UNIVERSITY OF CALIFORNIA, SAN DIEGO

# Processing of Pitch Height Information in Mandarin Tone Perception

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Psychology

by

Jing Shen

Committee in charge:

Professor Diana Deutsch, Chair
Professor Victor Ferreira
Professor F. Richard Moore
Professor Miller Puckette
Professor Keith Rayner
Professor Timothy Rickard

2013

The Dissertation of Jing Shen is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

---

---

Chair

University of California, San Diego

2013

TABLE OF CONTENTS

Signature Page.....iii

Table of Contents.....iv

List of Abbreviations.....vi

List of Figures.....vii

List of Tables.....viii

Acknowledgements.....ix

Curriculum Vitae.....xi

Abstract.....xiii

Chapter 1. Introduction.....1

1.1 Overview.....2

1.2 Lexical Tones.....2

1.3 Absolute Pitch.....9

1.4 Mental Template for Pitch Processing.....20

1.5 Studies.....24

1.5.1 Study 1: Overall Pitch Height as a Cue to Lexical Tone Perception.....25

1.5.2 Study 2: On-line Perception of Mandarin Tones 2 and 3:

Evidence from Eye Movements .....31

Chapter 2. Method .....43

2.1 Study 1.....,.....44

2.2 Study 2.....49

2.2.1 Experiment 1.....49

2.2.2 Experiment 2.....55

Chapter 3. Results.....60
3.1 Study 1.....61
3.2 Study 2.....68
3.2.1 Experiment 1.....68
3.2.2 Experiment 2.....77
3.2.3 Comparing the results of two experiments.....84
Chapter 4. Discussions.....87
4.1 Study 1.....88
4.2 Study 2.....92
4.3 General Conclusion.....98
4.4 Future Directions.....100
References.....105

LIST OF ABBREVIATIONS

AP: absolute pitch

EEG: electroencephalogram

ERP: event-related potential

fMRI: functional magnetic resonance imaging

MMN: mismatch negativity

MRI: magnetic resonance imaging

MTG: middle temporal gyrus

PT: planum temporale

STS: superior temporal sulcus

STG: superior temporal gyrus

LIST OF FIGURES

Figure 2.1: Pitch patterns of the four tones at their original height level.....46

Figure 2.2: Pitch patterns of tone stimuli in four conditions of Study 2.....52

Figure 2.3: Examples of the visual stimuli in the two experiments of Study 2....55

Figure 3.1: Study 1 correct response rate and reaction time data in detail for "yes" and "no" trials.....63

Figure 3.2: Study 1 summarized correct response rate and reaction time data for "yes" and "no" trials.....64

Figure 3.3: Study 1 scatter plots showing the linear regression lines with amount of pitch transposition as predictor and performance level.....66

Figure 3.4: Study 1 correct response rate and reaction time data as functions of musical training background.....68

Figure 3.5: Study 2 Experiment 1 proportion of fixations curves in 20 ms interval.....71

Figure 3.6: Study 2Experiment 1 bootstrapping curves in 2 ms interval.....76

Figure 3.7: Study 2 Experiment 2 proportion of fixations curves in 20 ms interval.....79

Figure 3.8: Study 2 Experiment 2 bootstrapping curves in 2 ms interval.....83

LIST OF TABLES

Table 2.1: Syllables used in Study 1 and their corresponding characters .....45

Table 2.2: Overall pitch height levels of sound stimuli in Study 1 .....46

Table 2.3: Sound stimuli used in Study 2 Experiment1.....50

Table 2.4: Pitch of onset, turning point, and offset in four conditions in Study 2.....51

Table 2.5: Sound stimuli used in Study 2 Experiment2.....57

Table 3.1: Study 1 means and standard deviations (in parenthesis) of correct response rate and reaction time data for the four tones.....62

Table 3.2: Study 2 Experiment 1 Means and standard deviations of proportion of fixations data in four 150 ms windows.....73

Table 3.3: Study 2 Experiment 2 Means and standard deviations of proportion of fixations data in four 150 ms windows.....80

ACKNOWLEDGEMENTS

There are a great many people who have provided me with support, advice, and friendship over the years, without whom I could not have done the work presented here. The following list is but a small sample:

Diana Deutsch, as my academic adviser and committee chair, has provided valuable guidance and advice on my development as a scientist. I am, and will always be, grateful for her patience and support through my multiple drafts and many long afternoons of work.

Keith Rayner, who generously opened up his lab for me to learn freely, introduced me to another new dimension of research. The valuable experience of working with him inspires me to explore novel questions in interdisciplinary fields.

My collaborator Jinghong Le, for being supportive to my dissertation projects and being dedicated to this line of research in general; Vic Ferreira, for providing advice and insight on innumerable occasions; Dick Moore, for many stimulating conversations and informative articles throughout the years; Members of Deutsch lab and Rayner lab, who has been an inexhaustible source of help and support; My family, especially my husband Hanbo and my parents Guilin and Shubing; who always support me on my decision and endeavor.

A report on Study 1 has been submitted for publication. The dissertation author was the primary investigator and author of this paper. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Jinghong Le.

Data of Study 2 is published in “On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements.” Shen, J., Deutsch, D., and Rayner, K., Journal of the Acoustical Society of America, 2013, 133, 3016-3029. The dissertation author was the primary investigator and author of this material. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Keith Rayner.

CURRICULUM VITAE

## Education

2013, Doctor of Philosophy, University of California, San Diego

2004, Master of Science, East China Normal University, Shanghai, China

2001, Bachelor of Arts, East China Normal University, Shanghai, China

## Publications

Shen, J., Deutsch, D., & Le, J. (under review). Overall Pitch Height as a Cue to Lexical Tone Perception.

Deutsch, D., Le, J., Shen, J., & Li, X. (in preparation). Large-scale Direct-test Study Reveals Unexpected Characteristics of Absolute Pitch.

Shen, J., Deutsch, D., & Rayner, K. (2013). On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements. Journal of the Acoustical Society of America, 133, 3016-3029.

Deutsch, D., Le, J., Shen, J., & Henthorn, T.(2009). The Pitch Levels of Female Speech in Two Chinese Villages. Journal of the Acoustical Society of America Express Letters, 125.

Shen, J. & Kong, K. (2004). The Progress and Prospect of the Research of Nature and Nurture Controversy in Personality Development, Psychological Science (Chinese), 27, 250.

Shen, J. (2003). A Review of Music Therapy and Its Related Psychological Research, Psychological Science (Chinese), 26, 176.

## Conference Presentations

Shen, J., Deutsch, D., & Le, J. (2011). Overall Pitch Height as a Cue for Lexical Tone Perception. Poster session presented at the 162nd meeting of Acoustical Society of America, San Diego, CA.

Shen, J., Deutsch, D., & Rayner, K. (2010). Processing of Endpoint Pitch in Mandarin Tone Perception: An Eye Movement Study. Paper presented at the 20th International Congress on Acoustics, Sydney, Australia.

Deutsch, D., Le, J., Shen, J., & Li, X. (2011). Large-scale Direct-test Study Reveals Unexpected Characteristics of Absolute Pitch. Paper presented at the 162nd meeting of Acoustical Society of America, San Diego, CA.

ABSTRACT OF THE DISSERTATION

Processing of Pitch Height Information in Mandarin Tone Perception

by

Jing Shen

Doctor of Philosophy in Psychology

University of California, San Diego, 2013

Professor Diana Deutsch, Chair

Communicating in tone language involves the use of *lexical tone* as a cue to determine the meaning of words. Lexical tones are defined both by their pitch height and also by their pitch contours. Given the height component, it has been hypothesized that experience in acquiring a tone language influences the development of a mental template that is used both for speech communication and also to facilitate the acquisition of absolute pitch in music (Deutsch, Henthorn, & Dolson, 2004). Evidence that has been observed in native speakers of tone languages includes consistent pitch height in speech production and high

prevalence of absolute pitch in music. My dissertation studies concern genesis of this pitch template and examine the question: what is the role of pitch height in lexical tone perception? The work presented here is the first to use both off-line and on-line measurements to examine *how* native speakers exploit various pitch height cues for identifying tones. The results suggest that native speakers utilize overall pitch height as a whole to guide immediate tone judgment, as well as keep updating their decisions directed by pitch height information that unfolds throughout the tone. Taken together, the findings demonstrate an instant and incremental association of pitch height with tone label in tone perception and lend support to the mental template for pitch processing across speech and music domains.

# Chapter 1. Introduction

## 1.1 Overview

In this chapter, I will first present background information about lexical tones and survey the literature on pitch cues in lexical tone perception. Secondly, I will present research findings on absolute pitch (AP), including its genesis, and evidence for an implicit form of AP. I will then explore the relationship between AP and speech processing, followed by an examination of the evidence from both speech and music domains suggesting a mental template for pitch processing that is possessed by tone language speakers. Lastly, I will introduce the present experiments comprising the chapters that follow, designed to investigate the role of pitch height cues in lexical tone perception.

## 1.2 Lexical Tones

A tone language is a language that utilizes pitch to contrast individual lexical items or words (McCawley, 1978). In tone languages, such as most East Asian and African languages, words with the same vowels and consonants have different meanings depending on the lexical tones in which they are enunciated. For example, the word 'ma' in Mandarin means 'mother' when spoken in the first tone, 'hemp' when spoken in the second tone, 'horse' in the third tone, and a reproach in the fourth tone. This contrasts with intonation languages such as English, in which pitch is used to convey emotion and the form of utterance (i.e., statement or question), but is not involved in determining the meaning of individual words. The principal feature of lexical tone is in the domain of pitch, whose primary acoustic correlate is fundamental frequency (F0, Howie, 1976).

When this primary cue of pitch is missing or ambiguous, other acoustic cues will be used for tone identification, such as amplitude contour (Whalen & Xu, 1992; Fu & Zeng, 2000) and duration (Blicher, Diehl, & Cohen, 1990).

There is a linguistic typological category of tone language called "register tone language" in which most tones can be described in terms of points within a pitch range (e.g. some African languages such as Yoruba, Mambila, etc.), whereas other tone languages called "contour tone languages" have some tones that are specified in terms of gliding pitch movements (e.g. most Southeast Asian languages such as Thai, Mandarin, Cantonese, etc., Pike, 1948). As register tone languages almost exclusively have only level tones, tone perception is realized here by dividing the perceptual space with respect to the pitch range of the tones (Connell, 2000). Connell also proposed that, to prevent overburdening of the auditory pitch discrimination function, when the number of tone goes over three in a language, pitch movement would begin to be another perceptual cue for tone perception. This idea is supported by the phenomenon that most Asian tone languages have more than three tones, which incorporates level tones and contour tones.

There is a longstanding debate concerning the underlying representation for lexical tones. Phonologists generally agree that contour tones should be represented as a sequence of high and low pitch registers (Anderson, 1978; Gandour, 1974; Yip, 1989, 2002; Zhang, 2002). However, experimental studies of contour tone perception and production argued against the view that contour tones should be decomposed into pitch sequences (Abramson, 1978; Gandour,

1981,1983; Gandour & Harshman, 1978; Xu, 2004). The alternative view of taking the dynamic movement of contour tones (i.e. direction and slope) as a perceptual feature has also been adopted by other researchers on tone perception (see, for example, Gandour, Wong, Hsieh, et al., 2000; Massaro, Cohen, & Tseng, 1982; Lin & Repp, 1989; Vance, 1977, among many others). On the other hand, a few perceptual studies on Thai tones provide new evidence supporting the hypothesis that the pitches of certain points within the syllable serve as major perceptual cue for identifying tones (Mixdorff, Luksaneeyanawin, Fujisaki, & Charnavit, 2002; Zsiga & Nitisaroj, 2004, 2007).

The multi-dimensional scaling studies of Gandour and colleagues (Gandour, 1983; Gandour & Harshman, 1978) reported major perceptual dimensions of pitch for tone perception that include average pitch height, direction of pitch change (rising vs. non-rising), onset/offset pitch height**,** and pitch contour (level vs. contour), in which the first two held primary importance. Using 50 native speakers of 4 different Southeast Asian tone languages (Cantonese, Mandarin, Taiwanese, Thai) and 50 native speakers of English, Gandour (1983) collected data of direct paired-comparison judgments of tone dissimilarity. The data analysis revealed that there are two main dimensions, which were interpreted as "height" and "direction". It was also found that listeners weighted these pitch cues differently according to their language background. Although this finding has been widely cited by researchers on tone perception (e.g., Lee, Vakoch, & Wurm, 1996; Cutler & Chen, 1997; Francis, Ciocca, & Ng, 2003), it was also criticized by others as not best representing the actual tone perception because

the findings were based on data of dissimilarity ratings of different pitch patterns that were obtained from a mix of speakers of tone languages and non-tone languages, instead of native speakers of a specific tone language (Zsiga & Nitisaroj, 2007, pp. 379).

***Overall pitch height***

Studies that investigated the cue of overall pitch height have been focused on perception of tones depending on pitch context, which is either provided by carrier sentence/phrase (Lin & Wang, 1985; Leather, 1983; Fox & Qi, 1990; Moore & Jongman, 1997; Wong & Diehl, 2003; Francis, Ciocca, Wong, Leung, & Chu, 2006; Huang & Holt, 2009) or inferred by speaker normalization (Wong & Diehl, 2003; Lee, 2009; Moore & Jongman, 1997). For example, Lin and Wang (1985) presented subjects with paired Mandarin tones in which the first tone, representing a high-level tone (Tone 1), was held at a constant of 115 Hz, while the second tone, representing a high-falling tone (Tone 4), had a onset F0 varied across a continuum of 110 to 140 Hz in 10 Hz steps and an F0 fall of 40 Hz. Subjects were asked to identify the first tone in each pair. They found that as the onset F0 of the second tone increased, identification of the first tone as a low-rising tone (Tone 2) increased. Thus the higher onset F0 of the second tone cued a wider pitch range and altered the relative F0 of the first tone to be perceived as low, demonstrating a contrastive context effect. Along the same line, Mandarin tone identification data from Moore and Jongman (1997) suggested a contrastive context effect from precursor phrase by showing identification shifts such that identical stimuli were identified as low tones for the high precursor condition, but

as high tones for the low precursor condition.

When the listeners do not have direct access to context pitch information (e.g., in the condition of isolated syllables), they are able to gauge speaker identity from other acoustic cues and exploit the information regarding pitch range of the speakers in perceiving the tones. For instance, Lee (2009) used as stimuli Mandarin syllables *sa* with all four tones produced by 16 speakers of each gender. He had 40 native listeners identify these tones that were digitally processed to only have the fricative and the first six pitch periods available. The results showed that tone identification accuracy exceeded chance level and tone classification based on pitch height also correlated with other voice quality measures across speakers, which suggested speaker identification is implicated in the process of tone perception. When Wong and Diehl (2003) tested subjects' performance of identifying the three Cantonese level tones in isolation without any context pitch information, they found subjects did significantly better when these tones were presented in blocks that each block only had one speaker compared to when they were mixed across speakers.

***Pitch change direction***

A number of studies have focused on pitch change direction as a cue to lexical tone identification. The results demonstrated that native speakers perceive tones in a categorical manner for contour tones but not level tones. Studies using level tones of Thai and Cantonese, both being tone languages with more than two level tones, reported a flat and high discrimination function unrelated to identification peaks and failed to meet the criteria for categorical perception

(Abramson, 1978; Francis, et al., 2003). Some authors (Francis et al.) proposed that this result could be explained by supposing that listeners would have difficulty placing the category boundaries for different level tones without the context needed for speaker normalization.

In several studies of Mandarin tone perception, a continuum based on high rising and high level tones was used for creating stimuli for identification and discrimination tasks (Chan, Chuang, & Wang, 1975; Wang, 1976; Xu, Gandour, & Francis, 2006). Results consistently showed that there was, for Chinese subjects, a categorical boundary located in the middle of the continuum space, which cannot be interpreted on purely psychoacoustically heightened sensitivity (Klatt, 1973) and thus has to be attributed to language experience of tone categories. Similar results were also reported in two other studies using different tone languages and subject groups. Data from Francis et al. (2003) demonstrated native speakers' categorical perception for Cantonese falling and rising tones. Halle, Chang, and Best (2004) showed a quasi-categorical perception (in the sense of existence of perceptual boundaries for discrimination/identification tasks) on Taiwan Mandarin speakers, but not French speakers.

***Pitch height at critical points***

As perceptual cues for tone identification, the pitch heights at critical points (i.e., onset, offset, midpoint) have only been investigated in a handful of experiments measuring correct identification rate and reaction time (Gottfried & Suiter, 1997; Lee, Tao, & Bond, 2008; Zsiga & Nitisaroj, 2007). In the study by Gottfried and Suiter (1997), silent-center syllables that comprised only the initial

six pitch periods and final eight pitch periods were played to native and non-native speakers of Mandarin. The percentage average correct identification of the tones by native speakers was about 94%, compared to 64.5% by non-native speakers. Findings from Lee et al., (2008) replicated this result with a larger sample size of native speakers (n = 40 compared with the earlier n = 6). Zsiga and Nitisaroj (2007) demonstrated, through a series of perceptual experiments that used both natural speech and digitally-altered speech, that pitch inflections at the syllable midpoint and offset point successfully categorized Thai tones in perceptual space. According to their tone identification data, falling tones were identified as having a high pitch height at the syllable midpoint; rising tones as having a low pitch height at syllable midpoint; and mid tones were identified by lack of any pitch inflection. Pitch direction and slope were also found to play a role in tone identification, particularly when cues of high/low location were ambiguous or conflicting.

***Pitch contour***

In the study of Gandour (1983), pitch contour was found to be one of the pitch cues for tone perception and it was defined as the difference between a level tone and a non-level tone (including tones with rising, falling, rising-falling, falling-rising pitch trajectories). Using a passive oddball paradigm, more recent cortical event-related potential (ERP) data were collected while Chinese and English speakers were presented with Mandarin tones. Three odd ball conditions were created using Mandarin Tone 1, 2 and 3, with two of them representing a contrast between a level and a contour tone (Tone 1/Tone2, Tone1/Tone3) and

one of them representing a contrast between two contour tones (Tone2/Tone3). Dissimilarity matrices were created based on the mean mismatch negativity (MMN) amplitude data and analyzed using multi-dimensional scaling (Chandrasekaran, Gandour, & Krishnan, 2007; Chandrasekaran, Krishnan, & Gandour, 2007). The results showed, for the Chinese group, the amplitude of MMN was significantly larger for processing the level/contour tone contrast, compared to the contrast between two contour tones. On the other hand, this difference depending on type of tone was not observed for English speaking subjects. After the data was analyzed using a two-dimensional space, namely height and contour, Chinese speaking subjects were found to assign more weight to the dimension of contour than English speaking subjects, while both subject groups weighed "height" as equal in importance.

### 1.3 Absolute Pitch

Absolute pitch (AP) is the ability to name or produce a musical note of a particular pitch without benefit of a reference note. People with AP can name a musical note quickly and effortlessly (Takeuchi & Hulse, 1993; Ward, 1999; Deutsch, 2013). AP has been reported to be a very rare faculty amongst intonation language speakers, with an estimated prevalence of less than one in ten thousand (Bachem, 1940; Profita & Bidder, 1988).

The experience of using AP is analogous to the instantaneity with which anyone with normal vision is able to name a color. When determining a red ball on a green grassy field, for example, the observer do not need to compare the ball

to the grass, or call on prior knowledge that the grass is green, to make an immediate and absolute judgment of the ball's color. Similar to this color-naming scenario, AP possessors perceive the quality of the pitch class of musical note in isolation, while non-possessors rely on relative pitch to perceive tones by abstracting the interval information between tones.

It has been suggested that an essential cognitive function of AP is conditional associative memory, which helps with constructing and retrieving the association of pitch categories with verbal labels (Deutsch, 2006; Zatorre, 2003). First, musicians with AP do not need to update working memory in pitch perception. Data from ERP studies have shown that, unlike listeners lacking AP, AP possessors had an absent or reduced electrical-evoked potential component to a pitch change (Klein, Coles, & Donchin, 1984; Wayman, Frisina, Walton, Hantz, & Crummer, 1992). Further supporting evidence came from functional neuroimaging data showing an area of the right inferior frontal cortex believed to be important for monitoring pitch information in working memory is more active in musicians lacking AP than in those who have AP (Zatorre, Perry, Beckett, Westbury, & Evans, 1998). Second, the posterior dorsolateral prefrontal cortex is a brain area that has been found to be involved in conditional associative memory, as demonstrated by the task of naming musical intervals by musicians (Zatorre et al. 1998), as well as identifying chords with arbitrary verbal labels by trained non-musicians (Bermudez & Zatorre, 2005). This area was observed to have more activation in listeners with AP than those did not have AP, which indicates AP possessors exploit a pitch-label association in musical tone perception (Zatorre et

al. 1998). More recent behavioral data also lend support to this view by showing an associative memory-retrieval system served as an efficient way for musicians with AP to outperform musicians lacking AP in pitch matching and production tasks (Hsieh & Saberi, 2008).

### *Genesis*

As AP is extremely rare in the Western world, there have been three types of hypotheses on its genesis. Firstly, it was thought that AP could be acquired at any time through intensive practice. A considerable amount of evidence has shown that, although AP is acquired in early childhood in an effortless and unconscious way, attempts to train adults through practice have had very little success (Cuddy, 1968; Heller & Auerbach, 1972; Brady, 1970).

Second, AP is an inherited trait and will manifest itself once given opportunity. The fact that this ability usually appears at a very young age, even when the child has had little or no formal musical training (Carpenter, 1951; Corliss, 1973; Takeuchi, 1989), has been used to argue for this genetic hypothesis. Further evidence supporting this view includes AP tends to run in families (Bachem, 1940, 1955; Baharloo, Johnston, Service, Gitschier, & Freimer, 1998; Baharloo, Service, Risch, Gitschier, & Freimer, 2000; Gregersen, Kowalsky, Kohn, & Marvin, 1999, 2001; Profita & Bidder, 1988; Theusch, Basu, & Gitschier, 2009). Based on data from 600 musicians, Barhaloo et al. (1998) found that AP possessors were four times more likely than non-possessors to report that a family member had AP. However, this argument does not hold too well due to the confounding factor of early musical training, which is also a strong predictor for

AP. Another finding that is in favor of genetic contribution to AP is the very different prevalence in various ethnic groups. Gregersen et al. (1999, 2001) carried out a survey on music students in the United States and found a large proportion of East Asian students reported having AP. However, Henthorn and Deutsch (2007) reanalyzed the data and found that for those students who spent their early childhood in America, the prevalence of AP was consistent across two ethnic groups of East Asian and Caucasian. While the prevalence was significantly higher among students had lived in East Asia during early childhood compared to their peers who did so in America. A more recent piece of evidence supporting the genetic argument was provided by Theusch et al. (2009) showing a genome-wide linkage on chromosome 8 in families with European ancestry that include AP possessors. This is a first step towards the ultimate goal of demonstrating the gene or genes that contribute to AP.

