How Native Chinese Listeners and Second-Language Chinese Learners Process Tones in Word Recognition: An Eye-tracking Study

By

Zhen Qin
M.Phil, Chinese University of Hong Kong, 2012
M.A., Chinese University of Hong Kong, 2009
B.A., Central South University, 2008

Submitted to the graduate degree program in Linguistics and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Co-Chair: Dr. Annie Tremblay ____________

Co-Chair: Dr. Jie Zhang ____________

Dr. Allard Jongman ____________

Dr. Joan Sereno ____________

Dr. Yan Li ____________

Date Defended: 14 June 2017

The dissertation committee for Zhen Qin certifies that this is the approved version of the following dissertation:

How Native Chinese Listeners and Second-Language Chinese Learners Process Tones in Word Recognition: An Eye-tracking Study

Co-Chair: Dr. Annie Tremblay ____________

Co-Chair: Dr. Jie Zhang ____________

Date Approved: 18 July 2017

Abstract

Chinese and English differ in the types of information they use to convey meaning in words: unlike English, Chinese uses lexical tones (i.e., pitch movement) to contrast word meanings (e.g., in Chinese, the word “ma” can mean *mother*, *hemp*, *horse*, or *scold* depending on its lexical tone). This difference between Chinese and English poses word recognition difficulties for English-speaking learners of Chinese in using lexical tones to recognize Chinese words. Existing research on L2 learners’ perception and processing of lexical tones has focused on whether native listeners of languages that do not have lexical tones can discriminate and identify lexical tones. To date, no study to our knowledge has examined how L2 learners use the fine-grained phonetic details of tonal information in the time course of spoken word recognition — that is, *as the speech signal unfolds over time*. In fact, little research has looked into the time course with which native listeners use the fine-grained phonetic details of tonal information in spoken word recognition. This doctoral dissertation examines how native Chinese listeners and highly proficient adult English-speaking learners of Chinese use tonal information in spoken word recognition as the speech signal unfolds in time.

More specifically, this research uses the visual-world eye-tracking paradigm to shed light on the precise time course with which native and non-native listeners use tonal information in online word recognition. The proposed research aims to investigate two potential differences between native listeners and highly proficient English-speaking L2 learners of Chinese in their use of tonal information as the speech signal unfolds: (i) their potentially different incremental use of the early pitch height before pitch contour information of the tone is available; (ii) their potentially different sensitivities to fine-grained within-category gradience of level and contour tones in the word recognition process.

Experiment 1 investigates whether or not native and non-native listeners make similar use of early between-category pitch height (T1-T4 with similar early pitch height vs. T1-T2 with different pitch height) before pitch contour information is available. A visual-world eye-tracking experiment in Chinese was conducted with two groups of participants: 36 native Chinese listeners and 26 highly proficient English-speaking L2 learners of Chinese. The target was either T1 or T2 word in T1-T2 condition whereas the target was either T1 or T4 word in T1-T4 condition. The auditory stimuli were natural tonal tokens. The time-window analyses on fixations showed that early pitch height constrained both Chinese and English listeners' lexical access. While Chinese listeners started using early pitch height in the time window in which pitch contour information was not available, English listeners started using early pitch height in the time windows in which pitch contour information had been available, and showed more tonal competition than Chinese listeners. The findings suggest that whether or not prosodic cues contribute to distinguishing among words in the L1, and how they do so, influence listeners' use of these cues in spoken word recognition.

Experiment 2 investigates whether native Chinese listeners and English-speaking L2 learners of Chinese differ in using the within-category gradience of level and contour tones to recognize spoken words. Another visual-world eye-tracking experiment in Chinese was conducted with the same participants. The target was a level tone (i.e., T1) and the competitor was a high-rising tone (i.e., T2), or vice versa. The auditory stimuli were manipulated such that the target tone was either canonical in the standard condition, acoustically more distant from the competitor in the distant condition, or acoustically closer to the competitor in the close condition. Growth curve analysis on fixations suggested that Chinese listeners showed a gradient pattern of lexical competition, with decreased competition in the distant condition and increased competition in the close condition than in the standard condition for the contour tone; English listeners, on the other hand, showed increased competition in both the distant and close conditions relative to the standard condition for the level tone. These findings suggest that Chinese listeners may show sensitivity to fine-grained tonal variability when this variability is along a dimension (i.e., pitch contour) that is meaningful for distinguishing tones whereas English listeners might show sensitivity to the fine-grained tonal variability along a dimension (i.e., pitch height) encoded in their L1 lexical representations. Moreover, native and non-native listeners, who potentially differ in the robustness of their representations of lexical tones, may adopt different strategies to deal with fined-grained tonal information to resolve the lexical competition.

The findings of this doctoral dissertation make a contribution to the understanding of how tonal information modulates lexical activation in native and non-native Chinese listeners. This research also has pedagogical implications for Chinese language teaching.

Acknowledgments

This dissertation would never have been completed in its entirety without the financial support of NSF Doctoral Dissertation Improvement Grant and KU Doctoral Research Fund.

I would like to express my deepest gratitude to my co-advisors, Dr. Annie Tremblay and Dr. Jie Zhang, for the various ways in which they have supported me during my academic career, and especially during the application of NSF grant, the proposal preparation, and the write-up of this dissertation. Their critiques, commentary, and genuine enthusiasm for this project kept me engaged and carried me forward in my difficult time. I owe you so much that goes beyond words, and I can only hope that I am able to pay it forward in the future and share with my own students the same enthusiasm in research and academia that you have transmitted to me.

I am tremendously grateful to my other committee members—Dr. Allard Jongman, Dr. Joan Sereno, and Dr. Yan Li —who asked me questions that made me think harder and produce better work. Throughout this process I felt challenged and supported in all the best ways. It was a true honor and privilege to work with you all. Special thanks goes to Dr. Jongman, who provided me with research assistantships, guided my research in tones, and taught me how to be a speech researcher in many ways.

In addition to the members of my committee, I would also like to thank all of the faculty members from the Linguistics Department at the University of Kansas. The excellent academic atmosphere in this department gave me a standard to which I will aspire in my future. I have truly enjoyed all of the courses I took here. In particular, I am thankful to Dr. Rob Fiorentino and Dr. Alison Gabriele for their Cognitive Neuroscience of Second Language Acquisition classes, which helped me develop my background in the fields. I also learned from them how to propose a second language research project and apply for NSF grants.

I would like to thank Dr. Xiaolin Zhou and my lab mates in the Center for Brain and Cognitive Sciences at Peking University, for providing me with the access to their lab, allowing me to collect data there, and helping me recruit participants for the study.

My colleagues in the Linguistics Department— Dr. Yu-Fu Chien, Dr. Kate Coughlin, Dr. María Teresa Martínez García, Dr. Jiang Liu, Dr. Steve Politzer-Ahles, Katrina Connell, Mingxing Li, Charlie Redmon, Seulgi Shin, Xiao Yang, Yuyu Zeng, along with other professors and my fellow students in the LING 850 group, offered critical support and feedback along the way, and helped me when I was wrestling with data, presentations, and write-ups. I am particularly grateful to Hongling Lao, Liu Yang, Cynthia Siew, Michael Swarts, Candice Tang, and Tong Xu for always being next to me, particularly when I needed them the most.

Lastly, I would like to thank my family for their continued support and their love with me for the past few years in the pursuit of my linguistics degrees. Especially, I thank my dad, Xuyan Qin, and my mom, Xiuyu Huang, who are always there to cheer me up and stand by me through the good times and bad.

All that is well done in this dissertation benefitted from the help of those I have mentioned, and any mistakes are mine alone.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

Mandarin Chinese (henceforth, Chinese) is becoming an increasingly important language on the international scene. Accordingly, it has become a commonly taught language in foreign language programs in the United States, and an increasingly large number of native English speakers are learning Chinese. Importantly, Chinese and English differ in the types of information they use to convey meaning in words: unlike English, Chinese uses lexical tones (i.e., pitch movement) to contrast word meanings (e.g., /pā/ 'eight' (Tone 1 (T1)), /pá/ 'to pull out' (T2), /pǎ/ 'to hold' (T3), /pà/ 'father' (T4); Yip, 2002). This difference between Chinese and English poses word recognition difficulties for English-speaking learners of Chinese (Sun, 2012; Wiener, 2015; Wiener, Ito, & Speer, 2016). This doctoral dissertation examines how native Chinese listeners and high-proficiency adult English-speaking second-language (L2) learners of Chinese use tonal information in spoken word recognition as the speech signal unfolds over time.

More specifically, this doctoral dissertation uses the visual-world eye-tracking paradigm to shed light on the precise time course with which native and non-native listeners use the fine-grained phonetic details of tonal information in online word recognition. The study aims to investigate two potential differences between native and non-native listeners in their use of tonal information in word recognition as the speech signal unfolds: (i) their potentially different sensitivities to the early pitch height of the tones; (ii) their potentially different sensitivities to within-category gradience in this early pitch height of level and contour tones. Two visual-world eye-tracking experiments were conducted to investigate these potential differences.

As the speech signal unfolds, lexical candidates that overlap the most with the input become partially activated and compete for recognition (e.g., Luce, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1989). For example, Allopenna, Magnuson, & Tanenhaus (1998) found that two words that share an onset (e.g., *beetle* and *beaker*) or a rhyme (e.g., *speaker* and *beaker*) with each other compete more for recognition than words that are phonologically unrelated. Moreover, a lexical competitor that shares its onset (i.e., *beetle*) with the target (i.e., *beaker*) causes more lexical competition than one that shares the rhyme (i.e., *speaker*). This suggests that spoken words are recognized incrementally, and lexical access takes place continuously.

Suprasegmental information (i.e., fundamental frequency (F0), duration, intensity) signals word identity and thus constrains lexical access in languages such as Chinese and English. This information has been shown to constrain lexical access for both native listeners and L2 learners (e.g., Cooper, Cutler, & Wales, 2002; Reinisch, Jesse, & McQueen, 2010). However, there is limited understanding of how suprasegmental information is encoded in word recognition. Theoretical models such as Cohort (e.g., Marlsen-Wilson, 1989), TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and other influential models have not satisfactorily incorporated the use of suprasegmental information. There have been some attempts at modeling the use of suprasegmental information such as lexical tones. For instance, Shuai and Malins (2016) attempted to incorporate lexical tones in their TRACE-T model. In the model, while segmental and tonal units alternate at the phoneme level, the pitch cues, that is, pitch height (i.e., 5 levels: from 1 to 5, 5 being the highest) and pitch contour (i.e., 3 levels: level, rising, and falling), are embedded simultaneously at the feature level. Nevertheless, it remains open for debate whether the TRACE-T model can account for the use of tonal information in word recognition. Furthermore, it is unclear how L2 learners use suprasegmental information during online word recognition. Since suprasegmental information plays an important role in word recognition (for discussion, see Cutler, 2012, Chapter 7), it is important to incorporate this information into current psycholinguistic models of native and L2 word recognition.

Suprasegmental information constrains lexical access differently and to different degrees across languages. For example, suprasegmental cues to word identity in English signal the presence or absence of stress on a particular syllable (e.g., *MYStic* vs. *misTAKE*) and tend to co-occur with segmental cues (i.e., vowel reduction; e.g., *ADjective* vs. *adVIsor*). As a result, English listeners show limited sensitivity to suprasegmental information in word recognition (e.g., Cooper et al., 2002). By contrast, suprasegmental information is extremely important for distinguishing words in Mandarin Chinese (henceforth Chinese), as segmentally identical words that contain different lexical tones differ in meaning. Figure 1 shows the suprasegmental realization of the four lexical tones in Chinese. In addition to having different pitch contours (i.e., different tone shapes), the four tones differ in their early pitch height (in F0): Some of the tones can be distinguished by their early pitch height difference from the very beginning of the tonal contour (e.g., T1-T2), whereas other tones have a similar pitch height early in the contour and can be distinguished only later (e.g., T1-T4). Hence, for some tone pairs (e.g., T1-T2), early pitch height can potentially constrain lexical access from the onset of the tone.

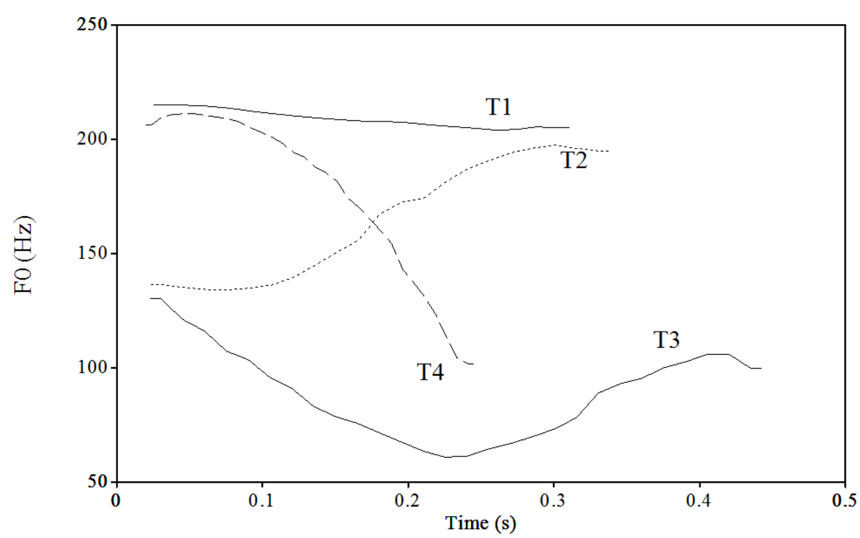

Figure 1: Tone contours of the four Chinese tones produced by a male native Chinese speaker

Since suprasegmental information plays a limited role in English word recognition, with tonal information *not* contrasting word meanings in English, native speakers of English who learn Chinese after the so-called "critical period" for language acquisition (e.g., Lenneberg, 1967) — also referred to as "late" learners — find it difficult to make accurate use of lexical tones to recognize Chinese words (Sun, 2012; Wiener, 2015). Existing research on late L2 learners' recognition of Chinese tones has found that L2 learners' discrimination and identification of Chinese tones is strongly influenced by whether or not the native language (L1) has lexical tones (e.g., Bent, 2005; Chandrasekaran, Sampath, & Wong, 2010; Francis, Ciocca, Ma, & Fenn, 2008; Hao, 2012; Sun & Huang, 2012; Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999). Other studies on the perception of segments (e.g., Francis & Nusbaum, 2002; Holt & Lotto, 2006; Ingvalson, Holt, & McClelland, 2011), however, suggest that whether or not cues contribute to distinguishing among words in the L1, and how they do so, are what determine L2 learners' use of these cues in the L2. Therefore, it is unclear from this literature whether L2 learners whose L1 does not have lexical tone would be able to use some cues (e.g., pitch height) to process lexical tones if they encode the cues as part of their native lexical representations (e.g., English stress), and whether they would use this tonal information as efficiently, indexed by the degree and timing of tonal competition, as native listeners in L2 word recognition (Braun, Galts, & Kabak, 2014; Lin, Wang, Idsardi, & Xu, 2014; Qin, Chien, & Tremblay, 2017).

Looking into the time course with which listeners use tonal information as the speech signal unfolds over time is important for understanding how native listeners and L2 learners of Chinese use tonal information in word recognition. The present study investigates the time course of native and non-native listeners' use of tonal information in word recognition. More specifically, first, it examines whether (and if so, how) native and non-native Chinese listeners use early pitch information incrementally; second, it seeks to determine whether (and if so, how) they use the fine-grained tonal information, that is, within-category gradience in level versus contour tones, over time. Given the Chinese and English listeners' different use of tonal cues found in previous studies, it is hypothesized that native and non-native Chinese listeners would have different time courses of use of the early pitch height information, with native Chinese listeners showing earlier use of this information than English-speaking L2 learners; moreover, native and non-native Chinese listeners are hypothesized to differ in the use of the fine-grained within-category gradience of contour and level tones due to their different sensitivity to pitch contour and pitch height information .

The findings of this doctoral dissertation research make an important contribution to the understanding of how tonal information modulates lexical activation in native and non-native Chinese listeners. Crucially, the findings of this research shed light on the models of native and non-native auditory word recognition, in which the use of suprasegmental information has not yet been satisfactorily incorporated. This research also has pedagogical implications for Chinese language teaching. Finding out what aspects of tonal information L2 learners of Chinese have difficulty using in online word recognition can help Chinese language-teaching practitioners develop focused L2 teaching materials for overcoming these difficulties in the classroom.

This dissertation is organized as follows: Chapter 2, which focuses on the incremental processing of tones, reviews the literature on native and non-native listeners' use of tonal information in lexical access; then, it presents the first experiment to investigate whether and how native Chinese listeners and English-speaking L2 learners of Chinese differ in their incremental use of early pitch height differences between tones to recognize Chinese words;

Chapter 3, which focuses on the use of fined-grained tonal information, reviews the literature on native and non-native listeners' use of fine-grained phonetic information of segments and tones, and then presents the second experiment to examine whether and how native Chinese listeners and English-speaking L2 learners differ in their use of fine-grained within-category gradience of level versus contour tones to recognize Chinese words. Chapter 4 provides a general discussion of the current findings, discusses the implications of our findings for native and non-native word recognition as well as Chinese pedagogy, and concludes this dissertation.

Chapter 2: Use of Early Tonal Information in Lexical Access

2.1 Introduction

Since lexical tones can contrast word meanings in Chinese, they play an important role in Chinese listeners' spoken word recognition (e.g., Ye & Connie, 1999; Zhao, Guo, Zhou, & Shu, 2011). Some research has looked into the time course with which native Chinese listeners use this tonal information in word recognition (c.f. Malins & Joanisse, 2010; Shen, Deutsch, & Rayner, 2013, discussed below). However, it remains unclear whether native Chinese listeners would show an incremental use of tonal information, and more specifically, it is unclear how they use early pitch height information before the pitch contour is heard in the time course of lexical activation, that is, as the speech signal unfolds over time.

In contrast to Chinese, English does not have lexical tones, but it encodes pitch height differences in its lexical prosody, for instance, in stress contrasts (e.g., Beckman, 1986; Fry, 1955; Lieberman, 1960). No study to our knowledge has examined whether L2 learners' L1 prosodic system influences the time-course with which they use tonal information in word recognition. Thus, it remains unclear whether late L2 learners of Chinese would show an incremental/continuous use of tonal information—that is, whether early pitch information would constrain L2 learners' lexical access before the pitch contour information of lexical tones is available in the signal.

This study used visual-world eye tracking to investigate whether (and if so, how) native and non-native Chinese listeners differ in their incremental use of tonal information *early on in the word recognition process*. Answering this question is important in order to understand how tonal information modulates early lexical competition from words that share the same segments but differ in tones (henceforth, tonal competitors), specifically for tone pairs that differ in their early pitch height (e.g., T1 vs. T2). By comparing late L2 learners of Chinese to native Chinese listeners, this research can provide important insights on whether L2 learners whose L1 does not have lexical tone but uses a pitch cue (e.g., pitch height) to distinguish words that differ in stress placement would be able to use this cue in the early processing of lexical tones, and whether their use of this cue would differ from that of native listeners.

2.2 Research Background

2.2.1 Native Chinese Listeners' Use of Tonal Information

Lexical tones are of great importance in the recognition of spoken Chinese words: If listeners do not use tonal information, they will not recognize the intended word, as there will be a great deal of homophony (Ye & Connine, 1999). A considerable body of research found that listeners use the phonetic details of pitch (e.g., pitch height, pitch direction, pitch turning point, duration, and intensity) to identify tones (Howie, 1976; Lin & Wang, 1984; Moore & Jongman, 1997). Although other acoustic cues (e.g., duration, intensity) can signal lexical tone, pitch is the primary cue in the perception of Chinese tones. For instance, pitch contour information, used to distinguish a rising tone from a falling tone, is found to be the primary cue for Chinese listeners in their perception of lexical tones. Moreover, Chinese listeners typically attend to pitch contour information more than pitch height differences when they discriminate and identify tones in offline tasks (e.g., Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Qin & Jongman, 2016; Qin & Mok, 2013).

In addition to pitch contour information, pitch height differences are also encoded in the Chinese tonal inventory. As was illustrated in Figure 1, the four tones in Chinese are crucially different in how they change dynamically over time. Some tone pairs (e.g., T1 vs. T2) can be distinguished by the pitch height information at syllable onset position. However, other tone pairs (e.g., T1 and T4) are similar in pitch height and can be distinguished only from the later information of pitch contour.

A growing number of studies have begun using online methods such as the visual-world eye-tracking paradigm to investigate how phonetic information modulates the time course of lexical activation and competition in spoken word recognition. Previous studies on spoken word recognition have aimed to explain how listeners match the input they hear from the continuous and variable speech signal to the stored representations of words (e.g., Luce, 1986; Luce & Pisoni, 1998; Marslen-Wilson, 1989). Many studies on spoken word recognition have used visual-world eye tracking to examine how phonemes constrain lexical activation and competition (for a review, see Huettig, Rommers, & Meyers, 2011). The linking hypothesis proposed between spoken word recognition and eye movements is that "the lexical activation of a name of any given object in the display determines the probability that a listener shifts attention to that object and thus makes eye movements to fixate it" (Tanenhaus, Magnuson, Dahan, & Chambers, 2000, p. 567). In other words, the time course of lexical access can be examined by analyzing the probability of fixations to each of the visual objects (e.g., target and competitor words) as the signal unfolds over time.

Recent eye-tracking studies used the linking hypothesis to test the incremental use of phonetic information during online word recognition. For instance, Allopenna et al. (1998) is a seminal eye-tracking study using the linking hypothesis to examine lexical activation and competition. More specifically, they tested the time course of activation of lexical competitors that shared an onset or a rhyme with the target word (e.g., *beetle* as an onset competitor for *beaker*; *speaker* as a rhyme competitor for *beaker*). On the one hand, the Cohort model of spoken word recognition (e.g., Marlsen-Wilson, 1989) predicts that only onset competitors should compete with the target word in recognition because the word onset activates a cohort of lexical candidates in parallel that compete for recognition; thus, according to this model, word onsets are crucial in word recognition. On the other hand, continuous mapping models, such as the TRACE model (McClelland & Elman, 1986) and the Shortlist model (Norris, 1994), predict that both onset competitors and rhyme competitors will become active when the target word is heard because the model assumes that lexical access takes place continuously. In other words, according to continuous mapping models, lexical activation is determined not only by prior information that has been processed, but also by the current degree of match between the input and the lexical candidates, so words may be activated even if their onset does not match the word onset that was heard in the input. The results showed greater proportions of fixations to both the cohort competitor and the rhyme competitor than to phonologically unrelated (distractor) items (e.g., *carrot*). This suggests that not only onset competitors but also rhyme competitors became active when the target word was heard, as predicted by continuous mapping models. The linking hypothesis between eye movement and word recognition thus makes it possible to study lexical competition between target and competitor words by analyzing the proportions of fixation to each of these items.

Relevant to the present research are several recent eye-tracking studies that investigated how native Chinese listeners use tonal information over the course of the word recognition process: Malins and Joanisse (2010), Shen et al. (2013), and Wiener and Ito (2015, 2016). Malins and Joanisse (2010) used the visual-world eye-tracking paradigm to examine the time course with which native Chinese listeners use tonal and segmental information in word recognition. Among several conditions, their experimental design included a condition in which the target and competitor words had the same segments but differed in tone (e.g., /tɕʰuáŋ/ ‘bed’ vs. /tɕʰuāŋ/ ‘window’), and a condition in which the target and competitor words had the same tone but differed in word-final segment (e.g., /tɕʰuáŋ/ ‘bed’ and /tɕʰuán/ ‘ship’). The stimuli in these two conditions had target and competitor words that disambiguated acoustically at a similar time. The results showed that target and competitor fixations diverged at a similar time in the two conditions and did not differ significantly between the two conditions. On the basis of these findings, the authors concluded that native Chinese listeners use tonal and segmental information concurrently, with the two types of information yielding a similar time course of lexical activation. What is unclear from this study, however, is whether the early pitch height information modulates early lexical activation.

Shen et al. (2013) also used the visual-world eye-tracking paradigm, but with the purpose of determining whether the fine-grained phonetic details of Chinese tones are utilized incrementally or holistically by manipulating the tonal offset information rather than the tonal onset information. Their study examined whether the perception of T2 and T3, the most confusable tone pair for native and non-native Chinese listeners (e.g., Hao, 2012; Huang, 2001; Moore & Jongman, 1997), was influenced by the offset of the pitch contour. The pitch contours of their stimuli, illustrated in Figure 2, were created such that the pitch height of T2 (the rising tone) would be identical to that of T3 (the dipping tone) at the onset, and the two tones would begin diverging at the turning point, by lowering the onset of T2 to match that of T3. The authors included two versions of each tone: the T2-T3 tones with their prototypical offsets (high-offset T2 and low-offset T3), and two corresponding non-prototypical tones that would be 1 semitone lower for T2 (low-offset T2) or 1 semitone higher for T3 (high-offset T3) at offset. The results indicated that Chinese listeners overall showed less competition when hearing T3 stimuli than when hearing T2 stimuli, suggesting a bias to T3 stimuli, likely due to the fact that the T2 stimuli began with the pitch onset of T3. In other words, native listeners tended to give T3 responses when they heard a low F0 onset (see also Lee, Tao, & Bond, 2008). This finding thus provides indirect evidence that the pitch height information at the onset of the tone constrained Chinese listeners' early lexical access. Furthermore, the results suggested that Chinese listeners showed more lexical competition when hearing the contours with the non-prototypical offsets (e.g., high-offset T3) than when hearing the contours with the prototypical offsets (e.g., low-offset T3). This indicates that the fine-grained phonetic details of the tone offsets also constrained Chinese listeners' lexical access. The authors concluded that lexical tones are used incrementally in word recognition. A few questions arise out of this study, however. For instance, the study did not compare different onset conditions, so it did not directly test the effect of early pitch height in word recognition. Furthermore, it is unclear whether the smaller amount of competition from T2 competitors (when recognizing T3 targets) as compared to that from T3 competitors (when recognizing T2 targets) is due to the T3 pitch onset used for the T2 stimuli or to other lexical factors (e.g., syllable frequency and tonal probability).

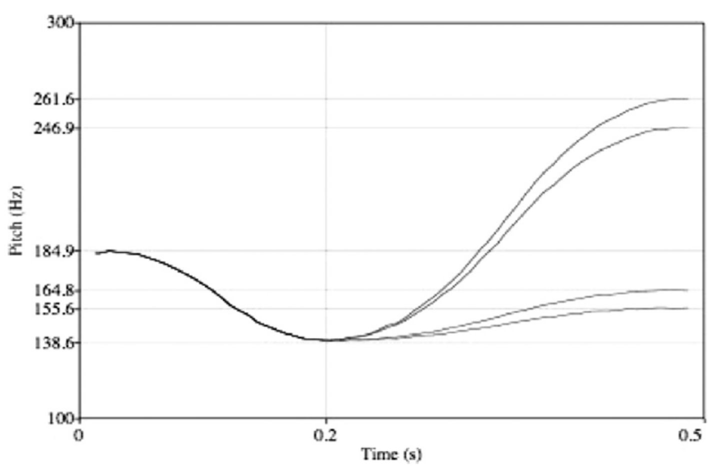

Figure 2: Tone contours of the T2-T3 stimuli used in Shen et al. (2013, p. 3020)

To address the effects of syllable frequency (i.e., syllable token frequency) and syllable-tone combination probability (i.e., the probability of a tone, relative to the other tones, co-occurring with a given syllable) on Chinese lexical access, Wiener and Ito (2015, 2016) used an eye-tracking and a gating task to investigate the influence of the lexical factors on native listeners' word recognition. Chinese monosyllabic words with higher versus lower syllable frequency as well as the most versus the least probable tone for a given syllable were used in the tasks. Wiener and Ito (2016) used a gating task to demonstrate that Chinese listeners immediately made use of syllable-tone combination probability information when hearing only the first 80 ms of the vowel. The effect of tonal probability disappeared after that. In other words, the tonal probability information likely had an effect early in the word recognition process, before sufficient pitch information became available in the signal. However, the effect of tonal probability was only found for infrequent syllables, not for frequent syllables. Likewise, Wiener and Ito (2015) used the visual-world eye-tracking paradigm to investigate how the interaction between syllable frequency and syllable-specific tonal probability guided the online lexical access of monolingual Mandarin listeners, bilingual Mandarin-Shanghainese listeners, and bilingual Mandarin-Cantonese listeners. While the mouse clicks indicated that all three groups were fastest for infrequent syllables with probable tones and slowest for infrequent syllables with improbable tones, the eye-tracking results showed that only monolingual Chinese listeners' fixations were fastest for infrequent syllables with probable tones and slowest for infrequent syllables with improbable tones. The two studies suggest that native listeners' processing of Chinese tones, at least for infrequent words, is initially guided by previously learned distributional tonal knowledge, that is, the tonal probability.

In summary, the research on Chinese listeners' use of lexical tone over the course of the word recognition process indicates that tonal information constrains native lexical access as early as segmental information. More importantly, Chinese listeners' word recognition is modulated by the later pitch contour information of lexical tones, and initially influenced by lexical factors such as tonal probability for infrequent syllables before enough pitch information is available in the signal. However, no known study has yet directly tested the question of whether early pitch height information constrains Chinese listeners' lexical access, which is important for testing the incremental use of tonal information in spoken word recognition.

