THE USE OF SEGMENTAL AND SUPRASEGMENTAL INFORMATION IN LEXICAL ACCESS: A FIRST- AND SECOND-LANGUAGE CHINESE INVESTIGATION

By
Katrina Connell

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.

Chairperson Annie Tremblay

Chairperson Jie Zhang

Allard Jongman

Joan Sereno

Michael Vitevitch

Date Defended: 21 June 2017

The Dissertation Committee for Katrina Connell

certifies that this is the approved version of the following dissertation:

THE USE OF SEGMENTAL AND SUPRASEGMENTAL INFORMATION IN LEXICAL ACCESS: A FIRST- AND SECOND-LANGUAGE CHINESE INVESTIGATION

Chairperson Annie Tremblay

Chairperson Jie Zhang

Date approved: 25th July 2017

Abstract

The present study investigated first language (L1) and second language (L2) Chinese categorization of tones and segments and use of tones and segments in lexical access. Previous research has shown that English listeners rely more on pitch height than pitch direction when perceiving lexical tones; however, it remains unclear if this superior use of pitch height aids English-speaking learners of Chinese in identifying the tones of Chinese that differ in initial pitch height. The present study aimed to investigate this issue to determine whether this pitch height advantage aids English-speaking Chinese learners in identifying the tones of Chinese by looking at the time course of categorization of Chinese tones that differed in initial pitch as well as segments. A norming study was first conducted to investigate the duration of acoustic input needed to hear tone and segment (rime) distinctions. In a gated AX discrimination task, native Chinese listeners and naïve English listeners heard increasingly large fragments of tonal pairs and segmental pairs that varied in the expected disambiguation point. The results of this norming study were used to select tonal and segmental stimulus pairs were controlled (as best as is possible) for the disambiguation timing in the next two experiments.

Experiment 1 investigated the time course of categorization of tones and segments using a forced-choice gating task designed to tap into listeners' identification of fragment categories taken from syllables that differ only in tones or only in segments. Native Chinese listeners and L1-English L2-Chinese listeners heard a single fragment of a Chinese word and identified either the tone or the rime of the heard fragment from two presented options. The results showed that the segmental contrasts had higher accuracy than tonal contrasts for both groups. The L2-

Chinese listeners performed comparably to the native listeners on both tonal and segmental contrasts, and L2 Chinese listeners showed no advantage over native listeners.

The second goal of this study was to investigate the time course of the use of tones and segments in lexical access. Previous work has shown that native Chinese listeners use tones and segments simultaneously in lexical access. Previous work on how second language learners of Chinese use tones in lexical access compared to segments showed that tones and segments are used at the same time; however, work in the segmental domain suggest that this should not be the case, and learners should struggle to use the new tones in online lexical access. As such, this work aimed to reinvestigate the timing of use of tones and segments in second language Chinese, as well as to compare learners' use of tones and segments to native listeners with a highly time-sensitive measure: visual-world eye-tracking.

Experiment 2 investigated the time course of use of tones and segments in online spoken word recognition for L1 and L2 groups. The same segmental and tonal pairs used in Experiment 1 were used in a visual-world eye-tracking experiment. Native Chinese listeners and L1-English L2-Chinese listeners saw two pairs of words displayed as corresponding images: one tonal pair and one segmental pair. Eye movements were recorded as participants heard a single target word in isolation and clicked on the corresponding picture. The eye movement data revealed that native Chinese listeners use tones and segments to begin constraining the lexical search at approximately the same time, and tonal information constrained the search more rapidly than did segments. The L2-Chinese learners showed segmental use comparable to that of native listeners; however, their tonal use was delayed by approximately 100 ms. In terms of speed, learners also showed more rapid use of tones in constraining the lexical search, although tones and segments were used to constrain the lexical search more slowly than they did for native listeners.

These results are discussed in relation to recent L1 studies on lexical access of tones and segments and computational modeling of suprasegmental information. The results of this research is in line with previous work that showed tones and segments are used to constrain lexical access simultaneously; however, the current work does not support the conclusion that tones and segments are used in the same way, with tones constraining the lexical search faster than segments. It is suggested that the cause of this tone speed advantage is the number of competitors removed from competition when the processor is certain of a tone as opposed to certain of a segment or even rime.

The present results also speak to the literature on the use of segmental and suprasegmental information in a second language and suggest that the timing of use of different cues to lexical identity is dependent on if that cue is used in the L1, since segments were processed at the same time as native speakers while tones were delayed. Speed of use seems to be independent of whether or not it is used in the L1, with both tones and segments being processed slower overall compared to native listeners.

Acknowledgements

A dissertation cannot be done alone, and so I have many people that I wish to thank for their help and support. First and foremost, I would like to thank my advisors, Annie Tremblay and Jie Zhang. Without their guidance and support I never would have made it here. They taught me not only how to conduct good, solid research, but also what it means to be a great professor and advisor. Thank you for always sticking with me and never giving up on me.

Additionally, I would like to thank my committee members, Joan Sereno, Allard Jongman, and Michael Vitevitch. Your feedback at every stage of this work strengthened this dissertation, and I am grateful for all of the time that you put in to help me improve my research.

I would also like to thank all of the student members of the 850 course for giving their feedback and suggestions on my work. Thank you for all that you did to teach and help me through these last four years.

Of the graduate students, there are two I would like to thank specifically. First, I would like to thank Kate Coughlin for always being the person I could come to for help and for being a wonderful friend. Second, I would like to thank Lauren Covey. From our Downton Abbey Sunday nights to celebrating my graduation, you have been the best friend anyone could ask for. Thank you for always being there for me, encouraging me, and keeping me sane. I truly would not have been able to do this without you.

I would also like to thank the KU Department of Linguistics for funding me during my time at KU, in particular for a semester of fellowship funding which allowed me to focus on writing my dissertation.

I am also grateful to the Fulbright organization for awarding me a Fulbright-Hays Doctoral Dissertation Abroad Fellowship, which allowed me to conduct my research in China.

While in Beijing, China I had the great fortune of working in the Psychology Department of Beijing University under the supervision of Xiaolin Zhou. I am grateful to Dr. Zhou and all of his students for hosting me during my time in China and for making me feel welcomed.

Last but not least, I would like to thank my family. My entire family has been so supportive of me this whole time, but no one has been as supportive as my parents, Susan and Robert Connell. They have supported me in following my dreams from the very beginning, even if it took me far from home. They have gone through so much on my account, and I could never thank them enough. You are the reason I have made it this far.

There is one additional person I would like to thank, my late grandfather, Robert Connell, Sr. His love of and passion for education lead to the amazing opportunity to attend college, where I discovered my love of language and linguistics. While I lost him long before I went to college, his love of learning has stayed with me my whole life and brought me to where I am today. He used to recite a poem to me called ‘It Couldn’t Be Done’ by Edgar Albert Guest, and that poem has given me strength so many times throughout my graduate career. So Pop-pop, while some may have said that it couldn’t be done, I just started to sing as I tackled the thing that could not be done and I did it. This one is for you Pop-pop.

Table of Contents

List of Tables

List of Figures

CHAPTER 1: INTRODUCTION

Auditory word recognition is the process by which listeners use acoustic information available in the speech signal to locate the intended word (and meaning) of the speaker in the mental lexicon. More specifically, lexical access is the process of using this information to locate the intended word of a speaker in the mental lexicon, to the exclusion of words with similar sounds and/or meanings. Segmental cues (e.g., consonants, vowels) and suprasegmental cues (e.g., tone, stress, prosody) can both signal lexical differences in language.

To illustrate, Mandarin Chinese (henceforth Chinese) differs from English in that both tonal and segmental information contribute to lexical identity: The word *ma* in Chinese can have four meanings depending on its tone (e.g., Tone 1 *mā* 'mother' [level tone] vs. Tone 2 *má* 'hemp' [rising tone] vs. Tone 3 *mǎ* 'horse' [dipping tone] vs. Tone 4 *mà* 'to scold' [falling tone]). Visual depiction of the tones and their corresponding tone numbers are presented in Figure 1 below on a 5-point pitch scale (Chao, 1930), with 5 being the highest pitch point and 1 being the lowest (figure adapted from Li, 2002).

Additionally, the word *mā* [ma1] 'mother' (examples are provided in Pinyin, followed by their phonetic transcription and translation when segments and tones are represented) can contrast segmentally with the word *mī* 'microphone,' in the change of the vowel, or with *bā* [pa1] 'eight,' in the change of the initial consonant. Research has shown that suprasegmental information, including not only tonal information, but also stress and prosody, is important for spoken word recognition (e.g., Cooper, Cutler, & Wales, 2002; Cutler & Chen, 1995; Reinisch, Jesse, & McQueen, 2010). How segmental information and suprasegmental information are integrated into the word recognition system is poorly understood, however.

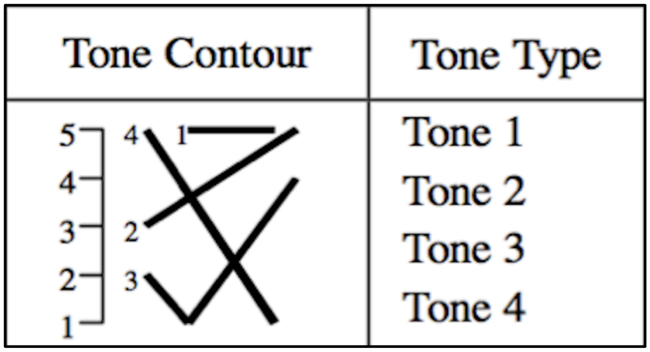

Figure 1: Mandarin Chinese tonal contours example with corresponding tone numbers (adapted from Li, 2002)

Research on how native Chinese listeners categorize the tones of Chinese has consistently shown that tones are disadvantaged compared to segments. For example, Taft and Chen (1992) showed that when native Chinese and Cantonese listeners judged whether two written words were homophonous in Chinese or Cantonese, respectively, they were significantly less accurate to say ‘no’ when the words differed only in tone as opposed to when they differed in a vowel. Similarly, Cutler and Chen (1997) found that, in an AX discrimination task, native Cantonese listeners made more errors when the two words heard contained a different tone (*ma3 – ma2*) than when the contrast was in segments (e.g., *ma3 – na3*).

Time-sensitive measures of the use of tones and segments in lexical access, however, all show that tones and segments are used in analogous ways. Both eye-tracking and neural imaging methods show that tones and segments are used to constrain the lexical search at the same time and to the same extent (Malins & Joanisse, 2010, 2012; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Zhao, Guo, Zhou, & Shu, 2011). These results appear to be in direct opposition to the earlier offline tasks that reported a tonal disadvantage.

The discrepancy between these two lines of studies could be attributed in part to the use of offline and online tasks, in part to the weaker versus greater emphasis of these tasks on lexical information, in part to the meta-linguistic versus unconscious measures, and in part to the different materials used in these studies. A study that directly compares the categorization of tonal and segmental information and the use of this information in lexical access with comparable materials is thus needed to elucidate the nature of this discrepancy.

In addition to native Chinese listeners' use of tones, English-speaking learners of Chinese need to learn to perceive and categorize the tones of Chinese. Existing research on non-native language acquisition of tones suggests that English listeners are able to learn to categorize the four tones of Chinese (e.g., Wang, Jongman, & Sereno, 2003; Wang, Spence, Jongman, & Sereno, 1999); however, additional research shows that naïve listeners differ from native Chinese listeners in the cues they use to perceive tones. Native English listeners use average pitch height to discriminate the tones, whereas native listeners who speak tone languages, such as Chinese and Thai, rely more on the direction of the pitch change (e.g., Gandour, 1983; Gandour & Harshman, 1978; Kaan, Wayland, Bao, & Barkley, 2007; Qin & Jongman, 2016).

What is unclear from this research, however, is whether the pitch height advantage seen for naïve English listeners can help English-speaking learners of Chinese in categorizing the tones. If these learners of Chinese can utilize this pitch height information to categorize the tones of Chinese, they may have a categorization advantage over native listeners when pitch height is the primary cue to tone identity.