Third, the potential of having AP is ubiquitous, but in order to realize this potential, exposure to pitch-label association during a critical period is indispensable. A large number of studies have found an association between AP and early onset of musical training (Bachem, 1940; Baharloo et al., 1998, 2000; Deutsch, Henthorn, Marvin, & Xu, 2006; Deutsch, Dooley, Henthorn, & Head, 2009; Deutsch, Le, Shen, & Li, 2011; Dooley & Deutsch, 2010, 2011; Gregersen et al., 1999; Lee & Lee, 2010; Levitin & Rogers, 2005; Miyazaki, 1988; Miyazaki & Ogawa, 2006; Profita & Bidder, 1988; Sergeant, 1969; Takeuchi, 1989; Takeuchi & Hulse, 1993). For example, Baharloo, et al., (1998) in a survey of 600 musicians, showed that 40% of those who had begun musical training by age 4

self-reported having AP, which contrasted with 27% in the age 4-6 group, 8% in the age 6-9 group, and 4% in the age 9-12 group. Although the data were likely to exaggerate the percentages of AP possession due to self-report of self-selected respondents, this finding demonstrated the high correlation between AP and early onset of musical training, which was later confirmed by large-scale direct-test studies (Deutsch, et al., 2006; Deutsch, Dooley, et al., 2009; Deutsch et al., 2011; Lee & Lee, 2010). For instance, Deutsch and colleagues administered a test of AP to 88 music school students at the Central Conservatory of Music in Beijing and 115 students at Eastman School of Music. In addition to the large effect of language, a systematic effect of age of onset of musical training was found. For example, for non-tone language speakers, 14% who had begun musical training at ages 4-5 met the AP criterion, while 6% who had begun training at ages 6-7 and none of those who had begun training at ages 8 or later did so.

### *Implicit AP*

Following the argument that exposure to pitch-label association during early childhood is critical for developing AP, it is not surprising to see that most non-tone language speakers cannot name the notes they are judging. Nevertheless, most people have an implicit form of AP, which has been demonstrated in several ways. One of them concerns the tritone paradox, a musical illusion in which people judge the relative heights of tones based on their positions along the pitch class circle. Furthermore, people who do not have AP can often judge whether a musical piece they know is being played in the correct key and can also reproduce their familiar melodies in a consistent pitch.

The tritone paradox (Deutsch, 1986) is produced by two sequentially presented tones that are half-octave apart (or tritone). These tones are generated employing Shepard tones, so that their note names (pitch classes) are clearly defined but they are ambiguous in terms of which octave they are in. When a pair of tritones is played (e.g., F# followed by C), some listeners hear an ascending pattern, while other hear a descending one. An important finding was that for any given listener, tones in one region of the pitch class circle are consistently heard as higher, while those in the other region are heard as lower. This pattern occurs even when spectral effects were controlled (Deutsch, 1987, 1992, 1994; Deutsch et al., 1987; Deutsch, Henthorn, & Dolson, 2004b; Giangrande, 1998; Repp & Thompson, 2010). When the listeners experience the triton paradox, they must be referring to the pitch classes of tones in judging their relative heights, so exploiting an implicit form of AP.

Further evidence in favor of an implicit form of AP came from pitch identification and production studies. A number of studies played musical excerpts, which were either in the original key or transposed to a different key, to musically literate subjects who were mostly AP non-possessors (Terhardt & Ward, 1982; Terhardt & Seewann, 1983; Vitouch & Gaugusch, 2000). The results showed that the subjects achieved significant identification performance across the board, even in conditions that had the amount of transposition of only one semitone. Data from experiments with musically untrained subjects are consistent with the previous findings in suggesting a good performance on pitch identification tasks. Schellenberg and Trehub (2003) asked unselected college

students to discriminate between six popular TV theme songs either presented in the original key or transposed up or down one or two semitones. Judgment accuracy for well-known theme songs was significantly above chance level, although it was not the case for unknown themes that were included as a control to rule out acoustic artifacts from the transposition process. Smith and Schmuckler (2008) evaluated the prevalence of implicit AP in non-musicians using telephone dial tone, which has not been changed for decades and consists of two tones at 350 and 440 Hz. It was demonstrated that these AP non-possessors could nevertheless make correct judgment when a tone had been transposed by three semitones.

Converging evidence from pitch production studies also supported the implicit form of AP in general population. For example, when singing the first note of a few different popular songs at separate sessions two days apart, unselected subjects showed a very low within-subject variability of pitch range (Halpern, 1989). Similarly, Levitin (1994) had subjects sing two different popular songs and their productions were compared to the actual pitches used in recordings of those songs. 40% of the subjects sang the correct pitch on at least one trial; 12% of the subjects hit the correct pitch on both trials, and 44% came within two semitones of the correct pitch on both trials. In a further study, Bergeson and Trehub (2002) asked mothers sing the same song to their infants in two sessions that were at least one week apart and found their pitch ranges across two sessions were quite consistent (less than a semitone apart, according to judges' estimate).

Overall, these findings suggest an implicit form of AP that is more widespread than explicit AP, which in addition requires the association of pitch with verbal label. This further lends support to the hypothesis that AP can be acquired for most people given they are exposed to the pitch-label association during early childhood.

### *Relationship between AP and speech*

Evidence from several lines of research suggests that absolute pitch is closely related to speech processing. First, when the listeners perceive the triton paradox, their identification of the pattern associates with the language or dialect they have been exposed, especially in early childhood. Second, critical periods for acquiring AP and speech share very similar timetables. Third, neuroimaging data indicates a shared brain structures that underlie AP and speech processing. Fourth, AP prevalence was found to be much higher among speakers of tone languages, in which pitch information is heavily involved in perceiving speech.

First, as described earlier, judgments of the pattern in perceiving the tritone paradox show a systematic relationship to the positions of the tones along the pitch class circle, although the listeners cannot name the tones. It was also demonstrated by Deutsch and colleagues' work that the arrangement of judgments along the pitch class circle depends on the language and dialect to which the listeners has been exposed (Deutsch, 1991, 1994; Deutsch, et al., 2004b), and pitch range of the listeners' speaking voice (Deutsch, North & Ray, 1990; Deutsch, et al., 2004a), which also varies depending on the speakers'

language or dialect (Dolson, 1994; Deutsch, et al., 1990; Deutsch, Le, et al., 2009).

It has been well documented that language acquisition has a critical period that extends up to puberty (Lennenberg, 1967; Doupe & Kuhl, 1999; Johnson & Newport, 1989; Newport, 1990; Newport, Bavelier, & Neville, 2001; Sakai, 2005) and children and adults learn a new language in qualitatively different ways (Lennenberg, 1967). Among several linguistic components of language learning (i.e., phonological, semantic, syntactic), the most difficult aspect is usually phonological. A second language learned after this critical period is usually spoken with a "foreign accent" (Scovel, 1969; Patkowski, 1990). Further supporting evidence for this critical period of speech acquisition comes from the findings that children who were socially isolated and later placed in a normal environment were not able to learn normal speech (Curtiss, 1977; Lane, 1976). Furthermore, recovery of speech following brain injury was shown to correlate with the age at which the injury occurred. The prognosis for recovery is most positive before age of 6 and becomes very poor after puberty (Bates, 1992; Dennis & Whitaker, 1976; Duchowny et al., 1996; Varyha-Khadem et al., 1997; Woods, 1983).

In parallel, studies have found AP is extremely difficult to learn in adulthood while can be acquired by children without effort. The prevalence of AP is very high among those musicians who began their musical training before age of 6 and become no higher than chance level after the age of 9 (Barhaloo, et al., 1998; Deutsch, et al., 2006; Deutsch, Dooley, et al., 2009; Deutsch et al., 2011;

Gregersen et al., 1999; Chin, 2003). Taken together, the observed overlap in the developmental trajectory for acquiring AP and speech suggests these two capacities may be subserved by a common brain mechanism.

Structurally, it has been demonstrated that there is a stronger leftward asymmetry of planum temporale (PT) in musicians with absolute pitch (Schlaug, Jancke, Huang, & Steinmetz, 1995; Keenan, Thangaraj, Halpen, & Schlaug, 2001). This brain area, which is in the temporal lobe that corresponds to the core of Wernicke's area, has been proposed to be a computational hub for segregation and matching of incoming auditory signal with learned intrinsic representations, such as verbal label and spatial map (Griffiths & Warren, 2002). In an MRI comparison of musicians and non-musicians, Schlaug et al., (1995) found a greater leftward PT asymmetry in the musicians, which was due to the subgroup if those with AP in the sample. Further examination replicated the increased PT asymmetry amongst AP possessors and attributed it to a decreased right PT instead of a enlarged left PT compared to non-possessors, suggesting pruning of the right PT and/or relocated resources to the left hemisphere (Keenan, et al., 2001). While these studies measured PT surface area, this conclusion has also been confirmed by an earlier study demonstrating a larger PT size in AP possessors compared to a musically unselected sample of subjects (Zatorre, et al., 1998). More resent functional imaging data further supported the link between AP and speech. Oechslin, Meyer, and Jäncke (2010) found that AP is associated with significantly different hemodynamic responses to complex speech sounds, compared to relative pitch processing. In the study, AP possessors showed

greater activation in the left PT and surrounding areas when they were engaged in segmental speech processing. In addition, fMRI data have shown a stronger activity of left PT was associated with perception of lexical tones by native listeners (Xu, et al., 2006). Moving beyond the region of left PT, AP possession was demonstrated to associate with elevated left superior temporal sulcus (STS) activity (Schultz, Gaab, & Schlaug, 2009) and heightened connectivity of white matters between left superior temporal gyrus (STG) to left middle temporal gyrus (MTG, Loui, Li, Hohmann, & Schlaug, 2011). These brain regions are within the left superior temporal lobe and have been proposed to be involved in the circuitry for perception and comprehension of speech sounds (Hickok & Poeppel, 2007; Möttönen, et al., 2006)

This analogy between AP and speech processing gains support as well from the findings that the prevalence of absolute pitch is strikingly higher among music students who speak a tone language fluently than their peers who do not do so, even when they started their musical training at the same age (Deutsch, et al., 2006; Deutsch, Dooley, et al., 2009; Deutsch, et al., 2011). It was also found that the results could not be explained by either ethnicity or country of early music education but should be attributed to early experience with a tone language (Henthorn, & Deutsch, 2007; Deutsch, Dooley, et al., 2009). These results pointed to a close relationship between AP and processing of lexical tones, which will be discussed in detail in the next section.

## 1.4 Mental Template for Pitch Processing

It has been argued that absolute pitch in its explicit form is analogous to the use of lexical tones in tone languages (Deutsch, 2006, 2013). When an AP possessor hears the note C# and attributed the label "C#", he or she is associating a particular pitch with a verbal label. Similarly, when a native speaker of Mandarin hears the word "ma" spoken in the first tone and recognizes it means "mother", he or she is associating a particular pitch (or a set of pitch cues) with a verbal label.

Different from processing of prosody and emotion information in speech, lexical tone processing has been identified to primarily involve left hemisphere. Dichotic-listening studies provided evidence for hemispheric lateralization of lexical tone processing by showing a right-ear advantage for native tone perception (Van Lanker & Fromkin, 1973; Wang, Sereno, & Jongman, 2001). Using English and Thai speakers as subjects, Van Lanker and Fromkin demonstrated a right-ear advantage for Thai speakers processing Thai tones. Studies on patients with left-side brain damage demonstrated impairments in lexical tone perception. Thai tone identification performance of left hemisphere damaged subjects was found to be worse than that of controls (Gandour & Dardarananda, 1983; Eng, Obler, Harris, & Abramson, 1996; Gandour et al., 1992; Moen & Sundet, 1996; Naeser & Chan, 1980; Packard, 1986). Another study showed that Mandarin-speaking right hemisphere damaged patients were impaired relative to normal controls in their perception and production of affective prosody; however tone identification ability remained intact (Hughes,

Chan, & Su, 1983). This leftward lateralization for lexical tone perception is further supported by results from neuroimaging studies (Gandour, Wong, & Hutchins, 1998; Gandour, et al., 2000; Klein, Zatorre, Milner, & Zhao, 2001; Hsieh, Gandour, Wong, & Hutchins, 2001). For example, it was found by Klein and colleagues (2001) in a Mandarin tone discrimination task, while English speakers showed activity in the right inferior frontal cortex, Mandarin group had activation in multiple regions of the left hemisphere. Using a design of before- and post- training comparison, Wang and colleagues found that after a two-week training program of Mandarin tones, English speakers showed an increased activation in left superior temporal gyrus (STG, Wang, Sereno, Jongman, & Hirsch, 2003).

Findings from these studies indicate that when tone language speakers exploit pitch cues and identify words, circuitry in the left hemisphere is involved. Converging evidence from developmental studies have shown infants experience a perceptual reorganization that is influenced by their exposure to speech tones in the first 10 months of life (Nazzi, Floccia, & Bertoncini, 1998; Harrison, 2000; Mattock & Burnham, 2006). For instance, Mattock and Burnham (2006) found that Chinese infants performed equally well at 6 and 9 months for speech and non-speech tone discrimination. Conversely, discrimination of lexical tone by English infants declined between 6 and 9 months of age, while their non-speech tone discrimination remained constant. Furthermore, production of lexical tones is usually acquired in a short period of time between age 1 and 2, which is earlier than acquisition of segmental phonemes (Li & Thompson, 1977; Tse, 1978).

Taken together, the evidence supports the conjecture that if pitch information is associated with meaningful words in infancy, the brain circuitry in the left hemisphere that supports the association between pitch and verbal label should develop starting from infancy. As described in the previous section, these brain regions appear to also subserve AP. Therefore, native speakers of a tone language would have a much higher probability of possessing AP, while non-tone language speakers who do not have this early experience would only have implicit AP but cannot acquire the pitch-label association later in life (see also Deutsch, 2013).

Following this rationale, it is further hypothesized that speaking a tone language, which indicates an early and extensive experience of constructing and retrieving the associations between pitches and lexical labels, could facilitate the development of a precise and stable mental template for processing pitch information. Evidence in favor of this hypothesis is observed in two lines of research across speech and music domains.

The first line of evidence concerns pitch consistency in speech. When native speakers of tone languages (Vietnamese and Mandarin) were given a list of words to read out in two separate days, it was demonstrated that compared with English speakers, individual native speakers of tone languages showed significantly more pitch consistency across days, with the majority of the subjects had averaged pitch differences of less than 0.5 semitones (Deutsch, et al., 1999, 2004a). Building upon the literature showing pitch of speech is influenced by language and/or dialect spoken by individuals (Deutsch, et al., 1990; Dolson, 1994), it was further hypothesized that people who live in the same linguistic

community throughout their lives would acquire a pitch of speech that is consistent within the community but different from another community. Aiming at testing this hypothesis, a more resent study (Deutsch, Le, et al., 2009) collected female speech data from two groups of Chinese speakers living in two different isolated linguistic communities. These two villages located in a relatively remote area of China. The dialects spoken in the two villages are similar to Standard Mandarin and all subjects had learned to read and speak Standard Mandarin at school. Subjects read out a passage of roughly 3.25 minutes in Standard Mandarin, and pitch values were obtained at 5 ms intervals. The overall pitch levels in the two villages differed significantly from each other while having little variation within each group. The results lend support to the hypothesis that pitch levels of speech are influenced by a mental template of pitch that is acquired through long-term exposure to the speech of others.

Secondly, a number of direct-test studies (Deutsch, et al., 2006; Deutsch, Dooley, et al., 2009; Lee & Lee, 2010; Deutsch, et al., 2011) and survey studies (Gregersen et al., 1999, 2001) have confirmed a very high prevalence of AP in the population of tone language speakers, relative to the English-speaking population. For instance, Deutsch et al. (2011) administered a test of AP to 160 first and second year students at the Shanghai Conservatory of Music. The data showed the overall level of performance was very high. Not allowing for semitone errors, those who had begun musical training at or before age 5 showed an average correct rate of 83% not allowing for semitone errors and 90% allowing for semitone errors, those who had begun training at ages 6-9 showed an average

correct rate of 67% not allowing for semitone errors and 77% allowing for semitone errors, those who had begun training at age 10 or over showed an average correct rate of 23% not allowing for semitone errors and 34% allowing for semitone errors. Overall these findings support the hypothesis of a shared mechanism of pitch processing across music and speech domains, in which pitch height is associated with verbal label in a precise and stable manner.

## 1.5 Studies

Given there are multiple pitch cues that can be exploited in identifying lexical tones, these are intriguing questions to ask: what is the role of pitch height among many other cues that native speakers of a tone language utilize for identifying tones? How native speakers exploit pitch height cues in tone perception? As the pitch information in absolute pitch is the overall pitch height of a musical note regardless of contour, this becomes a critical question with respect to the theoretical framework arguing for the link between absolute pitch and lexical tone perception. The following experiments were designed to investigate these questions by separately examining two pitch height cues, namely overall pitch height and pitch height at critical points, which have been identified by Gandour (1983) to be among the few major pitch cues in lexical tone perception.

It is worth noting that all my studies used Mandarin as an example of tone language. Mandarin Chinese (i.e., Standard Chinese) is a tone language that is based on the particular Mandarin dialect spoken in Beijing with some lexical and

syntactic influence from other Mandarin dialects. Mandarin has four lexical tones that can be described by their F0 patterns as high-level, high-rising, low-dipping, and high-falling (Chao, 1968). The reason that Mandarin is chosen for the present experiments is largely due to the tone typology. Mandarin has more "contour tones" (i.e., three tones with pitch contours) than level tone (i.e., only one level tone) and is typically characterized as an example of "contour tone language" (Pike, 1948). Compared with using other East Asian languages, for example Cantonese that has three level tones, it is a more stringent test to use Mandarin in examining the role of pitch height in tone perception, particularly under the circumstance that native speakers of Mandarin have been found to assign more perceptual weights to pitch contour than pitch height (Chandrasekaran, et al., 2007).

### 1.5.1 Study 1: Overall Pitch Height as a Cue to Lexical Tone Perception

The literature on pitch perception across the domains of speech and music has produced findings suggesting that two factors, namely context effects (e.g., Moore & Jongman, 1997; Wong & Diehl, 2003; Huang & Holt, 2009) and perception of absolute pitch level (Deutsch, 2002, 2013) could account for the sensitivity of tone language speakers to the cue of overall pitch height.

#### *Context effects involving pitch height*

A few studies investigating the cue of overall pitch height in lexical tone perception have provided evidence that this is influenced by context. For example, it has been found, using paired Mandarin syllables, that the overall F0 of one syllable influences the judgment of the tone of the other syllable (Lin &

Wang, 1985; Fox & Qi, 1990). Albeit inconsistent in terms of the magnitude and nature of this context effect, both studies have shown that judgments of Mandarin Tones 1 and 2 were affected by the pitch heights of preceding or following syllables.

Context effects of pitch height are even more evident using the paradigm of embedding syllables in context sentences that are manipulated in pitch, both using Cantonese level tones (Wong & Diehl, 2003; Francis, et al., 2006), and also using Mandarin contour tones (Leather, 1983; Moore & Jongman, 1997; Huang & Holt, 2009). For example, Wong and Diehl (2003) found that depending on the pitch height of the context, especially that of the portion of a sentence immediately adjacent to the target syllable, the sound token was identified as a different Cantonese level tone; i.e., as a low tone when the pitch of the sentence was high, and a high tone when the pitch of the sentence was low. Because Cantonese has multiple level tones, this result is not surprising. However for Mandarin, which has only one level tone and multiple contour tones, it is not evident *a priori* that an effect of pitch context would also occur. Data from several experiments have provided supporting evidence that this is indeed so. Moore and Jongman (1997) examined perception of syllables with Tone 2 (mid-rising) and Tone 3 (low-falling-rising). These two tones have somewhat similar pitch contours but different pitch heights when presented in isolation. When they are preceded by sentences recorded from two speakers with different mean F0s, sound tokens were identified as Tone 2 (i.e., the tone with a mid level pitch height) in a low F0 speaker context but as Tone 3 (i.e., the tone with a low pitch

height) in a high F0 context. Huang and Holt (2009), using Mandarin high-level and mid-rising tones (Tones 1 and 2), have demonstrated a similar contrast effect of the pitch height of context on tone judgment. They also found that speech and non-speech precursors produced similar results.

A context effect of pitch height, in a broader sense, was also found in experiments that employed stimuli consisting of isolated syllables produced by multiple speakers, both in tone language (Wong & Diehl, 2003; Lee, 2009; Moore & Jongman, 1997) and in nontone language (Honorof & Whalen, 2005). In one experiment, Wong and Diehl (2003) found that native speakers of Cantonese could identify isolated Cantonese level tones both when they were presented as blocked by speaker and also when the speakers were mixed within a block. However the subjects performed significantly better when the tokens were blocked by speaker, compared to when they were mixed across speakers. This finding suggests that pitch range of the speaker provides important context information in the identification of isolated lexical tones. In Honorof and Whalen (2005), the listeners, who were native speakers of English, listened to isolated syllables that were produced by several speakers, and were asked to judge where the pitch height of each token lay within a speaker's pitch range. They found a high correlation between assigned rankings and locations of the tones within the speaker-specific tessitura, which indicates that the listeners were able to locate an F0 reliably within a range without much prior exposure to a speaker's voice.

Based on these studies, we can surmise that native speakers of tone language exploit the cue of overall pitch height for tone identification by

normalization based on the pitch of the context sentence or the speaker's tessitura. However, although some findings have used isolated or segmented tones to examine the effect of pitch normalization across multiple speakers (Wong & Diehl, 2003; Lee, 2009; Moore & Jongman, 1997), no study has yet focused on the effect of overall pitch height on lexical tone perception by manipulating the pitch height of isolated syllables that are produced by a single speaker.

### *Absolute pitch and tone perception*

In tone language speakers without musical training, high pitch consistencies have been found for the production of lexical tones, both for the same individuals across time (Deutsch, et al., 2004a), and also across individuals who speak the same dialect (Deutsch, Le, et al., 2009. These findings support the hypothesized framework that speakers of tone languages acquire a mental template for processing speech-related pitch information through early and consistent exposure to speech in their linguistic communities (Deutsch, 2002, 2013; Deutsch et al., 2004a).