To directly investigate the effect of the early pitch height in word recognition (which was not directly tested in Shen et al., 2013), the present study used the visual-world eye-tracking paradigm to investigate whether (and if so, how) native Chinese listeners make incremental use of early pitch height differences (T1-T2 vs. T1-T4) in the recognition of Chinese words. In addition to investigating native listeners' use of tonal information, this study also tests English-speaking L2 learners' use of early pitch height information in spoken word recognition.

2.2.2 Chinese L2 Learners' Use of Tonal Information

Infants have the ability to learn all the sound contrasts they are exposed to, but this ability decreases as exposure to the L1 and age of L2 acquisition increase (e.g., Flege, 1991; Flege, Munro, & Mackay, 1995; Flege, Schmidt, & Wharton, 1996; Flege, Yeni-Komshian, & Liu, 1999; Johnson & Newport, 1989; Lenneberg, 1967). Although L2 learners have difficulty in distinguishing some L2 sounds, it does not imply that their ability to distinguish non-native sounds at an acoustic level has been lost (e.g., Best, McRoberts, & Sithole, 1988). A major source of L2 learners' difficulty in learning L2 sounds is the influence of the L1 sound system (e.g., Flege, 1995; Best, 1995). For example, studies have shown that L2 learners often differ from native listeners in their use of phonetic cues to sound contrasts, especially if these phonetic cues constrain lexical access in the L2 but not the L1 (e.g., Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008; Qin et al., 2017; Shen, 1989; Stagray & Downs, 1993). Another source of L2 learners' difficulty in learning L2 sounds is that the phonetic cues cannot be integrated efficiently, with more lexical competition or/and a slow-down of the word recognition process possibly due to the perceptual confusion of phonetic categories (e.g., Broersma, 2012; Broersma & Cutler, 2008, 2011).

L2 learners often have difficulty distinguishing a non-native sound contrast if the phonetic cues are used in the L2 but not in the L1. Prior studies found that Japanese listeners had great difficulty distinguishing between English /l/ and /ɹ/ (e.g., Bradlow, Pisoni, Yamada, & Tohkura, 1997; Lively, Logan & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Yamada, 1995). In their comparison of native English listeners and Japanese-speaking L2 learners of English, Iverson et al. (2003) showed that English listeners discriminated the native /l/-/ɹ/ contrast using the third formant (F3). In contrast, Japanese listeners were more sensitive to the second formant (F2) than to F3, potentially because the Japanese liquid is perceptually more similar to English /l/, which is often cued by a low F2, than to English /ɹ/, which is often cued by a low F3 (e.g., Aoyama, Flege, Guion, Yamada, & Akahane-Yamada, 2004). Since Japanese listeners did not pay sufficient attention to F3, a cue used in the L2 contrast, but not in the L1 contrast, they found it difficult to perceive the English contrast.

Specifically for lexical tones, English-speaking L2 learners of tonal languages have been shown to have difficulty distinguishing non-native tones, as English does not use lexical tones to contrast meanings (e.g., Chandrasekaran et al., 2010; Guion & Pederson, 2007; Hallé, Chang, &

Best, 2004; Wang et al., 1999). However, English listeners do encode pitch in their lexical representations. Stress is lexically contrastive English: Word pairs such as '*record* (noun) and *re'cord* (verb) differ with respect to stress placement. Both suprasegmental cues (i.e., pitch, duration, and intensity) and segmental cues (e.g., vowel quality) contribute to the realization of word-level stress and more generally to lexical identity (e.g., Beckman, 1986; Fry, 1955; Lieberman, 1960). For instance, stressed syllables have higher pitch than unstressed syllables. In the perception of lexical tones, English listeners typically attend to pitch height differences more than to pitch contour differences, as pitch height difference is an important cue to English stress (e.g., Beckman, 1986; Bolinger, 1958; Culter & Clifton, 1984; Fear, Cutler, & Butterfield, 1995; Lieberman, 1960). Thus, English-speaking L2 learners' difficulty in distinguishing Chinese tones does not mean that they are not sensitive to pitch differences, but rather that they may selectively attend to the pitch height differences instead of the pitch contour differences of Chinese tones (e.g., Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007).

The difficulty that English-speaking L2 learners of Chinese experience in the use of Chinese tones could also be attributed to their inefficiency in integrating tonal information in the word recognition process because of their perceptual confusion of lexical tones. Generally, L2 learners appear to integrate phonetic information less efficiently and thus experience more lexical competition in word recognition as compared to native listeners (Cutler, 2012). This inefficiency has been attributed to L2 learners' perceptual confusion of phonetic categories, which results in L2 learners' difficulty in ruling out irrelevant competitor words from their initial lexical search (e.g., Broersma, 2012; Broersma & Cutler, 2008, 2011). For L2 listeners, the set of potential word candidates is multiplied by the number of L2 words that L2 learners consider during recognition due to the confusion between two phonetic categories, thus increasing the time needed to retrieve the correct word and slowing down word recognition. For instance, Broersma and Cutler (2011) showed that for Dutch L2 learners of English, the confusion between two L2 vowels (English /æ/ vs. /ɛ/) results in lexical competition from "phantom words" (i.e., words that are not present in the acoustic signal but L2 learners hear). Dutch-speaking L2 learners of English often confuse English /æ/ vs. /ɛ/, as there is no such segmental contrast in Dutch. Because of this confusion, L2 learners experience competition from *deaf* when hearing the word *daffodil*, unlike English listeners. In an auditory lexical decision task, Dutch listeners accepted near-words (e.g., *daf*) as real English words (e.g., *deaf*) more often than English listeners did. Similarly, in a cross-modal priming experiment, near-words extracted from word or phrase contexts (*daf* from *DAFfodil*) induced activation of the corresponding real words (*deaf*) for Dutch listeners, but not for English listeners. Thus, the lack of a particular contrast in the L1 leads L2 learners to activate irrelevant words (e.g., *daffodil*) and/or non-words (e.g., *daf*), resulting in more lexical competition and making L2 word recognition less efficient.

Returning to lexical tones, if English-speaking L2 learners of Chinese find it difficult to map the acoustic signal onto different tonal categories due to potential perceptual confusion, they will also have difficulty in using tones to constrain lexical access (since English does not use lexical tones to contrast word meanings), and thus they will experience more lexical competition, as a result of activating irrelevant words and/or non-words, than native Chinese listeners would. Only two previous L2 studies have compared native listeners and English-speaking L2 learners of Chinese in their processing of tonal information in lexical access. Sun (2012) used an auditory lexical decision task to investigate how tonal neighborhood density influenced native Chinese listeners and English-speaking learners of Chinese in their processing of Chinese tones. The results showed that both native and non-native listeners correctly recognized fewer words from dense tone neighborhoods (i.e., Chinese syllables that can be associated with three tones) than from sparse tone neighborhoods (i.e., Chinese syllables that can be associated with only one tone), with the former being recognized more slowly than the latter. That is, words with more tone neighbors caused greater tonal competition for both native and non-native listeners, thus resulting in lower accuracy rates and long reaction times (RTs) for those words than for words with fewer tone neighbors. However, non-native listeners' performance was still inferior to that of native listeners, with non-native listeners having significantly lower accuracy and longer RTs than native listeners. This difference between native and non-native listeners was attributed to non-native listeners' difficulty in identifying tones accurately and in using tonal information to promptly activate and select the correct lexical item.

Unlike Sun (2012), who used a natural language, Wiener (2015) used an artificial language in which syllable frequency (higher vs. lower) and syllable-tone combination probability (a tone contour was the most probable vs. the least probable to occur with a given syllable) were manipulated in order to investigate how syllable frequency and probability of co-occurrence of syllables and tones in Chinese affect spoken word recognition by native and non-native listeners. Over a four-day training period, native Chinese listeners, English-speaking L2 learners of Chinese, as well as monolingual English listeners learned CV + tone nonce words, each paired with a black-and-white nonce symbol. The results of a visual-world eye-tracking task after the training revealed that all three groups showed greater early target fixations in the condition with infrequent syllables containing the most probable tones than in the condition with infrequent syllables containing the least probable tones, but the effect of tonal probability did not emerge in the conditions with frequent syllables. However, the three groups differed in the timing with which they integrated tonal probabilities: Whereas Chinese listeners showed their sensitivity to tonal probability the earliest, L2 learners showed a later sensitivity to this information than Chinese listeners, and English monolinguals showed their sensitivity the latest.

The findings of Sun (2012) and Wiener (2015) suggest that English-speaking L2 learners of Chinese showed a word recognition pattern similar to that of Chinese listeners, but their use of tonal information was not as efficient as that of native Chinese listeners, with L2 learners showing lower accuracy and later use of tonal information than Chinese listeners during online word recognition.

To further examine native and non-native listeners' incremental use of tonal information in their word recognition process, the present study investigates whether (and if so, how) native Chinese listeners and English-speaking L2 learners of Chinese differ in their use of early pitch height information, which distinguishes tones early on, as the speech signal unfolds during the word recognition process. Examining this question will help elucidate whether late L2 learners whose L1 does not have lexical tones but encodes pitch height in their lexical representations would be able to use this tonal information in early L2 word recognition. More specifically, this study will test whether L2 learners would be able to use the pitch height information in the early processing of lexical tones, and whether their use of this cue would differ from that of native listeners. Answering these questions will also help incorporate lexical tones into psycholinguistic models of native and L2 spoken word recognition, as well as models of L2 sound learning.

2.3 The Present Study

The present study used the visual-world eye-tracking paradigm to examine the processing of fine-grained phonetic information in lexical tones by native Chinese listeners and English-speaking L2 learners of Chinese. The visual-world eye-tracking paradigm provides a sensitive measure of spoken language processing in which listeners' eye fixations to objects in a display are closely time-locked to the speech signal without interrupting it (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This method has also proven to be effective in testing listeners' use of tonal information in word recognition (e.g., Malins & Joanisse, 2010; Shen et al., 2013; Wiener & Ito, 2015).

By including native Chinese listeners and English-speaking L2 learners of Chinese, the present study has two goals. First, it investigates whether Chinese listeners make incremental use of early pitch information in word recognition — that is, whether native lexical access is constrained by early pitch-height differences between tones. Two tone pairs will be included in this experiment. As illustrated in Figure 1, a level-rising tone pair, which has a different pitch height at the onset, and a level-falling tone pair, which has a similar pitch height at onset, will serve as stimuli to test listeners' sensitivity to early pitch height and their use of tonal information in online word recognition. Given the finding that word recognition is modulated by listeners' language-specific use of prosodic cues (i.e., pitch cues), and given previous findings about listeners' incremental intake of speech information (e.g., Allopenna et al., 1998), we predict that Chinese listeners' word recognition will be constrained by early pitch height information. Thus, Chinese listeners are predicted to show more competition at onset position for T1-T4 than for T1-T2.

Second, we investigate whether English-speaking L2 learners of Chinese differ from native Chinese listeners in how pitch height at onset position (i.e. early pitch height) constrains lexical access. Given that English listeners encode pitch height in their L1 stress system, and they should be able to perceive pitch height differences in the perception of non-native tones (e.g., Braun et al., 2014; Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Qin &

Jongman, 2016; Qin & Mok, 2013), we hypothesize that, like Chinese listeners, English listeners will also be sensitive to early pitch height differences during online word recognition. However, given English listeners' less efficient use of tonal information than Chinese listeners, either due to perceptual confusion or due to less robust representations of lexical tones (e.g., Sun, 2012; Wiener, 2015), it is hypothesized that English listeners will overall experience more tonal competition, and will also recognize words with tonal competitors more slowly than Chinese listeners in both T1-T4 and T1-T2 conditions.

2.4 Method

2.4.1 Participants

Thirty-six native Chinese speakers and twenty-six highly proficient Chinese learners who spoke English as L1 and who learned Chinese as L2 in college (i.e., adult L2 learners of Chinese) were recruited for this experiment. The testing took place at the Center for Brain and Cognitive Sciences at Peking University, China (with lab access provided by Dr. Xiaolin Zhou). The participants in both language groups were college students. The participants reported normal hearing and no history of speech or language disorders.

All of the L2 learners spoke English as their L1 (i.e., both parents were native English speakers, and English was the only language learned in the household), used English dominantly until the end of high school (i.e., the primary language of K-12 education was English), and learned Chinese after the age of 12. L2 learners of Chinese were not exposed to any tone languages other than Chinese.

A detailed language background questionnaire, which can be found in Appendix A, was used to collect information about the participants' language backgrounds. The participants' biographical, language background, and proficiency information is provided in Table 1. L2 learners' proficiency in Chinese was tested with a Chinese lexical decision task adapted from LexTALE (Lemhöfer & Broersma, 2012), as well as a Chinese cloze (i.e., fill-in-the-blank) test (Yuan, 2009). The Chinese LexTALE, which can be found in Appendix B, included a total of 120 items, 80 of which were words. The cloze test, which can be found in Appendix C, included a total of 40 missing words. The L2 learners' mean scores (converted into percentages) are provided in Table 1. Participants each received the equivalent of 30 US dollars in compensation for their time.

Table 1: Participants' biographical, language background, and proficiency Information

L1	Age (year)	Gender (F/M)	AOE (year)	Years of Chinese Instruction	LOR (month)	Lexical Decision Test (%)	Cloze Test (%)
Chinese n = 36	23.3 (1.7)	25/11	birth	–	–	–	–
English n = 26	22.3 (2.9)	9/17	17.6 (3.4)	4.1 (2.3)	13 (14.3)	66.5 (8)	82.1 (10)

Note. Mean (standard deviation); AOE = age of first exposure to Chinese; LOR = length of residence in Chinese-speaking countries

2.4.2 Materials

As shown in Table 2, two types of tone pairs served as experimental conditions: pairs that differ in the early pitch height (T1-T2), and pairs that have a similar early pitch height (T1-T4). Six word pairs that carry T1 (e.g., /jā/ 'duck') and T2 (e.g., /já/ 'tooth') and another six word pairs that carry T1 (e.g., /wā/ 'frog') and T4 (e.g., /wà/ 'sock') were selected. The two words in each pair shared the same segments, contrasted in tones, and were not semantically related. The target and competitor words in the test items had approximants (i.e., /j/, /w/, and /ɥ/) as the word-initial consonant (e.g., /jā/ ‘duck’) to make sure that tonal information would be available from the word onset. Which tone in the T1-T2 and T1-T4 pairs was heard as the target (e.g., T1 heard as target, henceforth, “T1 Target”) and which tone served as the competitor (e.g., T1 used as competitor, with either T2 or T4 heard as target, henceforth, “T2/T4 Target”) were counter-balanced using two lists.

The distracter words were phonologically and semantically unrelated to the target and competitor words, but like the target and competitor words, as a pair they shared the same segments and differed only in tones. Their similar phonological overlap thus prevented possible baseline effects in the results (i.e., the participants did not know ahead of time which pair would be the target and competitor). In the test trials, when the target and competitor words carried T1/T2, the two distracter words carried T3 and T4; and when the target and competitor words carried T1/T4, the two distracters carried T2 and T3.

Table 2: Target, competitor, and distracter words in test items

Condition	Target or Competitor		Distracter 1	Distracter 2
T1-T2	/jā/鸭 ‘duck’	/já/牙 ‘tooth’	/tɕǐŋ/井 ‘well’	/tɕìŋ/ 镜 ‘mirror’
	/jān/烟 ‘cigarette’	/ján/岩 ‘rock’	/tǎu/岛 ‘island’	/tàu/稻 ‘rice’
	/jāŋ/秧 ‘sprout’	/jáŋ/羊 ‘sheep’	/fěi/匪 ‘bandit’	/fèi/肺 ‘lung’
	/jī/衣 ‘clothes’	/jí/姨 ‘aunt’	/ʂǔ/鼠 ‘rat’	/ʂù/树 ‘tree’
	/jīŋ/婴 ‘infant’	/jíŋ/蝇 ‘fly’	/wěi/尾 ‘tail’	/wèi/胃 ‘stomach’
	/ɥāat 鸳 ‘Mandarin duck’	/ɥán/圆 ‘circle’	/t^{h}ǔ/土 ‘soil’	/t^{h}ù/兔 ‘rabbit’
T1-T4	/jāu/腰 ‘waist’	/jàu/药 ‘medicine’	/léi/雷 ‘thunder’	/lěi/蕾 ‘bud’
	/jɤ̄/椰 ‘coconut’	/jɤ̀/叶 ‘leaf’	/xú/湖 ‘lake’	/xǔ/虎 ‘tiger’
	/jīn/音 ‘sound’	/jìn/印 ‘seal’	/lí/梨 ‘pear’	/lǐ/礼 ‘gift’
	/wā/ 蛙 ‘frog’	/wà/袜 ‘sock’	/pí/鼻 ‘nose’	/pǐ/笔 ‘pen’
	/wān/湾 ‘bay’	/wàn/腕 ‘wrist’	/ɥú/鱼 ‘fish’	/ɥǔ/雨 ‘rain’
	/wū/屋 ‘house’	/wù/雾 ‘fog’	/líŋ/铃 ‘ring’	/lǐŋ/领 ‘collar’

Table 3 presents the properties of target and competitor words in test items. As illustrated in the table, the target and competitor words in the test trials were matched for log morphemic frequency within condition (T1-T2: $t < |1|$; T1-T4: $t < |1|$) and across tonal conditions ($t < |1|$) based on the SUBTLEX-CH database (Cai & Brysbaert, 2010). Moreover, the log morphemic frequencies of the two distractors in the test trials did not differ significantly from each other (T1-T2: $t < |1|$; T1-T4: t (12) = 1.47, p = .17) or from those of the target and competitor within each condition (T1-T2: t (24) = –1.50, p = .15; T1-T4: t (24) = –1.34, p = .19).

Moreover, the target and competitor words in the test trials were matched for tonal neighborhood (defined as the number of legal tones that can be associated with a given syllable; Sun, 2012; Yip, 2000, p. 140) within condition and across conditions, because the syllables for the target and competitor words carried four legal tones in Chinese.

Table 3: Properties of target and competitor words in test items

Condition	Tone	Log-Transformed Frequency	Tonal Neighborhood	Syllable-tone Combination Probability (%)	Image Rating (1-6 scale)
T1-T2	T1 (Level)	2.80 (1.19)	4 (0)	35.7 (10.5)	5.5 (0.8)
	T2 (Contour)	3.18 (0.41)	4 (0)	64.3 (10.5)	5.0 (1.3)
T1-T4	T1 (Level)	3.19 (0.70)	4 (0)	51.7 (23.2)	4.9 (1.3)
	T4 (Contour)	3.32 (0.52)	4 (0)	48.3 (23.2)	5.1 (1.4)

Note. Mean (standard deviation)

To tease apart the potential effect of syllable-tone combination probability (as found in Wiener & Ito, 2015) from the potential effect of early pitch height, the syllable-tone combination probability (i.e., the probability of a given syllable carrying a tone relative to the other tone in Chinese words) was calculated for target and competitor words in the test trials based on Neergaard’s Chinese phonological neighborhood density database (Neergaard, Xu, & Huang, 2016). Whereas the syllables selected for the T1-T2 condition were more likely to carry a T2 than a T1 (t (12) = –5.2, $p < .0001$),[1] the syllables selected for the T1-T4 condition did not differ in their probability of carrying a T1 or T4 ($t < |1|$).

In addition to the test trials, twelve filler trials with fricatives or affricates as word-initial consonants for the targets and the competitors (six T1-T2 and six T1-T4 filler trials), listed in Appendix D, were used to prevent approximant-initial words from standing out in the T1-T2 and T1-T4 pairs. An additional twelve T3-T4 pair trials and twelve T2-T3 pair trials, listed in Appendix D, were also created as fillers to prevent participants from focusing on the T1-T2 and T1-T4 pairs. In these filler trials, when the target and competitor words carried T3/T4, the distracter words carried T1 and T2; and when the target and competitor words carried T2/T3, the distracter words carried T1 and T4. Therefore, across all trials, the four words in each display carried four different tones and each tone was heard the same number of times. Targets and competitors in the filler trials overlapped segmentally, and they were either minimal pairs or near-minimal pairs that differed in tones. Half of the filler trials had minimal pairs, and the other half had near-minimal pairs that differed in tones as targets and competitors. A similar design was applied to the two distractors in the filler trials. The two distracters overlapped segmentally, with half of the trials having minimal pairs and the other half having near-minimal pairs that differed in tones.

The complete experiment thus included a total of 48 trials (12 test trials + 36 filler trials), in addition to 8 practice trials, and a total of 192 words/images, in addition to 32 words/images for the practice trials. Test and filler trials were pseudorandomized.

[1] The sonority of syllable-initial consonants is strongly correlated with tone type due to the historical evolution of tones. A smaller number of approximant-initial syllables appear with high-onset tones (e.g., Tone 1) than with low-onset tones (e.g., Tone2) (Norman, 1988).

All Chinese words used in the task were imageable monosyllabic nouns. In each trial of the experiment, images corresponding to the target and competitor words were presented together with images corresponding to two distracter words in the four cells of a (non-displayed) 2 x 2 grid, as illustrated in Figure 3. The relative locations of targets and competitors and of the different tones were counterbalanced across the experiment.

Figure 3: A visual display of T1-T2 trial used in the visual world paradigm (the orthographic transcriptions were not presented in the actual experiment)

To make sure that the images are representative of the words, twelve native Chinese speakers rated the goodness of each image in terms of representing each word using a 1-6 scale (with 1 meaning very bad and 6 meaning very good). As shown in Table 3, the images representing the target and competitor words in the T1-T2 and T1-T4 test trials showed high mean rating scores. The rating scores for the T1-T2 pairs did not differ significantly from those for the T1-T4 pairs ($t < |1|$).

2.4.3 Stimulus Manipulation

Natural tokens of Chinese spoken words were used in this experiment. One male native Chinese speaker was recorded producing all the stimuli in a quiet room. The speaker read a randomized list of words in isolation three times at a normal speech rate. One token was chosen for each word based on the recording quality.

The intensity of all stimuli was normalized to 70 dB. Given the different duration of the naturally produced tones in the test items (T1: 453 ms; T2: 524 ms; T4: 387 ms), the duration of all stimuli was normalized at 445 ms, which is the duration mean of the natural tokens in the T1-T2 and T1-T4 pairs. Using the "To Manipulation" function in Praat (Boersma & Weenink, 2015), the natural T2 tokens in the T1-T2 pairs were shortened (e.g., T2) as they were longer than the mean, which made the rising slope of normalized T2 tokens slightly steeper than that of the non-normalized tokens; by contrast, the natural T4 tokens in the T1-T4 pairs were lengthened (e.g., T4) as they were shorter than the mean, which made the falling slope of normalized T4 tokens slightly shallower than that of the non-normalized tokens.[2]

Since the word-initial approximants of the target-competitor words in our test trials carry pitch information, the tonal onset is also the word onset for these words. Figure 4 shows the pitch track of T1-T2 stimuli and T1-T4 stimuli (top and bottom panels, respectively). While the T1-T2 target-competitor words in the test items differ in their pitch height at the onset, the T1-T4 target-competitor words are similar in their pitch height at onset.

---

[2] The overlapping portion of the T1 and T4 tonal stimuli begins with the tonal onset and ends with the last crossing point of the T1 and T4 contours (e.g., 227 ms in Figure 4). This overlapping portion increases by 5.6% after duration normalization: Whereas T4 overlapped with T1 for 44.4% (169 ms) of its length in the natural tonal stimuli, T4 overlapped with T1 for 50% (227 ms) of its length in the normalized tonal stimuli.

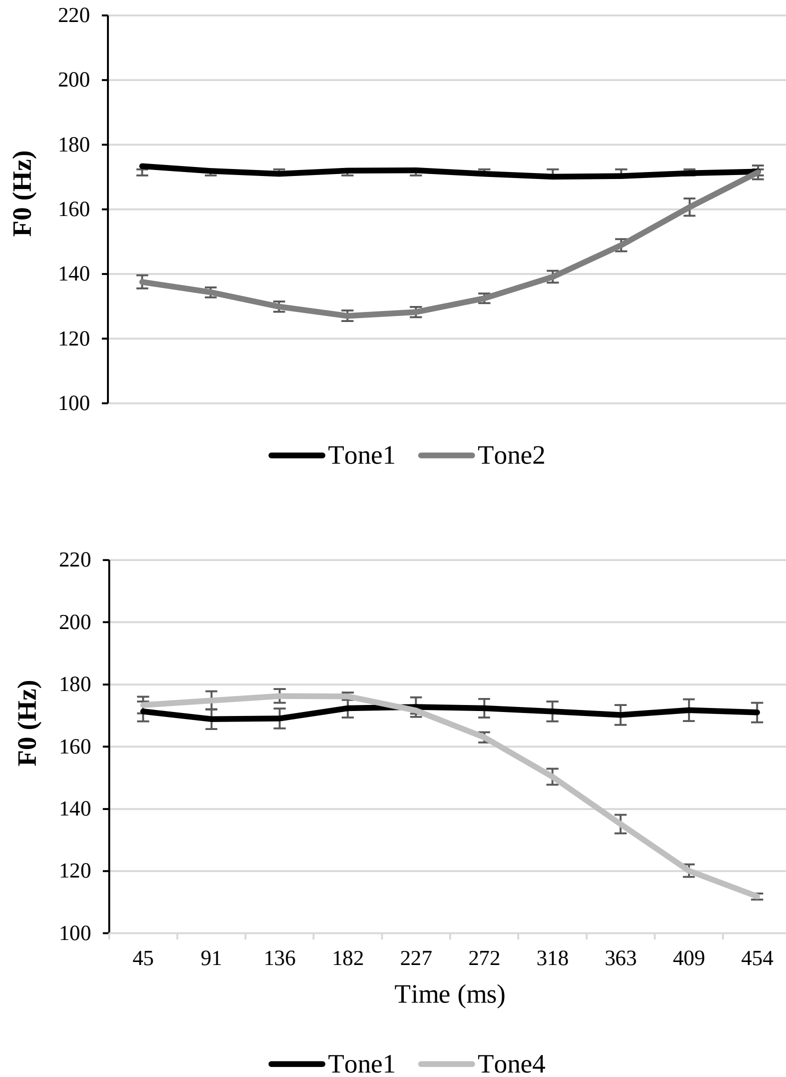

Figure 4: Pitch track using ten measurements (and standard deviations) of Tone 1-Tone 2 (top panel) and Tone 1-Tone 4 (bottom panel) target-competitor words in the test trials of Experiment 1

2.4.4 Procedures

To reduce participants' memorization burden (given that they had to learn the word-picture correspondences for all the trials before completing the eye-tracking experiment), a familiarity rating task, a training session, and an eye-tracking experiment were conducted over three sessions, with 1/3 of the words (and corresponding trials) administered over each session, and with at least two days between sessions, as shown in Table 4. Since the experiment included a total of 192 words, 64 words from Experiment 1 were used and tested within a block in each session. Only words that had been trained on the same day would appear in the eye-tracking word recognition task. The order of the different sessions was counterbalanced across participants.

Prior to completing the eye-tracking experiment, participants first completed a word familiarity task in which they rated their familiarity with the words corresponding to all the images in the experiment (including the distracter words in test trials and the words in filler trials) on a scale from 0 to 4 (0 = “I have never seen/heard this word”; 4 = “I have frequently seen/heard this word, I know what it means, and I can provide a definition for it”). In this task, a spoken stimulus was played, its printed name was presented in Pinyin and (both simplified and traditional) Chinese characters, and participants rated their familiarity with the word. English listeners’ familiarity rating *differences* between targets and competitors were not significantly different ($t < |1|$) between the T1-T2 conditions (Mean: 0.24; SD: 2.6) and the T1-T4 conditions (Mean: 0.02; SD: 2.0). The familiarity rating task in each session took about 10 minutes to complete.[3]

[3] English listeners’ familiarity ratings with the Chinese words showed the ratings between target and competitor words were also not significantly different from each other in either the T1-T2 condition ($t < |1|$) or the T1-T4 condition ($t < |1|$). And that the familiarity rating difference between the T1 and T2 words was not significantly different from that between the T1 and T4 words ($t < |1|$).

Table 4: Experimental procedures across the three sessions

	Total	Session A	Session B	Session C
Procedures		Familiarity rating Training Experiment 1 Experiment 2	Familiarity rating Training Experiment 1 Experiment 2	Familiarity rating Training Experiment 1 Experiment 2
Target-competitor Words	12 words for each tone pair	4 words for each tone pair	4 words for each tone pair	4 words for each tone pair
Training Words	192 words	64 words	64 words	64 words

Next, participants completed a word-picture association training: First, they went through a look-and-listen phase in which they heard a spoken word and saw the corresponding picture on the screen; second, they completed a picture selection test in which they heard a spoken word and selected the picture corresponding to the word from a large number of candidate pictures (including both the target and competitor pictures and two distracter pictures from test and filler trials). The target and competitor words from the same trial were displayed in the same set of pictures (22 or 20 pictures in each set) on the screen to make sure that the participants could distinguish the tonal contrast before the eye-tracking task. Participants received feedback on their responses, and the task ended only when they correctly identified all the pictures. This training was essential for four reasons: First, although all words were imageable, it was difficult to perfectly control the imageability of the words; second, L2 learners were not as familiar with the words as native listeners; third, and crucially, listeners' exposure to the auditory words allowed them to familiarize themselves with the pitch range of the speaker, which was crucial for them to be able to use early pitch height in word recognition; fourth, the training helped direct the participants' attention to pitch information, since the training stimuli also had their duration and intensity normalized. The training in each session took native listeners 15-20 minutes, and took L2 learners about 30-40 minutes to complete.