In terms of how Chinese learners incorporate tonal information in online spoken word recognition, very little is known. Recently, one priming study investigated native English speaking Chinese second language learners' use of tones and segments in lexical access by varying the inter-stimulus interval (ISI) to investigate the timing of use of tones and segments (Sun, 2012). The results showed that Chinese learners use tonal and segmental information in the same way as native Chinese listeners, but that their processing was overall slower compared to native listeners, for both tones and segments (Sun, 2012). Work in the segmental domain, suggests that sounds that do not exist in the learners' native language create difficulty for lexical access (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011). By analogy, one might expect that the non-existence of tonal categories in English would create difficulty for Chinese learners' use of tonal information in online word recognition. A time-sensitive measure of the use of tones in lexical access may be able to shed more light on how English-speaking learners of Chinese use tones in online spoken word recognition.

1.1. THE STRUCTURE OF THIS DISSERTATION

The studies presented in this dissertation used comparable materials in offline and online tasks to investigate the use of tones and segments in first language (L1) and second language (L2) Chinese categorization of tones and segments and the use of this information in lexical access. These studies investigated these issues by using both an offline measure, a forced-choice gating task, and an online and highly time-sensitive measure, a visual-world eye-tracking task. If the tone disadvantage is not task-dependent, then the same tone disadvantage seen in early work should be found in both offline and online tasks. Additionally, each task investigated issues specific to L1- and L2-Chinese listeners, such as the use of pitch height cues to identify and access tones.

This dissertation will be presented in two parts. The first part reports two experiments on the perception of tones and segments in L1 and L2 Chinese listener groups. A norming study was first conducted to select materials that were controlled as closely as possible for the duration of acoustic input needed to disambiguate the tonal and segmental pairs for the two experiments conducted in this study. Experiment 1 then investigated the time course of categorization of Chinese tones and segments (i.e., rime information) for L1- and L2-Chinese groups using a forced-choice gating task. The results show that despite their heightened sensitivity to pitch height, L2 learners do not have an early advantage in categorizing the tones of Chinese over native listeners, and instead performed similarly to native listeners for both tonal and segmental identifications.

Experiment 2 investigated the use of tones and segments in the online lexical access of L1- and L2-Chinese listeners with a visual-world eye-tracking task. Participants saw displays of four items corresponding to a tonal pair and a segmental pair. Either an item from the tonal pair or an item from the segmental pair was heard as the spoken target, and eye movements to each item on the screen were analyzed as participants heard the spoken word and clicked on the word's corresponding image. The results suggest that native listeners use tones and segments at the same time in lexical access, but tones and segments may constrain lexical access in different ways. Learners show a significant delay in the use of tonal information, but, like native speakers, the results showed that tones and segments might constrain the L2 lexical search in different ways. The results will be discussed in relation to previous L1- and L2-Chinese studies as well as current models of lexical access including tones.

Part I

CHAPTER 2: CATEGORIZATION OF TONES AND SEGMENTS

2.1. INTRODUCTION

Recent studies on the L2 acquisition of Chinese have shown that although English-speaking learners of Chinese initially struggle to learn the four tones of Chinese, with training they can improve their identification of the tones up to about 90% accuracy (Li, 2016; Wang et al., 2003; Wang et al., 1999). Additionally, perceptual work has shown that naïve English listeners rely on different aspects of the tones to distinguish them than native listeners who speak tone languages such as Mandarin, Cantonese, Yoruba and Thai. Studies using both dissimilarity ratings as well as electroencephalography (EEG) have shown that English listeners rely more on the average pitch height of the tone, whereas native listeners of tone languages such as Cantonese, Yoruba, Thai and Chinese, rely more on the direction and slope of the tone (Cantonese, Yoruba and Thai: Gandour & Harshman, 1978; Mandarin Chinese: Kaan et al., 2007).

2.1. NATIVE CHINESE LISTENERS VS. NAÏVE LISTENERS

There exists a large body of literature on the categorization of tones and segments across languages. The present study investigates possible differences in the categorization of tonal and segmental information and in the use of this information in lexical access. Because the present study seeks to directly compare tones to segments, it focuses on listeners' categorization of the rime portion of the syllable (see Section 3.1.5 for a justification of this choice). We therefore begin with a discussion of the research on vowel categorization.

Early work suggested that vowels were perceived much less categorically than consonants (Fry, Abramson, Eimas, & Liberman, 1962; Pisoni, 1973), in that although their identification was somewhat categorical, their discrimination still remained well above chance for within-category pairs. Fry et al. (1962) tested a range of vowels using synthetically produced /ɪ, ɛ, æ/ continua. Participants took part in both discrimination and identification tasks.[1] The discrimination task took the form of an AXB task, where participants were asked to decide if the stimulus X was identical to stimulus A or B. A and B stimuli were chosen so that they fell one, two, or three steps from each other at different points on the continuum. The identification task used the same stimuli as the discrimination task, but this time participants were asked to label each of the stimuli as /ɪ/, /ɛ/, or /æ/.[2] These results were compared to existing data on /b, d, g/ continua. The discrimination results show no peaks, with discrimination accuracy above chance for all possible pairings and steps. The identification results were semi-categorical, that is, much less categorical than for stop consonants in that the slope was shallower than for consonants, but not fully linear.

Pisoni (1973) later found slightly more categorical results for vowels. The author used a synthetically produced long and short /i-ɪ/ continuum and compared this vowel continuum to /bæ/-/dæ/ and /ba/-/pa/ continua. Native English listeners participated in discrimination and identification tasks. In the discrimination task, participants heard two stimuli two steps apart on the continuum, and made same-different judgments. In the identification task, participants heard a single stimulus and made a forced-choice identification specific to the continuum being heard (e.g., /b/ or /d/, /i/ or /ɪ/, etc.). The discrimination results showed a clear peak at the boundary, but discrimination of the stimuli at the end points remained well above chance. This is a clear departure from the stop consonant results, for which discrimination at the end points was at

---

[1] The native language of the participants was not reported; since the data were collected at the University of Connecticut, it is likely they were native English listeners.
[2] The original paper states that the participants were instructed to "label each stimulus as /ɪ/, /ɛ/, or /æ/" (Fry et al., 1962, p. 177), and makes no mention of whether participants were told what each phonetic symbol stood for, or whether English letters were used.

chance. The results of the identification task showed that both vowel continua were less categorical than the stop consonants; however, the results appeared much more categorical than those of Fry et al. (1962). These results of semi-categorical perception of vowels have been replicated in other behavioral work (Fujisaki & Kawashima, 1968, August; Fujisaki & Kawashima, 1969; Pisoni, 1971), as well as in neural imaging research using the M100 magnetoencephalography (MEG) component, which is sensitive to frequency (here F1 and F2) (Roberts, Flagg, & Gage, 2004).

Like with vowels, early work on Thai suggested that lexical tones are not perceived categorically, and are instead perceived in a gradient fashion (Abramson, 1979). Abramson (1979) used synthesized syllables with a 16-step continuum from the high to mid to low level tones of Thai. Participants completed discrimination and identification tasks.[3] The discrimination task took the form of a four-interval forced-choice task, where participants heard two pairs of stimuli: one identical pair and one pair with the stimuli differing along the continuum by one or two steps. The task was to choose the pair that differed. Identification data were collected as well, though how it was conducted is unclear from the paper. The discrimination results show above-chance accuracy on the within-category pairs and no clear discrimination peaks at the end points of the continuum. The identification results showed gradient shifts from high to mid to low tone identification. These results suggested that the Thai tones were not perceived categorically or even semi-categorically.

However, Hallé, Chang, and Best (2004) obtained different results for Chinese tones (Taiwan Mandarin): They found evidence of semi-categorical perception of tones similar to that found for vowels (Fujisaki & Kawashima, 1968, August; Fujisaki & Kawashima, 1969; Pisoni, 1971, 1973; Roberts et al., 2004), with categorical identification, but discrimination remaining

[3] The participants' native language was not reported, though it is assumed that they were native speakers of Thai.

above chance for within-category pairs. The authors tested native Taiwan Mandarin (tone language) and French (non-tone language) listeners on their categorization of Chinese tones. The stimulus tone pairs used were the Chinese tones of Tone 1 – Tone 2, Tone 2 – Tone 4 and Tone 3 – Tone 4 on three syllables, [pʰa], [pʰi] and [kʰuo]. Tone continua were created by synthesizing six intermediate steps between each of the tones in the pair. In Experiment 1, only the native Taiwan Mandarin listeners participated in discrimination and identification tasks. The discrimination task took the form of an AXB two-step task. In the identification task, participants were presented with a single stimulus, and were asked to make a forced-choice identification between two written characters that differed only in tone. The results of the identification task show steep slopes for all tone continua at the boundary. The discrimination results showed a weak peak at the boundary, with within-category discrimination still above chance. These results are similar to the pattern found for vowels (Fujisaki & Kawashima, 1968, August; Fujisaki & Kawashima, 1969; Pisoni, 1971, 1973; Roberts et al., 2004).

In Experiment 2, both native Taiwan Mandarin and native French listeners' participated in a task investigating the identification of Chinese tones using the same stimuli as Experiemnt 1 (with the exception of one syllable [kʰuo] to keep the length of the experiment reasonable). Since naïve French listeners could not label the tones of Chinese without training, an AXB "identification" task was used (Best, Morrongiello, & Robson, 1981).[4] In an AXB discrimination task, the A and B tokens are a number of steps from each other, whereas in the AXB "identification" task, they corresponded to the endpoints on the tone continua. In this way, the authors could have the naïve French listeners identify the intermediate stimuli as either endpoint, without having to train them how to identify the tones. The "identification" task results showed

[4] This task was discussed as an identification task in the original paper, even though AXB is classically discussed as a discrimination task.

that the slope at the boundary was steeper for native Taiwan Mandarin listeners compared to French listeners.

Experiment 3 tested both native Taiwan Mandarin and native French listeners' discrimination of Chinese tones using the same materials as in Experiment 2. French listeners showed no peak, indicating that discrimination at the tone boundary was not easier than that within-category. Native Taiwan Mandarin listeners, on the other hand, showed discrimination similar to that shown in Expirment 1, with a clear peak at the boundary, but with discrimination remaining above chance for within-category discriminations.

The results of these experiments revealed that native Taiwan Mandarin listeners' perception of the tones was more categorical (i.e., with better discrimination across categories than within and a more categorical "identification" of the boundary) than that of French listeners, with French listeners showing more psychophysical perception of the tones.

Similarly, Chang, Halle, Best, and Abramson (2008) also found that native tone-language listeners (Cantonese, Thai, Vietnamese) showed *more* categorical perception of tones than naïve listeners who did not know a tone language with a larger and more varied tone language groups. The authors tested native Cantonese, Thai, Vietnamese, Japanese, and English listeners on their categorization of Chinese tones. None of the participants had any knowledge of Chinese. Of these language groups, the authors describe only Japanese and English to be non-tonal languages. Participants participated in both discrimination and identification tasks using tone continua between Tone 1 and Tone 2, Tone 2 and Tone 4, and Tone 3 and Tone 4. The authors had participants complete AXB tasks both as discrimination and "identification" (described above) tasks so that all participants could complete the task without the need to be trained on the tones. In the discrimination task, the A and B tokens were two steps from each other, whereas in the “identification” task, they corresponded to the endpoints on the tone continua. The results of the two tasks revealed that tone-language listeners’ perception of the tones was more categorical (i.e., with better discrimination across categories than within and a more categorical “identification” of the boundary) than that of non-tone-language listeners.

Sun and Huang (2012) replicated these results with naïve English listeners and Taiwanese Southern Min listeners. Two tone continua were used: one ranging from a high-level tone to a mid-level tone, and the other ranging from the high-level tone to a high-falling tone. Participants completed an AX discrimination task with stimuli from these ranges. For native listeners, the results showed better discrimination of pairs spanning the tone categories than of those within the tone categories for native listeners; by contrast, for English listeners, the results showed no such pattern. Additionally, Sun and Huang (2012) found that native listeners perceived contour-level tone pairs (high-falling tone vs. high-level tone) more categorically than level-level tone (high-level vs. mid-level). This is likely the cause of the discrepancy between the results of Abramson’s (1979) and the original conclusion that tones were not perceived categorically, since the author only tested a continuum of level tones, which have been shown to be perceived less categorically. From this, we can conclude that contour-level tones are perceived semi-categorically.