Because absolute pitch possessors utilize the overall pitch of single musical tones in making tone identification judgments, the hypothesized relationship between absolute pitch and lexical tone perception would gain support from evidence showing that native speakers of tone language exploit the overall pitch of a lexical tone as a cue to retrieving its label, particularly when the tone is presented as an isolated syllable.

### *Musical training and tone perception*

The benefit of musical training to the processing speech sounds in nontone language speakers has been well documented (for reviews, see Kraus and Chandrasekaran, 2010; Patel, 2008). However, most musicians and non-musicians that have been studied were speakers of nontone languages and only a few studies have investigated the relationship between lexical tone perception and musical training in native speakers of tone languages (Lee & Lee, 2010; Mok & Zuo, 2012). Lee and Lee (2010) tested Mandarin-speaking musicians for absolute pitch and on lexical tone identification. In the tone identification task, subjects were asked to identify brief Mandarin stimuli that were devoid of dynamic F0 information, so that the subjects had to rely heavily on perception of pitch height in identifying tones. Their findings suggest that among native speakers of tone language, absolute pitch possessors do not necessarily have enhanced performance on lexical tone perception compared with nonpossessors. Further, using an AX discrimination task involving Cantonese tones, Mok and Zuo (2012) found that musicians who were native speakers of English discriminated tones more accurately than did their non-musician counterparts. However, among native speakers of Cantonese, those who were musically trained did not perform better or faster on this task than did those who lacked musical training.

Deutsch and colleagues (Deutsch, 2002, 2013; Deutsch et al., 2004) have hypothesized that, for musicians who are native speakers of tone language, the association between pitches and verbal meaning would have been acquired early in life as a feature of speech. However those who are native nontone language

speakers would not have acquired the appropriate circuitry during the speech-related critical period. For the latter group, musical training, if initiated early in life, would influence the development of pitch-label associations, and so enhance performance in lexical tone perception. The question then arises as to whether musical training also strengthens this association in tone language speakers, since for such speakers this association had already been developed in the process of acquiring a tone language. In other words, since the brain circuitry for identification of lexical tones would already be in place, it is not necessarily predicted from this hypothesis that musical training would further enhance performance for tone language speakers.

### *Motivations of the present study*

First, the present study examined the question: To what extent do native speakers of Mandarin use the cue of overall pitch height for identification of lexical tones? This question was examined using a new paradigm, in which tokens consisting of isolated Mandarin tones were transposed to different levels of pitch height (all within a speaker's tessitura) and native speakers of Mandarin were asked to identify them. This paradigm is novel in two respects. First, instead of embedding the tokens in the context of sentences or phrases, isolated syllables are presented at different levels of pitch heights so as to create a scenario that resembles the perception of single musical tones in tests of absolute pitch. Second, while most studies using isolated tones have focused on the effect of speaker normalization (e.g., Wong & Diehl, 2003; Lee, 2009), the present study instead limits the pitch range of the presented tones to a single speaker's tessitura.

In this way, issues concerning speaker identification and normalization are avoided.

The measures used in this study to evaluate the perception of lexical tones that had been transposed to different pitch levels were correct identification rate and reaction time. We note that reaction time paradigms have been used by others to study the perception of lexical tones (Chen & Cutler, 1996; Lee, et al., 2008). In particular, it has been argued that responses to tones made under time pressure reflect true perceptual processes, and are relatively free from response strategies or post-perceptual processes (Cutler, Dahan, & van Donselaar, 1997).

As a subsidiary issue, we examined whether musical training influences the pitch height component of lexical tone perception in native speakers of Mandarin. For speakers of nontone language, musical training would be expected to enhance perception of both pitch height and contour. On the other hand, if associations between pitches and their verbal labels have already been developed during the process of acquiring a tone language, an effect of musical training on the pitch height component is not necessarily predicted.

### 1.5.2 Study 2: On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements

The present study employed the "visual world" paradigm to investigate the hypothesis that fine-grained pitch cues are exploited in an incremental fashion in the pre-lexical perception of isolated lexical tones, in which pitch height at the critical points of the syllable (i.e. onset, turning point, and offset) serves as an important cue.

### *Influence of Endpoint and Midpoint Pitches on Tone Perception*

In several studies employing the paradigm of removing part of a tone (Gottfried & Suiter, 1997; Lee, et al., 2008; Liu & Samuel, 2004), it was found that native speakers of Mandarin were able to correctly identify most of the tones with which they were presented, even with only a small proportion of the sound signal available. Adopting the paradigm from the study of Strange, Jenkins, and Johnson (1983) on vowel identification, Gottfried and Suiter (1997) investigated tone perception by native and non-native speakers of Mandarin using four types of syllable: intact, center-only (without the initial six and final eight pitch periods), silent-center (only the initial six and final eight pitch periods are available), and onset-only (only initial six pitch periods are available). Native speakers outperformed non-native-speakers overall, especially for the silent-center syllables. The percentage average correct tone identification for the silent-center syllables was approximately 94% for native speakers, compared to 64.5% for non-native speakers. As a replication and extension of this study, Lee, et al., (2008) carried out a tone identification study using the same types of stimuli and a larger sample size consisting of 40 native speakers, and obtained a similar result. Using a different tone language, Thai, Zsiga and Nitisaroj (2007) demonstrated in a series of experiments that pitch inflections at the syllable midpoint and offset point successfully categorized Thai tones in perceptual space. The findings from these studies provide evidence that native speakers of tone languages use midpoint and endpoint pitches as perceptual cues for tone identification.

Several studies have investigated the cues of pitch change and the timing of the pitch turning point in Tones 2 (high-rising) and Tone 3 (low-falling-rising). X. Shen, Lin and Yan, (1993) found that the timing of the pitch turning point was an important cue for discriminating between Tone 2 and Tone 3. A turning point that occurred close to the tone onset served as a cue for Tone 2, and one that occurred late in the tone prompted more judgments of Tone 3. Findings by Moore and Jongman (1997) showed that both timing of pitch turning point and pitch difference between onset and turning point influenced perception of isolated tones. The stimuli were more likely to be identified as Tone 2 words when they had an early turning point and a small pitch difference between onset and turning point. Overall, perception of Tone 2 seemed to tolerate more variability, while Tone 3 required a late turning point and a large initial fall in F0.

The above studies all focused on the pitch information between onset and turning point for Tones 2 and 3, and the importance of the later parts of the tones has rarely been investigated. Liu and Samuel (2004) provided data on percentage correct identification for tone stimuli that had a portion of the syllable replaced by white noise (either from syllable onset to pitch turning point, or from pitch turning point to syllable offset). Their findings showed that for Tone 2, the information contained in the segment between the turning point and the offset point was critical for identification, while this phenomenon did not hold for Tone 3. Further, it was found that for Tone 3 the first half of the syllable was more important for identification than the second half. While this finding aligns with others in suggesting that some points or segments within the syllable contain

particularly salient pitch information for tone identification, it also raises the possibility of combining segments from different tones to create hybrid tones as stimuli for investigating the perception of tones.

### *Slope of Pitch Change*

The existing literature suggests that, along with pitch height, slope of pitch change is a perceptual cue used by native speakers for tone identification (Gandour, 1983; Massaro, Cohen & Tseng, 1985; Chandrasekaran, Gandour, et al., 2007). Research on lexical tone perception has consistently shown that native speakers perceive pitch glides categorically, with pitch patterns across the categorical boundary perceived as different tones, while those within the boundary perceived as the same tone (Francis, et al., 2003; Xu, Gandour, & Francis, 2006; Halle, et al., 2004). For example, modeling Mandarin high level and high rising tones, Xu et al. (2006) resynthesized a series of speech sound tokens carrying pitch glides with the same offset pitch height but different linear rising slopes. They found that native speakers identified sound tokens with slopes of pitch change larger than 0.047 Hz/ms as high rising tone, while those with shallow slopes as high level tone. Similar results were reported in two other studies using different tone languages and subject groups: Cantonese rising and falling tones by Francis et al., (2003), and Taiwan Mandarin level, rising, and falling tones by Halle et al., (2004).

In research on Mandarin tone perception, a continuum based on high level and high rising tones (Tones 1 and 2) has been frequently used for creating stimuli in identification and discrimination tasks (Chan, et al., 1975; Wang, 1976;

Xu, et al., 2006). However, there has not been any investigation of the rising slopes between pitch turning point and offset in Tones 2 and 3 to determine whether a steep slope of pitch change alone can prompt the judgment of Tone 2 and the lack of a steep slope can serve as a cue for Tone 3.

### *Motivations of the present study*

The existing tone perception literature has mostly employed paradigms in which subjects identified and/or discriminated between tones after the tokens were presented. By manipulating sound to control for different acoustic cues, this method can provide information regarding the acoustic cues that influence tone perception by native speakers. However, because these cues usually co-vary (e.g., pitch height at critical points can co-vary with the slope of pitch change), measuring only the final tone judgment has made it difficult to tease apart the effects of several different factors. For example, the experiment performed by Moore and Jongman (1997) involved three pitch cues: 1) pitch height at onset (with pitch height at turning point kept consistent), 2) pitch difference between onset and turning point, 3) the slope of pitch change between the onset and turning point. Changing one of these cues in such cases affects the other two, so making it difficult to know which cue is most influential in the final tone identification.

Another intriguing question is: how are these pitch cues utilized by native speakers of tone languages in perceiving lexical tones? Are these pitch cues used in an incremental way and change the cumulative evidence for identifying the tone as the speech sound unfolds? Or are they processed in a holistic manner so

as to make the judgment after the entire syllable has been presented? The traditional paradigm that only records off-line responses cannot provide an answer to this question.

Taken together, the problem and the question demand an on-line paradigm that can reflect any instant change of tone judgment before the syllable ends. The "visual world" paradigm is a good choice for tackling this problem. Because of the incremental fashion in which listeners recognize spoken words and comprehend speech, this eye tracking method has been used to study on-line speech perception (Rayner & Clifton, 2009). In a typical "visual world" study, the subject follows instructions to either complete certain tasks with one of a small set of objects presented in a visual workspace while receiving auditory stimuli (see Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) or simply hears spoken sentences and views the objects on a screen without performing any explicit task (see Huettig & Altmann, 2007). There are two basic components of eye movements in such a task. Fixation refers to the period of time when the eyes remain still at a location and extract visual information. Saccades are the movements themselves, during which there is no new information acquired because vision is suppressed (Matin, 1974). In this task, either eye fixation and covert attention are to the same location (i.e., during a fixation), or attention precedes the eyes to the next saccade location (Rayner, 2009). The subjects are pre-exposed to a set of visual objects before the sound stimuli are presented, and then make saccades and fixations on these objects while the speech sound is playing. When the data are aggregated across a large number of trials, there are,

overall, a high proportion of fixations on the object that corresponds to the auditory input at a certain time point. The timing and pattern of fixation to potential referents can be used to draw inferences about perception and comprehension, as the amount of fixation on a visual object is considered to be a reflection of the amount of neural activation associated with the word (Tanenhaus, Magnuson, Dahan, & Chambers, 2000). This paradigm has lately been adapted to study sub-phonetic processing in speech perception, so as to investigate how fine-grained acoustic differences can be used to map speech sounds into phonemes and words (McMurray, Tanenhaus, & Aslin, 2002; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). McMurray et al. (2002) used the "visual world" paradigm to demonstrate the gradient effect of voice onset time (VOT) on lexical activation. They used pairs of words that only have initial consonants which differed in VOT (e.g., /b/ vs. /p/) and synthesized a series of words with a nine-step VOT continuum. The subjects were instructed to click on the objects they heard while their eye movements were monitored. It was found that as VOT approached the categorical boundary, even though the subjects eventually selected the target objects, the fixations on the lexical competitor that differed in voicing increased. Dahan et al. (2001) created subcategorical mismatches by using cross-spliced words that contained mismatched acoustic information between vowels and consonants. The eye fixation data showed a significant effect on on-line lexical activation, coming from the fine-grained coarticulatory information at the onset of the target word. These studies demonstrated the effectiveness of the "visual world" paradigm for examining time-sensitive

phonetic and lexical processing in response to the fine-grained acoustic information in the speech sound.

Although this paradigm has been used intensively in studying spoken word recognition, only a few studies have so far examined on-line processing of lexical tones. Malins and Joanisse (2010) employed the "visual world" paradigm to examine how tonal versus segmental information influence spoken word recognition in Mandarin Chinese. By comparing the time course of viewing items that share the same segmental information versus items that share the same tonal information, they concluded that segmental and tonal information are accessed concurrently, and play a comparable role in Mandarin word recognition. Also using the "visual world" paradigm, Speer and Xu (2005) examined the effect of lexical tones on word recognition during the comprehansion of continuous speech. Listeners' eye movements were monitored as they listened to Mandarin sentences that contained ambiguous word sequences resulting from the operation of the third tone sandhi rule. Their findings suggest a very early use of tonal information in identifying the words. While extending the spoken word recognition literature by taking tonal information into account, these studies focused on the comparison of tonal and segmental processing as parallel mechanisms in spoken word recognition. However, no study has so far investigated on-line processing of pitch-tone mapping by manipulating the acoustic cues in speech sounds.

The motivation of the present study was two-fold. First, building on the literature concerning pitch cues for tone perception, we examined how native

speakers utilize the two acoustic cues (pitch height at critical points, and slopes of pitch change) for discriminating Mandarin Tones 2 and 3. Along with the manipulation of F0 at syllable onset, turning point and offset, the eye tracking paradigm enabled us to determine native speakers' response to acoustic information at specific time points within the syllable; this provided valuable information concerning how each of these pitch cues influenced tone judgments.

Second, the study extended research on spoken word recognition to investigate the processing of fine-grained acoustic information prior to lexical access. Lexical tone serves this goal nicely in the sense that the lexical meaning of syllables in a tone language differs only depending on pitch information given that vowels and consonants are held constant. As the "visual world" paradigm presents data on a timescale of milliseconds, it satisfies the need to examine on-line perceptual responses to instantaneous pitch changes in a tone as the sound unfolds. In the present study, the processing of the pitch information was revealed in a dynamic way by examining the time pattern of attention (i.e., measured by eye fixations) directed to potential referents, and how this changed with the unfolding of the speech sound.

Both experiments had the same design. The factor of offset pitch had four levels: original high, ambiguous high, original low, and ambiguous low. The four conditions were termed *High Tone 2* condition (with original high offset pitch), *Low Tone 2* condition (with ambiguous high offset pitch), *High Tone 3* condition (with ambiguous low offset pitch), and *Low Tone 3* condition (with original low

offset pitch). In all the conditions, the onset and turning point pitch were set to have the same low pitch values representing those pitch cues in Tone3.

The present study was motivated by two hypotheses. First, the data of Gottfried and Suiter (1997) and Lee et al. (2008) showed that native speakers were able to accurately identify tones with only onset and offset pitch information. It is predicted on this hypothesis that low onset pitch should give listeners an initial impression of Mandarin Tone 3. After the entire syllable is heard, a low offset pitch should confirm the earlier choice of Tone 3 while a high offset pitch should change most of the final judgments to Tone 2, given the finding that the second half of the tone is more important for identifying Tone 2 (Liu & Samuel, 2004).

Second, because the offset pitch values of High Tone 2 and Low Tone 3 were taken from the particular speaker's pitch of speech, these two conditions served as the original conditions. The two ambiguous conditions were created by increasing the offset pitch in the Low Tone 3 condition by one semitone, and by decreasing that of the High Tone 2 condition by one semitone. J. Shen et al. (2011) found that a difference of merely 1.5 semitones in overall pitch height significantly influenced how Tone 3 was identified. If the listeners in the present study responded to small differences in offset pitch height in the same way, the final tone judgment should differ for the original and ambiguous conditions.

Due to the close correspondence between orthographic and phonological information in alphabetic languages, studies using the "visual world" paradigm typically use pictures of objects as visual stimuli (e.g., Tanenhaus et al., 1995;

Huettig & Altmann, 2007). Experiment 1 of the present study followed this paradigm to obtain results comparable to those in the existing literature. Additionally, the stimuli in the present study were in Chinese, which are logographic in nature with highly arbitrary symbol-sound correspondences (Wang, 1973; Zhou, Shu, Bi, & Shi, 1999). As the written form of words (i.e., characters) do not correspond to lexical tones in Chinese, characters could then serve as visual stimuli in the "visual world" paradigm. Although Chinese characters and pictures of objects have both been used in eye tracking paradigms to study tone perception (Speer & Xu, 2005; Malins & Joanisse, 2010), no study so far has employed these two paradigms to investigate the same questions. Experiment 2 of the present study aimed at testing the paradigm of using characters rather than pictures.

**Acknowledgement**

A part of the material on Study 1 has been submitted for publication. The dissertation author was the primary investigator and author of this paper. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Jinghong Le.

A part of the material on Study 2 is published in "On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements." Shen, J., Deutsch, D., and Rayner, K., Journal of the Acoustical Society of America, 2013, 133, 3016-3029. The dissertation author was the primary investigator and author of this material. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Keith Rayner.

# Chapter 2. Method

## 2.1 Study 1

### *Subjects*

Twenty-eight native speakers of Mandarin (7 males, 21 females) were recruited from East China Normal University (ECNU) in Shanghai, China. They were all undergraduate or graduate students at ECNU. The average age of the subjects was 22.1 (standard deviation = 2.3). Based on findings that the overall pitch of an individual's speaking voice is influenced by the language and dialect in the speaker's linguistic community (Deutsch, et al., 1990; Deutsch et al., 2004a; Deutsch, Le, et al., 2009), it was considered important that all subjects should come from the same area. In the present study, all subjects were born in North/Northeast China (Beijing City, Heilongjiang Province, Jilin Province, and Liaoning Province) and had lived there until at least age 16. Among these subjects, 15 were musically untrained; 5 had 1-2 years of musical training; and 8 had 5-14 years of continuous musical training from ≤ age 9. No subject reported any hearing or speech disorders, and all subjects had normal or corrected to normal vision. Informed consent was obtained prior to the experiment, and the subjects were paid for their participation.

### *Stimuli*

The speech tokens consisted of 8 syllables, each instantiated in all four Mandarin tones, and with meanings in all the tones (see Table 2.1 for Pinyin, characters, and meanings of the syllables). They were initially recorded in a sound-attenuated booth by a female speaker from Heilongjiang Province.

**Table 2.1** Eight syllables used as stimuli and their corresponding characters.

| Syllable | Tone | Characters | English meanings |
|---|---|---|---|
| di | 1 | 滴 | drop |
| | 2 | 敌 | enemy |
| | 3 | 底 | bottom |
| | 4 | 弟 | brother |
| fu | 1 | 夫 | husband |
| | 2 | 福 | fortune |
| | 3 | 辅 | assist |
| | 4 | 附 | attach |
| jie | 1 | 街 | street |
| | 2 | 节 | festival |
| | 3 | 姐 | sister |
| | 4 | 界 | boundary |
| wan | 1 | 弯 | band |
| | 2 | 完 | finish |
| | 3 | 碗 | bowl |
| | 4 | 万 | ten thousand |
| wen | 1 | 温 | warm |
| | 2 | 闻 | hear |
| | 3 | 稳 | steady |
| | 4 | 问 | ask |
| yan | 1 | 烟 | smoke |
| | 2 | 言 | speak |
| | 3 | 眼 | eye |
| | 4 | 验 | examine |
| yi | 1 | 衣 | clothes |
| | 2 | 移 | move |
| | 3 | 以 | by |
| | 4 | 意 | meaning |
| yin | 1 | 音 | sound |
| | 2 | 银 | silver |
| | 3 | 引 | entice |
| | 4 | 印 | stamp |

The overall mean F0 of each initial speech stimulus was extracted using

Praat software (Boersma & Weenink, 2008) (see Figure 2.1 for the pitch patterns created by the four tones).

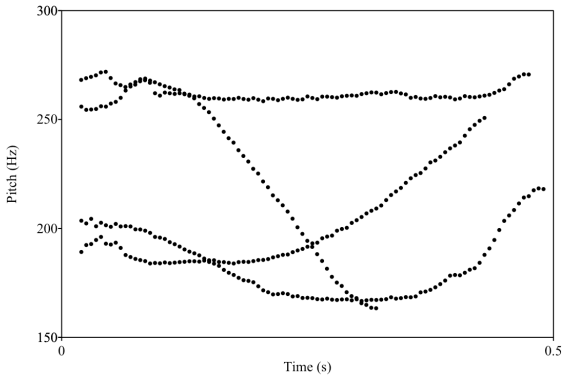


**Figure 2.1** Pitch patterns of the four tones at their original height levels

Five levels of pitch height, with 1.5 semitones between successive levels, were set for the resynthesized stimuli. For each of the four tones, there was one pitch height level that corresponded to the pitch height of the tone as it was originally recorded, and the remaining four levels were distributed within the speaker's pitch range (see Table 2.2).

**Table 2.2** Overall pitch height levels of sound stimuli.

| Pitch Height Level | Frequency (Hz) | Musical note | Original tone |
|---|---|---|---|
| High (H) | 261.6 | C4 | Tone1 |
| High-Mid (HM) | 240 | | |
| Mid (M) | 220 | A3 | Tone4 |
| Mid-Low (ML) | 201.7 | | Tone2 |
| Low (L) | 185 | F#3 | Tone3 |

The tokens were resynthesized from the natural speech token using the method of Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA, Moulines & Charpentier, 1990), keeping the formant patterns, amplitudes, and durations constant. All speech tokens were subjected to this procedure of pitch transposition and resynthesis, even those at their original pitch height.

To verify that the formant patterns were consistent across the conditions of original and transposed pitch height, F1, F2, and F3 data were obtained in 25 ms time windows and average formant frequencies were calculated for each of the sound tokens. Repeated-measures ANOVAs were carried out on each of the three average formant frequencies to test the effect of pitch height. All results were non-significant [F1: $F(4,28)=2.21$, $p>.1$; F2: $F(4,28)=0.9$, $p>.1$; F3: $F(4,28)=0.9$, $p>.1$].

***Procedure and Instrumentation***

The experimental software was developed using Matlab 7.8.0 (Mathworks, 2009). The subjects were tested individually in a quiet room using a HP Pavilion Elite HPE-170t desktop computer with a Creative Sound Blaster Audio PCI 128D sound card. The sounds were played at ~75 dB SPL through Creative HS400 headphones. The order of the trials was randomized for each subject. Each subject had 6 practice trials followed by 320 trials. A 10-minute break was placed half way through the experiment. On each trial, in order to establish attention on the center of the computer screen, the subject first viewed a fixation box (30 by 30 pixels) for 700ms. This was followed by the display of a Chinese character (300 by 300 pixels), which was kept on the screen until the subject made a

decision. The sound stimulus was played 300 ms after the onset of the visual display.