Following this training session, the visual-world eye-tracking word recognition task was administered. The experiment was conducted at the Center for Brain and Cognitive Sciences at Peking University, China. Eye movements were recorded using a desktop-mounted Eyelink 1000 (sampling rate: 1000 Hz), and the experiment was delivered using the software Experiment Builder (www.sr-research.com). In the task, participants were instructed to click on the picture corresponding to the monosyllabic Chinese word they heard through headphones. The visual display contained four black-and-white pictures (200 x 200 pixels) in a non-displayed 2 x 2 grid, as illustrated in Figure 3. On each trial of the experiment, participants first saw these four pictures for 2 seconds (preview phase). The pictures then disappeared and a fixation cross centered on the screen appeared and stayed on the screen for 500 ms. The fixation cross then disappeared and the four pictures reappeared on the screen (in the same location) and the auditory stimulus was simultaneously heard (through headphones). Participants' eye movements were recorded from the onset of the auditory stimulus, with the target word being presented in isolation, in each of the four regions of interest (300 x 300 pixels). The task began with 8 practice trials followed by the main experiment. The eye-tracking word recognition task in each session took 15-20 minutes to complete.

2.4.5 Data Analysis

Only trials in which participants clicked on the target word were analyzed, resulting in the exclusion of 1% of the data for Chinese listeners and 25% of the data for English listeners.

Proportions of fixations to the target, competitor, and distracter words were extracted in 8-ms time windows from the onset to the offset of the target word, with an adjustment of 200-ms delay (it takes approximately 200 ms for eye movements to reflect speech processing; Hallett, 1986). The dependent variable for the statistical analyses was the difference between the proportions of target and competitor fixations (i.e., the proportion of competitor fixations was subtracted from the proportion of target fixations) from the target word onset (i.e., 0 ms) to the word offset (i.e., 454 ms), with a delay of 200 ms. This dependent variable reflects the amount of lexical competition listeners experienced while factoring out overall processing-speed differences that are not due to lexical competition, thus making the data more comparable between native listeners and L2 learners.

The difference between participants' proportions of target and competitor fixations was plotted from the target word onset (i.e., tone onset) to the first 1000 ms. Since we are interested in the listeners' incremental use of tonal information, a time-window analysis, in which participants' differential fixations can be time-locked with early and late tonal information in separate time-windows, is appropriate to analyze the data collected from this experiment.[4] To investigate the use of early pitch height information, an early target-word window was defined on the basis of whether the early pitch portion of T1 and T4 words overlapped. Since the T1-T4 words in the test items had their early pitch overlap for an average duration of 180 ms (see Footnote 2 for details about the measurement of the overlapping portion) and the T1-T2 words did not overlap in the first 180 ms, the early target-word window was defined as the first 180 ms

[4] Growth curve analysis, which is used to analyze the eye fixations in Experiment 2, does not make it possible to test *when* effects emerge during processing.

from the word onset. For both T1-T4 and T1-T2 words, this early target-word time window also did not have any pitch contour information, as illustrated in Figure 4. The early target-word time window was then divided into two windows of 90 ms. The division of the early target-word window would allow a better understanding of listeners' incremental use of early pitch height information, which could be masked by the early tonal probability effect found in the first 80 ms of the vowel in Wiener and Ito (2016). In other words, listeners' tonal processing in the second half rather than the first half of the early target-word time window is less likely to be masked by the tonal probability effect, which often kicks in early in word recognition, before sufficient pitch information is available (Wiener & Ito, 2016). A late target-word time window was included to test whether early pitch height information continues modulating lexical access after the pitch contour information has been heard. A post target-word time window was also included in case the effect of early pitch height would emerge after the word has been heard. Finally, a baseline time window, the first 200 ms of the trial, was included to rule out baseline effects.

The difference between proportions of target and competitor fixations were analyzed in five time windows: a baseline time window (0-200 ms; time during which participants should not show any effect); a first half of early target-word time window (from 201 ms to the midpoint of the early time window); a second half of early target-word time window (from the midpoint to 380 ms, the endpoint of the early time window); a late target-word time window (from 381 ms to 654 ms, the target word offset); a post target-word time window (from 655 ms to 800 ms). The critical time windows for testing the listeners' early use of pitch height information are the early target-word time windows, in which the pitch height information but not the pitch contour is available in the acoustic signal. Moreover, the late and post target-word time windows could show an effect of early pitch height for listeners, especially for L2 learners, who may experience more tonal competition than native listeners. No effect should be found in the baseline time window.

Linear mixed-effects models were conducted on the differences between proportions of target and competitor fixations in each time window using the lme4 package in R (for discussion, see Baayen, 2008). For the sake of clarity, we first present the analysis of the individual language groups' results.[5] These analyses included Condition (T1-T2 vs. T1-T4, with T1-T2 as baseline) as a within-participant factor. Since the words in the experimental condition showed that T2 words had a higher probability of co-occurring with the syllables than T1 words in the T1-T2 condition (unlike T1 and T4 words in the T1-T4 condition, which were matched for this probability), a possible effect of tonal probability may need to be teased apart from the effect of early pitch height. Therefore, Tone (T1 heard as target vs. T2/T4 heard as target, with T1 Target items as baseline) was also included as a within-participant factor. A backward-fitting function from the package *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015) was used to identify the model that accounted for significantly more of the variance than all simpler models, as determined by log-likelihood ratio tests; only the results of the model with the best fit are presented, with *p* values being calculated using the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2016). Analyses yielding significant interactions between Condition and Tone were followed up by subsequent models conducted separately for T1 Target items and T2/T4 Target items, with the alpha level being adjusted to .025. All of the analyses included participant and item (i.e., target word) as crossed random variables.

[5] The dissertation targeted highly proficient L2 learners of Chinese and did not aim to investigate L2 development, so proficiency, which was not manipulated in our study, was also not included in the models of the L2 group.

To determine whether the L2 group differed from native listeners, we also conducted analyses that tested the interaction between Language (Chinese vs. English, with Chinese as baseline) and Condition (T1-T2 vs. T1-T4, with T1-T2 as baseline) separately for the T1 Target and T2/T4 Target items to simplify models (e.g., a full model with Group, Condition, and Tone which could yield a three-way interaction). These analyses followed the same procedure for model selection and included the same random effects as those described for the individual groups.

If Chinese listeners are able to use the early pitch-height information in word recognition, we should find a main effect of Condition (difference between proportions of target and competitor fixations: T1-T2 > T1-T4) in at least one of the early target-word time windows. This effect would be likely to emerge in the second half of early target-word window, at which point more pitch information is available and thus listeners' tonal processing is less likely to be masked by the information of tonal probability (compared to the first half of early target-word window). As shown in Table 3, however, T2 words have a higher syllable-tone combination probability than T1 words in the T1-T2 condition, but there is little difference between T1 and T4 words in the T1-T4 condition. Thus, we might find an interaction between Tone and Condition, that is, an effect of Tone for the T1-T2 condition, but not the T1-T4 condition (difference between proportions of target and competitor fixations: T2 > T1) in the first half of the early target-word window, at which point listeners' initial tonal processing might be masked by the tonal probability bias based on findings in Wiener & Ito (2015, 2016).

If English listeners are also sensitive to early pitch height information, the results would yield a significant effect of Condition for English listeners (difference between proportions of target and competitor fixations: T1-T2 > T1-T4). However, due to perceptual difficulties and/or due to less robust representations of lexical tones, English listeners might show a greater tonal competition and a slower word recognition process than Chinese listeners.[6] Then, the English listeners might display an effect of Condition only late in the word recognition process (i.e., a main effect of Condition found in the late or post target-word window). However, similarly to Chinese listeners, English listeners may show sensitivity to tonal probability information. Thus, English listeners may show an interaction between Tone and Condition, that is, an effect of Tone for the T1-T2 condition, but not the T1-T4 condition (difference between proportions of target and competitor fixations: T2 > T1), before showing an effect of early pitch height.

2.5 Results

As mentioned in the previous section, Chinese listeners performed at ceiling in identifying the target word (mean accuracy: 99.8%; SD: 4.3%), whereas English listeners were less accurate (mean accuracy: 75.1%; SD: 49%). Chinese listeners clicked on the target picture at an average of 1,568 ms (SD: 483 ms), whereas English listeners did so at an average of 2,701 ms (SD: 983 ms). The results of the mouse clicks are consistent with those of previous offline studies (e.g., Sun, 2012) in that Chinese listeners had a higher accuracy and shorter RTs than English listeners.

Figure 5 shows Chinese and English listeners' differential proportions of fixations in the T1-T2 and T1-T4 conditions (see Figure 14 in Appendix E for Chinese and English listeners'

---

[6] Perceptual confusion and weak/unstable representations of lexical tones may not be independent from each other (e.g., lexical representations can be unstable due to factors other than perceptual confusion, but perceptual confusion should cause unstable lexical representations). The present study was not designed to tease apart these possible explanations of the English listeners' performance.

separate proportions of target, competitor, and distracter fixations in the T1-T2 and T1-T4 conditions). Differential proportions of fixations above 0 mean that participants looked more at the target than at the competitor. Chinese and English listeners' differential proportions of fixations in different time-windows are presented in, respectively, the left and right panels of Figure 5. The time-window analysis was conducted separately for Chinese and English listeners to investigate the effect of Condition.

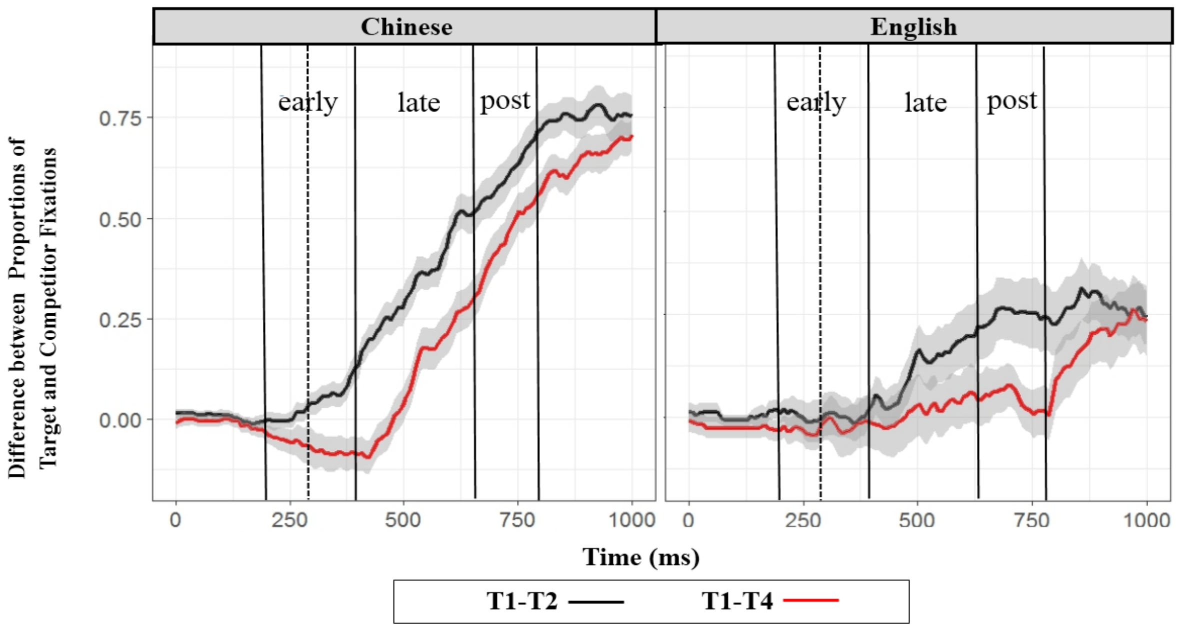

Figure 5: Difference between proportions of target and competitor fixations of T1-T2 (black) and T1-T4 (red) conditions by Chinese listeners (left panel) and English listeners (right panel) for the first 1,000 ms. The vertical solid lines show the timing of the different time windows (the first and third lines represent the onset and offset of the target word), with a 200-ms delay; the vertical dotted line represents the midpoint of the early target-word window, with a 200-ms delay

2.5.1 Chinese Listeners

Figure 6 shows Chinese listeners' differential proportions of fixations in the T1-T2 and T1-T4 conditions for T1 Target and T2/T4 Target items (see Figure 15 in Appendix E for Chinese listeners' separate proportions of target, competitor, and distracter fixations in T1-T2 and T1-T4 conditions for T1 Target and T2/T4 Target items).

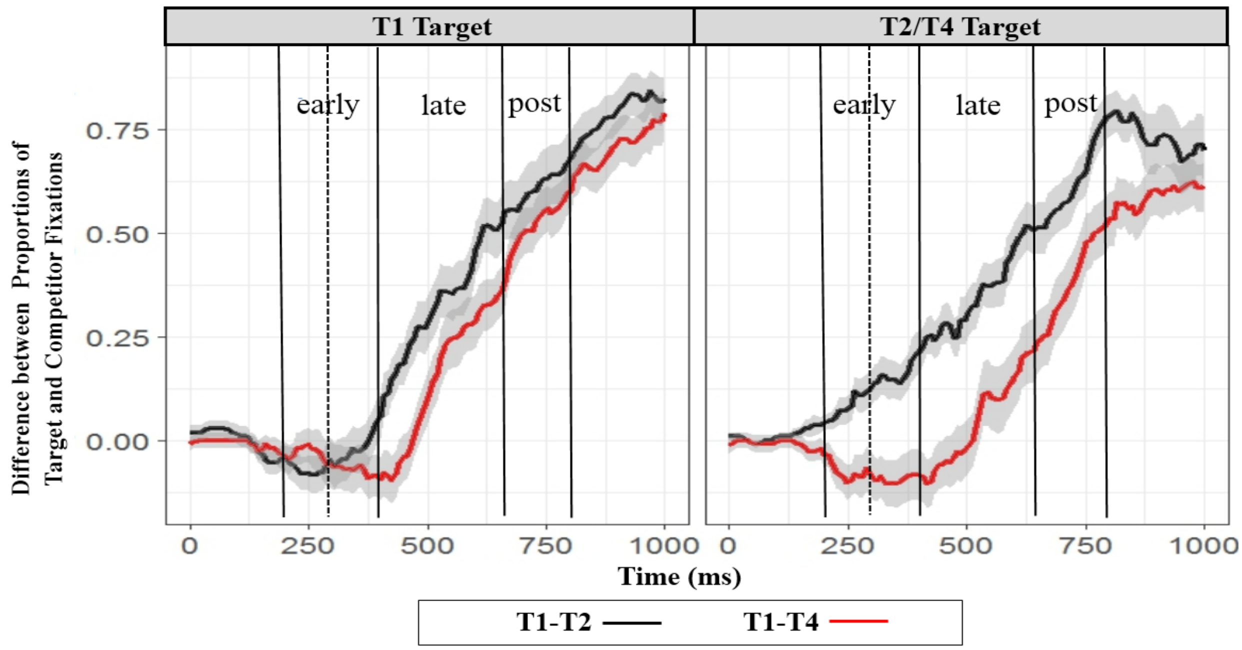

Figure 6: Chinese listeners' differential proportions of fixations of T1-T2 (black) and T1-T4 (red) conditions for T1 Target and T2/T4 Target items the first 1,000 ms. The vertical solid lines show the timing of the different time windows (the first and third lines represent the onset and offset of the target word), with a 200-ms delay; the vertical dotted line represents the midpoint of the early target-word window, with a 200-ms delay

Table 5 presents the results of the time-window analyses with the best fit on Chinese listeners' differential proportions of fixations in all conditions.

Table 5: Results of time-window analyses on Chinese listeners' differential proportions of fixations

Time window	**Variable**	**Estimate**	**Std. Error**	***t***	***p***
Baseline	(intercept)	-0.001	0.009	<|1|	.80
The first half of early target-word	(intercept)	–0.070	0.042	–1.67	.10
	Condition	0.049	0.058	<|1|	.39
	Tone	0.146	0.058	2.52	**.01**
	Condition × Tone	–0.208	0.082	–2.53	**.01**
The second half of early target-word	(intercept)	0.052	0.040	1.29	.21
	Condition	–0.139	0.055	–2.52	**.01**
Late target-word	(intercept)	0.308	0.052	5.91	**<.001**
	Condition	–0.222	0.052	–4.315	**<.001**
Post target-word	(intercept)	0.610	0.055	11.1	**<.001**
	Condition	–0.158	0.053	–2.98	**<.01**

Note. α = .05; significant results are in bold. Each model: *n* = 430 observations.

The effects in Table 5 can be summarized as follows. Although Chinese listeners did not show an effect of Condition in the first half of early target-word window, they showed an effect of Condition in the second half of early target-word window as well as in the late and post target-word windows, which indicate that Chinese listeners had a higher differential proportion of fixations in the T1-T2 condition than the T1-T4 condition in these windows. They also showed an effect of Tone in the first half of early target-word window, which indicates that Chinese listeners had a higher differential proportion of fixations for the T2/T4 Target items than the T1 Target items in the window. Moreover, an interaction between Condition and Tone was found in the first half of early target-word window, which indicates that the effect of Condition differed as a function of Tone in the window.

This two-way interaction in the first half of the early target-word window warranted additional analyses to test for the effect of Condition (T1-T2 vs. T1-T4, with T1-T2 as baseline) separately for items with T1 vs. T2/T4 as target. For these additional analyses, the alpha level was adjusted to .025. While Chinese listeners did not show an effect of Condition for T1 Target items [t (209) < |1|, p>.1], they showed an effect of Condition for T2/T4 Target items [t (215) = –3.20, p = .009]. Additional analyses also tested for the effect of Tone (Level tone vs. Contour tone, with Contour tone as baseline) separately for T1-T2 and T1-T4 conditions, with the alpha level being adjusted to .025. While Chinese listeners showed an effect of Tone for the T1-T2 condition [t (215) = –2.60, p = .01], they did not show an effect of Tone for the T1-T4 condition [t (209) = 1.0, p = .3].

In summary, the Chinese listeners' results revealed an interaction between Condition and Tone in the first half of the early target-word window, with an effect of Condition (i.e., a higher differential proportion of fixations in the T1-T2 condition than the T1-T4 condition) for T2/T4 Target items but not for T1 Target items. The follow-up analyses also indicated that Chinese listeners showed an effect of Tone in the T1-T2 condition, but not in the T1-T4 condition, with a higher differential proportion of fixations for T2 words than for T1 words. The absence of the effect of Condition in the first half of the early time window for items where T1 was the target thus appears to be due to syllable-tone combination probability, with the lower tonal probability of T1 than T2 in the T1-T2 condition resulting in the lack of effect for Condition. Similarly, the effect of Condition found in the first half of the early time window for items where T2/T4 were the targets may be caused in part by the tonal probability, with the higher tonal probability of T2 than T1 possibly inflating the effect of Condition. In other words, the effect of Condition, found for the T2/T4 Target items in the first half of the early time window, cannot be attributed straightforwardly to the early pitch height information. This suggests that Chinese listeners' use of early pitch height in the first half of the early target-word window might be masked by the early tonal probability effect (Wiener & Ito, 2015; 2016).

Crucially, these results indicate that only the main effect of Condition was found, and no interaction between Condition and Tone was found for the second half of the early, late, and post target-word time window. In other words, the early pitch height information accounts for Chinese listeners' differences between conditions in these windows, in which tonal probability does not matter as much given that enough pitch information of the lexical tones has been heard. This suggests that the early pitch height information started constraining Chinese listeners' lexical access before the pitch contour information had been heard, and continued its effect after the pitch contour information was heard. [7]

2.5.2 English-Speaking L2 Learners of Chinese

Figure 7 shows English listeners' differential proportions of fixations in the T1-T2 and T1-T4 conditions for T1 Target and T2/T4 Target (see Figure 16 in Appendix E for English listeners' separate proportions of target, competitor, and distracter fixations in T1-T2 and T1-T4 conditions for T1 Target and T2/T4 Target items).

[7] As in illustrated in Figure 4, T2 contour does not rise, and T4 contour does not fall yet in the early target-word window (i.e., the first 180 ms after tonal onset). Thus, the effects found in the first and second half of the early target-word window should not be due to the pitch contour information in the signal.

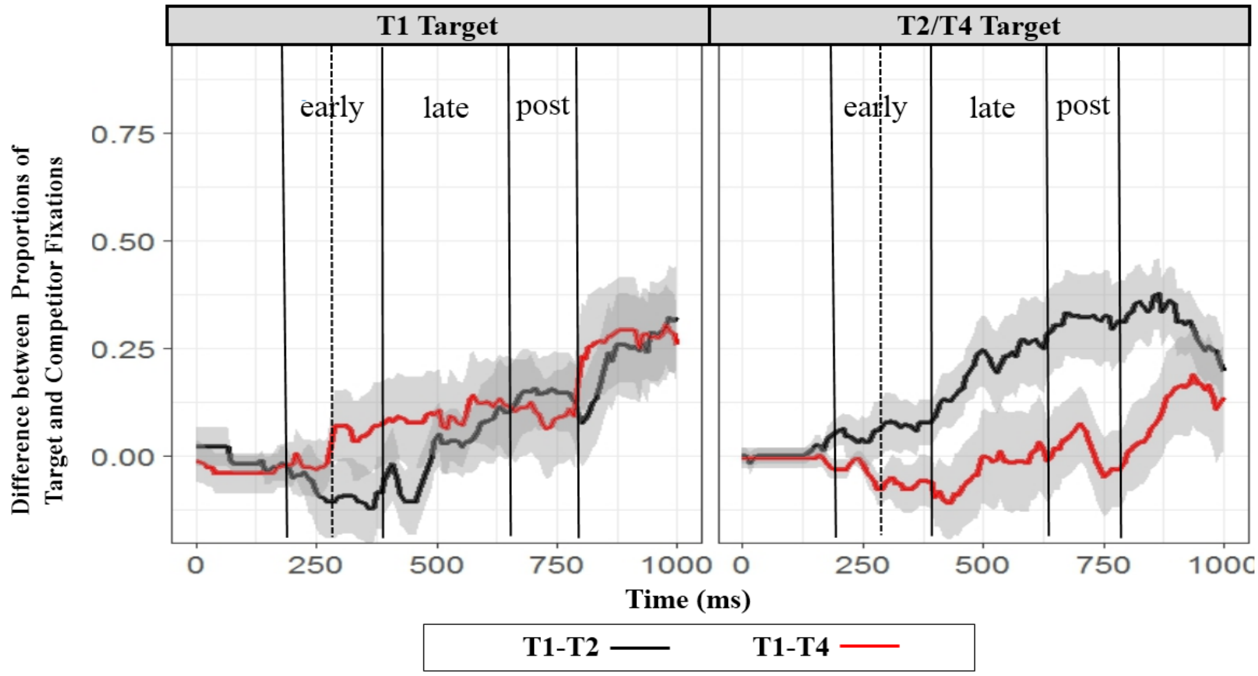

Figure 7: English listeners' differential proportions of fixations of T1-T2 (black) and T1-T4 (red) conditions for T1 Target and T2/T4 Target items the first 1,000 ms. The vertical solid lines show the timing of the different time windows (the first and third lines represent the onset and offset of the target word), with a 200-ms delay; the vertical dotted line represents the midpoint of the early target-word window, with a 200-ms delay

Table 6 presents the results of the time-window analyses with the best fit on English listeners' differential proportions of fixations in all conditions. The effects in Table 6 can be summarized as follows. English listeners did not show an effect of Condition in the early and late target-word windows. However, English listeners showed an effect of Condition in the post target-word windows, which indicates that they had a higher differential proportion of fixations in the T1-T2 condition than the T1-T4 condition in the window. They also showed an interaction of Condition and Tone in the late target-word window, which indicates that the effect of Condition differed as a function of Tone.

Table 6: Results of time-window analyses on English listeners' differential proportions of fixations

Time window	Variable	Estimate	Std. Error	*t*	*p*
Baseline	(intercept)	–0.001	0.018	<|1|	.90
The first half of early target-word	(intercept)	–0.017	0.025	<|1|	.60
The second half of early target-word	(intercept)	–0.009	0.049	<|1|	.84
	Condition	–0.027	0.068	<|1|	.69
Late target-word	(intercept)	0.017	0.076	<|1|	.82
	Condition	0.040	0.99	<|1|	.69
	Tone	0.187	0.098	1.91	.06
	Condition × Tone	–0.33	0.138	–2.42	**.02**
Post target-word	(intercept)	0.610	0.055	11.1	**<.001**
	Condition	–0.158	0.053	–2.98	**<.01**

Note. α = .05; significant results are in bold. Each model: *n* = 233 observations.

This two-way interaction in the late target-word window warranted additional analyses to test for the effects of Condition (T1-T2 vs. T1-T4, T1-T2 as baseline) separately for T1 Target items and T2/T4 Target items, with the alpha level being adjusted to .025. Although English listeners did not show an effect of Condition for T1 Target items [t (112) < |1|, p > .1], they showed an effect of Condition for T2/T4 Target items [t (100) = –3.0, p = .003]. Additional analyses to test for the effect of Tone (Level tone vs. Contour tone, Contour tone as baseline) were conducted separately for the T1-T2 and T1-T4 conditions in the late target-word window, with the same alpha-level adjustment. English listeners did not show an effect of Tone for the T1-T2 condition [t (116) = –1.82, p = .071] or for the T1-T4 condition [t (97) = 1.45, p = .15].

To sum up, the English listeners' results revealed an interaction between Condition and Tone in the late target-word window, with an effect of Condition (i.e., a higher differential proportion of fixations in the T1-T2 condition than the T1-T4 condition) found for T2/T4 Target items, but not for T1 Target items. Follow-up analyses indicated that English listeners did not show a significant effect of Tone either in the T1-T2 condition or in the T1-T4 condition in the late target-word window. Thus, the interaction between Condition and Tone cannot be attributed to a difference of tonal probability between T1 and T2. The effect of Condition found in the late time window for items with T2/T4 target words (and T1 competitor words) is thus better explained by the use of early pitch height.

Furthermore, the English listeners' results for the post-target-word time window revealed only a main effect of Condition and no interaction between Condition and Tone. These results showed that early pitch height information started constraining English listeners' lexical access in the late target-word window and continued its effect in the post target-word window for T2/T4 Target items. However, for T1 target items, it started constraining English listeners' lexical access only in the post target-word window. In other words, English listeners did not start using early pitch height information until they heard the pitch contour information of lexical tones. This suggests that English listeners experienced more tonal competition than Chinese listeners, resulting in a slower word recognition process, an effect likely due to their perceptual confusion among tones and/or their weak representations of lexical tones.

2.5.3 L2 Learners vs. Native Listeners

To determine whether L2 learners differ from native listeners in their use of pitch height information, time-window analyses were conducted on all listeners' differential fixations separately for the T1 Target and T2/T4 Target items on the basis of the interaction between Condition and Tone found in the previous analyses. This analysis included Condition (T1 Target vs. T2/T4 Target, T1 Target as baseline), Language (Chinese vs. English, Chinese as baseline), and their interaction as fixed effects. Table 7 shows the results of the time-window analysis with the best fit.

For T1 Target items, the time-window analysis yielded a significant effect of Language, with Chinese listeners having a higher differential proportion of fixations than English listeners, in the late and post target-word windows. For T2/T4 Target items, the time-window analysis yielded a significant effect of Condition, with a higher differential proportion of fixations in the T1-T2 condition than T1-T4 condition, in the early, late, and post target-word window. The analysis also yielded a significant effect of Language, with Chinese listeners having a higher differential proportion of fixations than English listeners, only in the post target-word window. No interaction between Condition and Language was found in any time window.

Table 7: Results of time-window analysis on Chinese and English listeners' differential proportions of fixations

Tone	**Time window**	**Variable**	**Estimate**	**Std. Error**	***t***	***p***
T1 Target	Baseline	(intercept)	–0.007	0.020	<|1|	.90
	The first half of early target-word	(intercept)	–0.048	0.045	–1.06	.40
	The second half of early target-word	(intercept)	–0.057	0.053	–1.08	.41
	Late target-word	(intercept)	0.215	0.064	3.34	**.01**
		Language	–0.174	0.061	–2.850	**<.01**
	Post target-word	(intercept)	0.558	0.058	9.60	**<.001**
		Language	–0.418	0.077	–5.38	**<.001**
T2/T4 Target	Baseline	(intercept)	0.003	0.013	<|1|	.90
	The first half of early target-word	(intercept)	0.701	0.028	2.55	**.01**
		Condition	–0.130	0.039	–3.55	**<.001**
	The second half of early target-word	(intercept)	0.119	0.041	2.85	**<.01**
		Condition	–0.021	0.054	–3.819	**<.001**
	Late target-word	(intercept)	0.284	0.052	5.49	**<.001**
		Condition	–0.291	0.056	–5.19	**<.001**
	Post target-word	(intercept)	0.635	0.060	10.54	**<.001**
		Condition	–0.367	0.071	–5.17	**<.001**
		Language	–0.254	0.063	–4.03	**<.001**

Note. α = .05; significant results are in bold. T1 Target items model: *n* = 327 observations; T2/T4 Target items model: *n* = 336 observations.

