*
SSC two-way RMANOVA	CATEGORY (AC vs. WC)	37.345	0.000*
	POSITION (LC vs. RC)	8.104	0.025*
	CATEGORY × POSITION	2.640	0.148
	significant level $p<0.05$		
		t(7)	p
paired t-test	SSC-LC-AC vs. SSC-RC-AC	-2.626	0.034
	INT-LC-AC vs. INT-RC-AC	-7.388	0.000*
	SSC-LC-WC vs. SSC-RC-WC	-2.668	0.032
	INT-LC-WC vs. INT-RC-WC	-1.665	0.140
	SSC-LC-AC vs. SSC-LC-WC	4.503	0.003*
	SSC-RC-AC vs. SSC-RC-WC	5.631	0.001*
	INT-LC-AC vs. INT-LC-WC	3.654	0.008*
	INT-RC-AC vs. INT-RC-WC	8.335	0.000*
	significant level $p<0.0125$		

Appendix Table 3.3. Statistical analysis on obtained discrimination scores on simple shifted (SSC), nonword sounds (NWS), and complex tones (CPT) continua

			F	p
Three-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	52.296	0.000*
	CONTINUUM (SSC, NWS, and CPT)	F(2,14)	2.104	0.159
	POSITION (LC vs. RC)	F(1,7)	26.004	0.001*
	CATEGORY × CONTINUUM	F(2,14)	4.322	0.035*
	CATEGORY × POSITION	F(1,7)	1.160	0.317
	POSITION × CONTINUUM	F(2,14)	0.441	0.652
	CATEGORY × CONTINUUM × POSITION	F(2,14)	1.208	0.328
significant level $p<0.05$				
SSC two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	37.345	0.000*
	POSITION (LC vs. RC)	F(1,7)	8.104	0.025*
	CATEGORY × POSITION	F(1,7)	2.640	0.148
NWS two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	103.431	0.000*
	POSITION (LC vs. RC)	F(1,7)	21.614	0.002*
	CATEGORY × POSITION	F(1,7)	0.033	0.861
CPT two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	6.574	0.037*
	POSITION (LC vs. RC)	F(1,7)	6.648	0.037*
	CATEGORY × POSITION	F(1,7)	0.519	0.495
significant level $p<0.05$				
		t(7)	p	
paired t-test	SSC-LC-AC vs. SSC-RC-AC	-2.626	0.034	
	SSC-LC-WC vs. SSC-RC-WC	-2.668	0.032	
	NWS-LC-AC vs. NWS-RC-AC	-4.167	0.004*	
	NWS-LC-WC vs. NWS-RC-WC	-4.124	0.004*	
	CPT-LC-AC vs. CPT-RC-AC	-2.077	0.076	
	CPT-LC-WC vs. CPT-RC-WC	-2.910	0.023	
	SSC-LC-AC vs. SSC-LC-WC	4.503	0.003*	
	SSC-RC-AC vs. SSC-RC-WC	5.631	0.001*	
	NWS-LC-AC vs. NWS-LC-WC	9.690	0.000*	
	NWS-RC-AC vs. NWS-RC-WC	5.530	0.001*	
	CPT-LC-AC vs. CPT-LC-WC	2.757	0.028	
	CPT-RC-AC vs. CPT-RC-WC	1.886	0.101	
significant level $p<0.008$				

APPENDIX 4 STATISTICAL ANALYSIS ON CHAPTER 3 EXPERIMENT IV

Appendix Table 4.1. Statistical analysis on obtained discrimination scores on language effect on real word continua

			F	p
Mix design	CATEGORY (AC vs. WC)	F(1,22)	48.903	0.000**
	CATEGORY × LANGUAGE (Cantonese and French)	F(2,22)	11.007	0.000**
	POSITION (LC vs. RC)	F(1,22)	25.451	0.000**
	POSITION × LANGUAGE	F(2,22)	0.856	0.439
	CATEGORY × POSITION	F(1,22)	2.342	0.140
	POSITION × CATEGORY × LANGUAGE	F(2,22)	0.692	0.511
significant level $p<0.05$				
CS two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	37.345	0.000**
	POSITION (LC vs. RC)	F(1,7)	8.104	0.025*
	CATEGORY × POSITION	F(1,7)	2.640	0.148
MS two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,8)	12.131	0.008*
	POSITION (LC vs. RC)	F(1,8)	15.201	0.005*
	CATEGORY × POSITION	F(1,8)	1.335	0.281
FS two-way RMANOVA	CATEGORY (AC vs. WC)	F(1,7)	2.305	0.173
	POSITION (LC vs. RC)	F(1,7)	4.198	0.080
	CATEGORY × POSITION	F(1,7)	0.004	0.954
significant level $p<0.05$				
			t	p
paired t-test	CS-LC-AC vs. CS-RC-AC	t(7)	-2.626	0.034
	CS-LC-WC vs. CS-RC-WC	t(7)	-2.668	0.032
	MS-LC-AC vs. MS-RC-AC	t(8)	-2.892	0.020
	MS-LC-WC vs. MS-RC-WC	t(8)	-5.361	0.001*
	FS-LC-AC vs. FS-RC-AC	t(7)	-1.775	0.119
	FS-LC-WC vs. FS-RC-WC	t(7)	-1.555	0.164
	CS-LC-AC vs. CS-LC-WC	t(7)	4.503	0.003*
	CS-RC-AC vs. CS-RC-WC	t(7)	5.631	0.001*
	MS-LC-AC vs. MS-LC-WC	t(8)	2.544	0.034
	MS-RC-AC vs. MS-RC-WC	t(8)	2.773	0.024
	FS-LC-AC vs. FS-LC-WC	t(7)	1.048	0.329
	FS-RC-AC vs. FS-RC-WC	t(7)	0.826	0.436
significant level $p<0.008$				