The subjects were informed that all the sound tokens had been produced by the same speaker, though they were not given examples of the original tones before they began making judgments. Their task was to respond as rapidly and accurately as possible, by pressing a key to indicate whether the sound token corresponded to the character on the screen. There were equal numbers of two types of trial in the experiment. In one type (termed "YES trials" hereafter, because correct responses were "yes" for these trials), the character displayed on the screen had the same tone (and the same syllable) as the sound token, the only manipulation being the overall pitch height of the tone. In the other type of trial (termed "NO trials" hereafter, because correct responses were "no" for these trials), the character had a different tone from the sound token, though the syllable was the same. The responses were collected on a HP mini numeric keypad. The keys numbered "1" and "3" were assigned as response keys, with the key-response assignment counterbalanced across subjects. The time allowed for response was limited to 2 seconds. The correct response rate and reaction time were recorded. In order to minimize any effect of outliers (Radcliff, 1993), any reaction time duration that was more extreme than ±2 standard deviations away from the mean reaction time (7.4% of all the trials) was excluded from all the data analysis. Furthermore, considering the fact that pitch information was not available during the consonant part of the syllable, the starting point for measuring reaction time was set as vowel onset, which was identified by

consulting both the waveform and the spectrogram generated by Praat software (i.e., at the point having the first detectable pitch value between 75-500 Hz, analyzed by the autocorrelation method).

## 2.2 Study 2

### 2.2.1 Experiment 1

#### *Subjects*

Twenty-four native Mandarin speakers (11 males, 13 females) were recruited from international and visiting students and scholars at the University of California, San Diego. The average age of the subjects was 26.8 (standard deviation = 4.1). All subjects were from Mainland China and had been living in the United States for a mean of 7.9 months and a range of 1 to 20 months. None of them had more than five years of musical training, and none had any recent musical training within the past ten years. No subject reported any hearing or speech disorders, and they all had normal or corrected to normal vision. Informed consent was obtained prior to the experiment and the subjects were paid for their participation. Three additional subjects were tested but their data were excluded from all the analyses due to excessive eye blinking (over 30% of trials).

#### *Stimuli and Instrumentation*

Speech tokens of 8 syllables, which have meanings for both Tones 2 and 3 in Mandarin, were recorded by a female native speaker in a sound attenuated booth using a Zoom H2 digital audio recorder, and saved as WAV files at a

sampling rate of 44.1 K and a 16 bit resolution. (See Table 3.1 for Pinyin of these syllables and name of the objects). None of these syllables had fricative consonants. The voiced portion of the sound tokens, which carried the pitch information, began no later than 30 ms into the syllable. The eye movement data corresponding to the unvoiced portion was discarded from the analysis, so the term "tone onset" is used hereafter to refer to the onset of the voiced portion of the syllable. All the syllables ended with either a vowel or a nasal consonant.

**Table 2.3** Sound stimuli used in Experiment1 (visual stimuli: object pictures).

| | Pinyin | Object |
|---|---|---|
| Tone2 stimuli | /bi/ | nose |
| | /lei/ | thunder |
| | /lian/ | curtain |
| | /ling/ | bell |
| | /liu/ | stream |
| | /wei/ | go (the game) |
| | /yan/ | rock |
| | /yu/ | fish |
| Tone3 stimuli (same syllables as Tone2 stimuli) | /bi/ | pen |
| | /lei/ | bud |
| | /lian/ | mask |
| | /ling/ | collar |
| | /liu/ | willow |
| | /wei/ | tail |
| | /yan/ | eye |
| | /yu/ | rain |
| Tone1 (distractors) | /che/ | car |
| | /dao/ | knife |
| | /deng/ | lamp |
| | /ding/ | nail |
| | /gou/ | hook |
| | /ji/ | chicken |
| | /shu/ | book |
| | /shua/ | brush |
| | /ti/ | ladder |
| | /zhong/ | clock |

(Table 2.3 Continued)

| | Pinyin | Object |
|---|---|---|
| Tone4 (distractors) | /chi/ | wing |
| | /dou/ | bean |
| | /jian/ | arrow |
| | /jing/ | mirror |
| | /ju/ | saw |
| | /mao/ | hat |
| | /pao/ | cannon |
| | /tu/ | rabbit |
| | /xiang/ | elephant |
| | /xie/ | crab |

Using Praat software (Boersma & Weenink, 2008), one pitch value of the speech stimulus was extracted every 10 ms. The onset, offset, and turning point F0s were set as the predefined values (see Table 3.2). All the other F0 values within the syllable were estimated by a parabolic interpolation method using Praat in order to retain the naturalness of the speech sound (Xu & Sun, 2001, see Figure 3.1). The tokens were resynthesized from the natural speech template using the method of Time-Domain Pitch-Synchronous Overlap-and-Add (TD-PSOLA, Moulines & Charpentier, 1990), keeping the formant patterns and amplitudes constant, and normalizing the duration to 500 ms (with pitch turning point at 200 ms).

**Table 2.4** Pitch of onset, turning point, and offset in four conditions (in Hz).

| Condition | Onset Pitch Height | Turning Point Pitch Height | Offset Pitch Height |
|---|---|---|---|
| Low Tone 3 | 184.9 | 138.6 | 155.6 |
| High Tone 3 | | | 164.8 |
| Low Tone 2 | | | 246.9 |
| High Tone 2 | | | 261.6 |

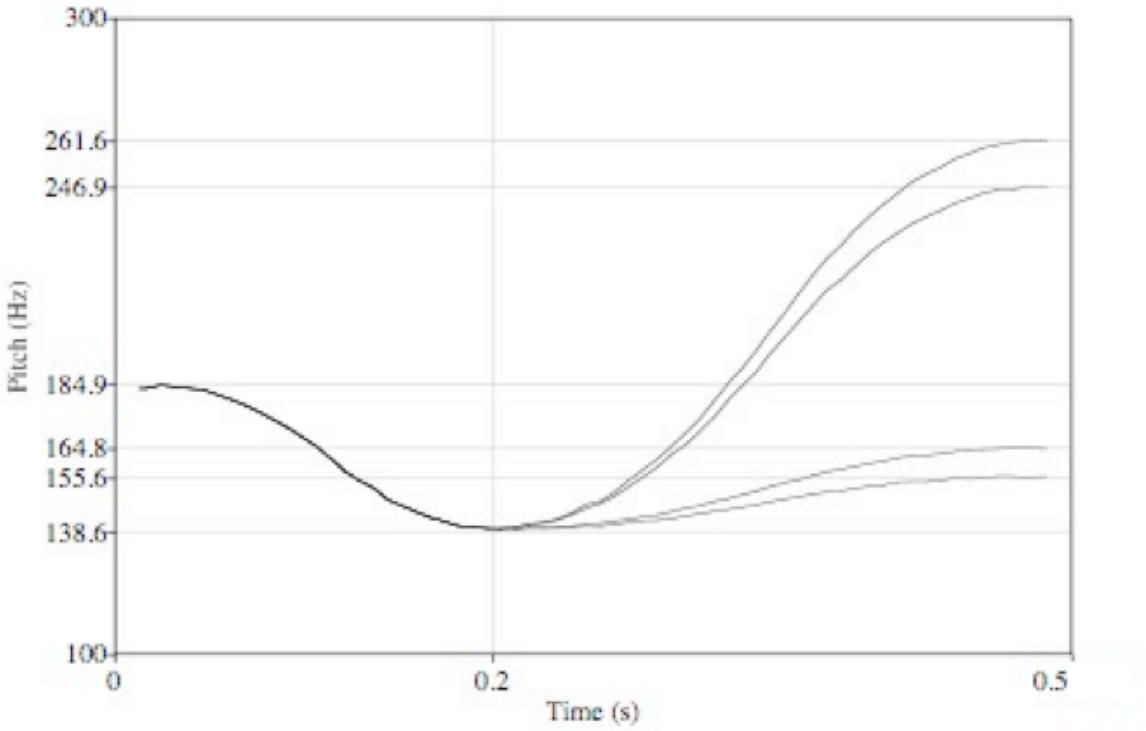


**Figure 2.2** Pitch patterns of tone stimuli in four conditions (offset pitch heights from top to bottom: *High Tone 2* condition, *Low Tone 2* condition, *High Tone 3* condition, *Low Tone 3* condition).

These speech tokens were first identified by 6 native speakers of Mandarin to ensure that they sounded natural enough to be identified as Mandarin Tones 2 and 3. These subjects identified tones with high offset pitches as Tone 2 in over 80% of trials, and stimuli with low offset pitches were identified as Tone 3 in over 90% of trials. No subject reported any unnaturalness in these stimuli.

Each sound token was presented twice during the experiment. To prevent subjects' anticipation of the experiment stimuli, the same number of filler trials

was created, in which speech tokens of 10 Tone 1 and 10 Tone 4 words were presented. None of the Tone 1 and Tone 4 words shared the same syllable as any of the Tone 2 and Tone 3 words and none of the Tone 1 words shared the same syllable with any of the Tone 4 words (See Table 1 for Pinyin of these syllables). All the words were normalized to the same peak intensity using the software Bias Peak Pro Version 5.2. They were then transferred to a Dell Precision 390 desktop computer and were presented at ~75 dB SPL through a BOSE Companion II speaker system. The visual stimuli were black line-drawings of objects that corresponded to these syllables. They were all resized to 200 by 200 pixels. All the visual stimuli were presented in black (rgb code 0-0-0) on a gray (rgb code 135-135-135) background on a 19-inch ViewSonic LCD monitor. The eye movement data were collected using a SR EyeLink1000 eye tracker (SR ltd., Canada) and a Dell Precision 390 desktop computer.

***Procedure***

The experimental software was developed using Matlab 7.8.0 (Mathworks, 2009). The subjects were tested individually. After arriving at the lab, the eye tracker in remote setup was calibrated with the standard 9-point calibration procedure. Viewing was binocular, but eye movements were recorded from the right eye only. The sampling rate of the eye tracker was set as 500 Hz. The order of trials was randomized for each subject.

The experiment began with two blocks of training trials to familiarize the subject with the names of the pictures. In the first block, each picture was displayed alone once together with its printed name. In the second block, a

printed picture name was displayed along with four candidate pictures. Subjects were required to click on the correct picture to advance to the next trial.

Ten practice trials that were identical to the test trials were given prior to beginning the formal experiment. Each trial started with a standard drift correction procedure, which measures how much the difference between a participant's fixation and a central point "drifts" over a short time period. Drift can occur because of factors such as fatigue and changes in body (head) position. Then a small black box (20 by 20 pixels) appeared at the center of the screen. The subjects were instructed to click on the box to activate the trial. Once the box was clicked on, four pictures, which were each 200 pixels (about 5 degrees of visual angle) in height and width, were displayed on the centers of the four quadrants of the screen. On each trial, the four pictures consisted of a pair of two pictures corresponding to Tones 2 and 3 objects, one Tone 1 object and one Tone 4 object. The Tone 2 and 3 objects shared the same syllable. A set of pictures of objects associated with the Tone 1 and Tone 4 words that were used in filler trials were presented as Tone 1 and Tone 4 objects. None of the Tone 1 and Tone 4 objects shared the same syllable as any of the Tone 2 and Tone 3 objects, and none of the Tone 1 objects shared the same syllable with any of the Tone 4 objects. (See Table 3.1 for Pinyin of these syllables and the names of the objects, and Figure 3.2 for an example of the visual stimuli). The locations of these pictures on the display were randomized. The sound stimulus, which consisted of the sound token preceded by a 700 ms prompt sentence of "Now click on __", was played simultaneously with the display of the pictures. The subjects were asked to

complete the task following the instruction given by the sound and to make their best decision with no time constraints.

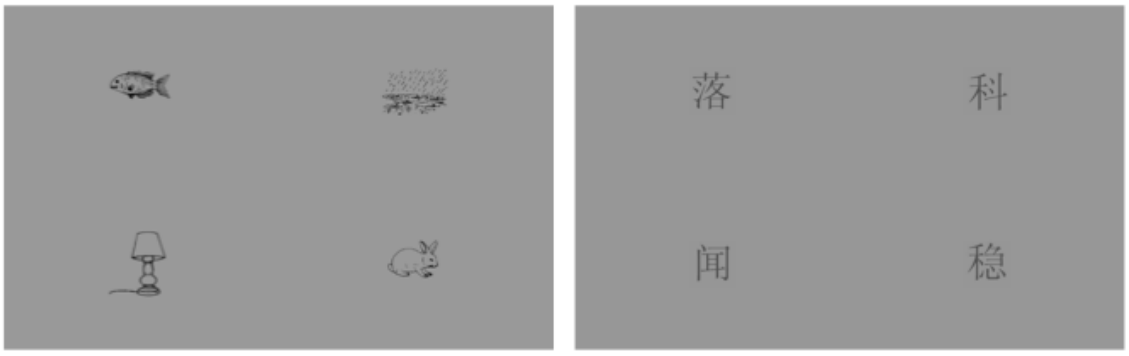


**Figure 2.3** Examples of the visual stimuli in the two experiments. (Left panel: 4 pictures (starting top-left clockwise) /yu2/, /yu3/, /tu4/, /deng1/ in Experiment1. Right panel: 4 characters (starting top-left clockwise) /luo4/, /ke1/, /wen3/, /wen2/ in Experiment2).

### 2.2.2 Experiment 2

#### *Subjects*

A different group of twenty-two native Mandarin speakers (13 males, 9 females) were recruited from international and visiting students and scholars at the University of California, San Diego. The average age of the subjects was 25.2 (standard deviation = 3.9). All were from Mainland China and had been living in the United States for a mean of 2.7 months and a range of 1 to 12 months. None of the subjects had more than five years of musical training, and none had had any recent musical training within the past ten years. None of them reported any hearing or speech disorders, and they all had normal or corrected to normal

vision. None of them had participated in Experiment 1. Informed consent was obtained prior to the experiment, and the subjects were paid for their participation. Two additional subjects were tested but their data were excluded from all the analyses due to excessive eye blinking (over 30% of trials).

### *Stimuli and Instrumentation*

The sound stimuli were created in the same way as in Experiment 1, except that 10 syllables were used to carry the Tone 2 and 3 tokens (see Table 3.4 for their Pinyin and the characters). The consonants of these syllables were controlled in the same way as in Experiment 1. The visual stimuli were black Chinese characters on a grey background. They included 10 Tone 3 characters, 10 Tone 2 characters that shared the same syllable with the Tone 3 ones, 10 Tone 1 characters, and 10 Tone 4 characters. None of the Tone 1 and Tone 4 characters shared the same syllable as any of the Tone 2 and Tone 3 characters, and none of the Tone 1 characters shared the same syllable with any of the Tone 4 characters. They were all controlled for visual complexity (i.e., number of strokes) and lexical frequency. The majority of the characters (85%) did not correspond to any object used in Experiment 1. They were typed in Chinese Song font and resized to 120 by 120 pixels.

The instrumentation was identical to Experiment 1 except that an Eyelink II head-mounted tracker (SR ltd., Canada) was used in this experiment.

**Table 2.5** Sound stimuli used in Experiment2 (visual stimuli: characters).

| | Pinyin | Character |
|---|---|---|
| Tone2 stimuli | /li/ | 离 |
| | /lian/ | 联 |
| | /mian/ | 棉 |
| | /miao/ | 苗 |
| | /wei/ | 围 |
| | /wen/ | 闻 |
| | /wu/ | 吴 |
| | /yan/ | 言 |
| | /yu/ | 鱼 |
| | /yuan/ | 员 |
| Tone3 stimuli (same syllables as Tone2 stimuli) | /li/ | 理 |
| | /lian/ | 脸 |
| | /mian/ | 免 |
| | /miao/ | 秒 |
| | /wei/ | 尾 |
| | /wen/ | 稳 |
| | /wu/ | 武 |
| | /yan/ | 眼 |
| | /yu/ | 雨 |
| | /yuan/ | 远 |
| Tone1 (distractors) | /chao/ | 超 |
| | /feng/ | 封 |
| | /jing/ | 京 |
| | /ke/ | 科 |
| | /mo/ | 摸 |
| | /pian/ | 偏 |
| | /qian/ | 铅 |
| | /tong/ | 通 |
| | /yin/ | 音 |
| | /zhuo/ | 桌 |

(Table 2.5 continued)

| | Pinyin | Character |
|---|---|---|
| Tone4 (distractors) | /dong/ | 洞 |
| | /gu/ | 固 |
| | /huo/ | 获 |
| | /jia/ | 架 |
| | /luo/ | 落 |
| | /na/ | 纳 |
| | /ruo/ | 弱 |
| | /sheng/ | 盛 |
| | /wang/ | 忘 |
| | /xi/ | 细 |

### *Procedure*

The experimental procedure was identical to that in Experiment 1, except that there was no name training block prior to the experiment.

**Acknowledgement**

A part of the material on Study 1 has been submitted for publication. The dissertation author was the primary investigator and author of this paper. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Jinghong Le.

A part of the material on Study 2 is published in "On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements." Shen, J., Deutsch, D., and Rayner, K., Journal of the Acoustical Society of America, 2013, 133, 3016-3029. The dissertation author was the primary investigator and author of this material. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Keith Rayner.

# Chapter 3. Results

## 3.1 Study 1

### *Overall Effect of Pitch Transposition*

Aggregated across all trials, the tokens at their original pitch height were responded to correctly on 98.6% of the trials, with a standard deviation of 2.5%. Tokens at transposed pitch heights were responded to correctly on 94.8% of the trials, with a standard deviation of 4.6%. To render these proportional data more suitable for analysis of variance, a rationalized arcsine transform (Studebaker, 1985) was performed to convert the proportions into R scores (in the unit of rau) before running them through the ANOVAs.

The overall difference between the original and transposed pitch height conditions was significant [$F(1,27)=76.1$, $p<.001$]. Detailed analyses examining each of the four tones separately showed that the differences were significant for Tone 1 [$t(27)=9.1$, $p<.001$], Tone 2 [$t(27)=5.7$, $p<.001$], and Tone 3 [$t(27)=6.6$, $p<.001$], but not for Tone 4 [$t(27)=1.4$, $p>.1$] (See Table 2.3 for the detailed data).

**Table 3.1** Means and standard deviations (in parenthesis) of correct response rate and reaction time data for the four tones.

| | Height level | Percent responded correctly | Reaction time of correct responses (ms) |
|---|---|---|---|
| Tone1 | Original | 99.4 (2.0) | 650 (87) |
| | Transposed | 95.0 (3.9) | 710 (97) |
| Tone2 | Original | 99.0 (2.4) | 671 (105) |
| | Transposed | 94.7 (5.2) | 692 (98) |
| Tone3 | Original | 97.9 (4.6) | 674 (112) |
| | Transposed | 91.9 (9.3) | 754 (106) |
| Tone4 | Original | 97.8 (3.7) | 640 (88) |
| | Transposed | 97.5 (3.6) | 647 (97) |

In order to minimize extraneous noise introduced by errors, many studies on lexical processing do not include trials with incorrect responses in the analysis of reaction time, if the correct response rate is high (e.g., Meyer & Schvaneveldt, 1971; Lee, et al., 2008). We followed this criterion in the present study and found that when their responses were correct, the subjects responded more rapidly to sound tokens at their original pitch heights (mean: 659 ms, standard deviation: 89 ms) compared to their responses to transposed tokens (mean: 701 ms, standard deviation: 97 ms, $F(1,27)=104.9$, $p<.001$). The differences were significant and substantial for Tone 1 [$t(27)=12.0$, $p<.001$], and Tone 3

[t(27)=7.3, p<.001], but were not significant for Tone 4 [t(27)=0.76, p>.1] or Tone 2 [t(27)=1.9, p>.05] (See Table III for the detailed data).

***"YES trials" versus "NO trials"***

Overall, the two types of trial (i.e., YES trials and NO trials) had comparable patterns of correct response and reaction time (see Figure 2.2 for graphs).

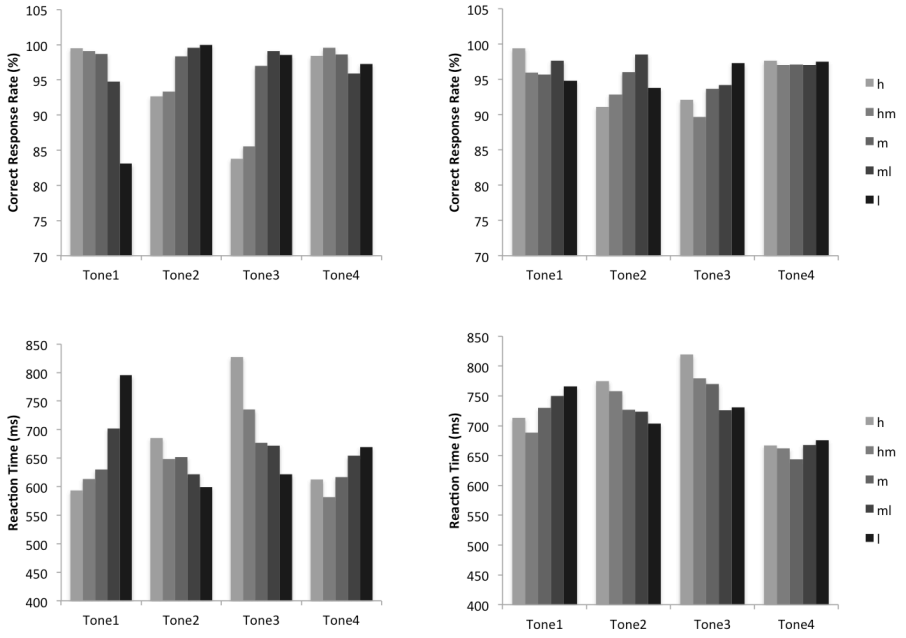


**Figure 3.1** Correct response rate and reaction time data in detail for "yes" and "no" trials (left top: correct response rate in YES trials; right top: correct response rate in NO trials; left bottom: reaction time in YES trials; right bottom: reaction time in NO trials)

For the case of correct response rate, both types of trial yielded comparable data. YES trials had a mean of 96.9%, standard deviation: 2.9%; NO trials had a mean of 96.4%, standard deviation: 4.8%; the difference between YES and NO trials was not significant statistically [F(1, 27)=0.85, p>.1]. However, consistent with other reaction time findings (e.g., Meyer & Schvaneveldt, 1971), YES trials had a much shorter average reaction time (mean: 643ms, standard deviation: 99ms) than NO trials (mean: 717ms, standard deviation: 91ms) [F(1, 27)=66.9, p<.001]. Interestingly, the difference between reaction times for YES and NO trials was significantly smaller in the transposed condition compared to the original condition [F(1, 27)=10.8,p<.01] . This is also shown in the top panel of Figure 2.3.