In summary, the results indicate that, for both T1 Target and T2/T4 Target items, the difference in the timing of the Condition effect due to the early pitch height information in the two language groups did not result in a significant interaction between Condition and Language.

Moreover, a main effect of Condition was found in all time windows when T2/T4 were heard as the target, but not when T1 was heard as the target. This suggests that it is more likely for English and Chinese listeners to show the effect of early pitch height information when T2/T4 were used as targets than when T1 was heard as target. The effect of Condition might be driven by the results of the Chinese listeners (who showed an effect of Condition for T2/T4 Target items, but not for T1 Target items), given the larger number of Chinese listeners than English listeners in our analysis. A main effect of Language was found in both the late and post target-word windows for T1 Target items as well as in the post target-word windows for T2/T4 Target items. This suggests that English listeners experienced more tonal competition between targets and competitors than Chinese listeners, especially after the pitch contour information was available in the acoustic signal.

2.6 Discussion

This study investigates whether Chinese listeners' lexical access is constrained by early pitch height information, comparing Chinese T1-T2 word pairs (non-overlapping early pitch height) with T1-T4 word pairs (overlapping early pitch height). This study also investigates whether English-speaking L2 learners of Chinese differ from native Chinese listeners in how this early pitch height constrains lexical access. Chinese and English listeners completed a visual-world eye-tracking experiment with images representing Chinese monosyllabic words. In the T1-T2 condition, they either heard T1 (a level tone) or T2 (a rising tone) as target, and in the T1-T4 condition, they either heard T1 (a level tone) or T4 (a falling tone) as target.

Unlike Shen et al.'s (2013) study, which focused on Chinese listeners' late use of tonal information and did not directly test the effect of the early tonal information, the current study tested Chinese listeners' early use of pitch height information at the tonal onset by comparing tone pairs with similar early pitch height versus those with different early pitch height. The results of the eye-tracking experiment showed that Chinese listeners used the early pitch height information in the second half of the early target-word window, and continued using it after the pitch contour information of the lexical tone was heard. These results suggest that native Chinese listeners have an incremental use of pitch information in that the early pitch height information constrains their lexical access before the pitch contour information is available. These findings complement those of previous eye-tracking studies on lexical tones (e.g., Malins & Joanisse, 2010; Wiener & Ito, 2015), and are consistent with current word recognition models (e.g., TRACE model, McClelland & Elman, 1986) in that lexical access takes place continuously and listeners are able to incrementally process phonetic information during online word recognition.

Previous studies on tone perception showed that Chinese listeners attended less to the pitch height differences of lexical tones than to the pitch contour differences when discriminating and identifying tones (e.g., Chandrasekaran et al., 2010; Francis et al., 2008; Guion & Pederson, 2007; Qin & Jongman, 2016). Although pitch height is not the main cue that Chinese listeners rely on in tone perception, our results showed that early pitch height information was used in word recognition before the pitch contour information of lexical tones was available in the signal. Unlike previous studies that used discrimination and identification tasks (e.g., Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Qin & Jongman, 2016) and focused on listeners' ultimate response (e.g., accuracy and reaction time), the present study used an eye-tracking experiment, a more sensitive measure of online spoken language processing, to capture listeners' time course of the early pitch height effect before the pitch contour information had been heard.

Unlike Sun's (2012) and Wiener's (2015) studies, which focused on the effects of lexical factors (tonal neighborhood in Sun, 2012; syllable frequency and tonal probability in Wiener, 2015) in non-native word recognition, the current study tested English-speaking L2 learners' incremental use of tonal information and compared them with native Chinese listeners in terms of the time-course with which they started using early pitch height information. The results of the eye-tracking experiment indicated that, like Chinese listeners, English listeners were able to use the early pitch height information to recognize spoken Chinese words. These findings are consistent with my predictions that English listeners encode pitch height differences in their native stress contrast, so they are able to use this early pitch height information to recognize Chinese words. These findings are also consistent with the claim that whether or not prosodic cues contribute to distinguishing among words in the L1, and how they do so, influences L2 learners' use of these cues in the L2 (Culter, 2012; Francis & Nusbaum, 2002; Holt & Lotto, 2006).

An important implication of these findings is that models of L2 word recognition should consider not only whether the L1 has lexical tones, but also whether and how cues to lexical tones are realized in an L1 lexical representations (cf. Bent, 2005; Chandrasekaran et al., 2010; Francis et al., 2008; Huang & Johnson, 2010). If specific cues to lexical tones can be mapped onto a different lexical representation (i.e., English stress) in the L1, then it is likely that English-speaking L2 learners will be able to use these cues to recognize Chinese words. In other words, L2 learners might be able to transfer a phonetic cue from an L1 representation (lexical stress) to an L2 representation (lexical tones), if that cue allows them to distinguish different words in their L1 (see also Qin et al., 2017; Shen, 1989; White, 1981). To corroborate these findings, future research should compare L2 learners of Chinese who speak non-tone languages in which suprasegmental cues play different roles in lexical access (e.g., English, which encodes pitch height in lexical stress contrasts, vs. French, which does not encode pitch height in lexical prosody).

Unlike Chinese listeners (who started using the early pitch height in the second half of the early target-word window), however, English listeners did not use the early pitch height to recognize T1 and T2/T4 targets until the post target-word window. These results are consistent with previous L2 Chinese word recognition studies that showed that English-speaking L2 learners appeared to integrate tonal information less efficiently and process it more slowly in word recognition as compared to native Chinese listeners (e.g., Sun, 2012; Wiener, 2015). The present study showed that English listeners experienced more tonal competition than Chinese listeners, with lower differential proportions between target and competitor fixations (in the late and post target-word window for T1 Target items; in the post target-word window for T2/T4 Target items). These results may be due to English-speaking L2 learners' perceptual confusion and/or unstable representations of lexical tones. Consequently, their use of tonal information was not as efficient as that of native Chinese listeners (e.g., Hao, 2012; Wang et al., 1999, 2003). In other words, the differences between the Chinese and English listeners found in the current study can be attributed to English-speaking L2 learners' difficulty in using tonal information efficiently to rule out irrelevant competitor words from their initial lexical search, which increased lexical competition in general and might have slowed down their use of early pitch height information.

An alternative interpretation of English listeners' results exists, however: English listeners may have shown an effect of pitch height late in the word recognition process due to an effect of the overall pitch contour of tones. Since English listeners had heard the pitch contour information of tones by the time their fixations showed an effect of condition (i.e., in the post target-word window), the pitch contour differences between T1-T2 and T1-T4 may have caused the observed effect of condition. In other words, the difference between a level tone and a rising tone might be more salient for English listeners than the difference between a level tone and a falling tone. As far as we know, however, no study has found a perceptual advantage for rising contours over falling contours in English listeners (e.g., if anything, the opposite pattern of results was found for English listeners in Liu, 2013). Hence, it is more likely that English-speaking L2 learners of Chinese showed an effect of pitch height late due to the fact that they integrate tonal information less efficiently and process it more slowly in word recognition as compared to native Chinese listeners.

English listeners' inefficiency at integrating tonal information, compared with Chinese listeners, may have also been due to their difficulty using the pitch contour information (i.e., rising vs. falling) of lexical tones, as pitch contour differences are not encoded in their L1 lexical representations (Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007). The current design did not allow us to directly compare English listeners' processing pitch height versus pitch contour information, because pitch contour is ultimately confounded with pitch height information in the Chinese tonal inventory. Further research, potentially using an artificial tone language where changes of pitch contour and average pitch height in tones are manipulated separately, thus needed to tease apart the effects of pitch height and pitch contour information to investigate how these cues contribute to English listeners' integration of tonal information in word recognition.

In addition, the results showed that the different syllable-tone combination probability (e.g., between T1 and T2 words) did not account for the effect of early pitch height found for both Chinese listeners (e.g., second half of the early target-word time window) and English listeners (e.g., post target-word time window). However, in the first half of early time window, Chinese listeners showed an effect of Tone in the T1-T2 condition, but not in the T1-T4 condition. In contrast, English listeners did not show such an effect in any condition. The effect of Tone for the T1-T2 condition for Chinese listeners was likely due to the different syllable-tone combination probabilities between T1 and T2 words, with T2 more likely to co-occur with the target syllables than T1 (Wiener & Ito, 2015, 2016), shown in Table 3. The main effect of Condition found for T2/T4 Target items (in the early and late target-word windows) but not for T1 Target items in the combined analysis of Chinese and English listeners' results may have been driven by the results of Chinese listeners given the fact that there were more Chinese listeners than English listeners in our analysis. Although the current study did not aim to investigate the influence of tonal probability on native and non-native listeners' word recognition, our results suggest that at least native Chinese listeners were sensitive to the tonal probability at the very beginning of the word recognition process. In other words, native Chinese listeners were able to integrate top-down knowledge, that is, previously learned distributional lexical knowledge, with the incoming acoustic phonetic information of lexical tones in early online word recognition. These findings are consistent with previous studies on Chinese word recognition (e.g., Wiener & Ito, 2015, 2016) as well as current word recognition models (e.g., TRACE model, McClelland & Elman, 1986), in that native listeners were able to use the early pitch height information and recognized spoken Chinese words incrementally. A possible direction of future research on Chinese word recognition would be to systematically manipulate the tonal probability and pitch information of lexical tones to investigate how native and non-native listeners' lexical knowledge as well as their use of pitch cues interact with each other at the initial stage of word recognition.

Previous studies on tonal perception found that Chinese and English listeners tend to rely on different pitch cues: Chinese listeners appear to be more sensitive to pitch contour than pitch height, as pitch contour is a dominant cue to differentiate Chinese tones (e.g., T2 vs. T4); by contrast, English listeners typically attend to pitch contour less than pitch height (e.g., Bent, 2005; Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Qin & Jongman, 2016; Qin & Mok, 2013). As a result, Chinese and English listeners might process the fine-grained information of level and contour tones differently. Recent studies on word recognition (e.g., McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray, Tanenhaus, & Aslin, 2002, 2009) have found that the fine-grained phonetic information of consonants and vowels modulates online word recognition, but it is unclear whether fine-grained tonal information similarly modulates native and non-native Chinese word recognition. Therefore, the next chapter investigates whether Chinese and English listeners use fine-grained phonetic information of level and contour tones differently in their spoken word recognition.

Chapter 3: Use of Fine-Grained Tonal Information in Lexical Access

3.1 Introduction

Previous research on the perception of lexical tones has shown that Chinese listeners have a quasi-categorical perception of level-to-contour tonal contrast, and thus do not show sensitivity to fined-grained phonetic variability in tones (e.g., Hallé et al., 2004; Leather, 1987). On the other hand, English listeners have been found to be sensitive to the fine-grained within-category information of pitch height, because non-native listeners do not have robust tonal categories, and thus rely on the acoustics of pitch height rather than pitch contour that is encoded in Chinese tones (e.g., Gandour, 1983; Leather, 1987). In light of the differences between Chinese and English, it would be interesting to examine whether (and if so, how) native Chinese listeners and English-speaking late L2 learners of Chinese exhibit different sensitivity to the fine-grained within-category pitch information of contour and level tones.

Recent studies on word recognition have found that fine-grained phonetic information modulates lexical activation and competition in word recognition (e.g., Dahan, Magnuson, Tanenhaus, Hogan, 2001; McMurray et al., 2008; McMurray et al., 2002, 2009; Salverda, Dahan, & McQueen, 2003). For instance, McMurray et al. (2002) found that listeners showed gradient sensitivity to within-category Voice Onset Time (i.e., VOT) variations during online word recognition. Thus, native spoken word recognition exhibits sensitivity to within-category gradience (e.g., Dahan et al., 2001; McMurray et al., 2002; McMurray et al., 2009). Given the findings that fine-grained phonetic information constrains native listeners' lexical activation as the speech signal unfolds (e.g., McMurray et al., 2002), using a time-sensitive method like eye-tracking, we may also find that even native Chinese listeners can use fine-grained tonal information such as within-category gradience in lexical tones (most likely, contour tones) to recognize words.

This chapter aims to address how native and non-native Chinese listeners use the within-category gradience of contour and level tones to recognize words, revealing whether native and non-native listeners use the fined-grained tonal information of different pitch cues (i.e., pitch contour vs. pitch height) to resolve lexical competition during online word recognition. By comparing native listeners and L2 learners' word recognition, this research can provide insights into the factors (e.g., the influence of the L1 prosodic system) responsible for late L2 learners' (in)ability in using tonal information in word recognition after the so-called "critical period" for language acquisition (e.g., Lenneberg, 1967). Furthermore, the findings of this research will have important implications for models of auditory word recognition, which currently have no explicit mechanisms for incorporating the use of suprasegmental information.

3.2 Research Background

3.2.1 Native and Non-Native Listeners' Perception of Lexical Tones

Studies on tone identification and discrimination found that Chinese listeners showed a "quasi-categorical perception" of tones,[8] more specifically, of level-to-contour or contour-to-contour tonal continua (e.g., Hallé et al., 2004; Peng, et al., 2010; Sun & Huang, 2012; Wang, 1976; Xu et al., 2006). In contrast to level-to-contour and contour-to-contour tonal continua, level-to-level

[8] "Quasi-categorical perception," a concept borrowed from Hallé et al. (2004), was used as a more liberal view of "categorical perception". The term in that study was intended to suggest that there is a gradient categoricity in the perception of different sound categories (e.g., consonants vs. tones): Whereas stop consonants showed a drastic increase in the discriminability of cross-boundary pairs versus within-category pairs, tones did not show such a drastic pattern (also see Schouten & van Hessen, 1992).

tonal continua did not exhibit such a pattern of "quasi-categorical perception" for either native or non-native listeners (Francis, Ciocca, & Ng, 2003).

Previous studies on the categorical perception of lexical tones (e.g., Burnham & Mattock, 2007; Hallé et al., 2004; Wang, 1976) suggested that whereas Chinese listeners rely more on their native tonal system and primarily perceive tonal contour differences categorically, English listeners rely more on the psychoacoustic differences of pitch height and perceive tones non-categorically. For instance, Wang (1976) claimed that Chinese listeners, but not English listeners, showed the typical pattern of categorical perception when perceiving a dynamic tonal continuum varying from a level tone to a rising tone. While Chinese listeners exhibited a linguistic boundary in distinguishing the level-contour tones, English listeners appeared to make judgments on the basis of the psychoacoustic properties of the stimuli. Later studies replicated Wang's (1976) study by using both speech and non-speech tones, and found that Chinese listeners showed a quasi-categorical perception of tonal contour differences (Hallé et al., 2004; Peng et al., 2010; Xu et al. 2006). Since Chinese listeners perceived level versus contour tones quasi-categorically, they used the cross-category differences in pitch contour to categorize tones, and did not show much sensitivity to within-category pitch variations; in contrast, since English listeners did not hear tones categorically, they detected the subtle differences in pitch height, and thus showed some sensitivity to within-category variations in lexical tones (Leather, 1987; Stagray & Downs, 1993).

In a mismatch negativity (MMN) study, Chandrasekaran, Krishnan, and Gandour (2009) corroborated these behavioral findings by showing that English listeners exhibited smaller and more delayed pre-attentive MMN response for a within-category tonal contrast between the canonical Tone1 (level contour) and contextual variant of Tone1 (rising contour) in the right hemisphere. However, Chandrasekaran et al.’s behavioral accuracy measures showed that English listeners were more accurate than Chinese listeners in discriminating the within-category pitch differences of the level tone: While Chinese listeners had more enhanced pitch sensitivity than English listeners at an early processing stage, they showed less sensitivity to within-category differences in the level tone than English listeners due to their tonal categorization ability at a later stage.

At this point, we may predict that whereas Chinese listeners will not show sensitivity to the within-category gradience of tones, English-speaking L2 learners of Chinese will. However, it is also possible that English-speaking L2 learners of Chinese differ from native Chinese listeners in their use of the fine-grained information of contour versus level tones, because previous studies on tone perception showed that Chinese listeners and English listeners used different aspects of tonal information to categorize lexical tones.

Chinese listeners typically attend to pitch contour information more than pitch height information when they discriminate and identify tones (e.g., Bent, 2005; Chandrasekaran et al., 2010; Guion & Pederson, 2007; Francis et al., 2008; Qin & Jongman, 2016; Qin & Mok, 2013). The variability of pitch contour may be more important than that of pitch height for Chinese listeners, as Chinese uses pitch contour differences as a primary cue to distinguish different contour tones from one another and from the level tone. Moreover, some contour tones (e.g., T2, a rising tone), which share much of their tonal space with other contour tones (e.g., T3, a dipping tone), may have less room for variability than the level tone, which is high up in the tonal space. For instance, studies on the categorical perception of the T1-T2 continuum showed that the boundary of the level-to-rising tonal continuum was closer to the rising tone than to the level tone end (Peng et al., 2010; Xu et al., 2006). In other words, Chinese listeners heard more of the tones in the continuum as level tones than as contour tones, because they needed to hear a steep rising slope to perceive the stimulus as a contour tone. Given the important role of pitch contour differences and the fact that tonal contours have less room for variability in the Chinese tonal inventory, it is likely that Chinese listeners will show more sensitivity to the fine-grained variability of contour tones than to that of the level tone.

In contrast, English does not use pitch contour to contrast meanings; instead, it uses pitch height, with stressed syllables usually having a higher pitch than unstressed syllables. Accordingly, English listeners typically attend to pitch height differences more than pitch contour differences in their tonal perception, as pitch height rather than pitch contour is an important cue in English stress (e.g., Bent, 2005; Chandrasekaran et al., 2010; Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Qin & Jongman, 2016; Qin & Mok, 2013). The variability along the pitch height dimension, encoded in English stress contrasts, may thus be more salient to English listeners than the variability along the pitch contour dimension. Hence, it is likely that English-speaking L2 learners of Chinese will show more sensitivity to the within-category gradience of the level tone than to that of contour tones.

Given the previous findings on Chinese and English listeners' tone perception, a key open question is whether these listeners will use the fine-grained within-category pitch information of contour and level tones differently to recognize spoken words. An increasingly important body of research has investigated the use of fine-grained phonetic information during online word recognition. This research is discussed next.

3.2.2 The Use of Fine-Grained Phonetic Information in Lexical Access

Although segments, especially consonants, have been shown to be perceived categorically by native listeners (e.g., Liberman, Harris, Hoffman, & Griffith, 1957; Stevens, 2002), the fine-grained within-category gradience is not lost in online word recognition. Recent studies on word recognition has made extensive use of the eye-tracking method to examine how native listeners used the fine-grained phonetic information located in consonants and vowels to recognize spoken words (e.g., Dahan et al., 2001; McMurray et al. 2002, 2008, 2009; Salverda et al., 2003).

McMurray and his colleagues conducted several studies to examine how within-category gradience in voice onset time (i.e., VOT) is preserved and used in word recognition (McMurray et al., 2002, 2008; Toscano & McMurray, 2012). Classic studies on consonant identification and discrimination (e.g., Liberman et al., 1957) had found that listeners showed categorical perception of voicing. The classic view of categorical perception assumes that speech perception and processing mechanisms discard variability to unmask the underlying phonemes (Stevens, 2002). However, later studies that used goodness ratings and priming tasks showed that the perception and processing mechanisms do preserve sensitivity to fine-grained acoustic differences and use them to recognize phonemes and words probabilistically (e.g., Allen & Miller, 2001; Andruski, Blumstein, & Burton, 1994; Miller, 1997). To address the crucial question of whether the effects of fine-grained phonetic differences found in previous studies (e.g., Andruski et al., 1994; Miller, 1997) are truly gradient or just limited to differences between exemplars that are near vs. far from the category boundary, McMurray et al. (2002) used the visual-world eye-tracking paradigm to test whether listeners show sensitivity to within-category VOT variations during online word recognition. English listeners heard twelve minimal pairs from a nine-step /b/-/p/ VOT continuum (e.g., *bear* vs. *pear*). After hearing a word, the participants had to click on the correct picture among four pictures in a display. Listeners' eye fixations showed gradient effects of VOT, such that proportions of fixation to competitors increased linearly as the VOT approached the category boundary. Importantly, a gradient effect of VOT was found even if the analyses did not include the VOT steps that were at the end points of the continuum or near the category boundary. These results suggest that the fine-grained within-category information available in the signal affects the degree of lexical activation, and the effect of within-category information on lexical activation are truly gradient, that is, they are not limited to differences between exemplars that are near vs. far from the category boundary. This study, consistent with others (e.g., Dahan et al., 2001; Salverda et al., 2003), provided evidence that fine-grained phonetic information constrains lexical activation.

Not only the fine-grained phonetic cues located in consonants, but also those located in vowels, were found to modulate lexical competition in word recognition (e.g., Beddor et al., 2013; Dahan et al., 2001; Salverda et al., 2003). To illustrate, Dahan et al. (2001) used the visual-world eye-tracking paradigm to test whether listeners used subcategorical matching or mismatching coarticulatory cues in a vowel during spoken-word recognition. The beginning CV portion (i.e., [nɛ]) of the stimuli was cross-spliced from a target word (e.g., *net*), a competitor word (e.g., *neck*), and a non-word (e.g., *nep*). The results showed that listeners fixated the target picture (e.g., *net*) less when the beginning CV sequence of the target word came from a competitor word (e.g., *ne(ck)t*) than when it came from a non-word (e.g., *ne(p)t*), regardless of whether or not the picture of the competitor word was present in the display. This suggests that the competitor word inhibits the recognition of the target word more when the acoustic realization of the CV sequence in the target word is consistent with that of the lexical competitor than when it is not consistent with any word.

Based on these findings, the authors argued that listeners are sensitive to matching/mismatching coarticulatory cues in a vowel and can use them in word recognition, resulting in the activation of competing words. Salverda et al. (2003) also used the visual-world eye-tracking paradigm to investigate listeners' use of duration cues to word boundaries. The researchers varied the duration of an ambiguous sequence (e.g., [hæm]) by replacing the first syllable of the target word (e.g., *hamster*) with a recording of the monosyllabic word (e.g., *ham*) or with a different recording of the target word (e.g., *ham-* from *hamster*). The acoustic analysis showed that the sequence (e.g., [hæm]) was longer in the monosyllabic word (e.g., *ham*) than in the disyllabic word (e.g., *ham-* from *hamster*). The displays contained pictures of the target (e.g., *hamster*), competitor (e.g., *ham*), and two unrelated distractors. The results showed that listeners fixated on the picture of the competitor more when the first syllable of the target word came from a recording of the competitor word than when it came from a different recording of the target word. Thus, the authors concluded that listeners systematically exploited fine-grained durational cues such as segmental lengthening in the online recognition of spoken words. The findings of the studies above indicate that the processing mechanism does not lose sensitivity to fine-grained phonetic differences and can use them to recognize words.

Research on word recognition provided evidence that native listeners can use fine-grained phonetic information to recognize spoken words as soon as it is available in the signal, though this was only shown for segments (e.g. /b/ vs. /p/ and /b/ vs. /w/; McMurray et al., 2008). Like segmental information, suprasegmental information also plays an important role in word recognition (for discussion, see Cutler, 2012, Chapter 7). Although several recent studies have begun to test the fine-grained phonetic effects of lexical stress (e.g., Reinisch et al., 2010) and lexical tone (e.g., Malins & Joannise, 2010; Shen et al., 2013) in word recognition, it is still unclear how listeners use fine-grained suprasegmental information to recognize words on a millisecond-by-millisecond basis, and how this information should be incorporated into current models of spoken word recognition. If this fine-grained information constrains lexical access, listeners should show more competition from tonal competitors when the pitch of the target word is acoustically closer to that of the tonal competitor word than when it is acoustically more distant from that of the tonal competitor.

Importantly, no previous study has looked at the moment-by-moment use of within-category gradience in lexical tones by non-native listeners. Thus, it remains unclear whether sensitivity to the within-category gradience of level versus contour tones is used to similar degrees by native and non-native Chinese listeners. Given the recent findings on the use of fine-grained phonetic variability, we might predict that Chinese listeners will not discard their sensitivity to the within-category gradience of tones, especially for contour tones, whereas English-speaking L2 learners may show their sensitivity to the within-category gradience of the level tone. To test this prediction, the present study investigates whether native Chinese listeners and English-speaking L2 learners of Chinese differ in using the within-category gradience of level and contour tones to recognize spoken Chinese words.

3.3 The Present Study

The goal of the current study is to use the visual-world eye-tracking paradigm to examine if fine-grained tonal information, that is, within-category gradience in lexical tones, is preserved in patterns of lexical activation and utilized to facilitate online word recognition, and whether native Chinese listeners and English-speaking L2 learners of Chinese differ in their sensitivity to this information. A second experiment was conducted to address these questions. Words from a single pair of level-contour tonal contrast that differ slightly in their pitch height at onset were used in this experiment. The design of the study is illustrated in the hypothetical example in Figure 8: Three within-category tonal items are used as stimuli for either T1 (T1-Distant, T1-Standard, and T1-Close), a level tone, or T2 (T2-Distant, T2-Standard, and T2-Close), a contour tone.

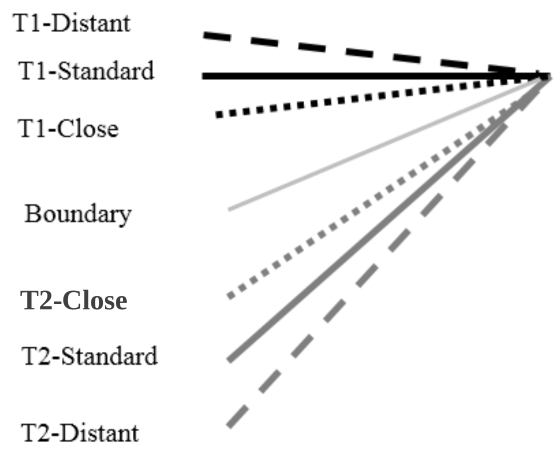

Figure 8: A hypothetical illustration of within-category tonal items (Distant, Standard and Close) of Tone1 (Level) and Tone2 (Contour)

When the target word in the auditory stimuli contains T1, the competitor word in the display contains T2. While a canonical tonal item (i.e.,T1-Standard) extracted from natural tokens was used for T1, two deviant tonal items (i.e., T1-Distant and T1-Close) were generated by manipulating the canonical tone and used for T1 as well. A similar design is used when the target word in the auditory stimuli contains T2, and the competitor in the display contains T1.

First, we investigate whether native Chinese listeners show sensitivity to fine-grained within-category gradience in lexical tones during online word recognition. If Chinese listeners preserve their sensitivity to fine-grained tonal information and use this information to recognize words, like the English listeners did in using the fine-grained phonetic information of VOT in McMurray et al. (2002), they should show a gradient pattern of lexical competition: when T1 is heard as target in the auditory stimuli, the words containing T1-Distant should show less competition from T2 words than those containing T1-Standard, as T1-Distant is acoustically more distant from the competing T2 than T1-Standard is; and the words containing T1-Close should show more competition from T2 words than those containing T1-Standard, as T1-Close is acoustically closer to the competing T2 than T1-Standard is; similarly, when T2 is heard as target in the auditory stimuli, the words containing T2-Distant should show less competition from T1 words than those containing T2-Standard; and the words containing T2-Close should show more competition from T1 word than those containing T2-Standard.

Second, the current study examines whether (and if so, how) native Chinese listeners and English-speaking L2 learners of Chinese differ in their use of the fine-grained information of level and contour tones. Given the findings that Chinese and English listeners have different sensitivities to pitch height versus contour in previous studies of tone perception (e.g., Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007), we predict that native Chinese listeners will be more sensitive to within-category pitch information of contour tones (i.e., the three T2 items) than level tones; in contrast, English-speaking L2 learners of Chinese will be more sensitive to within-category pitch information of level tones (i.e., the three T1 items) than contour tones during online word recognition. This will indicate that the two groups of listeners show sensitivity to fine-grained tonal variability, but along different dimensions (pitch height vs. pitch contour). Moreover, across the board, L2 learners will show more competition from tonal competitors than will native listeners based on previous research (e.g., Sun, 2012) as well as the findings of our first experiment.

3. 4 Method

3.4.1 Participants

Thirty-six native Chinese speakers and twenty-six highly proficient Chinese learners who spoke English as native language and who learned Chinese as L2 in college (i.e., adult L2 learners of Chinese) were tested for this experiment. These participants are the same as those tested in the first experiment. The testing took place at the Center for Brain and Cognitive Sciences at Peking University, China.