Additionally, several other studies have also found that speakers of lexical tone languages such as Mandarin, Taiwan Mandarin, Cantonese, Thai, and Vietnamese show stronger categorization of tones (including tones of tone languages not spoken by the participants) than speakers of non-tone languages such as Japanese, English, and French (Chan, Chuang, & Wang, 1975; Chang et al., 2008; Hallé et al., 2004; Huang & Johnson, 2010; Stagray & Downs, 1993; Sun & Huang, 2012). For example, it has been shown that Chinese listeners perform worse than

English listeners when discriminating frequency changes in a level tone (Stagray & Downs, 1993). Stagray and Downs (1993) had participants make same-different judgments on level tones of varying frequency. The results showed that English listeners were much more sensitive to small frequency changes than Chinese listeners, who needed larger differences to register that the two tokens were different. The authors argue that this effect comes from Chinese listeners perceiving pitch changes between linguistic categories.[5] The native listeners thus did not register the small changes. Since naïve English listeners did not have these linguistic categories, they were able to perceive more fined-grained pitch differences. The studies discussed in this section found that non-tone language listeners respond to tones in a more psychophysical way than native tone-language listeners, showing no categorical perception of the tones regardless of the tone pairs used, including perception of the Mandarin tone pairs of Tone 1 – Tone 2, Tone 2 – Tone 4, and Tone 4 – Tone 3 (Chang et al., 2008; Hallé et al., 2004; Sun & Huang, 2012; Xu, Gandour, & Francis, 2006).

The results discussed above clearly show that native Chinese listeners show semi-categorical perception of lexical tones in contour-level pairs. However, tonal information is only part of the information in the acoustic signal, and never arrives in natural speech without segmental information. When comparing how tones are used in relation to segments, a clear pattern emerges. Taft and Chen (1992) investigated native Chinese and Cantonese listeners' accuracy and speed at judging whether two written words were homophonous in Chinese or Cantonese, respectively. The results showed that both groups were significantly less accurate and slower to say 'no' when the words differed only in tone as opposed to when they differed in a vowel. Likewise, Ye and Connine (1999) showed that native Chinese listeners were slower and

[5] Though it is also possible that this is a result of using level-tone continua, which have been shown to be perceived less categorically by native listeners (Sun & Huang, 2012).

less accurate when monitoring speech for a tone than when monitoring speech for a segment. These results were interpreted as indicating that tonal information is inferior (i.e., less reliable) and more error prone than segments.

Similarly, in an AX discrimination task, the authors found that listeners responded more slowly and made more errors when the two words heard contained a different tone (*ma3 – ma2*) than when the contrast was in segments (e.g., *ma3 – na3*). Based on these results, the authors suggest that "the kind of perceptual decision involved in tone processing, even in its simplest form, requires a certain accumulation of evidence and may be more difficult than perceptual decisions about vowels" (Cutler & Chen, 1997, p. 177). The authors cite results from work conducted by Ritsma, Cardozo, Domburg, and Neelen (1965), who showed a direct positive relationship between the length of a complex tone stimulus and improvement in the accuracy of matching pitches. Cutler and Chen (1997) also mention the results of Robinson and Patterson (1995), who found that monophthong vowel identity is identified with less information than the note (i.e., the pitch value on a musical scale) of a vowel. They performed a gating study by increasing the number of cycles of the periodic waveform of the vowel and had participants report either the vowel identity or vowel note. The results showed that the vowel identity could be identified with much less information (duration) than the vowel note.

Taken together, these studies show that native Chinese listeners' perception of tones is semi-categorical in a similar way that vowels are perceived semi-categorically. Additionally, native Chinese listeners show a more categorical perception of tones than non-tone-language listeners, including English listeners. Furthermore, tones appear to be disadvantaged compared to segments, with tonal identification showing more errors and longer response times than segmental identification.

2.1.1. Second language learners of Chinese

From the studies in the previous section, it can be concluded that naïve non-tone-language listeners (e.g., Japanese, English, French) do not categorize tones as native listeners of tone languages do. In each case, the authors argued that tone-language listeners' categorization ability stems from their experience with the tone language; since naïve listeners do not have this experience, they would not have formed the appropriate tonal categories, thus relying on psychophysical aspects of the tones such as slight changes in pitch to perceive them. One question that arises from this research is whether prolonged exposure to a tone language, such as Chinese in an L2 setting, can improve non-native listeners' categorization of lexical tones.

Studies that have sought to train English listeners on the categorization of tones suggest that L1-English L2-Chinese listeners can improve their categorization of tones via multi-talker perceptual training. Wang et al. (1999) conducted a perceptual training study using the tones of Mandarin Chinese. They utilized the so-called high phonetic-variability training (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Lively, Logan, & Pisoni, 1993; Lively, Pisoni, Yamada, Tohkura, & Yamada, 1994; Logan, Lively, & Pisoni, 1991) to train English listeners with some experience with Chinese (approx. 7 months) to enhance their identification of tones. High phonetic-variability training involves training participants on the relevant sounds spoken by multiple talkers and in different phonetic contexts. In the segmental domain, this type of training has been shown to enhance the perception of sound contrasts not present in the L1: For example, if the target sounds were the /r/-/l/ contrast, this would mean training participants on /r/ and /l/ tokens spoken by multiple talkers in word-initial and word-final positions and in onset and coda clusters (e.g., Logan et al., 1991). This high phonetic variability has been shown to increase training effects as compared to training materials from the same talker or from a limited number of phonetic contexts (Bradlow et al., 1999; Bradlow et al., 1997; Lively et al., 1993; Lively et al., 1994; Logan et al., 1991).

Using a similar method, Wang et al. (1999) trained participants on the tones of Chinese spoken by multiple talkers in multiple phonetic contexts. Instead of varying the place in the syllable where the sound appears (which is not possible for tonal information, since it spans the whole syllable), the contexts were varied syllable types such as V, CV, CVN (CV + Nasal), and so forth. In the training session, participants were trained on tone pairs in blocks. Each trial would take the form of the participant hearing one item in the pair and making a forced-choice identification between the two tones in that block, with feedback provided on every trial. To measure the effectiveness of the training, this study used pre- and post-tests as well as a long-term retention post-test. The pre-, post-, and retention-tests all took the form of 100 randomized stimuli, and participants identified the tone of the stimuli from the four tones of Chinese. No feedback was given in these tests. This study also included a control group, who received no training.

The results showed that the trained group improved from 66% accuracy to almost 90% accuracy between pre- and post-tests, whereas the control group improved insignificantly from 57% to 63%. These results show that while English listeners begin with little to categorize the tones of Chinese ,with exposure and targeted training, they can learn to identify the tone categories to a relatively high degree, although arguably not to a native-level, in the post-test results.

The enhancing effect of tone training can be seen even at the cortical level. Using the training procedure of Wang et al. (1999), Wang et al. (2003) conducted pre-and post-tests using fMRI to investigate if the areas of the brain used to process tones changed with tone training.

The results showed that as participants progressed over eight training sessions, they showed an increase in left hemisphere activity and in activity area (size of the area in the brain where activation was seen) in language related regions, making them more similar to native Chinese listeners in the distribution and amount of activation.

Taken together, these studies have shown that high phonetic-variability training can improve how English listeners categorize the tones of Chinese. However, from these training studies, it remains unclear whether learners are tuning in to the same acoustic cues that native listeners use to categorize the tones: While English listeners' categorization of the tones improves, they may still differ from native listeners in the cues they rely on to categorize the tones.

2.1.2. Cues used to identify tones: native vs. naïve listeners

Previous research has shown that naïve listeners do not rely on the same cues as native listeners when perceiving Chinese tones. More specifically, both behavioral and neural imaging studies have shown that native Chinese listeners are more sensitive to pitch direction, whereas native English listeners are more sensitive to pitch height when discriminating tones (Gandour, 1983; Gandour & Harshman, 1978; Kaan et al., 2007; Qin & Jongman, 2016; Xu et al., 2006).

Gandour and Harshman (1978) began their work by investigating what cues to tonal identity native Thai, Yoruba, and English listeners rely on to tell tones apart. Both Thai and Yoruba have lexical tones (Thai has level and contour tones; Yoruba has only level tones), whereas English does not. These languages thus gave the authors the ability not only to compare listeners who speak two different tone systems (Thai and Yoruba), but also to compare them to the listeners who do not speak a tone language (English). Participants heard pairs of pitch contours superimposed on synthetic speech and rated the dissimilarity of the pitch contours on an 11-point scale from "no difference" to "extreme difference". A total of 13 tonal contours were included. Pairs of tones that differed in pitch height and/or pitch contour were selected. The stimuli included three level tones (using a 5-point scale, from 5 'highest pitch' to 1 'lowest pitch': 11, 33, 55), five falling tones (53, 31, 53-short, 31-short), and five rising tones (35, 13, 15, 35-short, 13-short).[6] The pairings included pairs with differing heights but the same slope (e.g. 53-short-31-short), as well as pairs with the same height but differing slope (e.g., 51-53). These pairings allowed the authors to investigate how onset vs. offset pitch differences in the tones influenced dissimilarity ratings, since several different tones began at the same pitch height but ended differently (e.g., 11-13), while others began differently but ended at the same pitch height (e.g., 53-33).

Using the participants' dissimilarity ratings and a multidimensional scaling procedure, Gandour and Harshman (1978) were able to determine the optimal set of cues that were shown to predict the rating scores for each group. The results showed that, overall, five dimensions best predicted the ratings: average pitch, direction (rising-level-falling), length (short-regular), extreme endpoint (where the endpoints of the tone were), and slope (whether the tone was a level tone or a contour tone). Average pitch was the most important dimension for all three groups. However, the English listeners relied on this dimension more than did either tone-language group. For direction, the results showed the opposite effect, with the tone-language listeners tested (Thai and Yoruba) weighting direction very high, and English listeners weighting it lower, with no differences between the tone-language listeners. Similarly, the dimension of slope was weighted more highly by the tone-language listeners than by the English listeners.

[6] All tones were long unless indicated to be short. The authors do not give any details on why this distinction was added.

These results were later replicated with a different multi-dimensional scaling procedure testing Cantonese, Mandarin, Taiwan Mandarin, Thai, and English listeners (Gandour, 1983). This study also showed that English listeners were more sensitive to pitch height than to pitch direction, and that they weighed pitch height more highly than any of the tone language listeners tested. These results can be attributed to the relevance of that cue in English: Pitch height distinctions are relevant to suprasegmental characteristics of English like lexical stress, with higher pitch indexing a stressed syllable (e.g., Beckman, 1986; Lieberman, 1960). This usefulness of recognizing pitch height in English is a likely cause of English listeners' heightened sensitivity to pitch height in lexical tones (e.g., Qin & Jongman, 2016).

More recent work with neural imaging has confirmed the results of Gandour and Harshman (1978) and Gandour (1983) by showing that English listeners are more sensitive to differences in tone-onset pitch height than native listeners, again showing a pitch height advantage for L1-English listeners. Kaan et al. (2007) used event-related potentials (ERP), specifically the mismatch negativity (MMN) response, to investigate how the L1 influences the discrimination of pitch contours. The MMN response is a negative deflection in the EEG waveform in response to a deviant stimulus. The authors tested native Thai-, Chinese-, and English-speaking subjects on their discrimination of Thai tones. The authors used the oddball paradigm — a task in which tokens are chosen to serve as either standards or deviants; a standard is repeated many times, and then a deviant is presented once, which is then followed by many standards and then another deviant and so on. It is this many-to-one ratio that elicits an MMN response to the deviant stimulus, but only if the deviant is actually perceived as different from the standard. If the deviant is not distinguished from the standard, no such response will be elicited.

In this task, the authors compared the Thai mid-level tone, high-rising tone, and low-falling tone. The stimuli consisted of naturally produced tokens of the Thai syllable [kʰa:] with all three tones by a single female native speaker of Thai.[7] The mid-level tone served as the standard, and the high-rising and low-falling tones served as deviants. It was expected that the larger the perceived difference between the tones, the larger the MMN response would be. Thai listeners were predicted to show MMN responses for both tone pairings since they are lexically distinctive in the L1. By contrast, Chinese and English listeners were predicted to show MMN responses only if they could discriminate the tones in pre-attentive processing, and if they did show the MMN response, that response were predicted to be smaller than the response seen by Thai listeners.