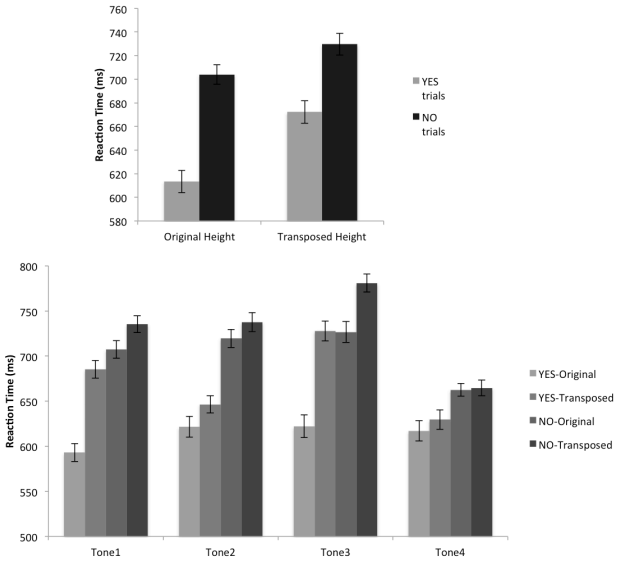


**Figure 3.2** Summarized correct response rate and reaction time data for “yes” and “no” trials (top panel: overall pattern, bottom panel: a breakdown by tones)

A more detailed examination of the data showed that this interaction was mainly attributable to responses to Tones 1 and 3. In YES trials, responses were slowed down dramatically for transposed Tones 1 and 3 [Tone 1: $t(27)=8.2$, $p<.001$, Tone 3: $t(27)=6.5$, $p<.001$] compared to nontransposed Tones 1 and 3 (see bottom panel of Figure 3). Transposition also slowed down the subjects' responses in NO trials significantly for Tones 1 and 3 [Tone 1: $t(27)=2.7$, $p<.05$, Tone 3: $t(27)=3.1$, $p<.01$], but the effects were less extreme than in YES trials. In contrast, transposition did not significantly affect reaction time for Tones 2 or 4, for either YES trials [Tone 2: $t(27)=1.3$, $p>.1$, Tone 4: $t(27)=0.98$, $p>.1$] or NO trials [Tone 2: $t(27)=1.7$, $p>.1$, Tone 4: $t(27)=0.17$, $p>.1$].

***Amount of Pitch Transposition***

The data revealed that responses were faster and more accurate when the pitch of the tone being judged was closer to its original pitch height, and gradually degraded as the pitch height of the tone was transposed (see Figure 2). To examine if performance level corresponded to degree of transposition, a linear regression line was fit on the data, with degree of transposition (measured by pitch difference from the original level of pitch height) as the predictor, and correct response rate, and reaction time as the dependent variables (see Figure 2.4 for the scatter plots).

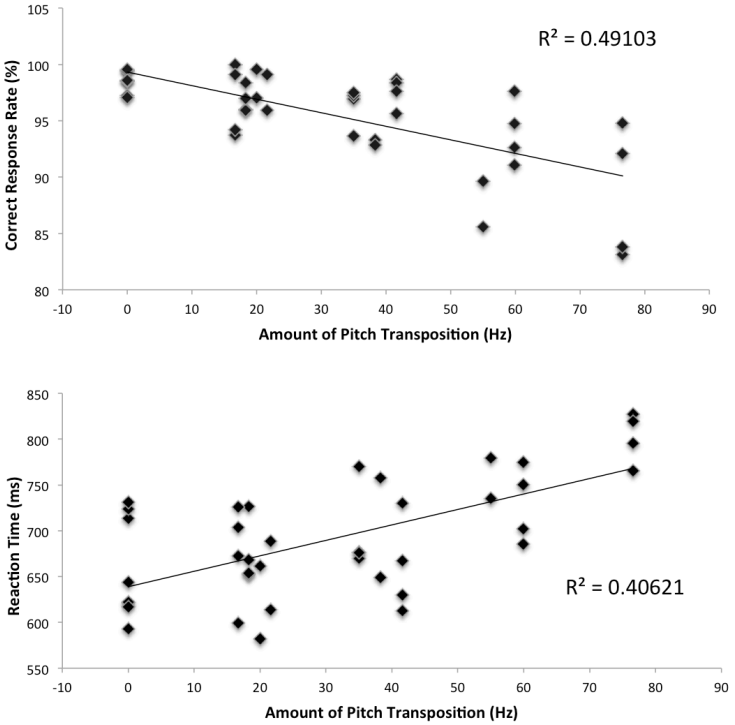


**Figure 3.3** Scatter plots showing the linear regression lines with amount of pitch transposition as predictor and performance level (top panel: correct response rate, bottom panel: reaction time) as dependent variables

The pattern of data demonstrated that the subjects' correct response rate was higher, and reaction time lower, when the degree of pitch transposition was smaller (i.e., the pitch height was closer to the original pitch height) and performance degraded with increase in the degree of transposition. A regression analysis showed that the degree of transposition explained a significant proportion of variance in both the correct response rate and the reaction time

data (correct response rate: $R^2$=0.491, F(1,39)=36.66, p<.001; reaction time: $R^2$=0.406, F(1,39)=25.99, p<.001).

### *Effect of Musical Training*

To examine whether musical training strengthens the association between pitch and lexical label, the percentage correct response and reaction time data were each evaluated for an effect of musical training, and its interaction with the effect of transposition. Because it has been shown that both age of onset of musical training is an important factor influencing the acquisition of absolute pitch (see Deutsch, 2013 for a review), only those subjects who had at least 5 years of formal training on a musical instrument or vocals, beginning ≤ age 9 were included in the "musically trained" group.

The difference between the musically trained and untrained subjects was not statistically significant for either correct response rate [F(1,21)=3.2, p>.05] or reaction time [F(1,21)=0.6, p>.1]. Furthermore, the interaction between musical training and transposition was nonsignificant for either correct response rate [F(1,21)=0.06, p>.1] or reaction time [F(1,21)=2.7, p>.1] (see Figure 2.5).

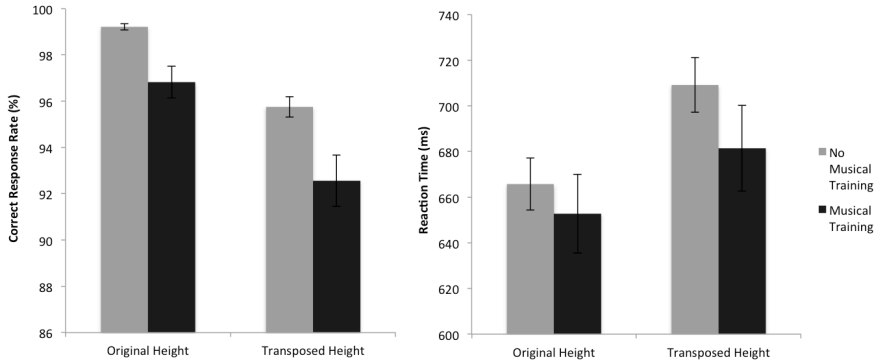


**Figure 3.4** Correct response rate and reaction time data (independent variables: musical training background, condition of pitch height)

## 3.2 Study 2

### 3.2.1 Experiment 1

#### *Tone Identification Data*

Analysis of the tone identification data showed that the subjects had an overall rate of correct response of 84.7% at the end of the syllables. The breakdown of the correct response rate for the 4 conditions were 99.7% (selecting Tone 3 in the *Low Tone 3* condition); 98.6% (selecting Tone 3 in the *High Tone 3* condition); 67.1% (selecting Tone 2 in the *Low Tone 2* condition); 74.2%

(selecting Tone 2 in the *High Tone 2* condition). For further analyses, the selected objects are termed "targets"; the objects associated with the same syllables as the targets but different tones are termed "competitors"; those Tones 1 and 4 objects that were presented on the same screen are termed "distracters".

To make these proportional data more suitable for analysis of variance, a rationalized arcsine transform (Studebaker, 1985) was performed to convert the proportions into R scores (in the unit of rau) before running them through the ANOVAs. In the two low offset conditions, subjects ultimately selected objects associated with Tone 3 more often than those associated with Tone 2, but in the two high offset conditions they selected more Tone 2 than Tone 3 objects. These differences were all significant [ps<.01], suggesting that the manipulation of offset pitch height and rising slope influenced tone identification. Furthermore, both tones were more frequently selected in the original conditions: Tone 3 was more frequently selected in the *Low Tone 3* condition compared with the *High Tone 3* condition [t(23)=2.04, p=.05], and Tone 2 was selected more often in the *High Tone 2* condition compared with the *Low Tone 2* condition [t(23)=1.68, p=.1]. This finding indicates native speakers can exploit subtle pitch cues of one semitone difference in making tone judgments.

### *Temporal Window Analysis*

The eye tracker recorded the subjects' fixation positions (i.e., coordinates on the screen) at 2 ms intervals. These coordinate data were then converted to one data point every 2 ms to show on which one of the four objects the eyes were fixated at that moment. To account for errors of calibration and drift in the eye

tracker, the four locations containing the objects were defined as four quadrants that were 640 pixels wide and 512 pixels high each.

Because the present study intended to examine the fixation data in those trials that the subjects made a tone judgment that was consistent with the pitch cue at the tone offset (i.e., selecting targets instead of competitors or distracters), those trials in which the subjects selected objects other than the targets (15.3%) were excluded from statistical analysis of the eye fixation data. The trials that contained blinks (8.3%) during playing of the sound tokens were also excluded from the analysis of the eye fixation data.

The number of fixations was counted in every 20 ms time window, which gives ten data counts per window to produce the total counts of eye fixations on the target, competitor, and the two distracters in each window. For each 20 ms window, the proportion of fixations (POF) data were then derived from the count data by dividing each of the four fixation count data (i.e., counts of fixation at target, counts of fixation at competitor, and counts of fixation at the distracters) by the sum of counts of four possible locations. The POF data were then collapsed over subjects and items to create four averaged trajectory graphs in 20 ms time windows (see Figure 3.3), for each of the four conditions.

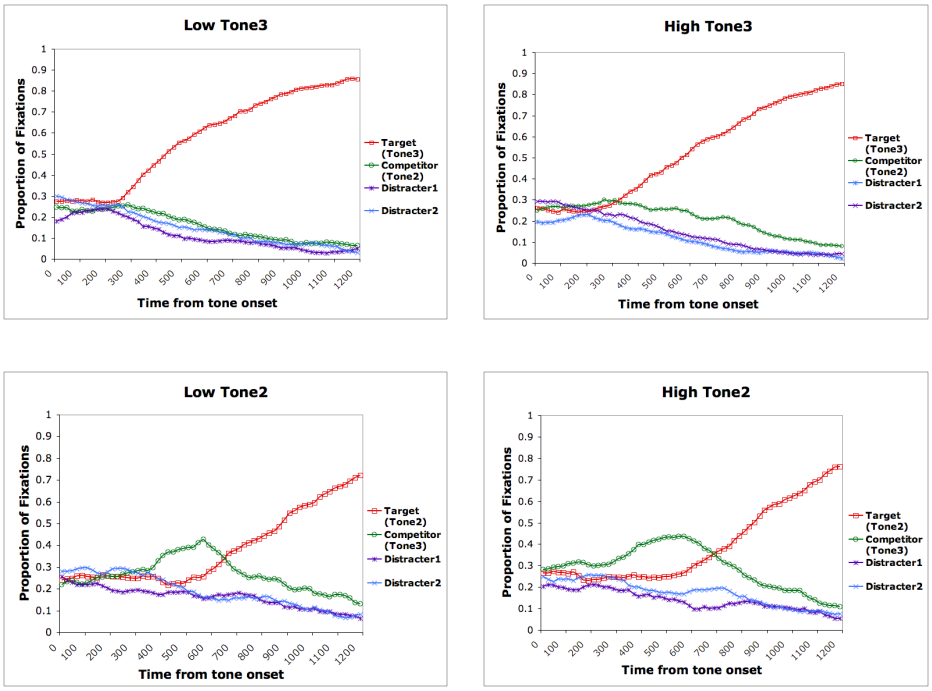


**Figure 3.5** Experiment 1 proportion of fixations curves in 20 ms interval (time scale in ms).

Prior eye tracking research suggests that, in the "visual world" paradigm, the time to program and execute a saccadic eye movement is approximately 150-200 ms or less (Rayner, 1998; 2009). To test the significance of the difference between fixation on targets and competitors while avoiding too many ANOVAs for each condition, the POF data were collapsed within four 150 ms time windows. The first window was taken as 301 ms to 450 ms from tone onset. The fixation data in this window should reflect the influence of the sound from tone

onset (i.e., 1-100 ms into the tone) to pitch turning point (i.e., till 250 ms into the tone) because of the 200 ms time delay for saccade planning and execution. Any effect from the pitch information up to 100 ms into the tone should fully reveal itself by the beginning of the first window. The last window was defined as 751-900 ms from tone onset. Considering the approximately 200 ms saccade planning and execution time, if pitch information at syllable offset serves as a cue for identifying the tones, the fixations should be significantly more likely to be on the target compared to the competitor by the beginning of the last window. The time spans of the 4 windows were: 301-450 ms; 451-600 ms; 601-750 ms; 751-900 ms. These 4 windows respectively reflect the processing of sound information during these time spans: 101-250 ms; 251-400 ms; 401-550 ms; 551-700 ms. The means and standard deviations of the POF data within these four windows are presented in Table 3.3.

**Table 3.2** Experiment 1 Means and standard deviations of proportion of fixations data in four 150 ms windows.

| | | Window 1 (301-450 ms) | Window 2 (451-600 ms) | Window 3 (601-750 ms) | Window 4 (751-900 ms) |
|---|---|---|---|---|---|
| Low Tone 3 condition | Target (tone3) | 0.426 (0.126) | 0.574 (0.128) | 0.665 (0.129) | 0.746 (0.146) |
| | Competitor (tone2) | 0.229 (0.106) | 0.178 (0.098) | 0.129 (0.087) | 0.101 (0.083) |
| | Distractor1 | 0.152 (0.068) | 0.010 (0.070) | 0.085 (0.077) | 0.069 (0.069) |
| | Distractor2 | 0.192 (0.095) | 0.148 (0.086) | 0.121 (0.073) | 0.083 (0.055) |
| High Tone 3 condition | Target (tone3) | 0.342 (0.101) | 0.463 (0.131) | 0.585 (0.144) | 0.690 (0.128) |
| | Competitor (tone2) | 0.279 (0.079) | 0.254 (0.092) | 0.216 (0.101) | 0.178 (0.097) |
| | Distractor1 | 0.168 (0.066) | 0.130 (0.056) | 0.085 (0.058) | 0.055 (0.058) |
| | Distractor2 | 0.211 (0.089) | 0.154 (0.096) | 0.113 (0.081) | 0.077 (0.055) |
| Low Tone 2 condition | Target (tone2) | 0.244 (0.130) | 0.246 (0.143) | 0.356 (0.152) | 0.458 (0.210) |
| | Competitor (tone3) | 0.318 (0.165) | 0.396 (0.162) | 0.317 (0.119) | 0.245 (0.153) |
| | Distractor1 | 0.184 (0.135) | 0.173 (0.122) | 0.173 (0.118) | 0.143 (0.116) |
| | Distractor2 | 0.254 (0.140) | 0.186 (0.109) | 0.154 (0.110) | 0.154 (0.143) |
| High Tone 2 condition | Target (tone2) | 0.247 (0.086) | 0.257 (0.132) | 0.343 (0.168) | 0.478 (0.173) |
| | Competitor (tone3) | 0.376 (0.128) | 0.430 (0.127) | 0.363 (0.154) | 0.250 (0.159) |
| | Distractor1 | 0.172 (0.107) | 0.140 (0.090) | 0.105 (0.090) | 0.125 (0.086) |
| | Distractor2 | 0.206 (0.109) | 0.173 (0.139) | 0.190 (0.122) | 0.146 (0.091) |

Repeated measures ANOVAs were carried out to examine if the POF (converted to R scores) on the four fixation locations (i.e., target, competitor, and

distracters) were different across the four windows. The main effect of the fixation locations and the windows were significant [location: $F(3,69)=124.07$, $p<.001$; window: $F(3, 69)=30.27$, $p<.001$]. Specifically, to test the difference between POF on targets and on competitors in each window, paired sample t tests were also undertaken on the basis of subject ($t_1$) and item ($t_2$) variability.

Results from the t tests showed that for the *Low Tone 3* condition, the POFs on the targets (objects associated with Tone 3 words) were significantly higher than those on the competitors (objects associated with Tone 2 words) for all four windows [Window 1: target mean = 0.426, competitor mean = 0.229, $t_1(23) = 4.6$, $p<.01$ and $t_2(7) = 3.38$, $p<.05$; Window 2: target mean = 0.574, competitor mean = 0.178, $t_1(23) = 9.14$, $p<.01$ and $t_2(7) = 5.31$, $p<.01$; Window 3: target mean = 0.665, competitor mean = 0.129, $t_1(23) = 11.49$, $p<.01$ and $t_2(7) = 5.98$, $p<.01$; Window 4: target mean = 0.746, competitor mean = 0.101, $t_1(23) = 11.81$, $p<.01$ and $t_2(7) = 11.07$, $p<.001$].

In the *High Tone 3* condition, a preference for targets over competitors started to be significant from Window 2 [Window 2: target mean = 0.463, competitor mean = 0.254, $t_1(23) = 5.2$, $p<.01$ and $t_2(7) = 3.82$, $p<.01$; Window 3: target mean = 0.585, competitor mean =0.216, $t_1(23) = 7.79$, $p<.01$ and $t_2(7) = 3.47$, $p=.01$; Window 4: target mean = 0.690, competitor mean = 0.178, $t_1(23) = 10.86$, $p<.01$ and $t_2(7) = 4.71$, $p<.01$].

In the *Low Tone 2* condition, a significant fixation difference between competitors and targets only began showing in Window 2 [Window 2: target mean = 0.246, competitor mean = 0.396, $t_1(23) = -2.34$, $p<.05$ and $t_2(7) = -1.77$,

p>.05]. The two curves crossed over in Window 3. In Window 4, judgments were securely locked on these later-decided targets, and the POF difference was once again significant, but in the opposite direction [Window 4: target mean = 0.457, competitor mean = 0.245, $t_1(23) = 3.11$, $p<.01$ and $t_2(7) = 5.24$, $p<.01$].

A pattern similar to the *Low Tone 2* condition was observed in the *High Tone 2* condition [Window 1: target mean = 0.247, competitor mean = 0.376, $t_1(23) = -3.79$, $p<.01$ and $t_2(7) = -2.05$, $p<.1$; Window 2: target mean = 0.257, competitor mean = 0.430, $t_1(23) = -4.17$, $p<.001$ and $t_2(7) = -2.49$, $p<.05$]. The two curves crossed over in Window 3. The POF on targets was significantly higher compared to the POF on competitors in Window 4 [Window 4: target mean = 0.478, competitor mean = 0.250, $t_1(23) = 3.73$, $p<.01$ and $t_2(7) = 4.26$, $p<.01$].

***Analysis of Diverging Points***

As these t-tests were carried out using data collapsed within 150 ms time windows, it is difficult to interpret the results in terms of how pitch information at specific time points influenced the POF data on targets and competitors. In order to reveal these effects in a fine-grained time scale, for each 2 ms time bin between 1-1200 ms from tone onset, 95% confidence intervals from each POF curve were obtained using a subjects-based bootstrap re-sampling procedure with 10,000 iterations (Efron & Tibshirani, 1993). The lower boundaries of the confidence intervals were then compared with chance (25%) to determine the time points at which POF on target or competitor was significantly above chance (see Figure 3.4).

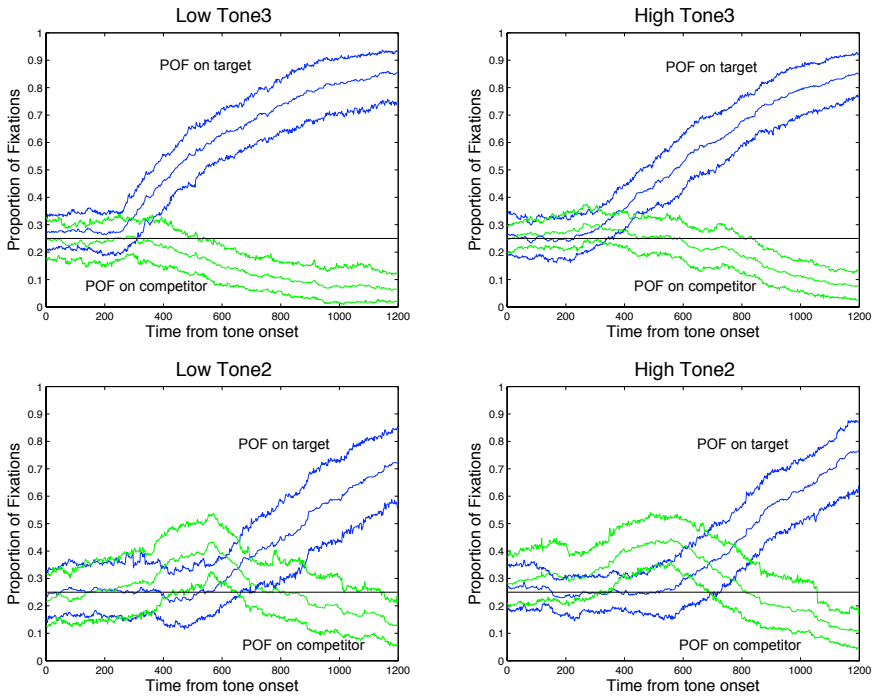


**Figure 3.6** Experiment 1 bootstrapping curves in 2 ms interval (time scale in ms). POF on target: proportion of fixations on targets; the middle line is the mean target POF; the other two lines are the upper and lower boundaries of the 95% confidence interval of target POF. POF on competitor: proportion of fixations on competitors; the middle line is the mean competitor POF; the other two lines are the upper and lower boundaries of the 95% confidence interval of competitor POF. Horizontal line is chance level (25%).

In the two low offset conditions, the lower boundaries of the confidence interval began to be above chance at 314 ms (in the *Low Tone 3* condition) and 346 ms (in the *High Tone 3* condition). In the two high offset conditions, the

lower boundaries of the confidence interval of POF on competitor began to be above chance at 434 ms (in the *Low Tone 2* condition) and 348 ms (in the *High Tone 2* condition). This data consistently demonstrate a significant above chance preference to objects associated with Tone 3 words, observed between 350-450 ms after tone onset. Considering the roughly 200 ms saccade latency before the fixations were observed, this suggests that the pitch information at 150-250 ms into the tone, which is around the turning point (around 200 ms into the tone), directed fixations to targets or competitors.