3.4.2 Materials

The twelve T1-T2 word pairs (e.g., T1, a level tone: /jā/ ‘duck’; T2, a contour tone: /já/ ‘tooth’) from Experiment 1 were reused in Experiment 2.[9] Using the same word pairs was necessary given the limited number of imageable monosyllabic noun pairs in Chinese. The auditory words with the two tones had their pitch height resynthesized such that their pitch contour would be either more similar to or more different from that of the competitor word. As illustrated in Figure 9, three levels of a within-category tonal continuum were created for T1 (T1-Standard, T1-Distant, and T1-Close) and for T2 (T2-Standard, T2-Distant, and T2-Close). T1-Standard

[9] The current experiment focuses on listeners’ use of the whole pitch contour in word recognition. As such, it examines listeners’ fixations in relation to the complete contour heard in the stimuli. By the time the pitch contour has been heard, the effect of tonal probability should be gone or have at least tapered off (recall that only the first 80 ms of the vowel in Wiener & Ito, 2016 showed an effect of tonal probability). Thus, the tonal probability information is less relevant to the current experiment.

and T2-Standard are standard (S) exemplars of T1 and T2 natural tokens. T1-Distant is acoustically more distant (D) from T2 than T1-S is, but T1-Close is acoustically closer (C) to T2 than the T1-S is. Likewise, T2-D is acoustically more distant from T1 than T2-S is, but T2-C is acoustically closer to T1 than the T2-S is.

The canonical (Standard condition) tokens were created by using the average pitch values of the natural tokens from Experiment 1. The acoustically closer and more distant tonal tokens were created as “non-canonical” tones by using the average pitch values of the natural tokens from Experiment 1, but raising or lowering the starting point of the contour by 10 Hz. The acoustically closer and more distant contours were obtained via incremental interpolation between the new starting points and the natural endpoint in 10 measurements. A manipulation value of 10 Hz was chosen, because it was small enough that the acoustically closer tone would not be close to the boundary (e.g., McMurray et al., 2002), but the difference between the acoustically closer (and more distant) tones and the canonical tones was larger than the Just Noticeable Difference (JND) of tone contour discrimination for both Chinese and English listeners (Liu, 2013). The resynthesized pitch contours illustrated in Figure 9 were superimposed on the duration-normalized stimuli from Experiment 1.

To ensure that all three levels of the two tonal continua would be within-category variations of the same tonal category and be far enough from the tonal boundary of T1-T2, six native Chinese listeners judged the naturalness of the resynthesized stimuli on a 6-point scale (1-6, 1 means “The word sounds really bad” and 6 means “The word sounds really good”) and categorized them into different tones. The stimuli were rated as natural (mean: 5.1; SD: 1.1) and had their target tonal category correctly identified (mean: 99.5%; SD: 6.8%). The three conditions, Standard (T1, mean accuracy: 99%, SD: 1.2%; mean rating: 5.3, SD: 1.2; T2, mean accuracy: 100%; mean rating: 5.1, SD: 1.1), Distant (T1, mean accuracy: 100%, SD: 1.2%; mean rating: 5.3, SD: 1.1; T2, mean accuracy: 100%; mean rating: 5.0, SD: 1.3), and Close (T1, mean accuracy: 100%; mean rating: 5.3, SD: 1.1; T2, mean accuracy: 100%; mean rating: 4.9, SD: 1.2), did not differ significantly in their naturalness rating ($t < |1|$) or in their identification accuracy ($t < |1|$).

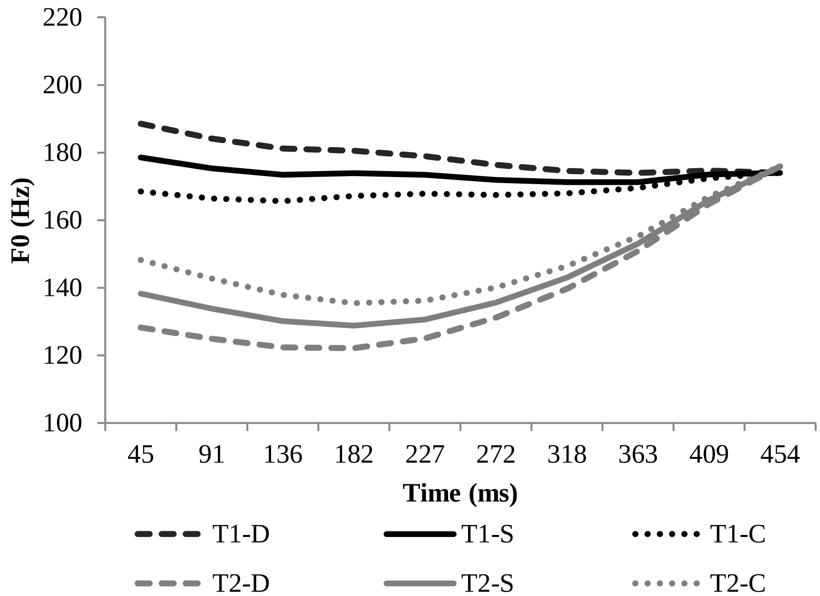

Figure 9: Tonal continua of Tone1 (Level) and Tone2 (Contour) used in Experiment 2

Like in Experiment 1, the target words in filler trials contained T3 or T4 to prevent participants from focusing on the T1-T2 pairs. The pitch contours of the T3 and T4 stimuli were resynthesized by raising or lowering the starting point of the canonical contour by 10 Hz, similarly to those of T1 and T2 stimuli (with standard and acoustically closer and more distant exemplars), so that the experimental stimuli would not stand out. Across all trials, the four words in each display carried four different tones, and each tone was heard the same number of times in the entire experiment. Unlike Experiment 1, the conditions in Experiment 2 were not counterbalanced across lists, as it was not possible to do so given the limited number of imageable monosyllabic noun pairs in Chinese. Each display was thus repeated three times for each word (e.g., one for T1-D, one for T1-S, and one for T1-C for T1 word; and the same for T2, T3, and T4 words). The complete experiment included a total of 72 (12 word pairs × 2 tones × 3 tokens) critical trials and 72 (12 word pairs × 2 tones × 3 tokens) filler trials, and a total of 48 words/images (24 word pairs).

The locations of target, competitor, and distracter pictures in the four cells of the (non-displayed) 2 x 2 grid were counterbalanced throughout the experiment. The relative locations of targets and competitors and of the different tones were also counterbalanced, and critical and filler trials were pseudorandomized.

3.4.3 Procedures

Experiment 2 was conducted immediately after Experiment 1. We kept the experiment order the same for all participants to ensure that all the trials in Experiment 2 had been repeated before. The procedures of Experiment 2 were identical to those of Experiment 1.

As with Experiment 1, to reduce participants' memorization burden, Experiment 2 was conducted over three sessions, with 1/3 of the words (and corresponding trials in the standard, close, and distant conditions) administered over each session, and with at least two days between sessions, as illustrated in Table 4, repeated below as Table 8 for convenience. Since the experiment includes a total of 48 words (and 144 trials), 16 words (and 48 trials) were used and tested in each session. Only words that had been trained on the same day would appear in the word recognition task. In other words, the training session led the participants to learn word-picture correspondences that were used in both Experiments 1-2 within a given session, with a third of the words and trials being used in each session. The order of the different sessions was counterbalanced across participants.

Table 8: Experimental procedures across the three sessions (repeated Table 4)

	Total	**Session 1 (Day A)**	**Session 2 (Day B)**	**Session 3 (Day C)**
Procedures		Training Experiment 1 (1 block) Experiment 2 (4 blocks)	Training Experiment 1 (1 block) Experiment 2 (4 blocks)	Training Experiment 1 (1 block) Experiment 2 (4 blocks)
Target Words	24 target words for each tone pair	8 target words for each tone pair	8 target words for each tone pair	8 target words for each tone pair
Training Words	192 words	64 words	64 words	64 words

Because Experiment 2 involved the repetition of words and of displays, trials were split into four different blocks, with each block containing only one repetition of each target word. The number of times each word and each tone is heard was counterbalanced within and across blocks. Ultimately, each experimental session, which included parts of Experiment 1 and 2, took approximately 60 minutes.

3.4.4 Data Analysis

Like in Experiment 1, only trials in which participants clicked on the target word were analyzed, resulting in the exclusion of 1% of the data for Chinese listeners and 21.5% of the data for English listeners. Proportions of fixations to the target, competitor, and distracter words were extracted in 8-ms time windows from the onset to the offset of the target word, with an adjustment of 200-ms delay. The dependent variable for the statistical analyses was the difference between proportions of target and competitor fixations (i.e., the proportion of competitor fixations was subtracted from the proportion of target fixations) from the target word onset (i.e., 0 ms) to the word offset (i.e., 454 ms), with a 200-ms delay.

Listeners' fixation differences were modeled using growth curve analysis (GCA; Mirman, 2014), a type of curvilinear regression that can model the linear (i.e., capturing the overall angle of a curve), quadratic (i.e., capturing a curve with a single inflection), and cubic (i.e., capturing a curve with two inflections) shapes of the differential fixation lines. GCAs are similar to mixed-effects models (Bates et al., 2015), but include time polynomials, thus enabling us to model participants' eye fixations over time rather than using specific time windows. Since this experiment focuses on how the participants use very fine-grained phonetic information of tonal contours over time as the speech signal unfolds, GCA is more appropriate than a time-window analysis for analyzing participants' fixations, because they can model subtle changes in the curvilinear patterns of eye fixations over time and capture the differences in the slope and curvature of the differential fixation lines, which traditional time-window analyses cannot do. Moreover, decisions regarding the onset and offset of the critical time windows are not required when using GCAs.

GCAs include orthogonal time polynomials, the fixed variables of interest, as well as random variables. The different time polynomials model the shape of the proportions of fixations over time. Our analysis included linear, quadratic, and cubic time polynomials. A significant *t* value for the linear time polynomial indicates an ascending slope (i.e., /, positive *t* value) or a descending slope (i.e, \, negative *t* value) of the differential fixation line in the baseline condition; a significant *t* value for the quadratic time polynomial indicates a convex shape (i.e., ∪, positive *t* value) or a concave shape (i.e., ∩, negative *t* value) of the differential fixation line in the baseline condition; a significant *t* value for the cubic time polynomial indicates a reverse 's' shape (i.e., ~, positive *t* value) or an 's' shape (i.e., ∽, negative *t* value) of the differential fixation line in the baseline condition. The time polynomials were centered and made orthogonal prior to entering the analyses because they would otherwise be highly correlated, which would make the results difficult to interpret (Mirman, 2014).

To be able to conclude that our tonal manipulation had an effect on participants' fixations, the GCAs must show *interactions* between Condition (Standard vs. Distant or Close) and the linear and/or quadratic time polynomials, and possibly the cubic time polynomial.[10] These interactions would indicate that as the speech signal unfolds over time, the *shape* of participants' differential fixation line changes differently across the different tonal conditions. Importantly, an effect of Condition without a significant interaction with the linear, quadratic, or cubic time polynomials indicates that fixation proportions are either higher or lower in one tonal condition than in another, but the *shape* of participants' differential fixation lines are similar among the different conditions; thus, such an effect cannot be attributed to the tonal manipulation in the speech signal, and is instead understood as a baseline difference between the two conditions (i.e., a difference that exists independently of the auditory stimuli; Barr, Gann, & Pierce, 2011).

The GCAs were conducted with the *lme4* package in R (Bates, Maechler, & Walker, 2015). For the sake of clarity, we first present the analysis of the individual language group's

[10] The cubic time polynomial tends to capture less relevant effects in the tail of the fixation line (e.g., an asymptote effect, and does not have a meaningful cognitive interpretation in many cases (Mirman, 2014). The interaction between Condition and the cubic time polynomial is thus less relevant than the interaction between the linear and/or quadratic time polynomials.

results. These analyses included the three time polynomials (linear, quadratic, and cubic), Condition (Standard, Distant, and Close; Baseline: Standard) and Tone of the target word (Level, Tone 1 vs. Contour, Tone 2; Baseline: Level) as fixed effects. A back-fitting function from the package *LMERConvenienceFunctions* in R (Tremblay & Ransijin, 2015) was used to identify the model that accounted for significantly more of the variance than simpler models, as determined by log-likelihood ratio tests; only the results of the model with the best fit are presented, with *p* values being calculated using the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2016). Analyses yielding significant interactions between Condition and Tone were followed up by subsequent GCAs conducted separately on the level and contour tones, with the alpha level being adjusted to .025. All analyses included participant as random intercept and the orthogonal time polynomials as random slopes for the participant variable, which allowed the analysis to model a line of a different shape for each individual participant. [11]

To determine whether the English-speaking L2 learners differed from native Chinese listeners in using fine-grained information of level and contour tones over time, we also conducted analyses that tested three-way interactions between the effects of Condition, Language (Chinese vs. English, Baseline: Chinese), and time separately for the level and contour tones items to simplify the presentation of the data as well as the interpretation of effects (i.e., four-way interaction). These analyses followed the same procedures for model selection and included the same random effects as those described for the individual groups.

---

[11] English listeners' familiarity ratings between the T1 and T2 words were not significantly different ($t < |1|$), so the familiarity ratings were not included in the models. Again, the dissertation targeted highly proficient L2 learners of Chinese and did not aim to investigate L2 development, so proficiency was not included in the models neither.

If listeners use within-category gradience in lexical tones, the GCAs should yield interactions between Condition and the linear time polynomial (with the differential fixation line showing a steeper ascending slope in the distant condition than in the standard condition, and a shallower ascending slope in the close condition than in the standard condition), and/or the quadratic time polynomial (with the differential fixation line being less U-shaped in the distant condition and more U-shaped in the close condition than in the standard condition), and possibly the cubic time polynomial (e.g., with the differential fixation line having a sharper 's' shape in the distant condition than in the standard condition as listeners may recognize the target words fast and show a tail effect of the fixation line in the distant condition, see also Footnote 8). While a differential fixation line with a steeper ascending slope and/or less U-shape in the distant condition would be indicative of faster word recognition due to a decreased tonal competition in that condition as a result of an acoustically greater tonal distance between targets and competitors, a differential fixation line with a potentially sharper 's' shape in the distant condition might be due to this line reaching an asymptote towards the end of the trial as a result of recognizing the target words faster in the distant condition than in the standard condition. A differential fixation line with a shallower ascending slope and/or more U-shape in the close condition would be indicative of slower word recognition due to an increased tonal competition in that condition as a result of an acoustically smaller tonal distance between targets and competitors.

If native Chinese listeners show more sensitivity to the within-category gradience of contour tones than that of level tones, or if English-speaking L2 learners of Chinese show more sensitivity to the within-category gradience of level tones than that of contour tones, the GCAs for either Chinese or English listeners should yield three-way interactions between Condition, Tone, and the time polynomials. The three-way interactions indicate that the conditions had a different effect on the shape of either Chinese or English listeners' differential fixation line for the contour tones as compared to the level tones. Chinese listeners should show an interaction between Condition and the linear/quadratic time polynomials for the contour tones (weaker or no effect for level tones), whereas English listeners should show an interaction between Condition and the linear/quadratic time polynomials for the level tones (weaker or no effect for contour tones).

In addition, if English listeners differ from Chinese listeners in their sensitivity to the within-category gradience of lexical tones, the GCAs for the level and contour tones should yield a significant three-way interaction between Language, Condition, and the linear and/or quadratic time polynomials, indicating that the shapes of the two language groups' differential fixation lines differed across the conditions as the speech signal unfolds. Given previous findings about the difference between native and non-native listeners in their difficulty processing tones in word recognition (e.g., Sun, 2012; Wiener, 2015), the GCAs should show a simple effect of Language, with English listeners showing a smaller difference between the proportions of target and competitor fixations than Chinese listeners, if English listeners experience more tonal competition.

3.5 Results

While Chinese listeners performed at ceiling in identifying the target word (mean accuracy: 99%; SD: 9.8%), English listeners were less accurate (mean accuracy: 78.5%; SD: 44.7%). Chinese listeners clicked on the target picture at an average of 1,521 ms (SD: 454 ms), whereas English listeners did so at an average of 2,380 ms (SD: 959 ms). These mouse click results are consistent with previous offline studies (e.g., Sun, 2012) in that Chinese listeners had a higher accuracy and shorter RTs than English listeners.

Recall that approximant-initial syllables carry pitch from the syllable onset, so the tonal onset is the word onset for our target words. GCAs were performed on the differential proportions of fixations corresponding to the tonal portion (454 ms) of the target word, separately for Chinese and English listeners. The best GCA on the difference between listeners' proportions of target and competitor fixations included the simple effects of Condition (Standard, Distant, Close; baseline: Standard), Tone (Level vs. Contour; baseline: Level), and the time polynomials (linear, quadratic and cubic). Because the time polynomials were made orthogonal, any effect of a fixed variable (e.g., Condition) is to be interpreted on the average differential fixations over time (Mirman, 2014).

3.5.1 Native Chinese Listeners

Figure 10 shows Chinese listeners' differential proportions of fixations in the standard, close, and distant conditions for the level and contour tones up to 1000 ms post target-word onset (Chinese listeners' proportions of target, competitors, and distracter fixations are shown in Figure 17 in Appendix F). Figure 11 shows Chinese listeners' differential proportions of fixations corresponding to the tonal portion of the stimuli (corresponding to the target word), over which the GCAs were conducted. The onset and offset of the tonal portion of the stimuli were adjusted with 200 ms in Figure 11. Differential proportions of fixations above 0 mean that participants looked more at the target than at the competitor.

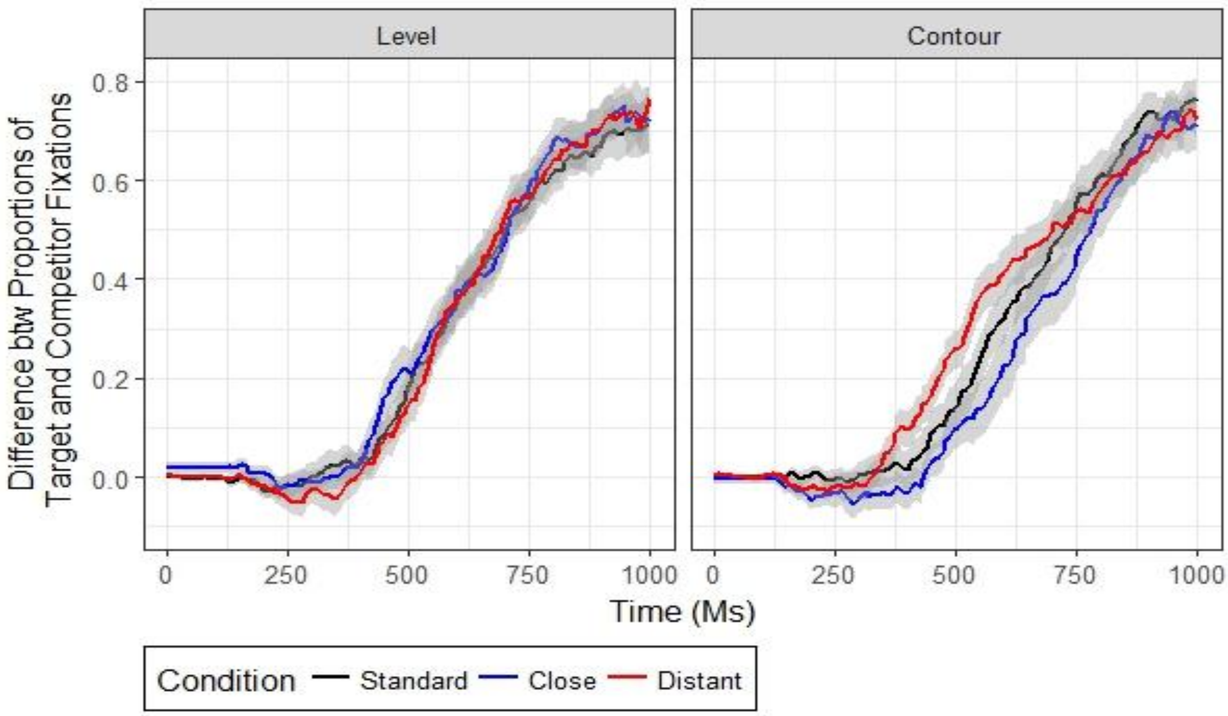

Figure 10: Chinese listeners' differential proportions of fixations in the standard (black), close (blue) and distant (red) conditions for the level and contour tones in the first 1000 ms; the shaded area represents one standard error above and below the mean

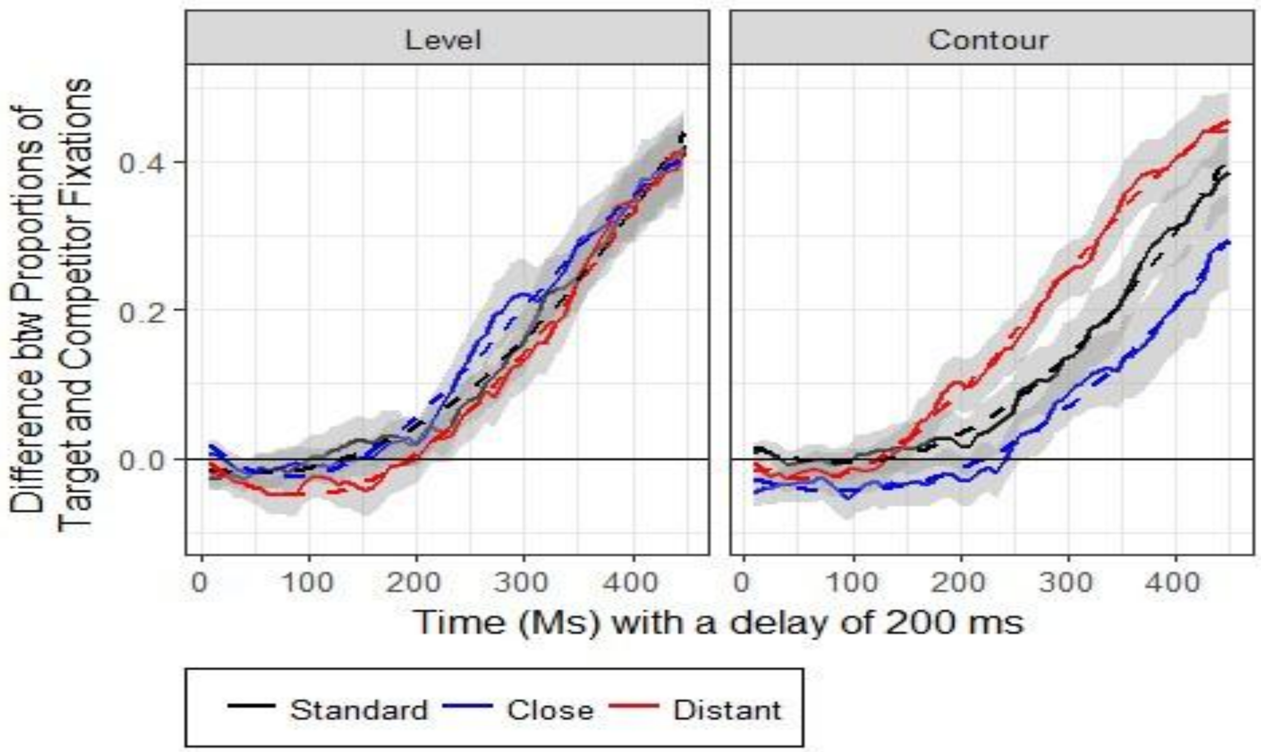

Figure 11: Chinese listeners' differential proportions of fixations in the standard, close and distant conditions for the level and contour tones; the solid lines represent listeners' data in the tonal portion; the dashed lines represent the predicted data based on the growth-curve analysis of Chinese listeners' data (Table 9); the shaded area represents one standard error above and below the mean

Table 9 presents the GCA with the best fit on Chinese listeners' data, which included all the simple effects and their interactions.

Table 9: Growth curve analysis on the difference between listeners' target and competitor fixations in different conditions by native Chinese listeners

Variable	**Estimate**	***t***	***p***
(intercept)	0.115	7.130	**<.001**
Time			
Linear	0.931	8.400	**<.001**
Quadratic	0.342	8.062	**<.001**
Cubic	–0.003	<|1|	.92
Condition (Distant)	0.016	3.991	**<.001**
Condition (Close)	–0.023	–5.484	**<.001**
Tone	–0.010	–1.702	.09
Time × Condition (Distant)			
Linear	–0.109	6.385	**<.001**
Quadratic	–0.035	–1.107	.26
Cubic	–0.096	–3.070	**<.01**
Time × Condition (Close)			
Linear	–0.054	–1.727	.08
Quadratic	–0.055	–1.756	.08
Cubic	–0.064	–2.057	**.04**
Time × Tone			
Linear	–0.130	–2.952	**<.01**
Quadratic	0.029	<|1|	.50
Cubic	–0.007	<|1|	.87
Tone × Condition (Distant)	0.076	9.099	**<.001**
Tone × Condition (Close)	–0.076	–9.118	**<.001**
Time × Tone × Condition (Distant)			
Linear	0.295	4.726	**<.001**
Quadratic	–0.224	–3.585	**<.001**
Cubic	–0.075	–1.209	.22
Time × Tone × Condition (Close)			

Linear	–0.214	–3.426	**<.01**
Quadratic	–0.008	<|1|	.89
Cubic	0.150	2.408	**.02**

Note. α = .05; significant results are in bold. *n* = 12096 observations.

The significant positive *t* value for the intercept means that Chinese listeners' differential proportion of fixations in the standard condition was higher than 0. The significant positive *t* values for the linear and quadratic time polynomials indicate that Chinese listeners' differential fixation line in the standard condition had an ascending slope and a U-shape. While the significant positive *t* value for Condition (Distant) means that Chinese listeners had a higher differential proportion of fixations in the distant condition than the standard condition, the significant negative *t* value for Condition (Close) means that Chinese listeners had a lower differential proportion of fixations in the close condition than the standard condition. The significant positive *t* value for the interaction between Condition (Distant) and the linear time polynomial indicates that Chinese listeners' differential fixation line in the distant condition had a more ascending slope than their differential fixation line in the standard condition, whereas the significant negative *t* value for the interaction between Condition (Distant) and the cubic time polynomial indicates that Chinese listeners' differential fixation line in the distant condition had a sharper 's' shape than their differential fixation line in the standard condition. The significant negative *t* value for the interaction between Condition (Close) and the cubic time polynomial indicates that Chinese listeners' differential fixation line in the close condition had a sharper 's' shape than their differential fixation line in the standard condition. The significant negative *t* value for the interaction between Tone and the linear time polynomial indicates that Chinese listeners' differential fixation line on the contour tone had a shallower ascending slope than their differential fixation line on the level tone. The significant positive *t* value for the interaction between Tone and the distant condition means that the effect of Condition (positive) increased (i.e., became more positive) for the contour tone as compared to the level tone, whereas the significant negative *t* value for the interaction between Tone and the close condition means that the effect of Condition (negative) increased (i.e., became more negative) for the contour tone as compared to the level tone. Finally, and importantly, the three-way interactions between Condition, Tone, and the time polynomials (linear, quadratic for the distant condition, and linear and cubic for the close condition) indicate that Condition (i.e., fine-grained tonal variability) modulated the shape of listeners' differential fixation line differently for the contour tone compared to the level tone.

In order to understand the directionality of the three-way interactions reported in Table 9, subsequent GCAs were performed on the differential proportions of fixations separately for each tone. Table 10 presents the results of the follow-up GCAs with the best fit.

Table 10: Growth curve analyses on the difference between listeners' target and competitor fixations in different conditions of the level tone and the contour tone by native Chinese listeners

Tone	**Variable**	**Estimate**	***t***	***p***
Level Tone	(Intercept)	0.120	7.251	**<.001**
	Time			
	Linear	1.032	8.736	**<.001**
	Quadratic	0.328	6.491	**<.001**
	Cubic	0.004	0.010	.99
	Condition			
	Distant	–0.021	–3.933	**<.001**
	Close	0.015	2.798	**<.01**
	Time × Cond (Distant)			
	Quadratic	–0.077	1.907	.06
	Cubic	–0.058	–1.432	.15
	Time × Cond (Close)			
	Quadratic	–0.050	1.252	.21
	Cubic	–0.139	–3.438	**<.01**
Contour Tone	(Intercept)	0.110	5.545	**<.001**
	Time			
	Linear	0.867	6.998	**<.001**
	Quadratic	0.358	5.939	**<.001**
	Cubic	–0.007	–0.149	.88
	Condition			
	Distant	0.055	9.852	**<.001**
	Close	–0.061	–10.99	**<.001**
	Time × Cond (Distant)			
	Linear	0.346	8.362	**<.001**
	Quadratic	–0.146	–3.531	**<.01**
	Cubic	–0.134	–3.221	**<.01**

Time × Cond (Close)			
Linear	–0.161	–3.878	**<.001**
Quadratic	–0.059	–1.419	.16
Cubic	0.011	0.264	.79

Note. α = .025; significant results are in bold. Level Tone, *n* = 6048 observations; Contour Tone, *n* = 6048 observations.