In addition to the ERP recordings, this study included a training component to see if training enhanced identification of tones differently for the two non-Thai groups. Participants first came in for an ERP recording session before any training had taken place. After the initial recording, the Chinese- and English-speaking participants returned twice for two days of training. Participants were trained on the low-falling tone to mid-level tone contrast only, which allowed the authors to test the effects of training carry-over on the high-rising to mid-level tone pair. They did not hear any tokens of the high-rising tone in the training phase.

In the introduction phase of the training, the participants heard mid-level and low-falling trials on every other trial and were instructed to press one button on even trials and another on odd trials. In this way, participants could implicitly learn to associate one tone with one button without explicit instruction. The participants were instructed to press the button that was associated with each tone as they heard it and to try to pay attention to the differences between the two tones. In the second phase of training, the low-falling and mid-level tones were not

[7] It is unclear from the paper if multiple tokens of each tone were used or if a single token of each tone was used.

presented in an every-other order and were instead randomized. The participants were instructed to press the button that matched each tone on each trial.[8] After the two days of training (approximately 30 minutes each day), the participants returned and completed a second ERP recording identical to the pre-test recording.

The results of the first ERP recording revealed that the Chinese listeners showed no MMN response to the high-rising deviant tone, whereas the English listeners showed a small MMN effect to the high-rising deviant tone, equal to that of the native Thai listeners. All groups showed an MMN effect for the low-falling tone, and the groups did not differ in the magnitude of the MMN effect. This difference in response (no/small MMN vs. MMN) between the high-rising and low-falling tones is explained by the physical differences between the onsets of the tones compared to that of the standard tone: The mid-level standard is more similar in onset pitch to the high-rising tone than to the low-falling tone. This causes a small MMN effect between the mid-level and high-rising tones, which are nearly identical at their onsets, and a larger effect between the more distinct mid-level and low-falling tones.

Training also had different effects on the Chinese and English groups. Training did not have an effect on the low-falling tone (the tone in the training) for either group. This is likely a ceiling effect, since discrimination of the mid-level and low-falling tones was already very high to begin with due to their different starting points. Although training did not seem to affect the tone pair that was trained, there were differences before and after training on the untrained high-rising tone to mid-level tone comparison. As a reminder, the high-rising deviant tone and the mid-level standard tone were similar in their initial pitch heights and differed from each other beginning with the second half of the tone. In the post-training ERP recordings, the English group showed an increased MMN response to the deviant high-rising tone compared to before

[8] The original paper does not mention whether or not feedback was given in this training phase.

training. The authors attribute this effect to English listeners' greater ability to tap into average pitch differences (Gandour & Harshman, 1978), in that they were able to tap into the small pitch difference in the early portion of the tones between the high-rising and mid-level tones, thus resulting in a larger MMN than for Chinese listeners. Chinese listeners were not as sensitive to these early pitch height differences, and so no improvement was seen for native Chinese listeners

This English listeners' advantage for using pitch height, potentially due to the use of pitch height as a cue to lexical stress, could have interesting implications for English-speaking L2 learners of Chinese. As the speech signal unfolds over time, the majority of Chinese tone pairs differ in early pitch height before they differ in pitch contour. In other words, of the six possible tone pairings, four begin with pitch values that differ drastically (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 2 – Tone 4, and Tone 3 – Tone 4). If English-speaking L2 learners of Chinese can make use of this early pitch information, they may use this information to help identify of the tones.

2.1.3. The present studies

The following two chapters aim to investigate the categorization of segments and tones for both native Chinese listeners and English-speaking L2 learners of Chinese. Given the research reviewed in this chapter, for tones that begin with different pitch heights, it is possible that the L1-English L2-Chinese learners will have an early advantage in categorizing tones compared to native listeners, since they have a heightened sensitivity to pitch height. In order to investigate this effect, a forced-choice gating task was used (Experiment 1), where participants heard increasingly longer fragments of words and were presented with two options, either two tones or two rimes. Gating allows for a time course investigation of the categorization of tones and segments. By choosing tonal pairs (the options presented for selection) that have drastically different onsets and investigating categorization of increasingly longer fragments, it is possible to see if learners have higher tonal identification accuracy compared to native listeners in the early portions of the tone. This early portion of the tone is where the tones differ the most in pitch height; as such, it provides an ideal test of whether or not English-speaking L2 learners of Chinese can utilize this pitch height sensitivity to identify the tones. Alternatively, given that English does not have lexical tones, L1-English L2-Chinese learners may have difficulty identifying tones even in the presence of an early pitch height difference, and as a result, they may be less accurate than native listeners at identifying the tones.

Segmental items that contrasted in the rime portion of the syllable were also included in the forced-choice gating task to investigate once again if tonal information is disadvantaged compared to segmental information when it comes to categorization. The use of gating to compare tones and segments gives the ability to investigate if tones are disadvantaged, and if so, how this disadvantage plays out over time. Questions of whether tones begin disadvantaged from the onset, and if so, whether this disadvantage disappears or grows over time can be answered only with a time-course measure. Therefore, this study will add to the literature comparing tones and segments with more detailed information; specifically, it will provide a time course of the categorization of segmental and tonal information, and consequently, how the comparison between the two types of categorization changes as the signal unfolds.

Before the time course of tonal and segmental categorization can be investigated, however, an additional concern needs to be addressed. At the level of psycho-acoustic perception, it is possible that tonal contrasts need a greater duration of acoustic input than segmental contrasts to disambiguate. Rime contrasts are signaled by a host of acoustic cues, such as F1, F2, F3, coarticulatory information from the surrounding segments, and many more. By contrast, tones are signaled by F0, and in some cases durational cues will aid in tone identification. Additionally, contour tones are defined as pitch change over time, and therefore cannot be identified with a single pitch point, whereas vowel quality can be identified by F1, F2, F3 etc. at one point in time. Thus, rime contrasts are signaled by many more cues than tone contrasts; this may have an effect on how much acoustic input (duration of input) is needed to distinguish tonal contrasts and segmental contrasts at a low, psycho-acoustic level. As Cutler and Chen (1997) suggested, tones may simply need more time to accumulate the relevant information before any kind of tonal judgment can be made. If tones need a greater duration of acoustic input to be perceived at a low level, this could have an effect on how tones and segments are categorized, with tones showing later and possibly less stable categorization due to the need for the relevant information to accumulate at a psycho-acoustic level.

To control for this possible difference between tones and segments, a norming study was first conducted to select tonal and segmental stimulus pairs that disambiguate psycho-acoustically with the same duration of acoustic input; the selected stimulus pairs were then used in the forced-choice gated category-identification task. We thus turn to this norming study before we present the forced-choice gated category-identification task.

CHAPTER 3: THE TIME COURSE OF DISCRIMINATION OF TONES AND SEGMENTS: A NORMING STUDY

3.1. Introduction

This chapter reports on a norming study, a gated AX discrimination task, which was used to select pairs that are optimally matched across conditions and groups for the duration of acoustic input needed to discriminate the pairs at a psycho-acoustic level for use in a gated category-identification task. The classic gating paradigm is a task where listeners hear increasingly longer fragments of words and decide what word they think they heard and how confident they are in their response. In addition to the participants' word identification and confidence ratings, this task gives an isolation point, or the point at which a participant selects the intended word and no longer changes their response. This isolation point is the point at which the word is considered to be recognized. This task has been used to study Chinese tones (e.g., Lai & Zhang, 2008).

Since the goal of this norming study was to find tonal and segmental pairs that would be matched in the duration of acoustic input needed to discriminate the pairs, the classic gating paradigm where participants respond with whole words was not used (i.e., a gating task that requires participants to respond with whole words would not tap into psycho-acoustic perception). Instead, a gated AX discrimination task was used. A gated AX discrimination task is a task in which stimuli consisting of increasingly longer word-pair fragments are compared. Participants hear fragments (of equal length) of two words and decide if the two fragments are the same or different. Because participants heard fragments rather than complete words, and because their task was to discriminate between the fragments rather than identify them, this gated AX discrimination experiment has the ability to by-pass lexical access (unlike a classic gating task), and is able to target discrimination at a low, psycho-acoustic level. The current task used a short ISI of 250 ms; thus, in addition to targeting psycho-acoustic perception, the current task also by-passed effects of L1-specific phonetic categories (Pisoni, 1973; Werker & Logan, 1985).

3.1. Methods

3.1.1. Participants

A total of 20 native Chinese listeners (14 female; mean age: 24.9; standard deviation (SD) 4.13) and 25 native English listeners (17 female; mean age 22.6; SD: 3.71) participated in this study. All Chinese listeners considered Standard Mandarin Chinese to be their native dialect and reported that their parents spoke Standard Mandarin Chinese or Taiwanese natively, with at least one parent speaking Standard Mandarin Chinese natively for each participant. All English listeners considered English to be their native language and reported that their parents spoke English, Spanish, French, Armenian, or Greek natively, with at least one parent speaking English natively for each participant.

3.1.2. Materials and design

The experiment included two conditions: a tonal condition and a segmental condition. The tonal condition included all of the possible tone pairs in Chinese with the exception of Tone 2 – Tone 3 (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 4 – Tone 2, Tone 4 – Tone 3, and Tone 1 – Tone 4). These pairs have been shown to be maximally distinct both as a whole (e.g., Gottfried & Suiter, 1997; Lee, Tao, & Bond, 2008; Wang et al., 1999) and at early time points (with the exception of T1-T4, which disambiguates late; Lai & Zhang, 2008). The pair of Tone 2 – Tone 3 was not included in this experiment since L2-Chinese learners, and even native speakers, classically have difficulty perceiving this tonal contrast (e.g., Gottfried & Suiter, 1997; Lee et al., 2008; Wang et al., 1999), thus, differences between Tone 2 and Tone 3 would likely not yield any useful data. T1 – T4 was included as a comparison case, since the tones do clearly disambiguate around halfway through the tones.

Four different monosyllabic word pairs were selected for each of the five tone pair types, creating five tonal sets and yielding a total of 20 word pairs in the tonal condition. All words began with voiceless initial consonants to control the timing of the tonal information; words that began with a voiced or sonorant consonant were not used as these onsets could possibly carry tonal information. Thus, the onset of tonal information could in theory be identified as early as the onset of the vocalic portion of the syllable. The words within a given tonal pair were identical segmentally (e.g.*, bā* [pa1] 'scar' – bǎ [pa2] 'target'). The syllables with the tones were produced naturally and only the duration and intensity of the syllables were manipulated (discussion to follow). All words from the tonal condition can be found in Appendix A.

For the segmental condition, a range of words were selected with four hypothesized timings of disambiguation of the segments, ranging from early disambiguation in monophthong vowels (*bi* [pi]-*ba* [pa]) to late disambiguation in the change of a vowel to a nasal coda (*sao* [sɑʊ] – *sang* [sɑŋ]), as illustrated in Table 1. A range of hypothesized disambiguation points was desired for multiple reasons. First, having a range in the syllable types allows for the best overall comparison of tones to segments, as opposed to limiting the selection to just one syllable type. Second, the majority of the tonal pairs were expected to disambiguate at around the same time point, since they all began with different tone onsets. If differences in discrimination timing between tones and segments were found, a range of syllable types would be needed to best match the tonal timing to a segmental timing to control for the difference.

In the segmental condition, only the segments differed between the pairs. In the *bì* [pi4] ‘to close’ – *bà* [pa4] ‘father’ example pair, the disambiguating segmental information is the difference in the monophthong vowel. In this way, the disambiguating information arrives during the last portion of the consonant (via co-articulatory cues) or soon after the end of the consonant. Slightly later in timing would be the disambiguation of pairs such as *táo* [tʰɑʊ2] ‘to lift’ – *tái* [tʰɑi2] ‘to escape,’ which have an allophonic difference in the vowel, as seen in the phonetic transcriptions in Table 1. Even later in timing would be the disambiguation of pairs such as *tiāo* [tʰiɑʊ1] ‘carry on a pole’ – *tīe* [tʰiɛ1] ‘to paste,’ which also differ in the vowel, but only after an onglide. Finally, the latest pairs to disambiguate are pairs whose only difference lies in the offglide/coda (with possible co-articulatory effect on the final portion of the vowel) such as *săo* [sɑʊ3] ‘to sweep’ – *săng* [sɑŋ3] ‘to push back’. This is the latest possible place where segmental information could disambiguate between the two words.