The lower boundary of POF on target (object associated with Tone 2 words) was above chance level at 668 ms (in the *Low Tone 2* condition) and 732 ms (in the *High Tone 2* condition). This result indicates the critical influence of the offset pitch on identification of the target, considering the approximately 500 ms tone duration in addition to the 200 ms saccade latency.

### 3.2.2 Experiment 2

#### *Tone Identification Data*

Analysis of the tone identification data showed that the subjects had an overall rate of correct response of 90.1% at the end of the syllables. The breakdown of the correct response rate for the four conditions were 95.5% (selecting Tone 3 in the *Low Tone 3* condition); 91.9% (selecting Tone 3 in the *High Tone 3* condition); 83.9% (selecting Tone 2 in the *Low Tone 2* condition); and 90.6% (selecting Tone 2 in the *High Tone 2* condition). For further analyses, the selected characters are termed “targets”; the characters associated with the

same syllables as the targets but different tones are termed “competitors”; the Tone 1 and 4 characters that were presented on the same screen are termed “distracters”.

Similar to the results of Experiment 1, Tone 3 characters were more frequently selected in the low offset conditions, and Tone 2 characters in the high offset conditions [ps<.01]. Tone 3 was also more frequently selected in the *Low Tone 3* condition compared to the *High Tone 3* condition [t(21)=2.26, p<.05], and Tone 2 was more often selected in the *High Tone 2* condition compared with the *Low Tone 2* condition [t(21)=5.09, p<.01].

***Temporal Window Analysis***

The eye fixation data were processed in the same way as in Experiment 1. Those trials in which the subjects selected characters other than the targets (9.9%) were excluded from further analysis of the eye fixation data. The trials that contained blinks (7.7%) during presentation of the sound tokens were also excluded from analysis of the eye fixation data. The POF data was collapsed over subjects and items, to create averaged trajectory graphs in 20 ms time windows (see Figure 3.5), for each of the four conditions.

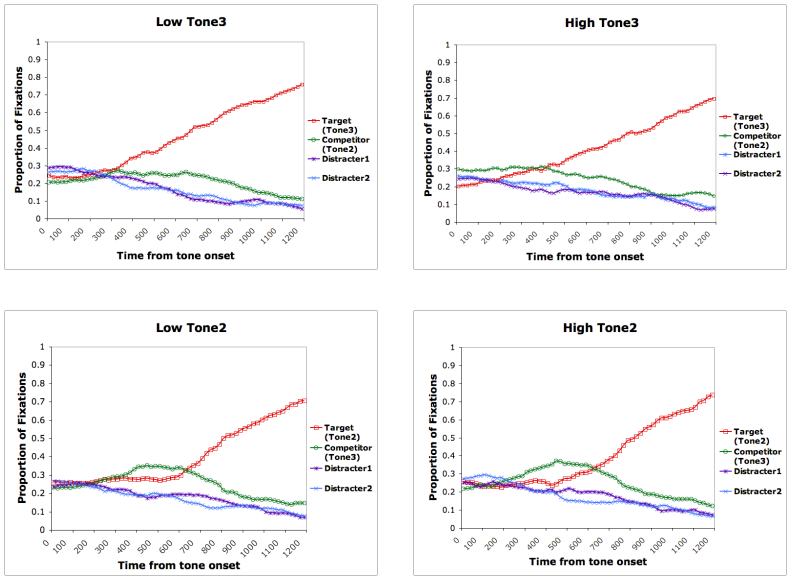


**Figure 3.7** Experiment 2 proportion of fixations curves in 20 ms interval (time scale in ms).

By visual inspection, the POF curves in this experiment appear to have a pattern similar to that of Experiment 1, so the same four windows were adopted for performing ANOVAs (The time intervals of these four windows are: 301-450 ms; 451-600 ms; 601-750 ms; 751-900 ms). The means and standard deviations of the POF data in these 4 windows for the 4 conditions are presented in Table 3.5.

**Table 3.3** Experiment 2 Means and standard deviations of proportion of fixations data in four 150 ms windows.

| | | Window 1 (301-450 ms) | Window 2 (451-600 ms) | Window 3 (601-750 ms) | Window 4 (751-900 ms) |
|---|---|---|---|---|---|
| Low Tone 3 condition | Target (tone3) | 0.321 (0.119) | 0.398 (0.122) | 0.495 (0.139) | 0.588 (0.153) |
| | Competitor (tone2) | 0.260 (0.088) | 0.250 (0.083) | 0.248 (0.094) | 0.210 (0.109) |
| | Distractor1 | 0.229 (0.087) | 0.082 (0.066) | 0.118 (0.059) | 0.091 (0.046) |
| | Distractor2 | 0.190 (0.078) | 0.169 (0.073) | 0.139 (0.082) | 0.111 (0.078) |
| High Tone 3 condition | Target (tone3) | 0.300 (0.089) | 0.359 (0.102) | 0.430 (0.127) | 0.502 (0.157) |
| | Competitor (tone2) | 0.304 (0.084) | 0.271 (0.070) | 0.249 (0.108) | 0.199 (0.096) |
| | Distractor1 | 0.217 (0.102) | 0.197 (0.108) | 0.158 (0.083) | 0.146 (0.097) |
| | Distractor2 | 0.180 (0.077) | 0.174 (0.086) | 0.164 (0.080) | 0.153 (0.093) |
| Low Tone 2 condition | Target (tone2) | 0.280 (0.074) | 0.280 (0.086) | 0.361(0.133) | 0.499 (0.151) |
| | Competitor (tone3) | 0.320 (0.108) | 0.343 (0.080) | 0.305 (0.081) | 0.221 (0.091) |
| | Distractor1 | 0.205 (0.070) | 0.187 (0.067) | 0.189 (0.078) | 0.154 (0.088) |
| | Distractor2 | 0.195 (0.085) | 0.190 (0.198) | 0.144 (0.072) | 0.127(0.080) |
| High Tone 2 condition | Target (tone2) | 0.252 (0.077) | 0.284 (0.085) | 0.361 (0.113) | 0.505 (0.156) |
| | Competitor (tone3) | 0.334 (0.097) | 0.356 (0.074) | 0.305 (0.072) | 0.214 (0.076) |
| | Distractor1 | 0.209 (0.086) | 0.205 (0.080) | 0.190 (0.077) | 0.144 (0.074) |
| | Distractor2 | 0.205 (0.075) | 0.155 (0.069) | 0.143 (0.079) | 0.137 (0.093) |

The same repeated measures ANOVAs on the transformed POF data showed that the POF on the four fixation locations (i.e., target, competitor, and two distracters) were different across the four windows [location: $F(3,63)=79.81$,

p<.001; window: F(3, 63)=37.93, p<.001]. Paired sample t tests that were identical to those in Experiment 1 were carried out on the basis of subject ($t_1$) and item ($t_2$) variability.

The results of t tests showed that for the *Low Tone 3* condition, POF on the targets (Tone 3 words) were significantly higher than those on the competitors (Tone 2 words) beginning on Window 2 [target mean = 0.399, competitor mean = 0.250, $t_1(21) = 3.80$, p<.01; $t_2(9) = 4.40$, p<.01] and continued significant thereafter [Window 3: target mean = 0.499, competitor mean = 0.248, $t_1(21) = 5.13$, p<.001, $t_2(9) = 3.67$, p<.01; Window 4, target mean = 0.587, competitor mean = 0.210, $t_1(21) = 6.31$, p<.001; $t_2(9) = 5.92$, p<.001].

The pattern in the *High Tone 3* condition was similar to that in the *Low Tone 3* condition [Window 2: target mean = 0.359, competitor mean = 0.271, $t_1(21) = 2.80$, p<.05; $t_2(9) = 1.32$, p>.05; Window 3: target mean = 0.430, competitor mean = 0.249, $t_1(21) = 3.81$, p<.01 and $t_2(9) = 2.41$, p<.05; Window 4: target mean = 0.502, competitor mean = 0.199, $t_1(21) = 6.00$, p<.001; $t_2(9) = 4.71$, p<.01].

In the *Low Tone 2* condition, the preference for competitors (Tone 3 words) over the targets (Tone 2 words) began to show in Window 2 [target mean = 0.280, competitor mean = 0.343, $t_1(21) = -2.21$, p<.05; $t_2(9) = -1.95$, p=.08]. The two curves then crossed over in Window 3 and significantly more fixations were then on targets than on competitors [Window 4: target mean = 0.499, competitor mean = 0.221, $t_1(21) = 5.93$, p<.001 and $t_2(9) = 7.15$, p<.001].

The pattern in the *High Tone 2* condition was again similar to the one in the *Low Tone 2* condition [Window 1: target mean = 0.252, competitor mean = 0.334, $t_1(21) = -2.80$, $p<.05$; $t_2(9) = -1.28$, $p>.05$; Window 2: target mean = 0.284, competitor mean = 0.356, $t_1(21) = -2.58$, $p<.05$; $t_2(9) = -1.79$, $p>.05$]. The two curves crossed over in Window 3. The POF on targets was significantly higher compared to the POF on competitors in Window 4 [Window 4: target mean = 0.505, competitor mean = 0.214, $t_1(21) = 6.25$, $p<.001$; $t_2(9) = 3.97$, $p<.01$].

### *Analysis of Diverging Points*

The same bootstrapping procedure as in Experiment 1 was used for Experiment 2 (see Figure 3.6 for the curves).

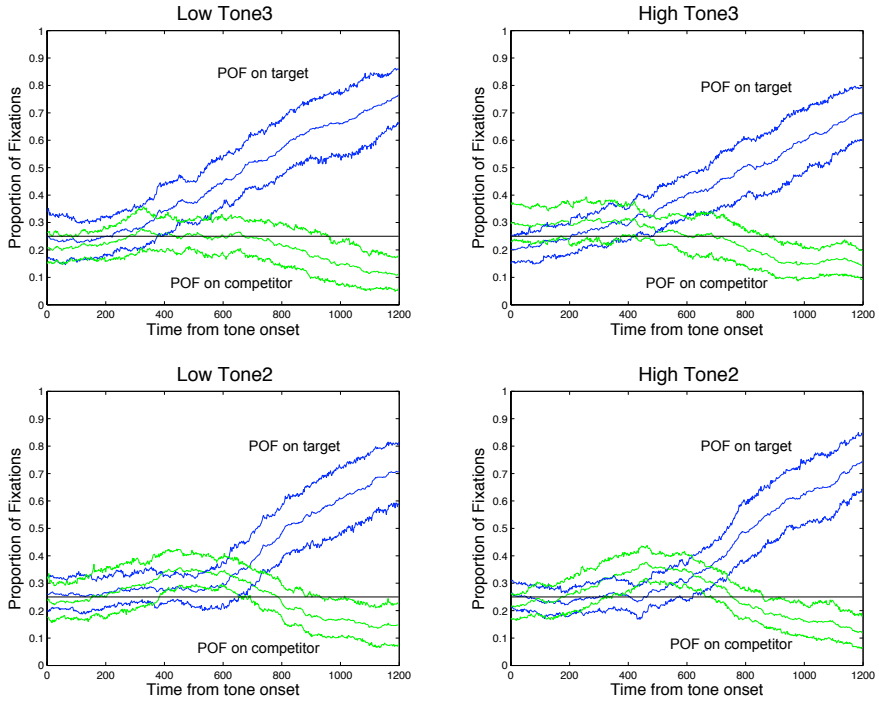


**Figure 3.8** Experiment 2 bootstrapping curves in 2 ms interval (time scale in ms). POF on target: proportion of fixations on targets; the middle line is the mean target POF; the other two lines are the upper and lower boundaries of the 95% confidence interval of target POF. POF on competitor: proportion of fixations on competitors; the middle line is the mean competitor POF; the other two lines are the upper and lower boundaries of the 95% confidence interval of competitor POF. Horizontal line is chance level (25%).

In the *Low Tone 3* and *High Tone 3* conditions, the POF curves that represent fixations on Tone 3 characters began rising by 300 ms into the tone.

The lower boundaries of the confidence interval began to be above chance from 382 ms and 416 ms, respectively for these two conditions, which indicates that the sound signal around the turning point had a critical influence on the preference towards the target.

In the *Low Tone 2* and *High Tone 2* conditions, the lower boundary of the confidence interval was above chance following 388 ms in the *Low Tone 2* condition, and 348 ms in the *High Tone 2* condition from tone onset. Approximately after 500 ms into the tone, the fixations on the competitor began to decrease and the ones on target to increase. The lower boundary of the confidence interval on target (Tone 2 words) was above chance starting from 656 ms (*Low Tone 2* condition) and 624 ms (*High Tone 2* condition) into the tone, which aligned with the time point of approximately 200 ms following tone offset.

### 3.2.3 Comparing the results of two experiments

Overall, the results of Experiment 2 were consistent with those of Experiment 1 in suggesting the influence of offset and turning point on the on-line tone judgment.

For each experiment, the diverging point analysis provided six critical time points on which the preference on the corresponding items (i.e., target or competitor that corresponds to the pitch information) was significantly above chance. A paired sample t-test was carried out and showed that the numerical difference of these critical points between the two experiments was not significant [$t(5)=0.87$, $p>.1$].

Visual inspection of the POF curves (see Figure 3 and 5) suggested that the differences between the POF on target and on competitor were larger in Experiment 1 compared with in Experiment 2, especially in the *Low Tone 3* and *High Tone 3* conditions. To examine this difference, an ANOVA was performed on the R difference score (i.e., transformed POF score on target minus transformed POF score on competitors) in four 150 ms windows. The result confirmed that the differences between the POF on target and on competitor were significantly larger in Experiment 1 than in Experiment 2 [$F(1,44)=6.26$, $p<.05$]. However this effect varied depending on conditions but not temporal windows [condition: $F(1,44)=12.25$, $p<.01$; window: $F(1,44)=1.39$, $p>.1$]. Further analysis using post hoc t-test was carried out to investigate the details of this effect. The differences between the POF on target and on competitor were significantly larger for Experiment 1 compared with Experiment 2 only in Window 2-4 of the *Low Tone 3* [ps<.01] and *High Tone 3* [ps<.05] conditions.

**Acknowledgement**

A part of the material on Study 1 has been submitted for publication. The dissertation author was the primary investigator and author of this paper. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Jinghong Le.

A part of the material on Study 2 is published in "On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements." Shen, J., Deutsch, D., and Rayner, K., Journal of the Acoustical Society of America, 2013, 133, 3016-3029. The dissertation author was the primary investigator and author of this material. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Keith Rayner.

# Chapter 4. Discussions

## 4.1 Study 1

The present study found that, for native speakers of Mandarin, changes in overall pitch height significantly influenced tone identification, even in the absence of immediate context information such as a carrier sentence or phrase, or familiarization with the speaker's voice.

This result adds to previous findings indicating that listeners can locate the pitch height of isolated syllables within a speaker's pitch range (Honorof & Whalen, 2005), and utilize this information in identifying lexical tones (Wong & Diehl, 2003; Lee, 2009; Moore & Jongman, 1997). We found that native speakers of Mandarin responded more slowly and less accurately to tones with transposed pitch heights, indicating that they had implicit knowledge of where the pitch height of a tone was located within the speaker's tessitura. Our results therefore add to the body of findings indicating that overall pitch height is involved as a cue in making judgments of lexical tone.

Furthermore, the data showed that the subjects performed better on transposed tones when their pitch heights were close to the original levels, and worse when they were transposed further away from the original. This result suggests that native speakers of Mandarin associate a certain pitch height with each of the tone labels (e.g., Tone 1 with a "high" pitch and Tone 3 with a "low" pitch). The subjects were therefore able to compare the pitch height of the tone with a mental template in retrieving lexical labels, analogous to the ability of absolute pitch possessors in identifying musical tones (Levitin & Rogers, 2005). This finding is consistent with the tone production data of Deutsch et al., (2004),

and supports the hypothesis that absolute pitch originally evolved as a feature of speech.

Detailed examination of the data reveal that transposition affected tone identification to different extents for the four tones. Identification of Tones 1 and 3 were influenced by transposition much more than were Tones 2 and 4. One potential explanation of this finding is that Tones 1 and 3 had their original pitch height at the highest and lowest ends of the speaker's tessitura, as well as the upper and lower boundaries of the pitch continuum in the present study. More specifically, Tone 1 is typically defined as a "high-level" tone and has its pitch height at the high range of the speaker's tessitura (Chao, 1968; Shih, 1987). Although Tone 3 has a more variable contour depending on dialect, it has the hallmark of a low pitch trough (Xu & Wang, 2001; Shen, Deutsch, & Rayner, 2010) and is therefore defined either as a "low-falling-rising" tone (Chao, 1968) or a "low" tone (Shih, 1987). The listeners in the present study may well have been searching for the highest overall pitch in the speaker's tessitura for identifying Tone 1, and the lowest one for identifying Tone 3.

A second possible explanation of the weaker effect on Tones 2 and 4 is that these two tones are more strongly contoured than Tones 1 and 3. Tone 2 is defined as a "rising" tone, and it has been found the rising part of the tone is critical for identifying this tone (Liu & Samuel, 2004). Further, Tone 4 is the only tone that has a distinctive "falling" pitch contour. Previous research investigating context effects on tone perception has shown that overall pitch height has a smaller influence on the perception of contour tones compared with level tones

(Moore & Jongman, 1997; Huang & Holt, 2009; Wong & Diehl, 2003; Francis, et al., 2006). The explanation was the multidimensional nature of the contour tones, such that overall pitch height and trajectory of pitch contour are two pitch cues that influence perception of the contour tones (Barrie, 2006). Nevertheless, our data has demonstrated that even judgments of the contour tones were affected by transposition, holding contour constant, albeit to a different extent than judgments of level tones. In addition, our finding suggests that the two pitch cues are weighted differently for contour tones versus level tones.

There was another interesting observation that for Tones 1 and 3, transposition had a larger inhibitory effect (i.e., in slowing down the response) on the process of confirming the correct tone label (by responding YES) than on rejecting a wrong label (by responding NO). In contrast, for Tones 2 and 4, transposition did not interact with type of response. This effect can potentially be explained by a difference in the weighting of cues for the different tones, and indicates that pitch height is a more important cue for identifying Tones 1 and 3 than Tones 2 and 4. It is worth noting that overall pitch height was utilized as a critical cue for Tone 3, which has usually been categorized as a “contour tone” (e.g., Chandrasekaran, Gandour, et al., 2007) and characterized by its contour nature (X. Shen, Lin & Yan, 1993; Moore & Jongman, 1997).

One distinctive aspect of the present experiment was that the overall range for pitch transposition was limited to corresponding to one female speaker’s pitch of speech in order to avoid speaker normalization effect in tone perception. This is different from previous studies in which speaker normalization was either

involved in the process or the focus of the study (Wong & Diehl, 2003; Lee, 2009; Moore & Jongman, 1997; Honorof & Whalen, 2005). The purpose of this manipulation was to prevent speaker normalization effect from being confounded with tone identification process and/or contributing to the variations in the performance measurements. However, this design inevitably introduced a possibility of range effect, which means, for example, original Tone 1 and 3 were identified better than transposed Tone 1 and 3 could be due to the reason that the original pitch level of these two tones were located at the highest and lowest ends of the speaker's tessitura, as well as the upper and lower boundaries of the pitch continuum in the present study. A potential direction in future research to examine this issue is to have sound stimuli produced by multiple speakers, which will cover a much larger and variable pitch range and use quantitative analysis (e.g., computational modeling) to further pinpoint the cognitive processes of speaker and tone identification respectively.

The effect of musical training was also examined in the present study. In contrast with findings on nontone language speakers, musical training was found not to have a significant effect on tone identification. No significant difference was found between musically trained and untrained subjects in responding to the presented tones, either in original or transposed form. In addition, it was found that musical training did not significantly interact with transposition, indicating that musically trained subjects did not appear to be more sensitive to either pitch height or contour compared to their non-musically trained counterparts. These results are in accordance with the theoretical framework (Deutsch et al, 1990;

Deutsch, 2002, 2013; Deutsch et al., 2004a) hypothesizing the absolute pitch and lexical tone perception share a common neural network and have an overlapping critical period for acquisition. In other words, our data indicate that musical training did not strengthen pitch-label associations in tone language speakers who had already developed such an association in the process of acquiring a tone language.

As predicted by the hypothesis that pitch perception across speech and music domains have a shared neural resource, musical training was found in this study not having a significant impact on both tone identification performance in general as well as identification of the pitch-transposed tones. Consistent with previous findings (Lee & Lee, 2010; Mok & Zuo, 2012), this result suggests that additional experience with musical tones does not give an edge to those amateur musicians who are native speakers of a tone language because the neural circuitry in the left hemisphere has already developed through earlier exposure to lexical tones. It would be an interesting follow-up study to test whether this result replicates with more intensive (i.e., professional) musical training and/or very early onset of musical training (e.g., < 5 years old).

## 4.2 Study 2

Using the "visual world" paradigm, the present findings demonstrated the dynamic process of lexical tone perception by revealing the effect of fine-grained pitch information at tone offset and turning point on on-line tone judgment for native Mandarin-speaking listeners. The eye fixation data in both experiments

yielded a very similar pattern. Low turning point pitch was exploited as a cue for on-line identification of Tone 3; offset pitch height was a critical cue for disambiguating Tones 2 and 3.

The temporal window and diverging point analyses both showed that the difference between the POFs on Tone 3 versus Tone 2 items did not become significant until around 400 ms; this revealed the pivotal influence of the low pitch trough of the turning point on Tone 3 perception. Prior studies on Mandarin tone perception and production have suggested that this low pitch target is a critical parameter of Tone 3 (Xu & Wang, 2001; Wong, Schwartz, & Jenkins, 2005). Taking a broad view, the importance of turning point pitch in Mandarin Tone 3 is also in line with data from another tone language, Thai (Zsiga & Nitisaroj, 2007), in suggesting that pitch inflection at the syllable midpoint could provide perceptual cues for tone identification.