Appendix Table 4.2. Statistical analysis on obtained discrimination scores on language effect on nonspeech continua

		F(1,14)	p
Mix design	CATEGORY (AC vs. WC)	1.721	0.211
	CATEGORY × LANGUAGE (Cantonese, French)	7.302	0.017*
	POSITION (LC vs. RC)	18.815	0.001*
	POSITION × LANGUAGE	0.240	0.632
	CATEGORY × POSITION	1.094	0.313
	POSITION × CATEGORY × LANGUAGE	5.211	0.039*
significant level $p<0.05$			
CN two-way RMANOVA	CATEGORY (AC vs. WC)	6.574	0.037*
	POSITION (LC vs. RC)	6.648	0.037*
	CATEGORY × POSITION	0.519	0.495
FN two-way RMANOVA	CATEGORY (AC vs. WC)	1.248	0.301
	POSITION (LC vs. RC)	29.920	0.001*
	CATEGORY × POSITION	10.545	0.014*
significant level $p<0.05$			
		t(7)	p
paired t-test	CN-LC-AC vs. CN-RC-AC	-2.077	0.076
	CN-LC-WC vs. CN-RC-WC	-2.910	0.023
	FN-LC-AC vs. FN-RC-AC	-6.173	0.000*
	FN-LC-WC vs. FN-RC-WC	-3.487	0.010*
	CN-LC-AC vs. CN-LC-WC	2.757	0.028
	CN-RC-AC vs. CN-RC-WC	1.886	0.101
	FN-LC-AC vs. FN-LC-WC	-1.909	0.098
	FN-RC-AC vs. FN-RC-WC	0.104	0.920
significant level $p<0.0125$			

Appendix Table 4.3. Statistical analysis on obtained discrimination scores on interaction of real word and nonspeech continua by Cantonese and French subjects

		F(1,14)	p
Mix design	CONTINUUM (SSC vs. CPT)	0.656	0.432
	CONTINUUM × LANGUAGE (Cantonese and French)	1.868	0.193
	CATEGORY (AC vs. WC)	18.670	0.001*
	CATEGORY × LANGUAGE	20.242	0.000*
	POSITION (LC vs. RC)	38.005	0.000*
	POSITION × LANGUAGE	1.334	0.267
	CATEGORY × CONTINUUM	9.464	0.008*
	CATEGORY × CONTINUUM × LANGUAGE	0.660	0.430
	POSITION × CONTINUUM	0.891	0.361
	POSITION × CONTINUUM × LANGUAGE	0.076	0.787
	CATEGORY × POSITION	4.994	0.042*
	CATEGORY × POSITION × LANGUAGE	0.254	0.622
	CONTINUUM × POSITION × CATEGORY	0.055	0.819
	CONTINUUM × POSITION × CATEGORY × LANGUAGE	3.186	0.096

significant level $p<0.05$

		F(1,7)	P
CS and CN 3-way RMANOVA	CATEGORY (AC vs. WC)	23.312	0.002*
	CONTINUUM (SSC vs. CPT)	2.535	0.155
	POSITION (LC vs. RC)	24.226	0.002*
	CATEGORY × CONTINUUM	9.096	0.019*
	CATEGORY × POSITION	1.602	0.246
	POSITION × CONTINUUM	0.120	0.739
	CATEGORY × CONTINUUM × POSITION	1.775	0.225

significant level $p<0.05$

		F(1,7)	P
FS and FN 3-way RMANOVA	CATEGORY (AC vs. WC)	0.048	0.833
	CONTINUUM (SSC vs. CPT)	0.145	0.714
	POSITION (LC vs. RC)	14.034	0.007*
	CATEGORY × CONTINUUM	2.193	0.182
	CATEGORY × POSITION	3.520	0.103
	POSITION × CONTINUUM	5.255	0.056
	CATEGORY × CONTINUUM × POSITION	1.413	0.273

significant level $p<0.025$

APPENDIX 5 PUBLICATION LIST

Zheng, H.-Y., Peng, G., Tsang, P. W.-M., & Wang, W. S.-Y. (2006). *Perception of Cantonese level tones influenced by context position*. Paper presented at the 3rd International Conference on Speech prosody, Dresden, Germany.

Zheng, H.-Y., Tsang, P. W.-M., & Wang, W. S.-Y. (2007). *Categorical perception of Cantonese tones in context: A cross-linguistic study*. Paper presented at the 8th Annual Conference of the International Speech Communication Association (Interspeech 2007), Antwerp, Belgium.

Zheng, H.-Y., & Wang, W. S.-Y. (2008). *An ERP study on categorical perception of lexical tones and nonspeech pitches*. Paper presented at the 9th Annual Conference of the International Speech Communication Association (Interspeech 2008) incorporating the 12th Australasian International Conference on Speech Science and Technology (SST 2008), Brisbane, Australia.

Zheng, H.-Y. (2008). *Categorical perception of lexical tones: Correlation between electric brain response and behavioral performance*. Paper presented at the 5th Postgraduate Research Forum on Linguistics: Contemporary Approaches to Linguistic Analysis.