The segmental condition included four word pairs in each of these four timing categories, creating four segmental sets, which yielded a total of 16 word pairs in the segmental condition. Only pairs where a sound had been changed, not added, were included in the present study in order to minimize durational differences between the two words of a pair in the natural production of the segmental contrasts (i.e., these durational differences would make duration normalization more difficult). The words within a pair were identical tonally. The segmental contrasts in each pair also existed in English. Such contrasts were selected in anticipation that the segmental condition in the forced-choice gated category-identification task (in Chapter 4) would be the condition where English L2 learners of Chinese are expected to pattern similarly to native Chinese listeners. All words from the segmental condition can also be found in Appendix A.[9]

---

[9] The tones were not specifically balanced for the segmental condition, since the crucial aspect was simply that the tones on the word pairs were the same tone in the segmental condition. That being said, the materials include four

Table 1: Example segmental contrasts

Discrimination Expected		Early ⟷ Late		
Name	Vowel	Allophonic Vowel	Post Onglide	Nasal Coda
IPA	pi4 – pa4	tʰai2 – tʰɑʊ2	tʰiɑʊ1 – tʰiɛ1	sɑʊ3 – sɑŋ3

The 36 word pairs were recorded by a male native speaker of Mandarin Chinese in the Anechoic Chamber of the University of Kansas with a cardioid microphone (Electrovoice, model N/D767a) and a digital solid-state recorder (Marantz, model PMD671) at a sampling rate of 22,050 Hz. All words were recorded at three speeds: slow, normal, and fast. The token that was closest to the average duration from all the normal-speed tokens (i.e., 524 ms) was selected and normalized for duration. For example, if, for a given word pair, the slowly produced token was 550 ms, the normally produced token 400 ms, and the rapidly produced token 375 ms, the slowly produced token was selected and manipulated. As a result, the tokens were manipulated as little as possible from a natural production, but all had the same duration so as to remove duration as a cue to tone identity. The word-initial consonant portion of the word was normed to 117 ms, and the rime portions were normed to 407 ms, as found from the average durations from the normal-speed productions.[10] Additionally, by norming the durations of all the word-initial consonant portions, it was possible to control the timing of the arrival of the tonal information as precisely as possible, with the onset of voicing in the rime being the same across all tokens. See Appendix

Tone 1 pairs, four Tone 2 pairs, three Tone 3 pairs and 5 Tone 4 pairs. Given that the tones are nearly balanced with the exception of one pair being swapped between Tone 3 and Tone 4, it is not believed that this will have an impact on the experiment.

[10] This means that fricatives, affricates and stops were normed to 117ms. For affricates and fricatives, the durations were simply extended or shortened to fit this timing. Since this was not possible for stops, a period of silence was added before the stop release. This was done so as to make sure that in all token, Gate1 was 117 and vowel/tone information arrived at the same time. Appendix A includes a follow-up analysis looking at whether the results of the norming study differed by this stop/non-stop contrast when possible.

B for acoustic measurements of the stimuli used including pitch contours for the tonal items and F1, F2, center of gravity measures for the segmental items.

After duration was normalized, the word-initial consonant and the *first half* of the rime portion of each word were divided into twelve gates; thus, participants never heard the complete word, and as such lexical effects on the results are not expected. The first gate was the initial consonant (117 ms). All eleven subsequent gates in the rime included 18 ms more information than the previous gate. All items were also normalized for intensity.

3.1.3. Procedures

Participants were seated at a computer with headphones in a quiet room. They read the instructions and began the task. In each trial, participants heard two fragments and were asked to decide if these fragments were the same of different by pressing mouse buttons that corresponded to the choices. Which button signaled 'same' and which button signaled 'different' was indicated on the screen at all times; hence, participants could always reference the labels to ensure that they were responding appropriately. Items with the same gate duration were blocked together in order from Gate 1 to Gate 12 and were randomized within the blocks. The order of appearance of items in each pair (i.e., which of the two items was presented first and which was presented second) was reversed in the next block. For example, if participants heard *bā-bá* in Block 1, they would hear *bá-bā* in Block 2, and so on. Participants were encouraged to respond as quickly and accurately as possible, though there was no time limit. An ISI of 250 ms was used in order to target psycho-acoustic perception (Pisoni, 1973; Werker & Logan, 1985).

Filler trials consisting of an exact repetition of a critical item were added to balance the number of 'same' and 'different' trials. Each block consisted of 36 critical trials and 36 filler trials. Filler trials were randomized with critical trials in each block.

3.1.4. Analysis of discrimination data and results

Below, we present the results of the trials where the two stimuli in the pair were *different*. Each trial yielded one of two potential responses: same or different. Every 'same' response was coded as 0, whereas every 'different' response was coded as 1. The mean 'different' responses at each gate are presented for native listeners in Figure 2 and for English listeners in Figure 3.

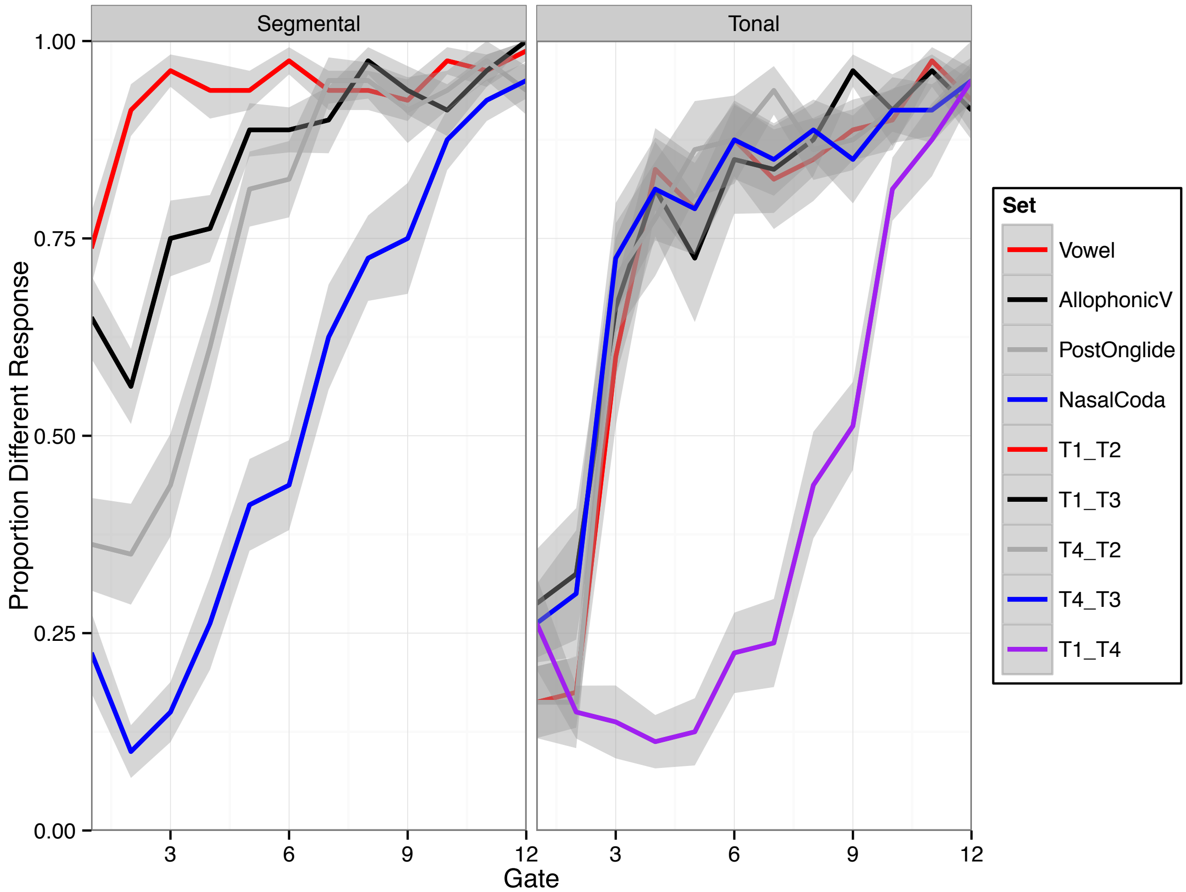

Figure 2: Native Chinese listeners' proportion of different responses over time for the segmental condition (left panel) and tonal condition (right panel). The gate number is presented on the *x*-axis and the proportion different responses are presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

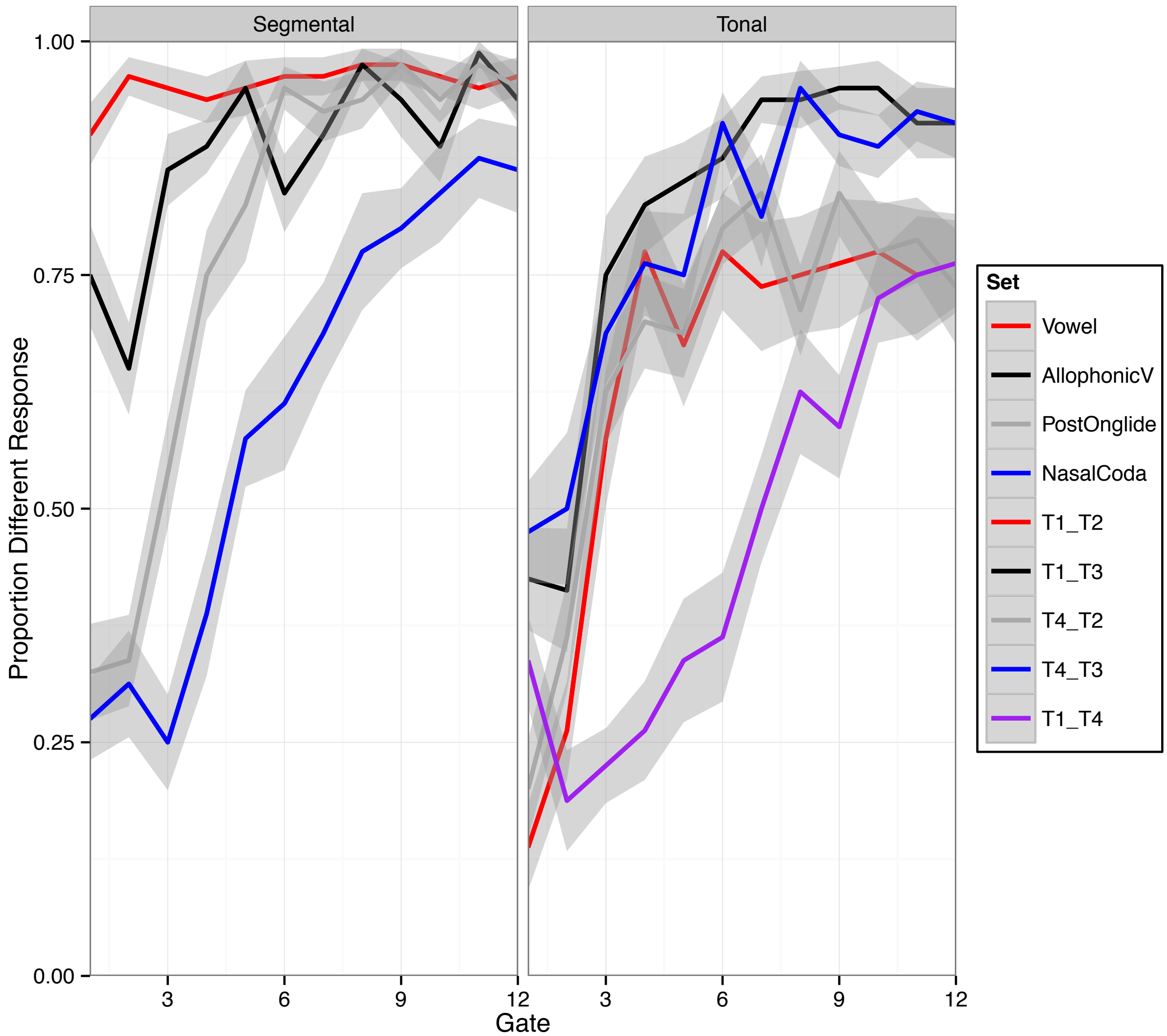

Figure 3: Native English listeners' proportion of different responses over time for the segmental condition (left panel) and tonal condition (right panel). The gate number is presented on the *x*-axis and the proportion different responses are presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

A visual inspection of the graphs shows that the hypothesized disambiguation differences between the sets appear to have an effect on discrimination timing. This is especially clear in the gradient effect seen with the segmental sets and the gradient change in improvement rates (i.e., the amount of accuracy improvement from one adjacent gate to the next). Both the vowel and allophonic onglide sets begin with discrimination scores above zero, indicating that participants were able to tap into co-articulatory information in the consonant to disambiguate them in the very first gate where no vowel information was present. Next, the post onglide set has responses above 50% at about Gate 4, with the latest segmental set having responses above 50% at about Gate 7. This progression is expected, as the sets were selected so as to have a range in discrimination timings in this order. For the tonal condition, we see that the early disambiguating tonal pairs (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 2 – Tone 4, and Tone 3 – Tone 4) pattern together, as predicted. These sets show responses above 50% by about Gate 3 for both groups. The late tonal set of Tone 1 and Tone 4, on the other hand, shows responses above 50% considerably later, by about Gate 10, as predicted.