In the present study, pitch information at tone onset (1-100 ms) initially directed more fixations to Tone 3 words but did not provide perceptual evidence strong enough to influence tone judgment significantly. At first sight, this finding appears inconsistent with the results of Gottfried and Suiter (1997) and Lee et al. (2008) since these authors used only the initial 6 pitch cycles (about 30 ms duration) of the syllable, and native speakers were able to correctly identify the tones in most cases. However, when examined more closely, the data from both the above studies also showed that if given only the tone onset (without offset), native speakers tend to confuse Tone 2 with Tone 3, because these two tones share the characteristic of low onset pitch. Since the present study focused on

word pairs consisting of Tones 2 and 3, and their onset pitches were very close to each other (around 185 Hz for Tone 3, 200 Hz for Tone 2), our results are consistent with these two studies in showing that native speakers had difficulty in identifying the sound token as a Tone 3 word based only on the cue of low onset pitch. These findings also align with the literature showing that Tones 2 and 3 are particularly difficult to differentiate, even for native speakers (Blicher et al., 1990; X. Shen et al., 1993; Whalen & Xu, 1992).

To tease apart the influence of the rising slope and the offset pitch height on on-line tone judgment, we examined the change in fixations on Tone 2 items across time. Because the steep slope of the rising pitch had fully unfolded by 400 ms into the tone, and on average it took 200 ms or less to program and execute a saccade, the influence of the rising slope should appear by 600 ms into the tone. This was indeed shown in both experiments by POF curves that represent fixations on Tone 2 items, which began rising around 500-600 ms. This result is in line with the earlier finding that a steep slope of pitch change is utilized as a cue for identifying Tone 2 (Xu et al., 2006). However, the diverging point analysis showed that a preference for target Tone 2 became significant only when information concerning offset pitch was available, which suggests that the cue of slope works better for tone identification when it is integrated with endpoint pitch height information.

The ultimate tone judgment data replicates findings (Gottfried & Suiter, 1997; Lee et al., 2008) showing that native speakers of Mandarin identify tokens with low onset and high offset pitch mostly as Tone 2 and those with low onset

and low offset pitch as Tone 3. Overall, the mean correct rate was higher for the two low offset conditions (96.4%) compared with the two high offset conditions (79%). This is likely due to the ambiguity in the two high offset conditions introduced by the hybrid pattern consisting of the Tone 3 onset mixed with Tone 2 offset. It was also found the listeners did respond to the difference between the original and the ambiguous conditions, which only had one semitone difference at the offset. Those conditions with original offset pitch height (i.e., *Low Tone 3* and *High Tone 2* conditions) were identified as the corresponding tones in 90% of the trials. This occurred significantly more often compared to the ambiguous conditions (i.e., *High Tone 3* and *Low Tone 2* conditions). Overall the sound tokens in the ambiguous conditions were identified as the corresponding tones in 85% of the trials. This finding is consistent with J. Shen et al. (2011) in showing native speakers' high sensitivity to small pitch differences in tone perception.

As reviewed by Dahan and Magnuson (2006), spoken word recognition is an incremental and constantly changing process, in which listeners continually evaluate the unfolding of the speech stream with fine-grained sensitivity, and activate certain lexical candidates that compete for recognition, even from the initiation of an utterance. The present study is in accordance with this line of research by showing that even short segments of pitch information can influence on-line tone judgment before the syllable ends. Because the listeners had a time interval of 700 ms to rapidly skim through the visually displayed object or character at the beginning of a trial during presentation of the prompt sentence, they already knew the location of the corresponding visual stimulus when the

sound token was played, and so were able to direct attention (and make a saccade) instantly to this item. In both experiments, sound stimuli were manipulated in such a way that the meaning of the spoken word depended on pitch cues at critical points in the syllable (e.g., a low turning point pitch indicates Tone 3 instead of Tone 2). In order to follow the instruction, the listeners had to exploit these pitch cues, make instantaneous saccadic eye movements to those lexical candidates before clicking on them. Aggregated over a large amount of trials, the listeners' eye fixation data (i.e., which object or character was fixated on at a certain time point) revealed influences of these fine-grained pitch cues on tone judgments.

By using different types of visual stimuli (object pictures versus Chinese characters) in the visual world paradigm, the two experiments reported here demonstrate a similar time trajectory of eye fixations on the tone targets and competitors. The diverging point analysis provides all the time points on which the preference for Tone 2 or Tone 3 items starts to be significant. These diverging points were not significantly different across the two experiments. This result suggests that when orthographic information is controlled (e.g., number of strokes that represents visual complexity), the written form of words in a logographic language may be used as visual stimuli in a visual world study that investigates lexical tones.

Comparing the results from the two experiments also provided an interesting observation for the two Tone 3 conditions (i.e., when the sound tokens had a low offset pitch that is consistent with the pitch information at the

onset). Compared with the competitors, the targets were more fixated on in the picture paradigm than in the character one. One possible reason that could explain this pattern is that pictures contain more novel visual information and are generally processed less automatically than words, which has been shown by the picture/color-word interference literature on both English and Chinese (Macleod, 1991; Smith & Kirsner, 1982; Lee & Chan, 2000). As a result, the subjects would move their eyes more actively to explore the visual stimuli and search for the target in the picture paradigm than in the character one. Therefore they made more fixations on the target picture for examining the object and confirming their tone judgment when the pitch information cumulatively and consistently provided cues for identifying the Tone 3 object as the target (i.e., in the Low Tone 3 and High Tone 3 Conditions). While in the character paradigm, the visual stimuli were processed more automatically and only a small amount of fixations can provide enough information for identifying the target.

As the above explanations are only based on the experimenter's observation during the experiment and largely speculative, new experiments should be designed to examine the nature of these phenomena. Because the two experiments were carried out using two different set of items and the two different groups of subjects, it is difficult to make direct comparison of the results from using the two types of visual stimuli. Future research is needed for further investigating this issue.

## 4.3 General Conclusions

While findings from the present studies address how native speakers of tone languages exploit various pitch height cues to identify lexical tones, it should be kept in mind that the big picture behind this line of research is the hypothesized connection between absolute pitch in music and lexical tone perception in speech (Deutsch, et al., 2004, Deutsch, 2013). For native speakers of tone languages, intensive experience of learning and using lexical tones, which happens early in their lives, could facilitate the development of the brain circuitry in the left hemisphere that supports the association between pitch height and verbal label. As a result, tone language speakers would acquire a precise and stable mental template for processing pitch information in both music and speech. Existing evidence that supports this hypothesis includes high pitch consistency in speech production (Deutsch, et al., 1999, 2004a; Deutsch, Le, et al., 2009) as well as high prevalence of absolute pitch in music (Deutsch, et al., 2006; Deutsch, Dooley, et al., 2009; Lee & Lee, 2010; Deutsch, Le, et al., 2011; Gregersen, et al., 1999, 2001), which can both be attributed to this mental template for pitch processing. The present studies further examined this hypothesis from the other angle concerning genesis of this mental template. The observed data demonstrated the importance of pitch height cues in lexical tone perception, which further lends support to the argument that early experience of speaking a tone language can facilitate development of this template for pitch processing that associates pitch height with verbal label (Deutsch, 2013).

More specifically, results of these two studies demonstrated precise

associations of pitch height and tone labels when all the pitch height levels were restricted within a single speaker's pitch range. Tone judgment data in Study 1 showed that the performance became worse with increase of the amount of pitch change. Perception of the tones was hindered if overall pitch height information was altered and in some conditions even only a small pitch change of 1.5 semitones from the original was sufficient for hurting the performance. Along the same line, in Study 2, a subtle 1-semitone deviation from the original pitch height at tone offset was found to affect identification of Tone 2 and 3. Taken together, this evidence indicates native speakers of tone languages are sensitive to very subtle pitch deviations when they retrieve tone labels, which suggests an implicit association of particular pitch height levels with tone labels. First, this is analogous to the case of AP possessors in music, who rely on conditional associative memory to construct and retrieve the association of pitch categories with verbal labels (Deutsch, 2006; Zatorre, 2003; Klein, et al., 1992; Zatorre, et al., 1998; Hsieh & Saberi, 2008). It is particularly worth noting that the precision demonstrated by the present data is comparable to AP perception in music, in which most responses (e.g., 90% in Deutsch, et al., 2011) are within one semitone from the correct note. In addition, these results are in accordance with previous pitch production data from native speakers of tone languages (Deutsch, et al., 1999, 2004a; Deutsch, Le, et al., 2009) and indicate a same precise and stable pitch template utilized by tone language speakers for tone production and perception.

Findings from these two studies also provide a comprehensive view

concerning how the two pitch height cues, namely overall pitch height and pitch height at critical points, are exploited for identifying lexical tones. It appears that native speakers not only utilize the overall pitch height cue as a whole to guide the immediate tone judgment (see Study 1), but also keep updating their decision while pitch height information unfolds throughout the tone (see Study 2). This is a new line of findings, considering all the previous research on tone perception had focused on what are the pitch height cues in tone perception (Gandour, 1983; Gandour & Harshman, 1978; Moore & Jongman, 1997; Gottfried & Suiter, 1997; Lee, et al., 2008; Zsiga & Nitisaroj, 2007) instead of how these cues are exploited. The fact that tone language speakers can utilize these pitch cues in an instant and incremental manner is in line with the literature on spoken language comprehension (see Rayner & Clifton, 2009; Dahan & Magnuson, 2006 for reviews), particularly those on sound-phoneme mapping during pre-lexical processing (McMurray, at el., 2002; Dahan, et al., 2001) in demonstrating high sensitivity and early utilization of fine-grained acoustical information. Collectively, the findings that pitch height information is constantly exploited, evaluated and associated with tone labels lend support to the hypothesized pitch template that is developed through early experience of using lexical tones (Deutsch, et al., 2004a, Deutsch, 2013).

## 4.4 Future Directions

These findings raise a number of testable follow-up questions regarding the mental template possessed by native speakers of tone languages for

processing pitch information in both speech and music.

As advanced by the theoretical framework (Deutsch et al., 2004a; Deutsch, Le, et al., 2009), native speakers of a tone language acquire a mental template of pitch through early and long-term exposure to the same pitch of speech used in the linguistic community. It was further argued that, for people who spent a long time living in the same community, the mental template of pitch should be fairly stable. Pitch of speech data collected from two remote villages in China support this hypothesis and demonstrate pitch ranges that are consistent across speakers in the same community while differ between two communities (Deutsch, Le, et al., 2009). Following this rationale, Study 1 controlled dialect backgrounds of the listeners to match up with that of the speaker. By keeping the overall pitch range consistent between speaker and listeners, this design should optimize the listeners' performance of tone judgment. On the other hand, a follow-up study can be designed to test whether this mental template of pitch varies depending on the listeners' linguistic experience. By using two different pitch ranges of speech (i.e., two different speakers), it is predicted that tone tokens within the pitch range that is consistent with the listener's pitch of speech will be identified faster and more accurate than those within a different pitch range.

As the present studies ultimately concern pitch processing across music and speech domains, it is quite intuitive to speculate that experience in these two domains would have an additive effect on performance of lexical tone perception. Results of Study 1 showed that this hypothesis does not hold up, at least for amateur musicians who are native tone language speakers. This finding is in

accordance with the theoretical framework (Deutsch, et al., 1990; Deutsch, 2002, 2013; Deutsch et al., 2004a) hypothesizing the absolute pitch and lexical tone perception share a common neural network and have an overlapping critical period for acquisition. Future studies can extend this line of research and seek replication of this finding by using professional musicians as subjects and also by examining performance of on-line processing of pitch height cues in tone perception.

Another intriguing aspect of this line of research that is deemed important is the developmental trajectory of this mental template. Specifically, there are a series of questions that can be asked pertaining to this issue. First, although it has been shown that native speakers learn to discriminate tones well before age of one year (Nazzi, et al., 1998; Harrison, 2000; Mattock & Burnham, 2006), it is unknown whether young children exploit those pitch height cues in a way similar to adult listeners. Second, as studies on AP in music demonstrated a degraded performance associated with advancing age (i.e., after age of 40-50; Ward, 1999; Athos, et al., 2007), it remains an open question whether the effect of aging also pertains to lexical tone perception. Last, building upon the literature on phonological attrition in bilinguals (Caramazza, Yeni-Komshian, Zurif, & Carbone, 1973; Yeni-Komshian, Flege, & Liu, 2000; Flege, Munro, & MacKay, 1995), further investigation is granted concerning whether prolonged experience of using a non-tone language as a second language has an effect on the mental template of pitch and thus influences perception and production of lexical tones. Research findings regarding these questions can be integrated into the theoretical

framework concerning the connection between AP and lexical tone perception and further shed light on the cognitive function of pitch processing across music and speech domains.

**Acknowledgement**

A part of the material on Study 1 has been submitted for publication. The dissertation author was the primary investigator and author of this paper. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Jinghong Le.

A part of the material on Study 2 is published in "On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements." Shen, J., Deutsch, D., and Rayner, K., Journal of the Acoustical Society of America, 2013, 133, 3016-3029. The dissertation author was the primary investigator and author of this material. The dissertation author acknowledges the following co-authors as collaborators on this manuscript: Diana Deutsch and Keith Rayner.

## References


Abramson, A. **(1978)**. "Static and dynamic acoustic cues in distinctive tones," Lang. Speech **23**, 319-325.

Anderson, S. **(1978)**. "Tone features," In *Tone: A linguistic survey,* edited by V. Fromkin (New York: Academic Press), pp.133-161.

Athos, E. A., Levinson, B., Kistler, A., Zemansky, J., Bostrom, A., and Freimer, N., et al. **(2007)**. "Dichotomy and perceptual distortions in absolute pitch ability," P. Natl. Acad. Sci. USA **104**, 14795-14800.

Bachem, A. **(1940)**. "The genesis of absolute pitch," J. Acoust. Soc. Am. **11**, 434-439.

Bachem, A. **(1955)**. "Absolute pitch," J. Acoust. Soc. Am. **27**, 1180-1185.

Baharloo, S., Johnston, P. A., Service, S. K., Gitschier, J., and Freimer, N. B. **(1998)**. "Absolute pitch: an approach for identification of genetic and nongenetic components," Am. J. Hum. Genet. **62**, 224-231.

Baharloo, S., Service, S. K., Risch, N., Gitschier, J., and Freimer, N. B. **(2000)**. "Familial aggregation of absolute pitch," Am. J. Hum. Genet. **67**, 755-758.

Barrie, M. **(2007)**. "Contour tones and contrast in Chinese languages," J. East Asian Linguist. **16**, 337–362.

Bates, E. **(1992)**. "Language development," Curr. Opin. Neurobiol. **2**, 180-185.

Bergeson, T. R., and Trehub, S. E. **(2002)**. "Absolute pitch and tempo in mothers' songs to infants," Psycol. Sci. **13**, 72-75.

Bermudez, P., and Zatorre, R. J. **(2005)**. "Conditional associative memory for musical stimuli in nonmusicians: implications for absolute pitch," J. Neurosci. **25**, 7718-7723.

Blicher, D. L., Diehl, R. L., and Cohen, L. B. **(1990)**. "Effects of syllable duration on the perception of the Mandarin Tone 2 / Tone 3 distinction: Evidence of auditory enhancement," J. Phonetics **18**, 37-49.

Boersma, P. and Weenink D. **(2008)**. "Praat: doing phonetics by computer (Version 5.0.42)" [Computer program]. Retrieved May 1, 2009, from http://www.praat.org/

Brady, P. T. **(1970)**. "Fixed scale mechanism of absolute pitch," J. Acoust. Soc. Am. **48**, 883-887.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., and Carbone, E. **(1973)**. "The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals," J. Acoust. Soc. Am. **54**, 421 – 428.

Carpenter, A. **(1951)**. "A case of absolute pitch," Q. J. Exp. Psychol. **3**, 92-93.

Chao, Y. R. **(1968)**. *A grammar of spoken Chinese* (University of California Press, Berkeley & Los Angeles), p.26.

Chan, S.W., Chuang, C-K., and Wang, W. S-Y. **(1975)**. "Cross-linguistic study of categorical perception for lexical tone," J. Acoust. Soc. Am. **58**, S119.

Chandrasekaran, B., Gandour, J. T., and Krishnan, A. **(2007)**. "Neuroplasticity in the processing of pitch dimensions: A multidimensional scaling analysis of the mismatch negativity," Restor. Neurol. Neuros. **25**, 195–210.

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

Chen, H.-C., and Cutler, A. **(1996)**. "Auditory priming in spoken and printed word recognition," In *Cognitive processing of Chinese and related Asian languages,* edited by Chen, H.-C. (Chinese University Press, Hong Kong), pp. 77–81.

Chin, C. S. **(2003)**. "The development of absolute pitch: A theory concerning the roles of music training at an early developmental age and individual cognitive style," Psychol. Music **31**, 155-171.

Connell, B. **(2000)**. "The Perception of Lexical Tone in Mambila," Lang. Speech **43**, 163-182.

Corliss, E. L. **(1973)**. "Remark on 'fixed-scale mechanism of absolute pitch'," J. Acoust. Soc. Am. **53**, 1737-1739.

Cuddy, L. L. **(1968)**. "Practice effects in the absolute judgment of pitch," J. Acoust. Soc. Am. **43**, 1069-1076.

Curtiss, S. **(1977)**. *Genie: A psycholinguistic study of a modern day 'wild child'* (Academic Press, New York, NY).

Cutler, A. **(1995)**. "Spoken word recognition and production," In *Speech, Language, and Communication,* edited by Miller, J. L. and Eimas, P. D., (Academic Press, San Diego), pp.97-136.

Cutler, A. and Chen, H.-C. **(1997)**. "Lexical tone in Cantonese spoken-word processing," Percept. Psychophys. **59**, 165-179.

Cutler, A., Dahan, D., and van Donselaar, W. **(1997)**. "Prosody in the comprehension of spoken language: A literature review," Lang. Speech **40**, 141–201.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., and Hogan, E. M. **(2001)**. "Subcategorical mismatches and the time course of lexical access: evidence for lexical competition," Lang. Cognitive Proc., **16**, 507-534.

Dahan, D., and Magnuson, J.S., **(2006)**. "Spoken word recognition," In *Handbook of Psycholinguistics (2nd ed.),* edited by Traxler, M.J., Gernsbacher, M.A., (Academic Press, Amsterdam), pp. 249–284.

Dennis, M., and Whitaker, H. A. **(1976)**. "Language acquisition following hemidecortication: linguistic superiority of the left over the right hemisphere," Brain Lang. **3**, 404-433.

Deutsch, D. **(1986)**. "A musical paradox," Music Percept. **3**, 275-280.

Deutsch, D. **(1987)**. "The tritone paradox: effects of spectral variables," Percept. Psychophys. **42**, 563-575.

Deutsch, D. **(1991)**. "The tritone paradox: an influence of language on music perception," Music Percept. **8**, 335-347.

Deutsch, D. **(1992)**. "Some new pitch paradoxes and their implications," Auditory Processing of Complex Sounds. Philos. T. R. Soc. B **336**, 391-397.

Deutsch, D. **(1994)**. "The tritone paradox: some further geographical correlates," Music Percept. **12**, 125-136.

Deutsch, D. **(2002)**. "The puzzle of absolute pitch," Curr. Dir. Psychol. Sci. **11**, 200-204.

Deutsch, D. **(2006)**. "The enigma of absolute pitch," Acoust. Today **2**, 11-19.

Deutsch, D. **(2013)**. "Absolute pitch," In *The psychology of music, 3rd Edition*, edited by D. Deutsch (Elsevier, San Diego), pp. 141-182.

Deutsch, D., Kuyper, W. L., and Fisher, Y. **(1987)**. "The tritone paradox: its presence and form of distribution in a general population," Music Percept. **5**, 79-92.

Deutsch, D., North, T., and Ray, L. **(1990)**. "The tritone paradox: correlate with the listener's vocal range for speech," Music Percept. **7**, 371-384.

Deutsch, D., Henthorn, T., and Dolson, M. **(1999)**. "Absolute pitch is demonstrated in speakers of tone languages," J. Acoust. Soc. Am. **106**, 2267.

Deutsch, D., Henthorn, T., and Dolson, M. **(2004a)**. "Absolute pitch, speech,

and tone language: some experiments and a proposed framework," Music Percept. **21**, 339-356.

Deutsch, D., Henthorn, T., and Dolson, M. **(2004b)**. "Speech patterns heard early in life influence later perception of the tritone paradox," Music Percept. **21**, 357-372.

Deutsch, D., Henthorn, E., Marvin, W., and Xu, H.-S. **(2006)**. "Absolute pitch among American and Chinese conservatory students: prevalence differences, and evidence for speech-related critical period," J. Acoust. Soc. Am. **119**, 719-722.

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. **(2009)**. "Absolute pitch among students in an American music conservatory: association with tone language fluency," J. Acoust. Soc. Am. **125**, 2398-2403.

Deutsch, D., Le, J., Shen, J., and Henthorn, T. **(2009)**. "The pitch levels of female speech in two Chinese villages," J. Acoust. Soc. Am. **125**, EL208-213.

Deutsch, D., Le, J., Shen, J., and Li, X. **(2011)**. Large-scale direct-test study reveals unexpected characteristics of absolute pitch. J. Acoust. Soc. Am. **130**, 2398.

Dolson, M. **(1994)**. "The pitch of speech as a function of linguistic community," Music Percept. **11**, 321-331.

Dooley, K., and Deutsch, D. **(2010)**. "Absolute pitch correlates with high performance on musical dictation," J. Acoust. Soc. Am. **128**, 890-893.

Dooley, K., and Deutsch, D. **(2011)**. "Absolute pitch correlates with high performance on interval naming tasks," J. Acoust. Soc. Am. **130**, 4097-4104.

Doupe, A. J., and Kuhl, P. K. **(1999)**. "Birdsong and human speech: common themes and mechanisms," Annu. Rev. Neurosci. **22**, 567-631.

Duchowny, M., Jayakar, P., Harvey, A. S., Resnick, T., Alvarez, L., and Dean, P., et al. **(1996)**. "Language cortex representation: effects of developmental versus acquired pathology," Ann. Neurol. **40**, 31-38.

Efron, B., & Tibshirani, R. **(1993)**. An introduction to the bootstrap. (Chapman & Hall, New York), pp. 168-177.

Eng, N., Obler, L. K., Harris, K., S., and Abramson, A. S. (1996). "Tone perception deficits in Chinese-speaking Broca's aphasics," Aphasiology **10**, 649-656.

Flege, J. E., Munro, M. L., and Mackay, I. R. A. **(1995)**. Factors affecting strength of perceived foreign accent in a second language. J. Acoust. Soc. Am. **97**,

3125-3134.

Fox, R., and Qi, Y. **(1990)**. "Contextual effects in the perception of lexical tone," J. Chin. Linguist. **18**, 261-283.

Francis, A. L., Ciocca, V., and Ng, B. K-C. **(2003)**. "On the (non)categorical perception of lexical tones," Percept. Psychophys. **65**, 1029-1044.

Francis, A., Ciocca, V., Wong, N., Leung, W., and Chu, P. **(2006)**. "Extrinsic context affects perceptual normalization of lexical tone," J. Acoust. Soc. Am. **119**, 1712-1726.