The GCA with the best fit on native Chinese listeners' data for the level tone included the three time polynomials, Condition, and the interactions between Condition and the quadratic/cubic time polynomials. The significant positive *t* value for the intercept means that Chinese listeners' differential proportion of fixations in the standard condition was higher than 0. The significant positive *t* values for the linear and quadratic time polynomials indicate that Chinese listeners' differential fixation line in the standard condition had an ascending slope and a U-shape. While the significant negative *t* value for Condition (Distant) means that Chinese listeners had a lower differential proportion of fixations in the distant condition than the standard condition, the significant positive *t* value for Condition (Close) means that Chinese listeners had a higher differential proportion of fixations in the close condition than the standard condition potentially due to a baseline effect. Crucially, the *t* value for the interaction between Condition (Distant) and the time polynomials was not significant; the negative *t* value for the interaction between Condition (Close) and the cubic time polynomial was significant (indicating that Chinese listeners' differential fixation line had a sharper 's' shape in the close condition than in the standard condition), but this effect appears to be due to a first curvature early on in the differential fixation line (early lexical competition) and a second curvature mid-way through the recognition process.

These results showed that the differential fixation lines of the distant versus standard condition as well as those of the close versus standard condition did not differ in their slope or U-shape. The interaction between Condition (Close) and the cubic time polynomial was significant, but the curvature differences in the fixation lines were not indicative of more lexical competition in the close condition than in the standard condition. These results suggest that Chinese listeners' word recognition was not modulated by the acoustic distance between the target and competitor when the target contained a level tone. In other words, Chinese listeners did not use the within-category gradience of the level tone to recognize Chinese words. The effects of Condition (i.e., the overall difference between the close and standard conditions) found for the level tone appear to be baseline effects (i.e., an effect present before listeners' intake of the speech signal), so these effects cannot be attributed to the tonal manipulation in the speech signal.

The GCA with the best fit on native Chinese listeners' data for the contour tone included all simple effects and all interactions, as shown in Table 10. The significant positive *t* value for the intercept means that Chinese listeners' differential proportion of fixations in the standard condition was higher than 0. The significant positive *t* values for the linear and quadratic time polynomials indicate that Chinese listeners' differential fixation line in the standard condition had an ascending slope and a U-shape. While the significant positive *t* value for Condition (Distant) means that Chinese listeners had a higher differential proportion of fixations in the distant condition than the standard condition, the significant negative *t* value for Condition (Close) means that Chinese listeners had a lower differential proportion of fixations in the close condition than the standard condition. Crucially, the significant positive *t* value for the interaction between Condition (Distant) and the linear time polynomial and the significant negative *t* value for the interaction between Condition (Distant) and the quadratic time polynomial suggest that Chinese listeners had a differential fixation line with a steeper ascending slope and a shallower U-shape in the distant condition than their differential fixation line in the standard condition. The significant negative *t* value for the interaction between Condition (Distant) and the cubic time polynomial means that Chinese listeners' differential fixation line in the distant condition had a sharper 's' shape than their differential fixation line in the standard condition, as the differential fixation line ascends more quickly and thus asymptotes earlier in the distant condition rather than in the standard condition as an artifact of the task (listeners have recognized the word, so they start looking elsewhere). Furthermore, the significant negative *t* value for the interaction between Condition (Close) and the linear time polynomial indicates that Chinese listeners' differential fixation line in the close condition had a shallower ascending slope than their differential fixation line in the standard condition.

These results showed that the differential fixation line of the distant condition had a steeper ascending slope and a less U-shaped than that of the standard condition. This indicates a decrease in tonal competition in that condition compared to the standard condition as a result of an acoustically greater tonal distance between the target and competitor. By contrast, the differential fixation line of the close condition showed a shallower ascending slope than that of the standard condition. This indicates an increase in tonal competition in that condition as a result of an acoustically smaller tonal distance between the target and competitor. The results suggest that Chinese listeners used the within-category gradience of the contour tone to recognize Chinese words.

3.5.2 English-Speaking L2 Learners

Figure 12 shows English listeners' differential proportions of fixations in the standard, close, and distant conditions for the level and contour tones in the first 1000 ms (English listeners' proportions of target, competitors, competitors, and distracter fixations are shown in Figure 18 in Appendix F). Figure 13 shows English listeners' differential proportions of fixations in the standard, close, and distant conditions for the tonal portion, over which the GCAs were conducted, of the words containing level and contour tones.[12]

[12] The GCA was conducted on English listeners' differential fixations for the tonal portion of the speech signal for the sake of simplification of the time polynomials. Since English listeners were slower than Chinese listeners in recognizing words, a GCA was also conducted on their differential fixations in a wider window, that is, the first 800 ms after tonal portion. The results did not differ from those of the current GCA except that the linear and cubic polynomials (rather than the linear and quadratic polynomials) were included in the model with the best fit.

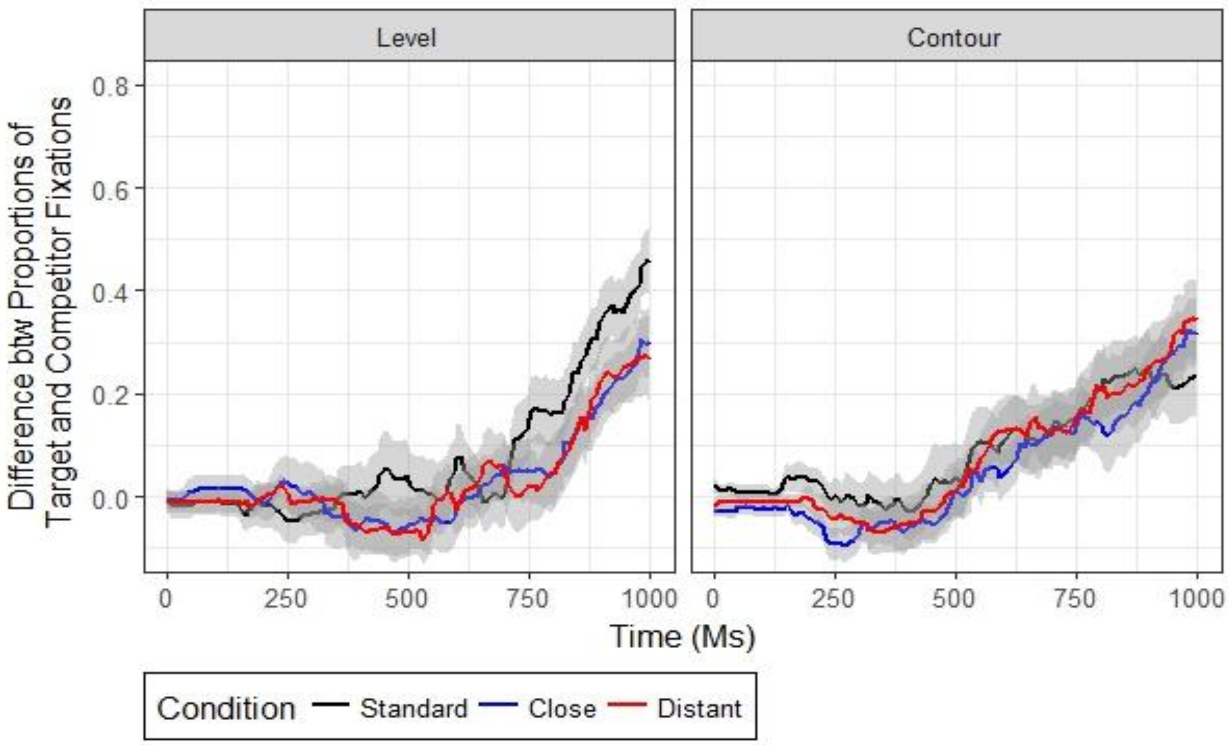

Figure 12: English listeners' differential proportions of fixations in the standard (black), close (blue) and distant (red) conditions for the level and contour tones in the first 1000 ms; the shaded area represents one standard error above and below the mean

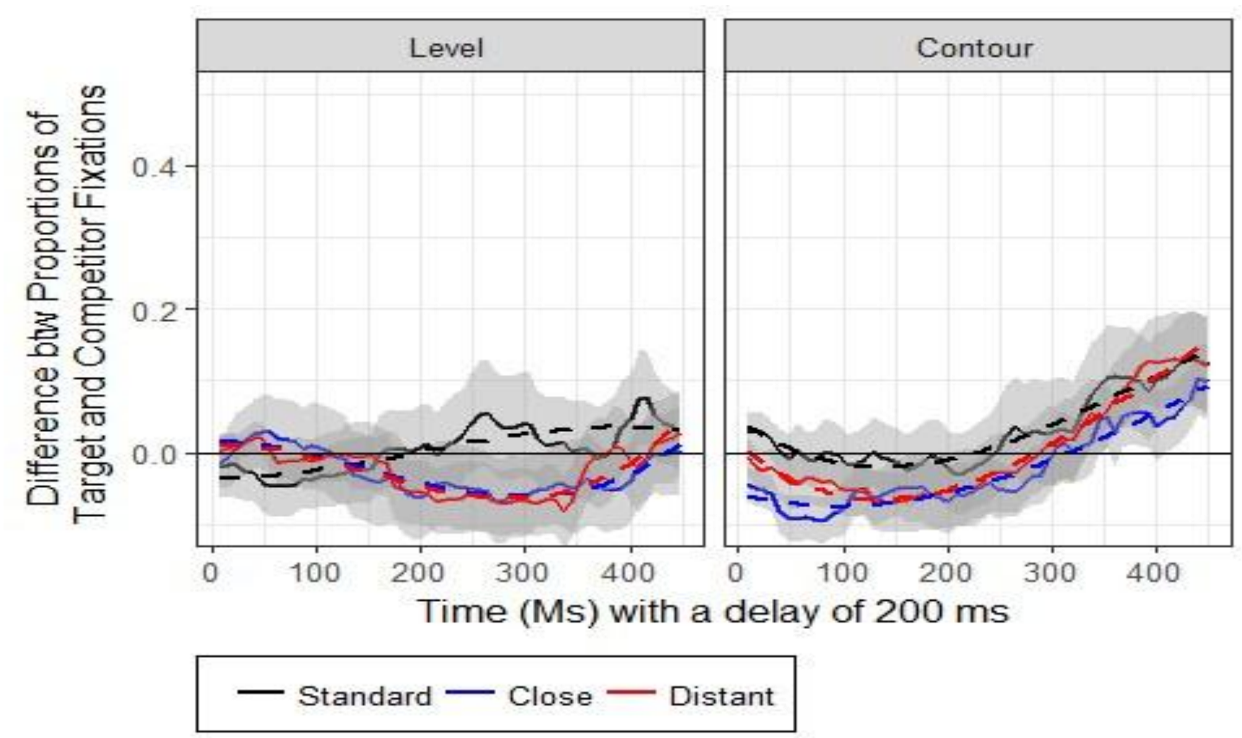

Figure 13: English listeners' differential proportions of fixations in the standard, close and distant conditions for the level and contour tones; the solid lines represent listeners' data in the tonal portion; the dashed lines represent the predicted data based on the growth-curve analysis of English listeners' data (Table 11); the shaded area represents one standard error above and below the mean

Table 11 presents the GCA with the best fit on English-speaking L2 learners' data, which included the linear and quadratic time polynomials, Condition, Tone as well as the interactions between the time polynomials and the other variables in the model. The significant positive *t* value for the linear time polynomial indicates that English listeners' differential fixation line in the standard condition had an ascending slope. The significant negative *t* values for the distant and close conditions indicate that English listeners had a lower differential proportion of fixations in the distant and close conditions than in the standard condition. The significant positive *t* value for Tone means that English listeners had a higher differential proportion of fixations for the contour tone words than for the level tone words. The significant positive *t* value for the interaction between Condition (Distant) and the quadratic time polynomial indicates that English listeners' differential fixation line was more U-shaped in the distant condition than in the standard condition, whereas the significant negative *t* value for the interaction between Condition (Close) and the linear time polynomial indicates that English listeners' differential fixation line had a shallower ascending slope in the distant condition than in the standard condition. Finally, and importantly, the three-way interactions between Condition, Tone, and the time polynomials (linear time polynomial for the distant and close conditions) indicate that Condition had a different effect on the shape of listeners' differential line for the contour tone as compared with the level tone.

Table 11: Growth curve analysis on the difference between listeners' target and competitor fixations in different conditions by English-speaking L2 learners

Variable	**Estimate**	***t***	***p***
(intercept)	0.018	1.224	.23
Time			
Linear	0.243	2.726	**.01**
Quadratic	0.089	1.246	.22
Condition (Distant)	–0.030	–4.472	**<.001**
Condition (Close)	–0.044	–6.941	**<.001**
Tone	0.027	2.985	**<.01**
Time × Condition (Distant)			
Linear	–0.062	–1.313	.19
Quadratic	0.141	3.001	**<.01**
Time × Condition (Close)			
Linear	–0.114	–2.426	**.02**
Quadratic	0.058	1.241	.21
Time × Tone			
Linear	0.106	1.598	.11
Tone × Condition (Distant)	0.002	<|1|	.86
Tone × Condition (Close)	–0.002	<|1|	.07
Time × Tone × Condition (Distant)			
Linear	0.358	3.794	**<.001**
Time × Tone × Condition (Close)			
Linear	0.361	3.829	**<.001**

Note. α = .05; significant results are in bold. *n* = 8736 observations.

In order to understand the directionality of the three-way interactions reported in Table 11, subsequent GCAs were performed on the differential proportions of fixations separately for the level and contour tones. Table 12 presents the results of the follow-up GCAs with the best fit.

Table 12: Growth curve analyses on the difference between listeners' target and competitor fixations in different conditions of the level tone and the contour tone by English-speaking L2 learners

Tone	**Variable**	**Estimate**	***t***	***p***
Level Tone	(Intercept)	0.002	0.076	.93
	Time			
	Linear	0.219	0.103	.10
	Quadratic	–0.033	–0.400	.69
	Condition			
	Distant	–0.309	–3.686	**<.001**
	Close	–0.322	–3.840	**<.001**
	Time × Cond (Distant)			
	Linear	–0.204	–3.830	**<.001**
	Quadratic	0.217	3.461	**<.01**
	Time × Cond (Close)			
	Linear	–0.295	–4.691	**<.001**
	Quadratic	0.175	2.786	**<.01**
Contour Tone	(Intercept)	0.029	1.841	.08
	Time			
	Linear	0.299	3.416	**<.01**
	Quadratic	0.200	2.900	**<.01**
	Condition			
	Distant	–0.0552	–6.666	**<.001**
	Close	–0.0288	–3.476	**<.01**

Note. α = .025; significant results are in bold. Level Tone, n = 4368 observations; Contour Tone, n = 4368 observations.

The GCA with the best fit on English-speaking L2 learners' data for the level tone included the linear and quadratic time polynomials, Condition, as well as the interactions between Condition and the time polynomials. The significant negative *t* values for the distant and close conditions mean that English listeners had a lower differential proportion of fixations in both the distant and close conditions compared with the standard condition. Crucially, the significant negative *t* value for the interaction between Condition (Distant) and the linear polynomial and the significant positive *t* value for the interaction between Condition (Distant) and the quadratic time polynomial suggest that English listeners' differential fixation line in the distant condition had a shallower ascending slope and a sharper U-shape than their differential fixation line in the standard condition. The more U-shaped differential fixation line in the distant condition appears to be due to an early baseline difference as well as an increased tonal competition mid-way through the tone in the conditions. Similarly to the distant condition, the significant negative *t* value for the interaction between Condition (Close) and the linear polynomial and the significant positive *t* value for the interaction between Condition (Close) and the quadratic time polynomial suggest that English listeners' differential fixation line had a shallower ascending slope and a sharper U-shape in the close condition than in the standard condition. Again, the more U-shaped differential fixation line in the close condition was perhaps due to an early baseline difference as well as an increased tonal competition mid-way through the tone in the conditions.

These results showed that the differential fixation lines in the distant and close conditions had a shallower ascending slope and a sharper U-shape than that of the standard condition. This indicates that English listeners showed a slower word recognition process in the distant and close conditions than in the standard condition of the level tone. These results suggest that English listeners used the within-category gradience of the level tone, but not in the way we predicted. The tones that were acoustically distant from or close to the tonal competitor might have disrupted their word recognition, which increased tonal competition in both the distant and close conditions.

The GCA with the best fit on English-speaking L2 learners' data for the contour tone included the linear and quadratic time polynomials and Condition, but no interaction between the two. The significant positive *t* values for the linear and quadratic time polynomial indicate that English listeners' differential fixation line in the standard condition had an ascending slope and a U-shape. The significant negative *t* values for the distant and close conditions suggest that English listeners had a lower differential proportion of fixations in both the distant and close conditions compared with the standard condition. The main effects of Condition could be due to baseline differences across the conditions.

These results showed that the differential fixation lines in the distant and close conditions did not differ in their slope or shape from that of the standard condition. This indicates that English listeners had a similar word recognition process in all three conditions of the contour tone. These results suggest that English listeners did not use the within-category gradience of the contour tone. The effects of Condition (i.e., the overall difference between the distant and close conditions and the standard conditions) found for the contour tone appears to be a baseline effect (i.e., an effect present before listeners' intake of the speech signal), so this effect cannot be attributed to the speech signal.

3.5.3 L2 Learners vs. Native Listeners

To determine whether English-speaking L2 learners differ from native Chinese listeners in their use of within-category gradience of level and contour tones, GCAs were conducted on all listeners' differential fixations separately for the level and contour tones, with Language, Condition, and the time polynomials as fixed effects. The results of the GCAs with the best fit for the level tones can be found in Table 13.

Table 13: Results of Growth curve analyses on all listeners' differential proportions of fixations separately for the level tones

Variable	**Estimate**	***t***	***p***
(intercept)	0.120	6.681	**<.001**
Time			
Linear	0.997	8.507	**<.001**
Quadratic	0.328	5.499	**<.001**
Cubic	0.001	<|1|	.99
Language	–0.115	–4.148	**<.001**
Condition (Distant)	–0.021	–3.437	**<.001**
Condition (Close)	0.015	2.446	**.01**
Time × Language			
Linear	–0.807	–4.462	**<.001**
Quadratic	–0.354	–3.846	**<.001**
Cubic	–0.032	<|1|	.66
Time × Condition (Distant)			
Linear	0.052	1.116	.26
Quadratic	0.077	1.667	.10
Cubic	–0.058	–1.252	.21
Time × Condition (Close)			
Linear	0.052	1.143	.26
Quadratic	–0.051	–1.094	.27
Cubic	–0.139	–3.004	**<.01**
Language × Condition (Distant)	–0.009	–1.011	.31
Language × Condition (Close)	–0.047	–4.596	**<.001**
Time × Language × Condition (Distant)			
Linear	–0.292	–4.086	**<.001**
Quadratic	0.140	1.960	**.05**
Cubic	0.181	2.526	**.01**
Time × Language × Condition			

(Close)			
Linear	–0.348	–4.859	**<.001**
Quadratic	0.226	3.155	**<.01**
Cubic	0.241	3.370	**<.001**

Note. α = .05; significant results are in bold. *n* = 10416 observations.

For the level tones, the GCA yielded a simple effect of Language, with English listeners having a lower differential fixation line than Chinese listeners. The analysis yielded a significant interaction between Language and Condition (Close), with the (positive) effect of Condition in the close condition decreasing (i.e., becoming less positive) for English listeners as compared to Chinese listeners. The GCAs also revealed a significant three-way interaction between the linear time polynomials, Condition (Close and Distant), and Language, with English listeners showing a differential fixation line with a shallower ascending slope than that of Chinese listeners in the close and distant conditions. They also revealed a significant three-way interaction between the quadratic time polynomials, Condition (Close and Distant), and Language, with English listeners showing a differential fixation line with a sharper U-shape than that of Chinese listeners in the close and distant conditions. Finally, the GCAs revealed a significant three-way interaction between the cubic time polynomials, Condition (Close and Distant), and Language, with English listeners showing differential fixation lines which had a less sharp ‘s’ shape than that of Chinese listeners in the close and distant conditions.

These results confirm that, for the level tone, L2 learners differed from native listeners in the effect of Condition they showed: First, English listeners showed more tonal competition and took more time than Chinese listeners in recognizing the target over the competitor word in the close condition as compared to the standard condition; moreover, while English listeners showed an effect of Condition (Distant and Close vs. Standard conditions) as the speech signal unfolded over time (see Section 3.5.2 for more details), Chinese listeners did not show such a pattern (see Section 3.5.1 for more details).

Table 14 presents the results of the GCAs with the best fit for the contour tones. For the contour tones, the GCA yielded a simple effect of Language, with English listeners having a lower differential fixation than Chinese listeners. The analysis yielded a significant interaction between Language and Condition (Distant), with the (positive) effect of Condition in the distant condition decreasing (i.e., becoming less positive) for English listeners as compared to Chinese listeners. The GCAs also revealed significant three-way interactions between the linear time polynomials, Condition (Close and Distant), and Language, with English listeners showing a differential fixation line with a steeper ascending slope than that of Chinese listeners in the close condition and a differential fixation line with a shallower ascending slope than that of Chinese listeners in the distant condition. Finally, there is a significant three-way interaction between the quadratic time polynomials, Condition (Distant), and Language, with English listeners showing a differential fixation line with a sharper U-shape than that of Chinese listeners in the distant condition.

Table 14: Results of Growth curve analyses on all listeners' differential proportions of fixations separately for the contour tones

<table>
<tr><th>Variable</th><th>Estimate</th><th>t</th><th>p</th></tr>
<tr><td>(intercept)</td><td>0.106</td><td>6.158</td><td><.001</td></tr>
<tr><td>Time</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>0.797</td><td>8.046</td><td><.001</td></tr>
<tr><td>Quadratic</td><td>0.355</td><td>5.647</td><td><.001</td></tr>
<tr><td>Language</td><td>–0.078</td><td>–2.928</td><td><.01</td></tr>
<tr><td>Condition (Distant)</td><td>0.055</td><td>8.784</td><td><.001</td></tr>
<tr><td>Condition (Close)</td><td>–0.061</td><td>–9.798</td><td><.001</td></tr>
<tr><td>Time × Language</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>–0.562</td><td>–3.673</td><td><.001</td></tr>
<tr><td>Quadratic</td><td>–0.154</td><td>–1.590</td><td>.12</td></tr>
<tr><td>Time × Condition (Distant)</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>0.347</td><td>7.455</td><td><.001</td></tr>
<tr><td>Quadratic</td><td>–0.146</td><td>–3.148</td><td><.01</td></tr>
<tr><td>Time × Condition (Close)</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>–0.161</td><td>–3.457</td><td><.001</td></tr>
<tr><td>Quadratic</td><td>–0.059</td><td>–1.265</td><td>.21</td></tr>
<tr><td>Language × Condition (Distant)</td><td>–0.083</td><td>–8.688</td><td><.001</td></tr>
<tr><td>Language × Condition (Close)</td><td>0.006</td><td><|1|</td><td>.55</td></tr>
<tr><td>Time × Language × Condition (Distant)</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>–0.229</td><td>–3.198</td><td><.01</td></tr>
<tr><td>Quadratic</td><td>0.212</td><td>2.949</td><td><.01</td></tr>
<tr><td>Time × Language × Condition (Close)</td><td></td><td></td><td></td></tr>
<tr><td>Linear</td><td>0.227</td><td>3.160</td><td><.01</td></tr>
<tr><td>Quadratic</td><td>0.001</td><td><|1|</td><td>.99</td></tr>
</table>

Note. $\alpha = .05$; significant results are in bold. $n = 10416$ observations.

These results indicate that, for the contour tones, L2 learners differed from native listeners in the effect of Condition they showed: First, English listeners showed more tonal competition and took more time than Chinese listeners in recognizing the target over the competitor word in the distant condition as compared to the standard condition; additionally, whereas Chinese listeners showed an effect of Condition (Distant and Close vs. Standard conditions) as the speech signal unfolded over time (see Section 3.5.1 for more details), English listeners did not show such a pattern (see Section 3.5.2 for more details).

3.6 Discussion

This study investigates whether native Chinese listeners show sensitivity to within-category gradience in lexical tones, and whether native Chinese listeners and English-speaking L2 learners of Chinese differ from each other in their use of the fine-grained tonal information of level versus contour tones. Participants completed a visual-world eye-tracking experiment with pictures representing Chinese words contrasting in lexical tones. In the experimental condition, the target was a level tone (i.e., T1) and the competitor was a high-rising contour tone (i.e., T2), or vice versa. The auditory stimuli were manipulated such that the target tone was either canonical in the standard condition, acoustically more distant from the competitor in the distant condition, or acoustically closer to the competitor in the close condition. Chinese and English listeners' lexical competition patterns in the distant and close conditions were compared with those in the standard condition separately for the level and contour tones.

The results of the Chinese listeners' differential proportions of fixations showed that the within-category gradience of the contour tone, but not that of the level tone, modulated their word recognition: For the contour tone, decreased tonal competition was found in the distant condition as a result of an acoustically greater tonal distance between the target and competitor, whereas increased tonal competition took place in the close condition as a result of an acoustically smaller tonal distance between the target and competitor. However, the same pattern of results was not found for the level tone. Unlike the results of native Chinese listeners, English listeners' differential proportions of fixations demonstrated that the within-category gradience of the level tone rather than that of the contour tone modulated their word recognition. For the level tone, increased tonal competition took place in both the distant and close conditions, but the same pattern of results was not found for the contour tone.

Consistent with previous findings on English listeners' sensitivity to within-category gradience in VOT (McMurray et al., 2002, 2008, 2009), the results of the Chinese listeners indicate that the within-category gradience of the contour tone influenced listeners' word recognition as the speech signal unfolded. The use of the eye-tracking method, which is highly sensitive to lexical activation over time and can reveal differences elicited by fine-grained acoustic information, made it possible to capture these fine-grained effects of within-category gradience of lexical tones (e.g., Allopenna et al., 1998; Tanenhaus et al., 1995).

The main contribution of this experiment is in demonstrating that, under specific circumstances, native Chinese listeners were able to use the within-category gradience in contour tones, but not in the level tone, during online word recognition. These findings suggest that some but not all of the fine-grained tonal information is kept by Chinese listeners during their online word recognition. One possibility is that Chinese listeners do not discard the fine-grained tonal variability when this variability is along a dimension (i.e., pitch contour) that is meaningful for distinguishing tones, whereas they do discard the tonal variability when it is along a dimension (i.e., pitch height) that is not contrastive in their native tonal inventory. This explanation is consistent with tone perception studies in that pitch contour is more important than pitch height for Chinese listeners, as the Chinese tones are mainly cued by differences in pitch contour (e.g., Francis et al., 2008; Guion & Pederson, 2007). This explanation is also consistent with the VOT studies conducted by McMurray and colleagues (McMurray et al., 2002; 2008; 2009) in that VOT is a meaningful dimension that contrasts voiced and voiceless stops in English.

An alternative explanation of the results that is related to the account of meaningful dimensions is that the contour tone T2 may have less room for variability, as it is similar to another Chinese contour tone, T3, whereas the level tone, which is high up in the tonal space, may have more room for tonal variability (e.g., Huang, 2001; Moore & Jongman, 1997; Xu et al., 2006). The hypothesis that T2 may have less room for tonal variability than T1 can also account for the present findings that Chinese listeners used the fine-grained variability of the contour tone, but not that of the level tone, though such an account does not explain English listeners' results. Further research is needed to determine whether this finding generalizes to the other contour tones of Chinese. Moreover, future research, potentially using artificial tone languages where either pitch contour or pitch height is contrastive and the tones all have similar room for variability, is suggested to tease the account of meaningful dimensions apart from the account of variability in the tonal space.

To the best of our knowledge, this study was the first to examine whether (and if so, how) L2 learners of Chinese show different time courses of lexical activation in their use of the fine-grained tonal information of level versus contour tones—that is, *as the speech signal unfolds over time*. Echoing previous behavioral and ERP findings (Chandrasekaran et al., 2009; Leather, 1987; Stagray & Downs, 1993), the results of English-speaking L2 learners of Chinese showed that the within-category gradience of the level tone, but not that of the contour tone, influenced L2 learners' word recognition. English listeners are more sensitive to pitch height than to pitch contour in their perception of lexical tones, potentially due to that pitch height is encoded in English stress contrasts (e.g., Braun et al., 2014; Francis et al., 2008; Gandour, 1983; Guion & Pederson, 2007; Shen, 1989; White, 1981). This can potentially explain why they were able to distinguish the stimuli in the distant and close conditions from those in the standard condition for the level tone, but they had a similar word recognition process in all the three conditions of the contour tone.

Interestingly, when native and non-native listeners showed an effect of within-category gradience in lexical tones (contour tones for native listeners and level tones for non-native listeners), they did so differently: While native listeners showed a gradient pattern of lexical competition based on the phonetic distance (or closeness) of the target tones (in relation to the tones represented for the competitor word), L2 learners were only able to distinguish the two "non-canonical" tonal tokens from the canonical tonal token by treating both of them as worse tokens. This different pattern of results for native and non-native listeners indicate that they might rely on different mechanisms to process tones during online word recognition. Specifically, Chinese listeners may have more robust representations of the tones, and thus may be better at organizing the "non-canonical" tokens relative to the canonical tokens in the tonal space. Consequently, they are better at mapping the tokens they hear as closer to or more distant from prototypical tokens. On the other hand, English listeners may not have been exposed to enough tonal input to develop robust representations; as a result, they may only be able to distinguish "non-canonical" tokens from the canonical ones.[13] In other words, English listeners may not be good at organizing the "non-canonical" tokens relative to the prototypical tokens in the tonal space, which may have increased tonal competition when the pitch contour of the tokens they heard did not exactly match the pitch contour represented for the prototypical tokens.