Based on the visual inspection of the graphs, in order to control for discrimination between tones and segments and for both groups, it would be best to use the early disambiguating tonal pairs (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 2 – Tone 4, and Tone 3 – Tone 4) and the post onglide segmental pair, given the similar shapes of the response curves and discrimination timings across groups and conditions. In order to see the degree of fit between the selected tonal sets and the optimally match segmental timing, these selected sets were graphed together. Figure 4 shows the average across all of the early tonal sets and the post onglide segmental timing by group. As can be seen, the response curves are extremely similar across conditions for both groups. This pairing of tonal and segmental sets therefore appears to provide the optimal control of discrimination timing and response improvement rates between the tonal and segmental conditions. Statistical analyses could not be conducted to confirm this since the two conditions had unequal numbers of items (16 tonal items but only 4 segmental items). While these matching are not perfect, given the differences between the conditions, particularly around the 50% mark, they are as close as is possible

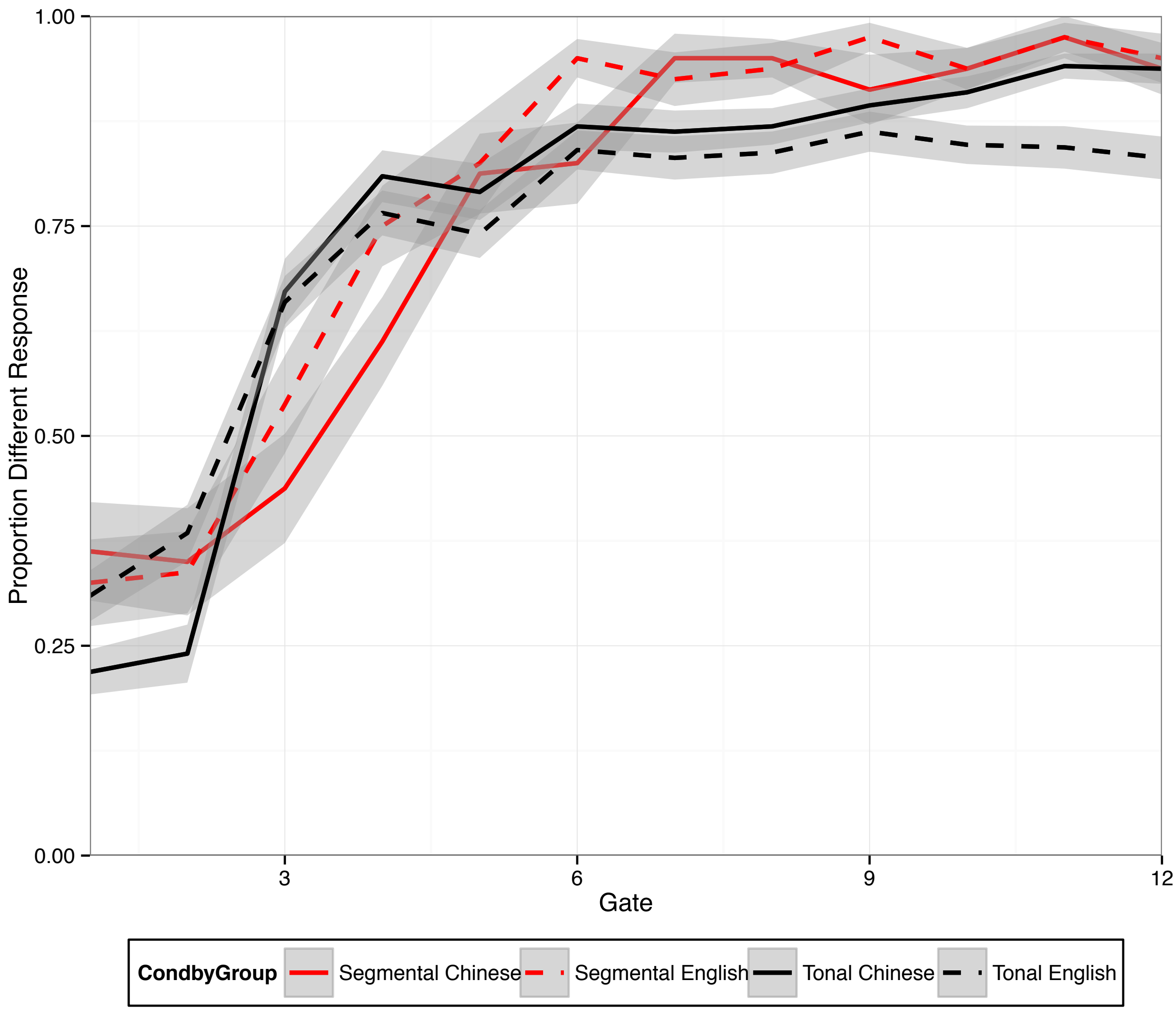

Figure 4: Averages of selected tonal and segmental sets by group and condition. The gate number is presented on the *x*-axis and the proportion different responses are presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean. The segmental condition includes only the post-onglide set. The tonal condition includes the tone pairs Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 4 – Tone 2 and Tone 4 – Tone 3.

3.1.5. Discussion

The results of the norming study show that the segmental and tonal pairs compared in the gated AX discrimination task differ in the timing of correct responses, such that for pairs that would be expected to disambiguate later, listeners needed to hear more of the pairs to tell them apart (based on a visual inspection of the graphs). This was seen in the progression of segmental responses, with vowel and allophonic pairs starting with responses above 50%, then the post-onglide pairs, and then finally the nasal coda pairs. In terms of tones, it was shown that all pairs that disambiguate at the beginning of the tone disambiguate with the same duration of acoustic input. For the late tonal pair (i.e., Tone 1 and Tone 4), participants needed to hear more of the pair in order to tell them apart (again based solely on a visual inspection of the graphs), due to their similar onsets.

In order to ensure that the pairs selected for the gated forced-choice category-identification task (in Chapter 4) were controlled as closely as possible for low-level psycho-acoustic discrimination between tones and segments for both groups, optimally matched sets were chosen. The results showed that all the early disambiguating tonal pairs (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 2 – Tone 4, and Tone 3 – Tone 4) best matched the post-onglide segmental timing in terms of when responses rose above 50% accuracy. This was true for both native Chinese listeners as well as naïve English listeners. Therefore, in order to ensure the psychoacoustic discrimination timing was matched across conditions for both groups, the stimuli in the category-identification task included pairs of words from the early tonal pairs (Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 2 – Tone 3, and Tone 3 – Tone 4) and from the post-onglide segmental pairs.

CHAPTER 4: THE TIME COURSE OF CATEGORIZATION OF TONES AND SEGMENTS: EXPERIMENT 1

4.1. Introduction

This chapter reports on a forced-choice gating category-identification task designed to tap into listeners' identification of categories from fragments taken from syllables that differ only in tones or only in segments. The aim of this study was to investigate the time course of category identification for native Chinese listeners and English-speaking L2 learners of Chinese. By looking at the time course of categorization of tones that disambiguate at their onsets, it will be possible to reveal whether English listeners have an early pitch height advantage over Chinese listeners. The work that has shown that English listeners have superior sensitivity to pitch height when perceiving tones was strictly with naïve listeners (Gandour, 1983; Gandour & Harshman, 1978; Kaan et al., 2007; Qin & Jongman, 2016; Xu et al., 2006). To my knowledge, no study has directly investigated if this pitch height sensitivity extends to L2 categorization of the tones of Chinese. To investigate this, a time-course study is needed to see if L2 listeners have an advantage in identifying the tones of Chinese when pitch height is the primary cue to tonal identity.

Additionally, this task will once again investigate if tones are disadvantaged compared to segments, but will do so with a time course measure to gain a better understanding of this disadvantage. It is possible that tones are consistently disadvantaged compared to segments, or that tones and segments begin with similar categorization abilities, and then this tonal disadvantage appears as the words progress. Given that tonal identity relies on a single primary cue (i.e. pitch) and vowel identity (crucial to rime identity) relies on many cues (i.e., F1, F2, and F3), it is possible that, with increasing duration of acoustic information, segmental contrasts have more cues, causing the segmental advantage to appear later in the word. A time course analysis comparing tonal and segmental contrasts would then clarify the underlying cause of a tonal disadvantage, if found.

To achieve these goals, a forced-choice gating task was conducted to compare tonal and segmental contrasts for both L1 and L2 Chinese listeners. Participants heard fragments of Chinese words and identified either the tone or the rime of the presented fragment. The stimulus pairs presented were selected so as to match the discrimination timing between the tonal and segmental conditions, as determined from the norming study (Chapter 3). This ensured that the results of the forced-choice gating category-identification task would reflect differences due to category identification and not differences at the level of psycho-acoustic perception. This control removes a source of timing variation not due to categorization, which allows for a more careful investigation of the time course of tonal and segmental categorization.

4.1. Methods

4.1.1. Participants

A total of 24 native Chinese listeners (14 female; mean age: 25.2; standard deviation (SD) 3.5) were recruited from Beijing University, China and the surrounding area. Additionally, a total of 22 native English listeners (17 female; mean age 22.6; SD: 3.71) with advanced levels of L2 Chinese proficiency were recruited from Beijing University, China and the surrounding area (*n*=17), the University of Maryland (*n*=1), and the University Kansas (*n*=4). All native Chinese listeners considered Standard Mandarin Chinese to be their native dialect and reported that their parents spoke Standard Mandarin Chinese natively, except for one participant who had one parent speaking Kazakh as their native language.[11] All English listeners considered English to be their native language and reported that their parents spoke English, Tamil, Yoruba, Tagalog, Sinhalese, Amharic, Danish, or a Chinese dialect natively, with at least one parent speaking English natively for all but one participant.[12] Participants who had a parent speaking a Chinese dialect or Yoruba reported that these languages were not used in the home in early childhood.

Seven native Chinese participants were excluded from the analyses and thus were not included in the above report: One for not knowing the numbers corresponding to the tones of Chinese (necessary for Experiment 1), and six for reporting that they were native speakers of a dialect other than Mandarin or for having exposure from a parent in early childhood to a dialect of Chinese other than Mandarin. Four L2 learners of Chinese participants were excluded from the analyses and thus were not included in the above report: Two for their low proficiency (not being able to complete all of the tasks) and two for having substantial exposure to a tone language in early childhood (Taiwan Mandarin and Yoruba[13]).