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

Fu Q. J. and Zeng F.G. **(2000)**. "Identification of temporal envelope cues in Chinese tone Recognition," Asia Pac. J. Speech Lang. Hearing **5**, 45-57.

Gandour, J. **(1974)**. "On the representation of tone in Siamese". In *Studies in Thai linguistics in honor of William J. Gedney.,* edited by J. G. Harris and J. R. Chamberlain, (Central Institute of English Language, Bangkok), pp.170-195.

Gandour, J. **(1981)**. "Perceptual dimensions of tone: Evidence from Cantonese," J. Chin. Linguist. **9**, 20-36.

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

Gandour, J., and Harshman, R. **(1978)**. "Cross language differences in tone perception: a multidimensional scaling investigation," Lang. Speech, **21**, 1-33.

Gandour, J., and Dardarananda, R. **(1983)**. "Identification of tonal contrasts in Thai aphasic patients," Brain Lang. **18**, 98-114.

Gandour, J., Ponglorpisit, S., Khunadorn, F., Dechongkit, S., Boongird, P., Boonklam, R., and Potisuk, S. **(1992)**. "Lexical tones in Thai after unilateral brain damage," Brain Lang. **43**, 275-307.

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

Gandour, J., Wong, D., Hsieh, L., Weinzapfel, B., Van Lancker, D., and Hutchins, G. D. **(2000)**. "A Crosslinguistic PET Study of Tone Perception," J. Cognitive Neurosci. **12**, 207-222.

Giangrande, J. **(1998)**. "The tritone paradox: effects of pitch class and position of

the spectral envelope," Music Percept. **15**, 253-264.

Gottfried T.L., and Suiter T.L. **(1997)**. "Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones," J. Phonetics. **25**, 207-231.

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. **(1999)**. "Absolute pitch: prevalence, ethnic variation, and estimation of the genetic component," Am. J. Hum. Genet. **65**, 911-913.

Gregersen, P. K., Kowalsky, E., Kohn, N., and Marvin, E. W. **(2001)**. "Early childhood music education and predisposition to absolute pitch: teasing apart genes and environment," Am. J. Med. Genet. **98**, 280-282.

Griffiths, T.D. and Warren, J.D. **(2002)**. "The planum temporale as a computational hub," Trends Neurosci. **25**, 348-353.

Halpern, A. R. **(1989)**. "Memory for the absolute pitch of familiar songs," Mem. Cognition **17**, 572-581.

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

Harrison, P. **(2000)**. "Acquiring the phonology of lexical tone in infancy," Lingua **110**, 581-616.

Heller, M. A., and Auerbach, C. **(1972)**. "Practice effects in the absolute judgment of frequency," Psychon. Sci. **26**, 222-224.

Henthorn, T., and Deutsch, D. **(2007)**. "Ethnicity versus early environment: Comment on 'Early Childhood Music Education and Predisposition to Absolute Pitch: Teasing Apart Genes and Environment' by P. K. Gregersen, E. Kowalsky, N. Kohn, and E. W. Marvin [2000]," Am. J. Med. Genet. **143**, 102-103.

Hickok, G., and Poeppel, D. **(2007)**. "The cortical organization of speech processing," Nat. Rev. Neurosci. **8**, 393-402.

Honorof, D. N., and Whalen, D. H. **(2005)**. "Perception of pitch location within a speaker's F0 range," J. Acoust. Soc. Am., **117**, 2193-2200.

Howie, J. **(1976)**. *Acoustical studies of Mandarin vowels and tones* (Cambridge Univ. Press, London), pp.138-246.

Huang, J., and Holt, L. L. **(2009)**. "General perceptual contributions to lexical tone normalization," J. Acoust. Soc. Am., **125**, 3983-3994.

Huettig, F. and Altmann, G. T. M. **(2007)**. "Visual-shape competition during language-mediated attention is based on lexical input and not modulated by contextual appropriateness," Vis. Cogn. **15**, 985-1018.

Huettig, F., Rommers, J., and Meyer, A. S. **(2011)**. "Using the visual world paradigm to study language processing: A review and critical evaluation," Acta Psychol. **137**, 151-171.

Hsieh, L., Gandour, J., Wong, D. and Hutchins, G. D. **(2001)**. "Functional heterogeneity of inferior frontal gyrus is shaped by linguistic experience," Brain Lang. **76**, 227-252.

Hsieh, I-H., and Saberi, K. **(2008)**. "Dissociation of procedural and semantic memory in absolute-pitch processing," Hearing Res. **240**, 73–79.

Johnson, J. S., and Newport, E. L. **(1989)**. "Critical periods in second language learning: the influence of maturational state on the acquisition of English as a second language," Cognitive Psychol. **21**, 60-99.

Kamide Y. Altmann G.T.M. and Haywood S.L. **(2003).** "The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements," J. Mem. Lang. **49**, 133-159.

Keenan, J. P., Thangaraj, V., Halpern, A. R., and Schlaug, G. **(2001)**. "Absolute pitch and planum temporale," NeuroImage **14**, 1402-1408.

Klatt, D. H. **(1973)**. "Discrimination of fundamental frequency contours in synthetic speech: implications for models of pitch perception," J. Acoust. Soc. Am. **53**, 8-16.

Klein, M., Coles, M. G. H., and Donchin, E. **(1984)**. "People with absolute pitch process tones without producing a P300," Science **223**, 1306-1309.

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

Kraus, N., and Chandrasekaran, B. **(2010)**. "Music training for the development of auditory skills," Nat. Rev. Neurosci. **11**, 599–605.

Lane, H. L. **(1976)**. *The wild boy of Aveyron* (Harvard University Press, Cambridge, MA).

Leather, J. **(1983)**. "Speaker normalization in perception of lexical tone," J. Phonetics **11**, 373–382.

Lee, C.-Y. **(2009)**. "Identifying isolated, multispeaker Mandarin tones from brief acoustic input: A perceptual and acoustic study," J. Acoust. Soc. Am. **125**, 1125–1137.

Lee, C-Y., Tao, L., and Bond, Z. S. **(2008)**. "Identification of acoustically modified Mandarin tones by native listeners," J. Phonetics **36**, 537-563.

Lee, C.-Y., and Lee, Y.-F. **(2010)**. "Perception of musical pitch and lexical tones by Mandarin-speaking musicians," J. Acoust. Soc. Am. **127**, 481-490.

Lee, Y-S., Vakoch, D. A., and Wurm, L. H. **(1996)**. "Tone perception in Cantonese and Mandarin: A cross-linguistic comparison," J. Psycholinguist Res. **25**, 527-542.

Lee, T.M.C. and Chan, C.C.H. **(2000)**. "Stroop Interference in Chinese and English," J. Clin. Exp. Neuropsyc., **22**, 465-471.

Levitin, D. J. **(1994)**. "Absolute memory for musical pitch: evidence for the production of learned melodies," Percept. Psychophys. **56**, 414-423.

Levitin, D. J., and Rogers, S. E. **(2005)**. "Absolute pitch: Perception, coding, and controversies," Trends Cogn. Sci. **9**, 26-33.

Lennenberg, E. H. **(1967)**. *Biological foundations of language* (Wiley, New York, NY).

Li, C. N., and Thompson, S. A., **(1977)**. "The acquisition of tone in Mandarin-speaking children," J. Child Lang. **4**, 185-199.

Lin, T., and Wang, W. **(1985)**. "Tone perception," J. Chin. Linguist. **2**, 59-69.

Lin, H.-B., and Repp, B. **(1989)**. "Cues to the perception of Taiwanese tones," Lang. Speech **32**, 25-44.

Liu, S., and Samuel, A. G. **(2004)**. "Perception of Mandarin lexical tones when F0 information is neutralized," Lang. Speech, **47**, 109 – 138.

Loui, P., Li, H., Hohmann, A., and Schlaug, G. **(2011)**. "Enhanced cortical connectivity in absolute pitch musicians: a model for local hyperconnectivity," J. Cognitive Neurosci. **23**, 1015-1026.

MacLeod, C. M. **(1991)**. "Half a Century of Research on the Stroop Effect: An Integrative Review," Psychol. Bull. **109**, 163-203.

Malins, J. and Joanisse, M. F. **(2010)**. "The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study," J. Mem. Lang. **62**, 407-420.

Massaro, D.W., Cohen, M. M., and Tseng, C. **(1985)**. "The Evaluation and Integration of Pitch Height and Pitch Contour in Lexical Tone Perception in Mandarin Chinese," J. Chin. Linguist. **13**, 267-290.

Matin, E. **(1974)**. "Saccadic suppression: A review and an analysis," Psychol. Bull. **81**, 899–917.

Mattock, K., and Burnham, D. **(2006)**. "Chinese and English Infants' Tone Perception: Evidence for Perceptual Reorganization," Infancy **10**, 241-265.

McCawley, J. C. **(1978)**. "What is a tone language?" In *Tone: A Linguistic Survey,* edited by Fromkin, V., (Academic Press, New York), pp. 113-131.

McMurray, B., Tanenhaus, M. K., and Aslin, R. N. **(2002)**. "Gradient effects of within-category phonetic variation on lexical access," Cognition **86**, B33–B42.

Meyer, D. E., and Schvaneveldt, R. W. **(1971)**. "Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations," J. Exp. Psychol. **90**, 227-234.

Mixdorff, H., Luksaneeyanawin, S., Fujisaki, H., and Charnavit, P. **(2002)**. "Perception of tone and vowel quality in Thai," Paper presented at the 7th International Conference on Spoken Language Processing, Denver, Colorado.

Miyazaki, K. **(1988)**. "Musical pitch identification by absolute pitch possessors," Percept. Psychophys. **44**, 501-512.

Miyazaki, K., and Ogawa, Y. **(2006)**. "Learning absolute pitch by children: a cross-sectional study," Music Percept. **24**, 63-78.

Moen, I., and Sundet, K. **(1996)**. "Production and perception of word tones (pitch accents) in patients with left and right hemisphere damage," Brain Lang. **53**, 267-281.

Mok, P.K.P. and Zuo, D. **(2012)**. "The separation between music and speech: Evidence from the perception of Cantonese tones," J. Acoust. Soc. Am. **132**, 2711-2720.

Moore, C. B., and Jongman, A. **(1997)**. "Speaker normalization in the perception of Mandarin Chinese tones," J. Acoust. Soc. Am. **102**, 1864-1877.

Möttönen, R., Calvert, G. A., Jääskeläinen, I. P., Matthews, P. M., Thesen, T., and Tuomainen, J., and Mikko, S. **(2006)**. "Perceiving identical sounds as speech or non-speech modulates activity in the left posterior superior temporal sulcus," Neuroimage **30**, 563-569.

Moulines, E. and Charpentier, F. **(1990)**. "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones," Speech Commun. **9**, 453-467.

Naeser, M. A., and Chan, S. W.-C. **(1980)**. "Case study of a Chinese aphasic with the Boston diagnostic aphasia exam," Neuropsychologia **18**, 389-410.

Nazzi, T., Floccia, C., and Bertoncini, J. **(1998)**. "Discrimination of pitch

contours by neonates," Infant Behav. Dev. **21**, 779-784.

Newport, E. L. **(1990)**. "Maturational constraints on language learning," Cognitive Sci. **14**, 11-28.

Newport, E. L., Bavelier, D., and Neville, H. J. **(2001)**. "Critical thinking about critical periods," In *Language, brain, and cognitive development: Essays in honor of Jacques Mehler*. edited by E. Dupoux, (MIT Press, Cambridge, MA).

Oechslin, M. S., Meyer, M., and Jäncke, L. **(2010)**. "Absolute pitch: functional evidence of speech-relevant auditory acuity," Cereb. Cortex **20**, 447-455.

Packard, J. L. **(1986)**. "Tone production deficits in nonfluent aphasic Chinese speech," Brain Lang. **29**, 212-223.

Patel, A. D. **(2008)**. *Music, Language, and the Brain* (Oxford University Press, New York), pp. 3-411.

Patkowski, M. S. **(1990)**. "Age and accent in a second language: a reply to James Emil Flege," Appl. Linguist. **11**, 73-89.

Pike, K. L. **(1948)**. "Tone Language," Ann Arbor: University of Michigan Press.

Profita, J., and Bidder, T. G. **(1988)**. "Perfect pitch," Am. J. Med. Genet. **29**, 763-771.

Ratcliff, R. **(1993)**. "Methods of dealing with reaction time outliers," Psychol. Bull. **114**, 510–532.

Repp, B. H., and Thompson, J. M. **(2010)**. "Context sensitivity and invariance in perception of octave-ambiguous tones," Psychol. Res. **74**, 437-456.

Ramadoss, D., and Smolensky, P. **(2011)**. "Tone perception cues: Pitch targets, trajectories, or both?" J. Acoust. Soc. Am. **129**, 2420.

Rayner, K. **(1998)**. "Eye movements in reading and information processing: 20 years of research," Psychol. Bull. **124**, 372-422.

Rayner, K. **(2009)**. "Eye movements and attention in reading, scene perception, and visual search," Q. J. Exp. Psychol. **62**, 1457-1506.

Rayner, K., and Clifton, C. **(2009)**. "Language processing in reading and speech perception is fast and incremental: Implications for event-related potential research," Biol. Psychol., **80**, 4-9.

Sakai, K. L. **(2005)**. "Language acquisition and brain development," Science **310**, 815-819.

Schellenberg, E. G., and Trehub, S. E. **(2003)**. "Good pitch memory is

widespread," Psychol. Sci. **14**, 262-266.

Schlaug, G., Jäncke, L., Huang, Y., and Steinmetz, H. **(1995)**. "In vivo evidence of structural brain asymmetry in musicians," Science **267**, 699-701.

Schulze, K., Gaab, N., and Schlaug, G. **(2009)**. "Perceiving pitch absolutely: comparing absolute and relative pitch possessors in a pitch memory task," BMC Neurosci. **10**, 1471-2202.

Scovel, T. **(1969)**. "Foreign accent, language acquisition, and cerebral dominance," Lang. Learn. **19**, 245-253.

Sek, A., and Moore, B. **(1999)**. "Discrimination of frequency steps linked by glides of various durations," J. Acoust. Soc. Am. **106**, 351-359.

Sergeant, D. **(1969)**. "Experimental investigation of absolute pitch," J. Res. Musical Educ. **17**, 135-143.

Shen, J., Deutsch, D., and Rayner, K. **(2010)**. "Processing of Endpoint Pitch in Mandarin Tone Perception: An Eye Movement Study," Paper presented at the 20th International Congress on Acoustics, Sydney, Australia.

Shen, J., Deutsch, D., and Le, J. **(2011)**. "Overall pitch height as a cue for lexical tone perception," Poster session presented at the 162nd meeting of Acoustical Society of America, San Diego, CA.

Shen, J., Deutsch, D., and Rayner, K. **(2013)**. "On-line Perception of Mandarin Tones 2 and 3: Evidence from Eye Movements," J. Acoust. Soc. Am. **133**, 3016-3029.

Shen, X., Lin, M., and Yan, J. **(1993)**. "F0 turning point as an F0 cue to tonal contrast: A case study of Mandarin tones 2 and 3," J. Acoust. Soc. Am., **93**, 2241-2243.

Shih, C. **(1987)**. "The phonetics of the Chinese tonal system," Bell Laboratories Technical Memorandum.

Smith, M. C. and Kirsner, K. **(1982)**. "Language and Orthography as Irrelevant Features in Color-Word and Picture-Word Stroop Interference," Q. J. Exp. Psychol., **34**, 153-170.

Smith, N. A., and Schmuckler, M. A. **(2008)**. "Dial A440 for absolute pitch: absolute pitch memory by non-absolute pitch possessors," J. Acoust. Soc. Am. **123**, EL77-EL84.

Speer, S. R., and Xu, L. **(2005)**. "Ambiguous lexical tone process during Mandarin sentence comprehension-evidence form eye-movement experiment," Paper presented at the 11th Annual Conference on Architectures and Mechanisms for Language Processing, Ghent, Belgium.

Strange, W., Jenkins, J. J., & Johnson, T. L. **(1983)**. "Dynamic specification of coarticulated vowels," J. Acoust. Soc. Am. **74**, 695–705.

Studebaker, G. A. **(1985)**. "A 'Rationalized' Arcsine Transform," J. Speech Lang. Hear. Res. **28**, 455-462.

Takeuchi, A. H. **(1989)**. *Absolute pitch and response time: The processes of absolute pitch Identification* (Unpublished master's thesis) Johns Hopkins University, Baltimore, MD.

Takeuchi, A. H., and Hulse, S. H. **(1993)**. "Absolute pitch," Psychol. Bull. **113**, 345-361.

Tanenhaus, M. K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. **(1995)**. "Integration of visual and linguistic information in spoken language comprehension," Science, **268**, 1632-1634.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D. and Chambers, C. **(2000)**. "Eye Movements and Lexical Access in Spoken-Language Comprehension: Evaluating a Linking Hypothesis between Fixations and Linguistic Processing," J. Psycholinguist. Res. **29**, 557-580.

Terhardt, E., and Ward, W. D. **(1982)**. "Recognition of musical key: exploratory study," J. Acoust. Soc. Am. **72**, 26-33.

Terhardt, E., and Seewann, M. **(1983)**. "Aural key identification and its relationship to absolute pitch," Music Percept. **1**, 63-83.

Theusch, E., Basu, A., and Gitschier, J. **(2009)**. "Genome-wide study of families with absolute pitch reveals linkage to 8q24.21 and locus heterogeneity," Am. J. Hum. Genet. **85**, 112-119.

Tse, J. K.-P. **(1978)**. "Tone acquisition in Cantonese: A longitudinal case study," J. Child Lang. **5**, 191-204.

Van Lancker, D., and Fromkin, V. **(1973)**. "Hemispheric specialization for pitch and 'tone': Evidence from Thai," J. Phonetics **1**, 101-109.

Varyha-Khadem, F., Carr, L. J., Isaacs, E., Brett, E., Adams, C., and Mishkin, M. **(1997)**. "Onset of speech after left hemispherectomy in a nine year old boy," Brain **120**, 159-182.

Vitouch, O., and Gaugusch, A. **(2000)**. "Absolute recognition of musical keys in non- absolute-pitch-possessors," In *Proceedings of the 6th International Conference on Music Perception and Cognition*. edited by C. Woods, G. Luck, R. Brochard, F. Seddon, & J. A. Sloboda, (Dept. of Psychology, Keele University, Keele, UK).

Wang, W. S.-Y. **(1973)**. "The Chinese language," Sci. Am. **228**, 50-60.

Wang, W. S.-Y. **(1976)**. "Language change," Ann. NY Acad. Sci. **280**, 61-72.

Wang, Y., Sereno, J. A., and Jongman, A. **(2001)**. "Dichotic perception of Mandarin tones by Chinese and American listeners," Brain Lang. **78**, 332-348.

Wang, Y., Sereno, J. A., Jongman, A., and Hirsch, J. **(2003)**. "fMRI Evidence for Cortical Modification during Learning of Mandarin Lexical Tone," J. Cognitive Neurosci. **15**, 1019-1027.

Ward, W. D. **(1999)**. "Absolute pitch". In *The psychology of music (2nd edition)*, edited by D. Deutsch, (Academic Press, San Diego), pp. 265-298.

Wayman, J. W., Frisina, R. D., Walton, J. P., Hantz, E. C., and Crummer, G. C. **(1992)**. "Effects of musical training and absolute pitch ability on event-related activity in response to sine tones," J. Acoust. Soc. Am. **91**, 3527-3531.

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

Wong, P. C-M., and Diehl, R. L. **(2003)**. "Perceptual Normalization for Inter and Intratalker Variation in Cantonese Level Tones," J. Speech Lang. Hear R. **46**, 413-421.

Wong, P., Schwartz, R.G., and Jenkins, J. J. **(2005)**. "Perception and production of lexical tones by 3-year-old, Mandarin-speaking children," J. Speech. Lang. Hear. R. **48**, 1065-1079.

Woods, B. T. **(1983)**. "Is the left hemisphere specialized for language at birth?" Trends Neurosci. **6**, 115-117.

Xu, Y. and Sun, X. **(2001)**. "Maximum speed of pitch change and how it may relate to speech," J. Acoust. Soc. Am. **111**, 1399-1413.

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

Xu, Y. **(2004)**. "Understanding tone from the perspective of production and perception," Lang. Linguist. **5**, 757-797.

Xu, Y., Gandour, J., Talavage, T., Wong, D., Dzemidzic, M., Tong, Y., Li, X. and Lowe, M. **(2006)**. "Activation of the left planum temporale in pitch processing is shaped by language experience," Hum. Brain Mapp. **27**, 173-183.

Xu, Y., Gandour, J. T., and Francis, A. L. **(2006)**. "Effects of language experience and stimulus complexity on the categorical perception of pitch direction," J. Acoust. Soc. Am. **120**, 1063-1074.

Yeni-KomshianN, G. H., Flege, J. E., and Liu, S. **(2000)**. "Pronounciation proficiency in the first and second languages of Korean-English bilinguals," Biling-Lang Cogn. **3**, 131-149.

Yip, M. **(1989)**. "Contour tones," Phonology **6**, 149-174.

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

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., and Evans, A. C. **(1998)**. "Functional anatomy of musical processing in listeners with absolute pitch and relative pitch," P. Natl. Acad. Sci. USA **95**, 3172-3177.

Zatorre, R. J., **(2003)**. "Absolute pitch: a modal for understanding the influence of genes and development on neural and cognitive function," Nat. Neurosci. **6**, 692-695.

Zatorre, R. J., Perry, D. W., Beckett, C. A., Westbury, C. F., and Evans, A. C. **(1998)**. "Functional anatomy of musical processing in listeners with absolute pitch and relative pitch," P. Natl. A. Sci. USA **95**, 3172-3177.

Zhang, J. **(2002)**. *The effects of duration and sonority on contour tone distribution: A typological survey and formal analysis* (Routledge, New York).

Zhou, X., Shu, H., Bi, Y., & Shi, D. **(1999)**. "Is there phonologically mediated access to lexical semantics in reading Chinese?" In *Reading Chinese script: A cognitive analysis,* edited by J. Wang, A. Inhoff, & H.-C. Chen, (Erlbaum, Hillsdale, NJ), pp. 135–171.

Zsiga, E., and Nitisaroj, R. **(2004)**. "Perception of Thai tones in citation form and connected speech," J. Acoust. Soc. Am. **116**, 2628.

Zsiga, E., and Nitisaroj, R. **(2007)**. "Tone features, tone perception, and peak alignment in Thai," Lang. Speech **50**, 343-383.