The present findings raise questions about the different mechanisms underlying native and non-native listeners' encoding of prosodic cues in spoken word recognition. These findings imply that native and non-native listeners, who potentially differ in the robustness of their representations of lexical tones, may adopt different strategies to deal with fined-grained tonal information to resolve the lexical competition during online word recognition. Future research recruiting L2 learners of Chinese at near-native proficiency level, who have been exposed to substantial tonal input in Chinese-speaking countries and thus who should have robust representations of lexical tones, would be required to investigate whether their strategies to deal with the fine-grained tonal information will still differ from native listeners' use of this information.

More importantly, the current findings suggest that while native Chinese listeners were more likely to use the variability of tonal contour, English-speaking L2 learners were more likely to use tonal height differences to process tones. The findings imply that not only whether the L1 has lexical tones, but also whether prosodic cues contribute to distinguishing among words in the L1, and how they do so, influence L2 learners' use of these cues in their word

[13] Given the L2 learners' previous exposure to the speaker's natural tokens in the training phase and in Experiment 1, they should have stored some form of canonical tones for the speaker when they participated in Experiment 2.

recognition (Cutler, 2012). To further investigate whether the use of prosodic cues in the L1 prosodic system influences L2 learners' use of fine-grained tonal information of level versus contour tones, a future direction of research would be compare L2 learners whose L1s differ with respect whether prosodic cues signal lexical identity (e.g., Cantonese, which encodes both pitch height and pitch contour in lexical tones, English, which encodes pitch height in lexical stress contrast, vs. French, which does not encode pitch in lexical prosody) (cf. Braun et al., 2014). Such research would contribute to our understanding of the mechanism that underlies non-native listeners' use of prosodic information, thereby explaining L2 learners' (in) ability to process lexical prosody.

Chapter 4: General Discussion

This dissertation research examined how native listeners and advanced English-speaking L2 learners of Chinese used tonal information as the speech signal unfolded over time in word recognition. More specifically, using two visual-world eye-tracking experiments, the current research investigated two potential differences between native Chinese listeners and English-speaking L2 learners of Chinese in their use of tonal information in word recognition: (i) their potentially different incremental use of the early pitch height of the tone; and (ii) their potentially different sensitivities to the within-category gradience of contour versus level tones.

Experiment 1 was a visual-world eye-tracking experiment in Chinese with 36 native Chinese listeners and 26 highly proficient English-speaking L2 learners of Chinese. It examined Chinese and English listeners' incremental use of early pitch height differences between T1-T2 and T1-T4 pairs. The results showed that whereas Chinese listeners used the early pitch height information before the pitch contour information was available, and continued using it after the pitch contour information of lexical tones had been heard, English listeners were also able to use the early pitch height information to recognize spoken Chinese words, but they used this information later than native Chinese listeners, that is, after the pitch contour information of the tone had been heard.

Experiment 2, which used the same experimental paradigm and tested the same participants, was explicitly designed to test native and non-native listeners' sensitivities to the within-category gradience (Standard, Distant, and Close condition) of the level versus contour tone in the T1-T2 pair. The results showed that native Chinese listeners showed sensitivities to the within-category gradience of the contour tone (i.e., T2) rather than that of the level tone (i.e., T1), whereas L2 learners showed some sensitivities to the within-category gradience of the level tone rather than that of the contour tone. Moreover, L2 learners treated tonal variability (i.e., Close/Distant stimuli) differently from native listeners, potentially due to the degree of robustness of their representations of lexical tones.

The following sections further discuss these results and their implications for native and non-native word recognition as well as for Chinese pedagogy. This chapter ends by outlining the main contributions of this dissertation on the use of tonal information in spoken word recognition.

4.1 The Role of Tone in Native and Non-Native Word Recognition

Our findings have important implications for models of native auditory word recognition (e.g., Marlsen-Wilson, 1989, McClelland & Elman, 1986, Norris, 1994). Our findings suggest that native listeners used tonal information incrementally, which is consistent with models where the mapping between the speech signal and lexical representations is continuous (e.g., TRACE model). Shuai and Malins (2016) attempted to incorporate lexical tones in their TRACE-T model, in which pitch height (i.e., from 1 to 5, 5 being the highest) and pitch contour (i.e., level, rising, vs. falling) are available simultaneously from the beginning of the tone at the feature level. First, the results of our first experiment indicate that the time-course with which listeners use pitch cues is important for spoken word recognition, as lexical tones often have pitch dynamic changes. For instance, we found that, for the T1-T2 pair in this study, native Chinese listeners' word recognition was constrained by pitch height differences before the pitch contour information of lexical tones had been heard. Our findings suggest that pitch height and pitch contour are used with different time-courses (i.e., when tones differ in early pitch height, this information is used before pitch contour information is heard and integrated). Hence, a model that treats pitch contour, together with pitch height, as a feature that is available right from the beginning of the tone (e.g., TRACE-T) would be likely unable to simulate what the participants did in our first experiment. Future models that incorporate lexical tones in word recognition would need to incorporate the time-course information with which pitch height and pitch contour are used to simulate listeners' use of tonal information during online word recognition.

Furthermore, the results of our second experiment suggest that the fine-grained tonal information of contour tones influenced native listeners' word recognition process. A model that does not have a mechanism for capturing the fine-grained tonal variability (e.g., TRACE-T) would likely be unable to simulate what the participants did in our second experiment. Based on the findings of our second experiment, we suggest that future models that incorporate lexical tones in word recognition would need to incorporate the nuanced changes of pitch height (e.g., increasing pitch height values in TRACE-T) over time to capture the effect of the fine-grained tonal information, especially for contour tones.

The findings of the current research also have important implications for non-native word recognition. The results of our current research suggest that English-speaking L2 learners of Chinese, whose L1 does not have lexical tone, were able to use pitch height and its fine-grained variations to process lexical tones, likely because they encode pitch height cues as part of their native lexical representations (e.g., English stress). In other words, if specific cues to lexical tones in the L2 are mapped onto a different lexical representation in the L1, listeners are able to use these cues to recognize words in the L2. Therefore, the current models of L2 word recognition and speech perception (e.g., Best, 1995; Cutler, 2012; Flege, 1995) need to consider what prosodic cues signal word identity in the L1, and how they are realized in L1 versus L2 lexical representations.

Additionally, the results of our second experiment indicate that native and non-native listeners might adopt different strategies to deal with fine-grained tonal information to resolve lexical competition during online word recognition. While native listeners have robust representations of lexical tones and were able to map the tone contour of "non-canonical" tokens (i.e., Distant and Close conditions) onto that of the prototypical representation (Standard condition) based on the acoustic distance between the target and competitor words, non-native listeners do not have as robust representations as Chinese listeners and thus have difficulty mapping the tone contour of both "non-canonical" tokens with that of prototypical representation (they treated both distant and close tokens of the level tone as worse tokens than the canonical one). These findings have implications for research on the effect of tonal variability, which has been examined in L2 tonal training studies (Chang & Bowles, 2015; Leather 1990; Li, 2016; Liu & Zhang, 2016; Logan et al., 1991; Wang et al., 1999, 2003). Non-native listeners' mapping difficulty between tone contour of "non-canonical" tokens and prototypical representations need to be considered in future training studies that focus on tonal variability caused by different speakers, tonal contexts, and speech rates. Therefore, the current models of L2 word recognition and speech perception need to consider non-native listeners' difficulty in dealing with fine-grained tonal variability, especially along the pitch contour dimension.

Last but not least, our results indicate that non-native listeners, who always showed more tonal competition than native listeners, were not able to make efficient use of tonal information (e.g., pitch height and its fine-grained variability) efficiently to recognize Chinese words. The findings are consistent with studies on segments (e.g., Broersma, 2012; Broersma & Cutler, 2008, 2011) in that the tonal information could not be integrated efficiently in the word recognition process due to non-native listeners' perceptual confusion of tonal categories and/or their possibly unstable representations of lexical tones. The hypotheses of what could cause non-native listeners' greater tonal competition, that is, perceptual confusion and/or unstable lexical representation of tones, should be assessed in further studies of L2 spoken word recognition.

4.2 Chinese Pedagogy

In addition to implications for native and non-native word recognition, this research has pedagogical implications for Chinese language teaching. The results of our current research showed that English-speaking L2 learners of Chinese had several difficulties using tonal information.

First, L2 learners in the classroom may have more difficulty using the fine-grained tonal information of contour tones rather than the fine-grained tonal information of level tones in online word recognition. This difficulty likely stems from the fact that English-speaking L2 learners of Chinese rely more on height information and may have difficulty integrating the late pitch contour information. Second, L2 learners in the classroom may have difficulty mapping the tonal contour of "non-canonical" tones onto that of the prototypical representation based on the tonal distance between the target and competitor tone. This difficulty is potentially due to the fact that the L2 learners did not have sufficient exposure to Chinese tones.

Given the L2 learners' difficulties using tones, Chinese language teachers are encouraged to develop L2 teaching/training materials that focus on the use of pitch contour (e.g., rising vs. falling) and its fine-grained variability. One approach to help L2 learners overcome their difficulty using pitch contour information would be to have them complete training in a laboratory setting, in which learners need to learn to pay more attention to pitch contour changes than to pitch height differences (Chandrasekaran et al., 2010; Francis et al., 2008; Holt & Lotto, 2006; Liu & Zhang, 2016). To help L2 learners overcome their difficulty dealing with the tonal variability, acoustically variable tonal stimuli produced by different speakers (female and male), in different tonal contexts (preceding/following the same and different tones), and in different words can be used in a such training (Chang & Bowles, 2015; Leather 1990; Li, 2016; Liu & Zhang, 2016; Logan et al., 1991; Wang et al., 1999, 2003). This high-variability perceptual training would help L2 learners improve their use of pitch contour and its fine-grained variability. Consequently, it would result in more robust representations of tonal categories, thus more efficient integration of the pitch contour over time for L2 learners.

4.3 Concluding Remarks

This doctoral dissertation research investigated the processing of Chinese lexical tones in spoken word recognition by native Chinese listeners and highly proficient English-speaking L2 learners of Chinese. More specifically, this research used the visual-world eye-tracking paradigm to shed light on the precise time course with which native and non-native listeners used tonal information in online word recognition. Its findings contribute to the understanding of how tonal information modulates lexical activation in native and non-native Chinese listeners and have important implications for models of spoken word recognition and for Chinese language teaching. Future research should use different methods to examine the phenomenon of tonal contrasts from other tonal languages (e.g., level-level and contour-contour tonal contrasts in Cantonese) as well as other prosodic categories (e.g., pitch accent or intonation categories) to further test native and non-native listeners’ use of suprasegmental information during online word recognition.

References

Allen, J. S., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: a distinction between lexical status and speaking rate. *Perception and Psychophysics*, 63, 798–810.

Allopenna, P. D, Magnuson, J.S., & Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition: evidence for continuous mapping models. *Journal of Memory and Language*, 38, 419–439.

Andruski, J., Blumstein, S., & Burton, M. (1994). The effect of subphonetic differences on lexical access.*Cognition*, 52, 163–187.

Aoyama, K., Flege, J. E., Guion, S., Yamada, T., & Akahane–Yamada, R. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /r/ and /l/. *Journal of Phonetics*, 23, 233–250.

Barr, D.J., Gann, T.M., & Pierce, R.S. (2011) Anticipatory baseline effects and information integration in visual world studies. *Acta Psychologica*, 137(2), pp. 201–207.

Bates, D., Maechler, B., Bolker, B., & Walker, S. (2015). Fitting linear mixed–effects models using lme4. *Journal of Statistical Software*, 67, 1–48. doi:10.18637/jss.v067.i01.

Beckman, M. E. (1986). *Stress and Non–stress Accent*. Dordrecht: Foris.

Beddor, P. S., McGowan, K. B., Boland, J. E., Coetzee, A. W., & Brasher, A. (2013). The time course of perception of coarticulation. *Journal of the Acoustical Society of America.* 133, 2350–2366.

Bent, T. (2005). *The perception and production of non–native prosodic categories.* Unpublished Ph.D. thesis, Department of Linguistics, Northwestern University, Evanston, IL.

Best, C. T. (1995). A direct realistic view of cross–language speech perception. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross–language research* (pp.171–204). Baltimore, MD: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English–speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 345–360.

Boersma, P., & Weenink, D. (2012). *Praat: Doing phonetics by computer* (computer program), Retrieved from http://www.praat.org.

Bolinger, D. (1958). A theory of pitch accent in English, *Word*, 14, 109–149.

Bradlow, A. R., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America*, 101, 2299–2310.

Braun, B., Galts, T., & Kabak, B. (2014). Lexical encoding of L2 tones: the role of L1 stress, pitch accent and intonation. *Second Language Research*, 30, 323–350.

Broersma, M. (2012). Increased lexical activation and reduced competition in second–language listening. *Language and Cognitive Processes*, 27, 1205–1224.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System: An International Journal of Educational Technology and Applied Linguistics*, 36, 22–34.

Broersma, M., & Cutler, A. (2011). Competition dynamics of second–language listening. *Quarterly Journal of Experimental Psychology*, 64, 74–95.

Burnham, D., & Mattock, K. (2007). The perception of tones and phones. In O.–S Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp.259–280). Amsterdam: John Benjamins.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX–CH: Chinese word and character frequencies based on film subtitles. *Plos ONE*, 5(6), e10729.

Chandrasekaran, B., Krishnan, A., & Gandour, J. T. (2009). Sensory processing of linguistic pitch as reflected by the mismatch negativity. *Ear and Hearing*, 30(5), 552–558.

Chandrasekaran, B., Sampath, P. D., & Wong P. C. M (2010). Individual variability in cue–weighting and lexical tone learning. *Journal of Acoustical Society of America*, 128, 456–465.

Chang, C. B., & Bowles, A. R. (2015). Context effects on second-language learning of tonal contrasts. *Journal of the Acoustical Society of America*, 138(6), 3703–3716.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non–native listeners. *Language and Speech*, 45, 207–228.

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*, MIT Press.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition, in H. Bouma and D. G. Bouwhuis (Eds), *Attention and performance X*, 183–196, Erlbaum.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. *Language and Cognitive Processes*, 16, 507–534.

Dupoux, E., Sebastián–Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress "deafness": The case of French learners of Spanish. *Cognition*, 106, 682–706.

Fear, B. D., Cutler, A., & Butterfield, S. (1995). The strong/weak syllable distinction in English. *Journal of the Acoustical Society of America*, 97, 1893–1904.

Flege, J. E. (1995). Second–language speech learning: findings, and problems. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–273). Baltimore, MD: York Press.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 89, 395–411.

Flege, J. E., Munro, M. & MacKay, I. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication*, 16, 1–26.

Flege, J. E., Schmidt, A., & Wharton, G. (1996). Age of learning affects rate–dependent processing of stops in a second language. *Phonetica*, 53, 143–161.

Flege, J. E., Yeni–Komshian, G., & Liu, S. (1999). Age constraints on second language learning. *Journal of Memory and Language*, 41, 78–104.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non–tonal language listeners. *Journal of Phonetics*, 36, 268–294.

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

Francis, A. L., & Nusbaum, H. C. (2002). Selective attention and the acquisition of new phonetic categories. *Journal of Experimental Psychology: Human Perception and Performance*, 28, 349–366.

Fry, D. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustic Society of America*, 27, 765–768.

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

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning: In honor of James Emil Flege* (pp. 57–77). Amsterdam: John Benjamins.

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

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.11–10.112). New York: Wiley.

Hao, Y. C. (2012). Second language acquisition of Mandarin Chinese tones by tonal and non–tonal language speakers. *Journal of Phonetics*, 40(2), 269–279.

Holt, L.L., & Lotto, A.J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition. *Journal of the Acoustical Society of America*, 119, 3059–3071.

Howie, J. M. (1976). *Acoustical studies of Mandarin vowels and tones*. Cambridge, UK: Cambridge University Press.

Huang, T., (2001). The interplay of perception and phonology in Tone 3 sandhi in Chinese Putonghua. In Hume, E. and K. Johnson (eds.), *Studies on the interplay of speech perception and phonology* (The Ohio State University Working Papers in Linguistics, No. 55), pp. 23-42.

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2011). Can native Japanese listeners learn to differentiate /r–l/ on the basis of F3 onset frequency? *Bilingualism: Language and Cognition*, 15, 255–274.

Iverson, P., Kuhl, P. K., Akahane–Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non–native phonemes. *Cognition*, 87, B47–B57.

Johnson, J., & Newport, E. (1989). Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language. *Cognitive Psychology*, 21, 60–99.

Kuznetsova, A., Brockhoff, B., & Christensen, H. (2016). Tests in linear mixed effects models. Retrieved from https://cran.r–project.org/web/packages/lmerTest/index.html.

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

Leather, J. (1987). F0 pattern inference in the perceptual acquisition of second language tone. In A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 59–81). Dordrecht: Foris Publications.

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

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for Advanced Learners of English. *Behavior Research Methods*, 44, 325–343.

Lenneberg, E. H. (1967). *Biological Foundations of Language*. New York: Wiley.

Li, Y. J. (2016). Effects of high variability phonetic training on monosyllabic and disyllabic Mandarin Chinese tones for L2 Chinese learners. *Unpublished PhD Dissertation*. Department of Linguistics, the University of Kansas, KS.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 54, 358–368.

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *Journal of the Acoustic Society of America*, 32, 451–454.

Lin, C., Wang, M., Idsardi, W. J., & Xu, Y. (2014). Stress processing in Mandarin–English and Korean–English bilinguals. *Bilingualism: Language and Cognition*, 17, 316–346.

Lin, T., & Wang, W. S.–Y. (1984).The perception of tones. *Journal of Chinese Linguistics*, 2, 59–69.

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/: II. The role of phonetic environment and talker variability in learning new perceptual categories. *Journal of the Acoustical Society of America*, 94, 1242–1255.

Liu, C. (2013). Just noticeable differences of tone pitch contour change for English– and Chinese–native listeners. *Journal of the Acoustic Society of America*, 134, 3011–3020.

Liu, J., & Zhang, J. (2016). The effects of talker variability and variances on incidental learning of lexical tones. In proceedings of *the 5th International Symposium of Tonal Aspects of Languages*, DOI: 10.21437/TAL.2016.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, 89, 874–886.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Percept Psychophysics*, 39, 155–158.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.

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

Marslen–Wilson, W. (1989). Access and integration: Projecting sound onto meaning. In W. Marslen–Wilson (Eds.), *Lexical representation and process* (pp. 3–24). Cambridge, MA: MIT Press.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, 1–86.

McMurray, B., Clayards, M., Tanenhaus, M., & Aslin, R. (2008) Tracking the timecourse of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin and Review*, 15, 1064–1071.

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

McMurray, B., Tanenhaus, M., & Aslin, R. (2009) Within–category VOT affects recovery from "lexical" garden paths: Evidence against phoneme–level inhibition. *Journal of Memory and Language*, 60, 65–91.

Miller, J. L. (1997). Internal structure of phonetic categories. *Language and Cognitive Processes*, 12, 865–869.

Mirman, D. (2014). *Growth Curve Analysis and Visualization Using R.* Chapman and Hall / CRC.

Moore, C.B., & Jongman, A. (1997). Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, 102, 1864–1877.

Neergaard K., Xu H., & Huang C.R. (2016) Database of Mandarin Neighborhood Statistics. In Proceedings of *Language Resources and Evaluation Conference 2016,* Slovenia.

Norman, J. (1988). *Chinese*. Cambridge: Cambridge University Press.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition,* 52, 189–234.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, W. S–Y. (2010) The influence of language experience on categorical perception of F0 contour–levels. *Journal of Phonetics*, 38, 616–624.

Qin, Z., Chien, Y., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics*, 38 (3), 541–570.

Qin. Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech*, 59, 318–338.

Qin. Z., & Mok, P. (2013). Discrimination of Cantonese tones by speakers of tone and non–tonal languages. *Kansas working papers in Linguistics,* 34.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *Quarterly Journal of Experimental Psychology*, 63(4), 772–783.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition,* 90, 51–89.

Schouten, M., & van Hessen, A. (1992). Modeling phoneme perception. I: Categorical perception. *Journal of the Acoustical Society of America*, 92, 1841–1855.

Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 24(3), 27–47.

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.

Shen, X. (1989). Toward a register approach in teaching Mandarin tones. *Journal of the Chinese Language Teachers Association*, 24, 27–47.

Shuai, L., & Malins, J. G. (2016). Encoding lexical tones in jTRACE: A simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behavior Research Methods*, 1–12. DOI:10.3758/s13428–015–0690–0.

Stagray, J. R., & Downs, D. (1993). Differential sensitivity for frequency among speakers of a tone and a nontone language. *Journal of Chinese Linguistics*, 21, 143–163.

Stevens, K. N. (2002). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of the Acoustical Society of America*, 111, 1872–1891.

Sun, K. C. (2012). The role of lexical tone in L2 spoken word recognition. *Unpublished PhD Dissertation*. Department of Linguistics, the University at Buffalo, State University of New York, NY.

Sun, K. C., & Huang, T (2012). A cross–linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguist*, 21, 305–327.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. G. (2000). Eye movements and lexical access in spoken language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research*, 29, 557– 580.

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

Toscano J. C., & McMurray B. (2012). Cue–integration and context effects in speech: Evidence against speaking–rate normalization. *Attention, Perception, and Psychophysics*, 74, 1284–1301.

Tremblay, A., & Ransijn, J. (2015). Model selection and post–hoc analysis for (G)LMER models. Retrieved from https://cran.r–project.org/web/packages/LMERConvenienceFunctions/

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Sciences,* 28, 61–72.

Wang, Y., Jongman, A., & Sereno, J. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *Journal of the Acoustical Society of America*, 113, 1033–1043.

Wang, Y., Spence, M., Jongman, A., & Sereno, J. (1999). Training American listeners to perceive Mandarin tones. *Journal of the Acoustical Society of America*, 106, 3649–3658.

White, C. (1981). Tonal pronunciation errors and interference from English intonation. *Journal of the Chinese Language Teachers Association*, 16(2), 27–56.

Wiener, S. (2015). *The representation, organization, and access of lexical tone by native and non–native Mandarin speakers.* Unpublished Ph.D. thesis, Ohio State University.

Wiener, S., & Ito, K. (2015). Do syllable–specific tonal probabilities guide lexical access? Evidence from Mandarin, Shanghai and Cantonese speakers. *Language, Cognition, and Neuroscience,* 30(9), 1048–1060.

Wiener, S., & Ito, K. (2016). Impoverished acoustic input triggers probability–based tone processing in mono–dialect Mandarin listeners. *Journal of Phonetics,* 56, 38–51.

Wiener, S., Ito, K., & Speer, S. R. (2016). Individual variability in the distributional learning of L2 lexical tone. In J. Barnes, A. Brugos, S. Shattuck-Hufnagel, and N. Veilleux (Eds.), *Proceedings of the 8th International Conference on Speech Prosody* (pp. 538–542). Boston, MA.

Xu, Y. S., Gandour, J., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of F0 direction. *Journal of the Acoustical Society of America,* 120, 1063–1074.

Yamada, R.A. (1995). Age and acquisition of second language speech sounds: perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross–language research* (pp. 305–320). Baltimore, MD: York Press.

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

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

Yuan, B. (2009). Non–permanent representational deficit and apparent target–likeness in second language: Evidence from wh–words used as universal quantifiers in English and Japanese speakers' L2 Chinese. In Snape, N., Leung, Y.–K. L., & Sharwood Smith, M. (Eds.).

Representational Deficits in SLA: In honour of Roger Hawkins, 69–103, Amsterdam: John Benjamins Publishing.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011).Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses. *Neuropsychologia*, 49, 1761–1770.

Appendix A: Language Background Questionnaire

Name:____ Age:_____________

Gender:___________ Major:____

Musical experience (e.g., did/do you play a musical instrument; how long):

____

What university year are you? Year ___ of ☐ undergraduate ☐ graduate studies.

What is **your** native language?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

What is your **first parent's** native language?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

Other (specify) _________________________

What is your **second parent's** native language?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

Other (specify) _________________________

What language(s) were used in your home from **birth to 5 years of age**?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

Other (specify) _________________________

What language(s) were used in your home from **6 to 11 years of age**?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

Other (specify) _________________________

What language(s) were used in your home from **12 to 17 years of age**?

English: American Australian British Canadian S. African

Chinese: Standard Mandarin Taiwan Mandarin Taiwanese Beijing

Northeast Other Dialects (specify) __________

Other (specify) _________________________

In what **country/countries** did you live…

…as a child?	…as a teenager?	…as an adult?

Excluding language classes, in what language were you taught (e.g., math, history, etc.) in…

… elementary school?	… middle school?	… high school?

Please list all languages you know **in order of dominance**.

1)	2)	3)	4)	5)

Please estimate your **global proficiency** in all the languages you know (beginner, intermediate, advanced, near-native, native).

Language	Mandarin	English	Taiwanese	Other (specify)	
Proficiency					

Please give the **percentage of time** you currently **use** each language (your percentages should add to 100%).

Language	Mandarin	English	Taiwanese	Other (specify)	
Percent					

If a text were available in all your languages, what **percentage of the time** would you **choose to read it** in each language (assume the original language of the text was a language you do not know)?

Language	Mandarin	English	Taiwanese	Other (specify)	
Percent					

When speaking a language with someone who is equally fluent in all your languages, **what percent of the time** would you choose to **speak** each of your languages?

Language	Mandarin	English	Taiwanese	Other (specify)	
Percent					

If Chinese is not your native language…

How many years of Chinese instruction have you received?

____

What Chinese dialects did your instructors speak (circle all that apply)?

Standard Mandarin (Putonghua)   Taiwan Mandarin (Guoyu)   Other________

Were a majority of your instructors native Chinese speakers? ☐ Yes ☐ No

At what **age** did you begin…

… learning Chinese at school?	… listening to Chinese?	…interacting with native Chinese speakers?

Please provide information about your **experiences in an Chinese speaking environment**.

Country	Age during visit	Length of visit (in months)	Context (study abroad, vacation, etc.)

How would you estimate your **proficiency in Chinese** (beginner, intermediate, advanced, near-native) for…

…reading?	… writing?	… listening?	…speaking?

Please describe the **circumstances** in which you **currently use Chinese** (e.g., Chinese class, with friends, listening to music, watching movies, etc.) and how often you do so (e.g., daily, frequently, sometimes, rarely).

Activity/circumstances	Frequency

In your perception of your own Chinese, **how much of an accent** would you say you have on a scale from 1-10 (1 being nearly indistinguishable from native Chinese speakers)? _______________________

Appendix B: Chinese LexTALE

Instructions

This test consists of about 120 trials, in each of which you will see a disyllabic word. Your task is to decide whether this is an existing Chinese word or not. If you think it is an existing Chinese word, please click on "yes", and if you think it is not an existing Chinese word, please click on "no". If you are sure that the word exists, even though you don't know its exact meaning, you may still respond "yes". But if you are not sure if it is an existing word, you should respond "no". In this experiment, simplified (left) and traditional (right) Chinese characters are provided for each word (e.g., 学会/ 學會) if they are different. You have as much time as you like for each decision. This part of the experiment will take about 10 minutes. If everything is clear, you can now start the experiment.