For L2 learners, Chinese proficiency was established by years of instruction, length of stay in China, and other factors self-reported in a language background questionnaire, provided in Appendix E. In addition to filling out a language background questionnaire, all learners completed a Chinese proficiency test (Qin, Connell, & Tremblay, in prep), which was based on the design of the English proficiency test LexTALE (Lemhofer & Broersma, 2012). This test is a lexical decision task with nonce words and real words of varying frequencies. Participants decided whether or not 120 Chinese disyllabic sequences were real Chinese words. These words

[11] Given that Kazakh is not a tone language, Mandarin was the only tone language spoken in the home of this participant. Therefore, this participant was included in the study.

[12] While Yoruba is a tone language, the participant reported that their parents spoke this language and English, and that English was the primary language in the home during early childhood.

[13] This refers to a separate participant from that discussed in Footnote 12. This participant was excluded for reporting that their exposure to Yoruba in early childhood was substantial, and that they still spoke Yoruba on a regular basis.

were selected to include 40 'difficult' words (with an average frequency of 0.87 words per million) as well as 40 relatively 'easy' words (with an average frequency of 41.3 words per million). In addition, 40 nonce words were created by pairing two syllables together in a way that created nonexistent but semantically plausible words of Chinese. Participants saw each word written in simplified characters in the middle of the screen and used the left and right arrow keys marked as 'word' and 'not a word' to give their response. Participants were instructed to respond with 'word' *only* if they personally knew it to be a word of Chinese. Correct answers were balanced between left and right responses. This task was implemented with Psychopy (Peirce, 2007) on a MacBook Pro. The test was scored using a weighted score: Correct responses to words and non-words received 1 point, incorrect word responses received no point, and there was a penalty of −1 point for incorrect non-word answers (i.e., saying a non-word is a word). This penalty attempts to account for learners having a bias to say that items are words. These scores were then averaged to give a final LexTALE weighted average correct. All words for this task can be found in Appendix F. Since this task is still being tested for validity, participants also completed a Chinese cloze test (Yuan, 2009), provided in Appendix F. Measures such as years of instructions, age of acquisition, age, and months spent in a Chinese speaking country were established based on a language background survey, provided in Appendix E. A summary of the participants' language background information and proficiency results is presented in Table 2.

Table 2: Language background and proficiency information

	Age	Years of Inst.	Age of Acq.	Months in China	% Use of Chinese	LexTale Weighted Accuracy	Cloze Score (% accuracy)
Mean	23.00	4.29	17.82	21.66	35.14	62%	70%
SD	3.24	1.87	2.68	26.89	25.80	9%	15%

Note: Years of Inst. = year of instruction; Age of Acq. = age of acquisition

The native Chinese-speaking participants were paid 150 RMB (approx. $20) for their participation upon completing all portions of the study. The English-speaking L2 learners of Chinese were paid 200 RMB (approx. $30) in China or $30 in the US for their participation upon completing all portions of the study

4.1.2. MATERIALS AND DESIGN

The experiment consisted of two conditions: a tonal condition and a segmental condition. For the tonal condition, all early tone pairs used in the norming study were included (i.e., Tone 1 – Tone 2, Tone 1 – Tone 3, Tone 4 – Tone 2, and Tone 4 – Tone 3). Four different monosyllabic word pairs consisting of a target and a tonal competitor were selected for each of the four tonal comparisons. This yielded 16 critical trials in the tonal condition. All words began with a voiceless initial consonant to control the timing of the onset of tonal information; words that begin with a voiced or sonorant consonant were not used as they could possibly carry tonal information on the consonant portion. The words within a given tonal pair were identical segmentally and only differed in their tone (e.g*., bā* [pa1] 'scar' – *bǎ* [pa2] 'target'). All words from the tonal condition can be found in Appendix C.

The segmental condition included only word pairs with the post-onglide timing, as established from the norming study. Sixteen pairs of words, including a target and a segmental competitor with the post-onglide timing, were selected. All words began with voiceless initial consonants to control the timing of the onset of tonal information. The words within a pair were identical tonally and varied only in their segments (e.g., *guī* [kuei1] 'turtle' – *guō* [kuo1] 'pot'). All words from the segmental condition can be found in Appendix C.

The same male native speaker of Mandarin Chinese who recorded the stimuli for the norming study recorded the stimuli for Experiment 1. It was not possible to use the recordings from the norming study since new words were added to the stimuli list. The post-onglide timing set from the norming study only included 4 pairs of words. An item count of 16 pairs was desired, so a substantial set of new pairs needed to be added. Additionally, some words were changed in anticipation of the eyetracking study, so that all words were imagable. But as stated above, all newly selected stimulus pairs fit the post-onglide timing, which provides the best match with the tone pairs, as determined from the norming study. For consistency, all recordings were done in a single new session. The stimuli were recorded in the Anechoic Chamber of the University of Kansas with a cardioid microphone (Electrovoice, model N/D767a) and a digital solid-state recorder (Marantz, model PMD671), using a sampling rate of 22,050 Hz. All words were normalized for duration and intensity, and were gated following the same method as in the norming study as described in Section 3.1.2. Acoustic analyses are presented in Appendix D, which includes pitch contours for tonal items and F1 and F2 over the 12 gates, as well as the center of gravity, variance, skewness and kurtosis on the initial consonants in the pairs.

4.1.3. Procedures

For each trial, participants were instructed that they would hear a single fragment and see two options on the screen, either a tone pair presented as numbers (e.g., 1 or 2) or a rime pair present in Pinyin (e.g., _ai or _ang). All participants reported being familiar with both Pinyin and the tone numbers before beginning the task. Their task was to select the option that corresponded to the auditory stimulus heard. The options were written on the bottom left and right sides of the screen, and participants used the arrow keys to select either the left or right option. The options were presented for 1,000 ms before the stimulus was heard to give participants time to recognize whether they should respond with a tone or a rime. The task included the stimuli from Gates 2-12 from each word and pair described in Section 3.1.2, with the first block containing all Gate-2 items, the second block containing all Gate-3 items, and so forth. Thus, each block contained all tonal and all segmental items for that gate. The tonal and segmental stimuli trials were randomly intermixed. This means that a tonal trial could be followed by a segmental trial or vice versa. The presented options on the screen indicated to the participants which response they should be giving.

For example, in the segmental condition, if the target word was *bai1* and the alternative *bang1*, the options presented on the screen would be "_ai" and "_ang," and the response "_ai" would be coded as correct. Both items in each pair were included in the task: On another trial, *bang1* would be the target word and bai*1* the alternative, and the options presented on the screen would again be "_ai" and "_ang," but for this trial, selecting "_ang" would be coded as correct. Likewise, in the tonal condition, if the target word was *shu3* with the alternative *shu4*, the options presented would be "3" or "4," and a response of "3" would be coded as a correct response; on another trial, *shu4* would be the heard target word and *shu3* the alternative, but for this trial, selecting "4" would be coded as correct. No filler items were included. Figure 5 below shows the progression of a single trial.

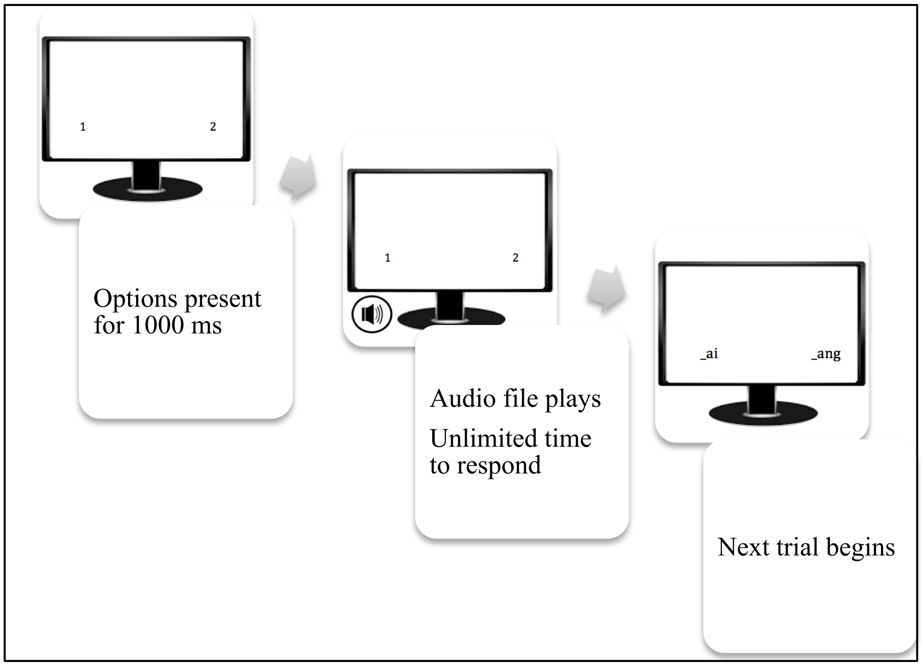

Figure 5: Experiment 2 trial procedure

A practice session of 20 items preceded the experiment to ensure that participants understood the task. All instructions were visually presented in English and simplified Chinese and were also explained verbally by the main experimenter in the language of preference of the participant. This task was implemented with PsychoPy (Peirce, 2007) on a MacBook Pro. There was no time limit to respond, but participants were encouraged to respond with their first intuition as quickly as possible. An equal number of left and right responses were correct responses.

4.1.4. Data analysis

Each trial yielded one of two potential responses: the target or the alternative. Target responses were coded as 1 (correct) and alternative responses were coded as 0 (incorrect).

As a brief introduction, GCA has recently been suggested as an improved method for analyzing time course data (e.g., Mirman, 2014; Mirman, Dixon, & Magnuson, 2008), and has been used in several recent eye-tracking studies (e.g., Kukona, Fang, Aicher, Chen, & Magnuson, 2011; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Malins & Joanisse, 2010; Tremblay, Broersma, Coughlin, & Choi, 2016) and analyses of tonal contours (e.g., Li & Chen, 2016; Zhang & Meng, 2016).

To determine whether the numerical trends are statistically reliable, both GCA (otherwise known as hierarchical regression) (Mirman, 2014) and linear mixed-effects models (LME) were used. In a GCA model, time-series data are modeled with third-order orthogonal polynomials, with fixed effects of the chosen conditions and their interaction on all time terms. Orthogonal polynomials remove the co-linearity between the terms that exists in natural polynomials. For this reason, the time terms presented here (linear, quadratic, and cubic) can be interpreted independently of one another. In GCA, the intercept corresponds to the overall average of the time polynomial, which also corresponds to the mid-point of this polynomial on the *x*-axis; as a result, any effect of condition on the intercept is an effect that can be observed halfway through the *x*-axis. The linear polynomial models the overall linear trend of the data over time; thus, an interaction between condition and the linear term indicates that the slopes of the data in each condition are different. The quadratic and cubic terms model the quadratic and cubic trends of the data respectively; hence, they show differences in the U- and S-shapes of the data. Likewise, interactions with condition and the quadratic or cubic time terms would indicate that the U-shape or S-shape of the curves differed between the conditions (Mirman, 2014).

4.1.5. RESULTS

Average target responses for each group and condition at each gate are presented in Table 3 below.

Table 3: Target response averages by group and condition

	Gate	2	3	4	5	6	7	8	9	10	11	12
L1	Segmental	0.63	0.65	0.72	0.79	0.86	0.87	0.91	0.91	0.95	0.96	0.96
	Tonal	0.55	0.61	0.62	0.64	0.64	0.68	0.71	0.72	0.78	0.82	0.88
L2	Segmental	0.59	0.64	0.72	0.82	0.86	0.91	0.91	0.93	0.95	0.94	0.97
	Tonal	0.57	0.65	0.67	0.69	0.70	0.69	0.70	0.74	0.79	0.83	0.86

Figure 6 shows gate number on the *x*-axis and the proportion of target responses on the *y*-axis. This figure shows these proportions for each group, with L1 listeners in the left panel and L2 listeners in the right panel. Solid lines represent the actual data; dashed lines represent the predicted data based on the growth curve analysis (GCA) of all listeners' response data (Table 4). The shaded regions represent one standard error above and below the mean. Figure 7 shows the same results as in Figure 6 but re-plotted by condition for ease of comparison.