Word List

	Simplified	Traditional			Simplified	Traditional	
1	方圈	方圈	nonce	61	卷尺	卷尺	word
2	欢敢	歡敢	nonce	62	知音	知音	word
3	美秀	美秀	nonce	63	小鹿	小鹿	word
4	兴重	興重	nonce	64	奸商	奸商	word
5	怪后	怪后	nonce	65	校庆	校慶	word
6	宝定	寶定	nonce	66	猛增	猛增	word
7	保棍	保棍	nonce	67	吟诗	吟詩	word
8	采术	採術	nonce	68	划定	划定	word
9	舰船	艦船	nonce	69	问诊	問診	word
10	电牌	電牌	nonce	70	念经	念經	word
11	引光	引光	nonce	71	宝剑	寶劍	word
12	果春	果春	nonce	72	雨林	雨林	word
13	合区	合區	nonce	73	改写	改寫	word
14	极面	極面	nonce	74	上山	上山	word
15	服包	服包	nonce	75	学历	學歷	word
16	结开	結開	nonce	76	新型	新型	word
17	害鬼	害鬼	nonce	77	谅解	諒解	word
18	亏警	虧警	nonce	78	老年	老年	word
19	掉脑	掉腦	nonce	79	惊恐	驚恐	word
20	女击	女擊	nonce	80	电器	電器	word
21	批会	批會	nonce	81	小鸟	小鳥	word
22	血纸	血紙	nonce	82	暑假	暑假	word
23	适海	海物	nonce	83	聊天	聊天	word
24	幻手	幻手	nonce	84	好奇	好奇	word
25	数事	數事	nonce	85	阳光	陽光	word
26	天该	天該	nonce	86	学会	學會	word
27	映镜	映鏡	nonce	87	放心	放心	word
28	往过	往過	nonce	88	新鲜	新鮮	word
29	下态	下態	nonce	89	教育	教育	word
30	电巫	電巫	nonce	90	打败	打敗	word
31	自宝	自寶	nonce	91	好看	好看	word
32	座质	座質	nonce	92	效果	效果	word
33	错娘	錯娘	nonce	93	环境	環境	word
34	火反	火反	nonce	94	年级	年級	word
35	平决	平決	nonce	95	神奇	神奇	word
36	师行	師行	nonce	96	蓝色	藍色	word

37	随共	隨共	nonce	97	惩罚	懲罰	word
38	衣动	衣動	nonce	98	巨大	巨大	word
39	圆完	圓完	nonce	99	孤独	孤獨	word
40	查听	查聽	nonce	100	魔鬼	魔鬼	word
41	药厂	藥廠	word	101	政治	政治	word
42	国语	國語	word	102	姐妹	姐妹	word
43	报国	報國	word	103	大约	大約	word
44	采煤	採煤	word	104	变化	變化	word
45	师弟	師弟	word	105	后悔	後悔	word
46	省长	省長	word	106	空气	空氣	word
47	远近	遠近	word	107	护士	護士	word
48	晨练	晨練	word	108	相处	相處	word
49	水车	水車	word	109	石头	石頭	word
50	品茶	品茶	word	110	太阳	太陽	word
51	孤苦	孤苦	word	111	箱子	箱子	word
52	浑浊	渾濁	word	112	唱歌	唱歌	word
53	豪放	豪放	word	113	爷爷	爺爺	word
54	姑父	姑父	word	114	女生	女生	word
55	票务	票務	word	115	旅馆	旅館	word
56	书柜	書櫃	word	116	蛋糕	蛋糕	word
57	校花	校花	word	117	许多	許多	word
58	近景	近景	word	118	前面	前面	word
59	重负	重負	word	119	健康	健康	word
60	宿敌	宿敵	word	120	干净	乾淨	word

Appendix C: Chinese Cloze Test

Instructions

In the following texts, some of the words have been replaced by blanks numbered 1 through 40. First, read the complete texts in order to understand it. Then reread it and choose the correct word to fill each blank from the answer sheet. Mark your answers by circling your choice on the answer sheet, not by filling in the blanks in the text. The texts in simplified and traditional Chinese characters are provided.

Text 1

Simplified

有一(1) 人在路上遇到一个神仙（fairy），这个神仙以前是他(2) 朋友。他告(3) 神仙，现在他的情况越(4) 越不如从(5)，生活很困(6)。神仙听完(7) 的话，用手一(8) 路旁的一块小石头，那块石头立刻变(9) 了金子，神仙把这块金子(10)了他。这个(11) 得到金子，还不满意。神(12) 又用手一指，把一块大石头(13) 变成了金子，又给了(14)。这个人(15) 是不满意。神仙(16) 他："怎么样你(17) 满意呢? "这个人回(18) 说："我想……我(19) 要你的(20)。"

Traditional

有一(1) 人在路上遇到一個神仙（fairy），這個神仙以前是他(2) 朋友。他告(3) 神仙，現在他的情況越(4) 越不如從(5)，生活很困(6)。神仙聽完(7) 的話，用手一(8) 路旁的一塊小石頭，那塊石頭立刻變(9) 了金子，神仙把這塊金子(10)了他。這個(11) 得到金子，還不滿意。神(12) 又用手一指，把一塊大石頭(13) 變成了金子，又給了(14)。這個人(15) 是不滿意。神仙(16) 他："怎麼樣你(17) 滿意呢? "這個人回(18) 說："我想……我(19) 要你的(20)。"

Text 2

Simplified

有一个老人和他的儿子赶着一头驴（donkey）(21) 集市上去卖，没走(22) 远，遇到一群人。其(23)

一个姑娘说："看! (24) 两个人真傻，有驴不骑，倒要走路。"老人(25)

到这些话，就让儿子骑到驴背上。过了一会儿，遇到一位(26) 大爷，老大爷说："这个年轻人，(27)

老人太不尊敬，老人走路，(28) 却骑驴! " 于是，父亲 (29) 儿子下来，自己骑了上去。

又走了一会儿，前面来了一个女人，她说："你这个(30) 真狠心，自己骑驴，却让这个可怜(31)

孩子跟在驴后面走。"老人只好也(32)

他的儿子拉上了驴，两人一起骑驴。刚走一会儿，又遇到一个行人，行人说："(33) 个人骑一头驴，它能(34)

得了吗? "这下老人可为难(35)，他只好把驴腿捆起来，(36) 儿子一起抬着驴。驴可(37) 干了，他们过一座

(38) 的时候，这头驴挣脱了绳子，父子俩(39) 这头驴都掉到河(40) 去了。

Traditional

有一個老人和他的兒子趕著一頭驢（donkey）(21) 集市上去賣，沒走(22) 遠，遇到一群人。其(23)

一個姑娘說："看! (24) 兩個人真傻，有驢不騎，倒要走路。"老人(25)

到這些話，就讓儿子騎到驢背上。過了一會兒，遇到一位(26) 大爺，老大爺說："這個年輕人，(27)

老人太不尊敬，老人走路，(28) 卻騎驢! " 於是，父親 (29) 兒子下來，自己騎了上去。

又走了一會兒，前面來了一個女人，她說："你這個(30) 真狠心，自己騎驢，卻讓這個可憐(31)

孩子跟在驢後面走。"老人只好也(32)

他的兒子拉上了驢，两人一起騎驢。剛走一會兒，又遇到一個行人，行人說：" (33) 個人騎一頭驢，它能(34)

得了嗎? "這下老人可爲難(35)，他只好把驢腿捆起來，(36) 兒子一起擡著驢。驢可(37) 干了，他们過一座

(38) 的時候，這頭驢掙脫了繩子，父子倆(39) 這頭驢都掉到河(40) 去了。

Answer key: Cloze Test

The answer key is in the first column. Traditional characters are provided on the right slide if they are different from simplified characters. The order of the four options were randomized in the real test.

Text1	Correct	Incorrect	Incorrect	Incorrect
1	个/個	穷/窮	各	条/條
2	的	虎	得	地
3	诉/訴	去	这/這	吓/嚇
4	来/來	得	跳	穷/窮
5	前	来/來	这/這	的
6	难/難	了	呢	的
7	他	金	人	我
8	指	个/個	条/條	下
9	成	代	得	化
10	给/給	看	送	把
11	人	块/塊	子	神
12	仙	还/還	人	他
13	也	再	更	还/還
14	他	我	神	仙
15	还/還	又	真	也
16	问/問	骂/罵	说/説	打
17	才	就	不	会/會
18	答	打	他	来/來
19	想	不	可	才
20	手	指	金	爱/愛
Text2				
21	到	在	从/從	去
22	多	路	长/長	大
23	中	他	实/實	余/餘

24	这/這	你	有	他
25	听/聽	说/說	想	见/見
26	老	个/個	好	他
27	对/對	每	一	给/給
28	他	还/還	她	我
29	叫	和	使	给/給
30	人	父	话/話	爸
31	的	多	得	儿/兒
32	把	让/讓	给/給	说/說
33	两/兩	那	这/這	一
34	受	坐	跑	到
35	了	看	想	听/聽
36	跟	他	使	给/給
37	不	太	捆	要
38	桥/橋	山	位	路
39	跟	把	为/為	带/帶
40	里/裏	出	回	边/邊

Appendix D: Target, Competitor, and Distracter Words in Filler Items

	/tʂ^{h}ā/叉 ‘fork’	/tʂ^{h}á/茶 ‘tea’	/ɕɥě/雪 ‘snow’	/ɕiàu/ 笑 ‘smile’
	/tʂuāŋ/窗 ‘window’	/tʂuáŋ/床 ‘bed’	/mǎ/马 ‘horse’	/mài/麦 ‘wheat’
T1-T2	/ʂī/ 狮 ‘lion’	/ʂí /食 ‘food’	/xǔo/火 ‘fire’	/xàu/号 ‘trumpet’
	/ɕjōŋ/胸 ‘chest’	/ɕjóŋ/熊 ‘bear’	/tɕjǎn/剪 ‘scissor’	/mjàn/面 ‘noodle’
	/tɕ^{h}jāŋ/枪 ‘gun’	/tɕ^{h}jáŋ/墙 ‘well’	/kǔ/鼓 ‘drum’	/k^{h}ù/裤 ‘pants’
	/tɕ^{h}ī/妻 ‘wife’	/tɕ^{h}í/旗 ‘flag’	/wǎn/碗 ‘bowl’	/tsuàn/钻 ‘drill’
	/ɕīŋ/星 ‘star’	/ɕìŋ/杏 ‘apricot’	/tɕ^{h}jáu/桥 ‘bridge’	/tɕjǎu/脚 ‘feet’
	/ʂān/山 ‘mountain’	/ʂàn/扇 ‘fan’	/xóu/猴 ‘monkey’	/tʂǒu/肘 ‘elbow’
T1-T4	/xuā/花 ‘flower’	/xuà/画 ‘painting’	/nán/男 ‘male’	/njǎu/鸟 ‘bird’
	/ɕīn/心 ‘heart’	/ɕìn/ 信 ‘letter’	/ʂə́/蛇 ‘snake’	/ʂǒu/手 ‘hand’
	/fēŋ/风 ‘wind’	/fèŋ/凤 ‘phoenix’	/njoú/牛 ‘cattle’	/nǎu/脑 ‘brain’
	/ɕjāŋ/箱 ‘case’	/ɕjàŋ/象 ‘elephant’	/lóu/楼 ‘building’	/kǒu/狗 ‘dog’
	/fǔ/斧 ‘ax’	/fù/父 ‘father’	/tʂū/猪 ‘pig’	/tʂú/竹 bamboo’
	/jǎn/眼 ‘eye’	/jàn/燕 ‘swallow’	/tɕ^{h}jōu/秋 autumn’	/tɕ^{h}jóu/球 ‘ball’
	/tjǎn/点 ‘dot’	/tjàn/电 ‘electricity’	/t^{h}āŋ/汤 ‘soup’	/t^{h}áŋ/糖 ‘candy’
	/mjǎu/秒 ‘second’	/mjàu/庙 ‘temple’	/kuō/锅 ‘pot’	/kuó/国 ‘country’
	/kuǐ/鬼 ‘ghost’	/kuì/柜 ‘closet’	/t^{h}ī/梯 ‘ladder’	/t^{h}í/蹄 ‘hoof’
T3-T4	/mǐ/米 ‘rice’	/mì/蜜 ‘honey’	/pɤ̄/波	/pɤ́/脖 ‘neck’

			‘wave’	
	/luǎn/卵 ‘eggs’	/ljàn/链 ‘chain’	/fān/帆 ‘sail’	/lán/篮 ‘basket’
	/kuǎn/管 ‘tube’	/kùn/棍 ‘stick’	/tʂʰə̄/车 ‘car’	/tʂʰóŋ/虫 ‘bug’
	/suǒ/锁 ‘locker’	/xuò/货 ‘goods’	/kōŋ/弓 bow’	/lóŋ/龙 dragon’
	/pjǎu/ 表 ‘watch’	/pʰjàu/ 票 ‘frog’	/tāu/刀 ‘knife’	/tí/笛 ‘flute’
	/jǐ/椅 ‘chair’	/tɕʰì/气 ‘air’	/ʂū/书 ‘book’	/ʂáu/勺 ‘spoon’
	/pǎn/板 ‘board’	/pèi/被 ‘back’	/kɤ̄/歌 ‘song’	/xɤ́/盒 ‘box’
	/lján/帘 ‘curtain’	/ljǎn/脸 ‘face’	/māu/猫 ‘cat’	/màu/帽 ‘hat’
	/ljóu/流 ‘stream’	/ljǒu/柳 ‘willow’	/tɕjān/肩 ‘shoulder’	/tɕjàn/剑 ‘sword’
	/tʂʰí/池 ‘pond’	/tʂʰǐ/尺 ‘ruler’	/tēŋ/灯 ‘light’	/tèŋ/凳 ‘stool’
	/tʂʰáŋ/肠 ‘intestines’	/tʂʰǎŋ/厂 ‘factory’	/pīŋ/兵 ‘soldier’	/pìŋ/病 ‘illness’
	/xái/孩 ‘child’	/xǎi/海 ‘sea’	/pāu/包 ‘bag’	/pàu/豹 ‘leopard’
T2-T3	/jóu/油 ‘oil’	/jǒu/友 ‘friend’	/pjān/鞭 ‘whip’	/pjàn/辫 ‘braid’
	/tʰíŋ/亭 ‘pavillion’	/tʰǒŋ/筒 ‘barrel’	/ tʂī/枝 ‘branch’	/tʂʰì/翅 ‘wing’
	/tʂʰuí/锤 ‘hammer’	/tuǐ/腿 ‘leg’	/pēi/杯 ‘cup’	/lèi/泪 ‘tear’
	/mén/门 ‘door’	/mǔ/母 ‘mother’	/pʰɤ̄/坡 ‘slope’	/pʰàu/炮 ‘cannon’
	/tjé/ 蝶 ‘butterfly’	/tɕjě/ 姐 ‘older sister’	/kū/菇 ‘mushroom’	/lù/鹿 ‘deer’
	/pʰán/盘 ‘plate’	/sǎn/伞 ‘umbrella’	/tɕī/鸡 ‘chicken’	/tɕià/架 ‘shelf’
	/wán/丸 ‘meatball’	/wǔ/舞 ‘dance’	/tīŋ/钉 ‘nail’	/tòu/豆 ‘bean’

Appendix E: Listeners' Proportions of Target, Competitor, and Distracter Fixations in Experiment 1

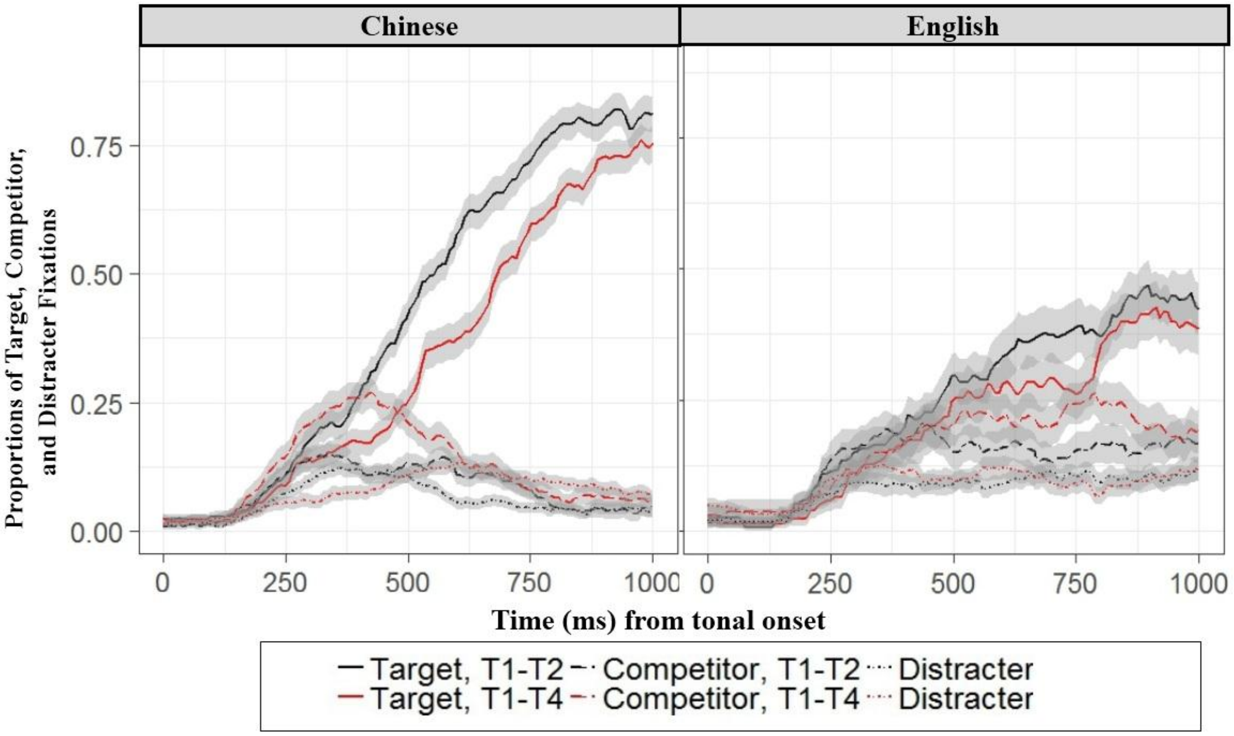

Figure 14: Chinese (left) and English (right) listeners' proportions of target, competitor, and distracter fixations in the T1-T2 (black) and T1-T4 (red) conditions for the first 1,000 ms

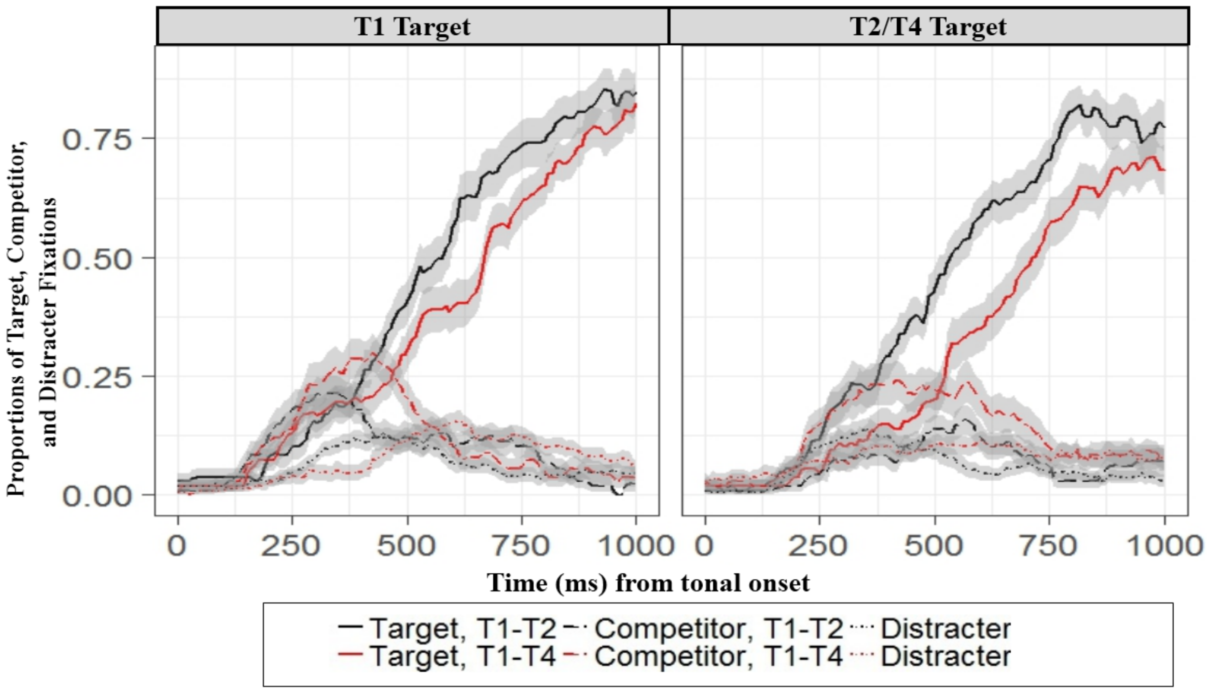

Figure 15: Chinese listeners' proportions of target, competitor, and distracter fixations in the T1-T2 (black) and T1-T4 (red) conditions for T1 Target (left) and T2/T4 Target (right) items for the first 1,000 ms

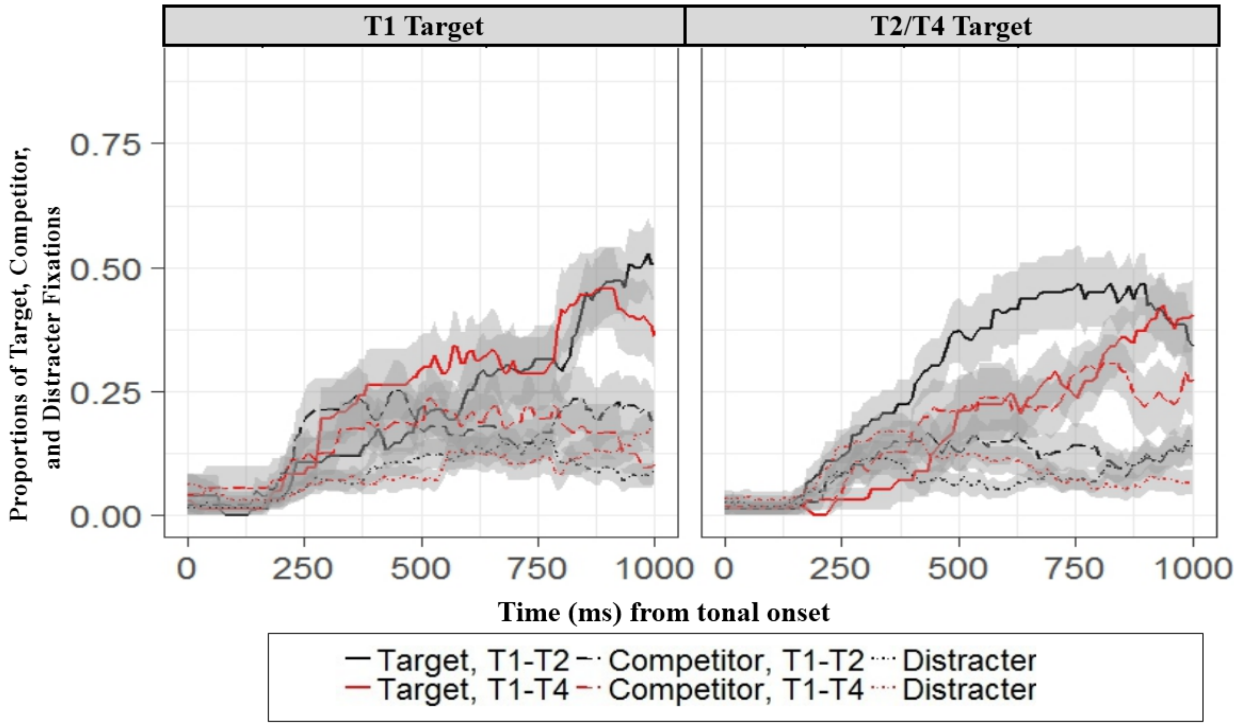

Figure 16: English listeners' proportions of target, competitor, and distracter fixations in the T1-T2 (black) and T1-T4 (red) conditions for T1 Target (left) and T2/T4 Target (right) items for the first 1,000 ms

Appendix F: Listeners' Proportions of Target, Competitor, and Distracter Fixations in Experiment 2

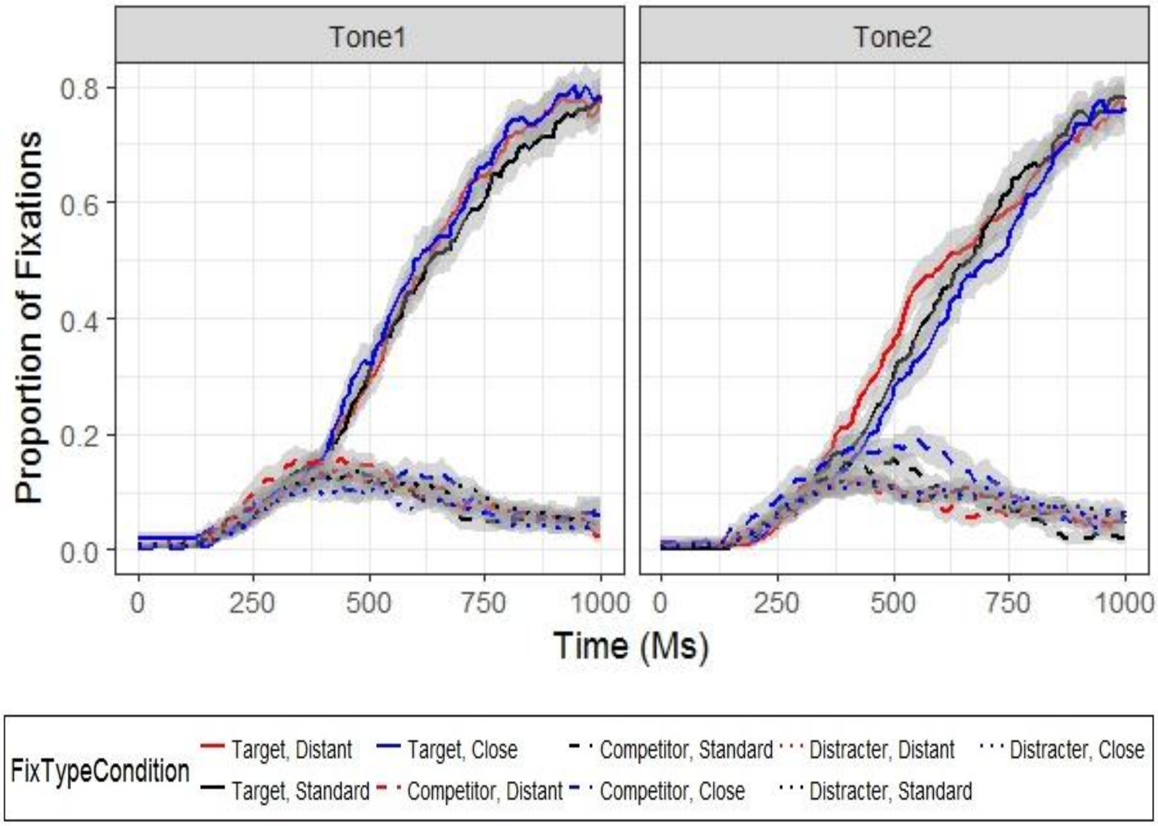

Figure 17: Chinese listeners' proportions of target, competitor, and distracter fixations in the standard, close and distant conditions for the level and contour tones in the first 800 ms; the shaded area represents one standard error above and below the mean

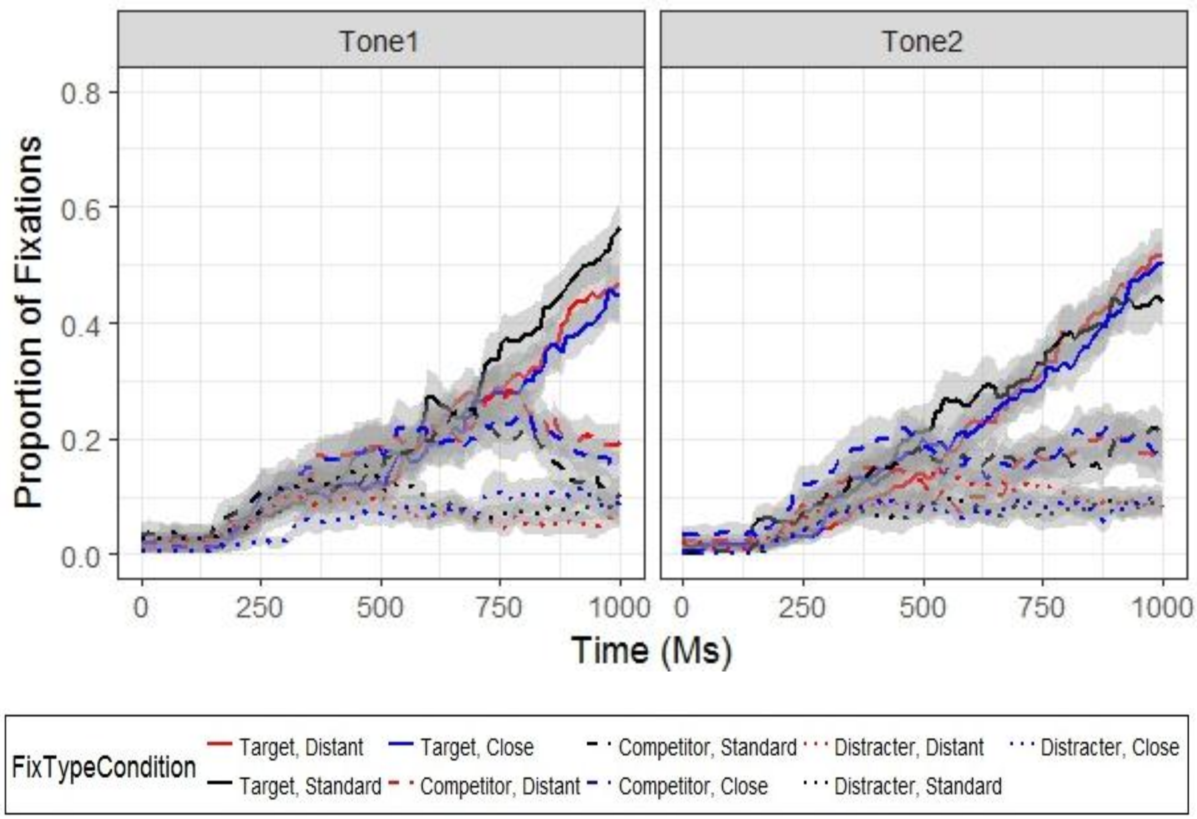

Figure 18: English listeners' proportions of target, competitor, and distracter fixations in the standard, close and distant conditions for the level and contour tones in the first 800 ms; the shaded area represents one standard error above and below the mean