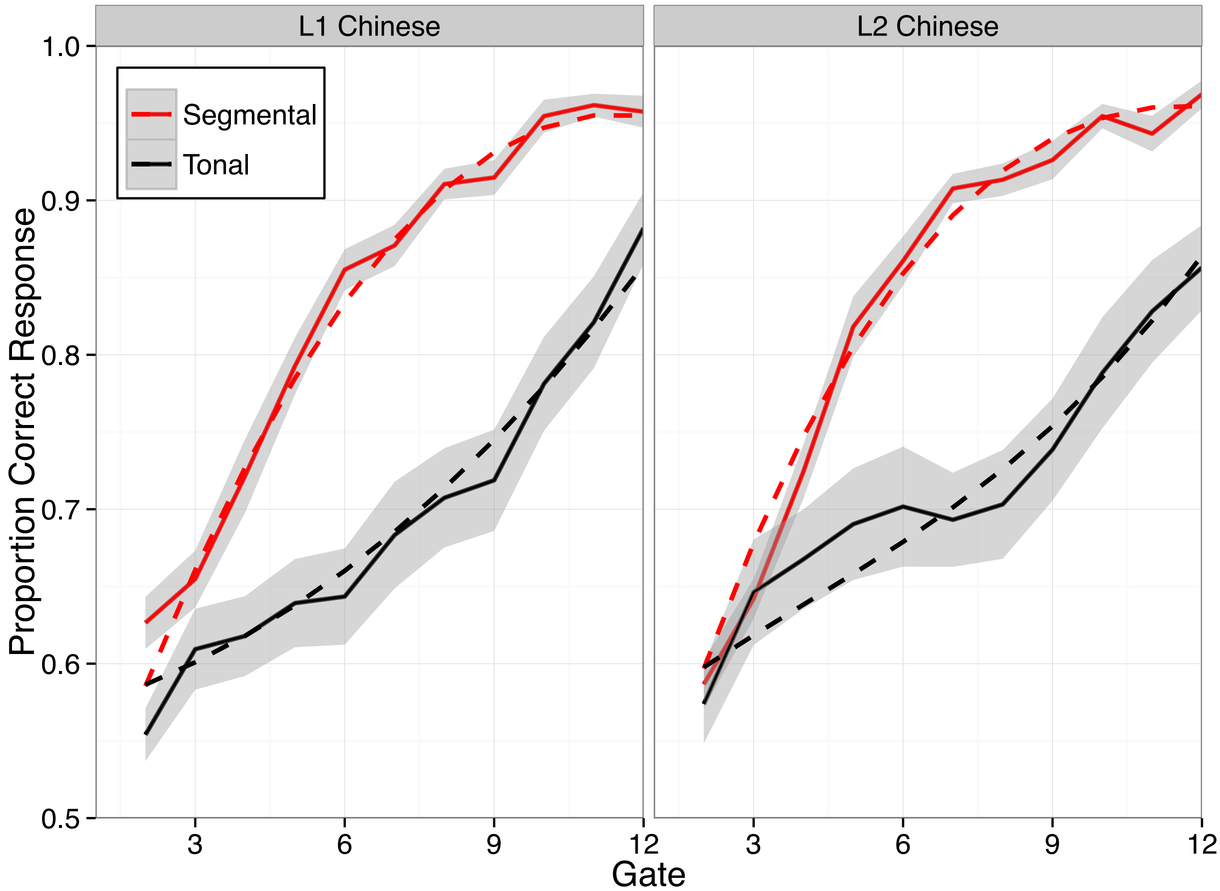

Figure 6: Actual and predicted proportion correct responses, with actual data in solid lines and predicted data (from the Growth Curve Analysis presented in Table 4) in dashed lines. L1-Chinese listeners (left panel) and L2-Chinese listeners' (right panel) each show segmental responses in red and tonal responses in black. Gate number is presented on the *x*-axis and proportion correct responses are presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

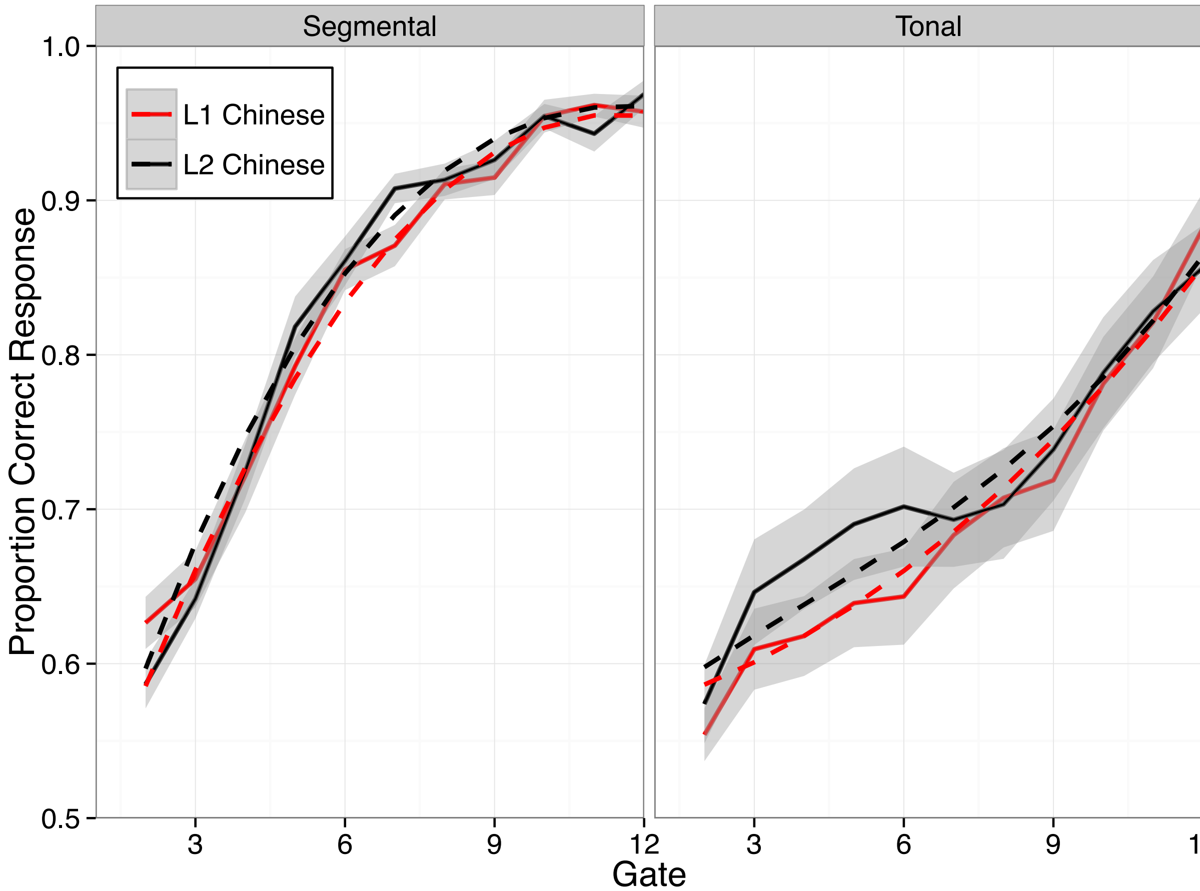

Figure 7: Actual and predicted proportion correct responses with actual data in solid lines and predicted data (from the Growth Curve Analysis presented in Table 4) in dashed lines. The segmental condition (left panel) and tonal condition (right panel) each show L1-Chinese listeners' responses in red and L2-Chinese listeners' responses in black. Gate number is presented on the *x*-axis and proportion correct responses are presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

A visual inspection of the results in Figure 6 reveals three observations. A first observation is that as listeners heard more of the stimulus, their ability to identify the tone and the rime increased. A second observation is the striking difference between the tonal and segmental items in terms of the proportion target responses in both groups: The rate at which target responses increase (i.e., the 'improvement rate,' for convenience) is larger in the segmental condition than in the tonal condition. Finally, a third observation is that both groups in both conditions appear to correctly identify the tone at above-chance levels from the very first gate (in this case gate 2, since gate 1 was not included due to time constraints), which includes the consonant and the first 18 ms of the rime.

Figure 7 directly compares L1-Chinese and L2-Chinese listeners on each condition. As can be seen in Figure 7, the L1- and L2- Chinese listeners are nearly identical in the segmental condition, as expected. In the tonal condition, we see a slight numerical advantage for L2-Chinese listeners for the first six gates, which disappears around Gate 7. Finally, as can be seen in both Figure 6 and Figure 7, the target responses in the tonal condition never reach as high a proportion as those in the segmental condition, with the proportion of target responses reaching almost 1 in the segmental condition and just under .9 in the tonal condition.

The proportions of target responses were analyzed with GCA (Mirman, 2014) using the *lme4* (Bates, Bolker, & Walker, 2015) and *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015) packages in R (R Core Team, 2013). The proportions of target responses were modeled with third-order orthogonal polynomials, with the fixed effects of condition, group, and condition by group on all time terms. Chinese listeners' performance on the segmental condition was the baseline. The model also included random effects of participant on all time terms. The most complex model was evaluated using the backward fit function bfFixfLMER_F.fnc of the *LMERConvenienceFunctions* package (Tremblay & Ransijn, 2015) using log-likelihood ratio tests to determine whether removing any effect from the model adversely affected the ability of the model to predict the data. The results of the simplest model with the best fit are presented in Table 4 below, with *p* values being calculated using the *lmerTest* package (Kuznetsova, Brockhoff, & Christensen, 2016). The model with the best fit included the effect of condition as well as the interactions between condition and the linear and quadratic terms.

Table 4: Results of GCA on native and learner target responses GCA

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.83	0.01	76.60	<.001
Linear	0.39	0.02	24.52	<.001
Quadratic	–0.12	0.01	–8.44	<.001
Condition	–0.13	0.01	–24.53	<.001
Linear : Condition	–0.10	0.02	–5.72	<.001
Quadratic : Condition	0.17	0.02	9.22	<.001

The results in Table 4 show that the significant effect of the linear term with a positive estimate indicates that native listeners' proportion target responses in the segmental condition had a positive linear trend. The significant effect of the quadratic term with a negative estimate indicates that native listeners' proportion of target responses in the segmental condition had a concave (i.e., ∩) shape. The effect of condition significantly improved the model and had a negative estimate, indicating that collapsing across all gates there were fewer target responses in the tonal condition than in the segmental condition. The significant interaction between condition and the linear term with a negative estimate indicates that the slope of the response data was shallower in the tonal condition than in the segmental condition. This means the proportion of target responses increased more slowly with gate in the tonal condition than in the segmental condition. The significant interaction between condition and the quadratic term with a positive estimate indicates that the proportion of target responses was less concave in the tonal condition than in the segmental condition. This means the proportion of target responses increased more slowly and thus did not asymptote in the tonal condition compared to the segmental condition.

The effect of group and its interaction on any time term did not significantly improve the model. The effect of group not significantly improving the model indicates that, across all gates, there were no differences in the proportion of target responses between the L1- and L2-Chinese listeners on the segmental condition. These results confirm the general conclusions made from Figure 6 and Figure 7 that native Chinese listeners and English-speaking L2 learners of Chinese pattern similarly on the segmental condition. Additionally, the effect of group and its interaction on the time terms did not significantly improve the model. This suggests that the two groups had similar rates of target response improvement over time in the segmental and tonal conditions.

However, our visual inspection of Figure 7 suggested that L2 learners showed a higher rate of correct tonal responses than native listeners at Gates 3-6, which would have been as predicted given English listeners' sensitivity to pitch height. If this effect were reliable, it would have been seen in the GCA either as an overall condition-by-group interaction or as a condition-by-group-by-time interaction(s), indicating that the shapes of the tonal lines were different between the two groups. This lack of difference indicates native Chinese listeners and English-speaking L2 learners of Chinese did not perform significantly differently in either condition.

4.1.6. Discussion

This study reported on a forced-choice gating experiment investigating the time course of categorization of tones and segments for native Chinese listeners and English-speaking L2 learners of Chinese. Listeners' proportions of correct responses were calculated, and GCA models were conducted to investigate how the two groups categorized tones and segments of Chinese as the speech signal unfolds. The results presented in Section 4.1.4 revealed several important effects. First, comparing the Chinese and English groups, the predicted early advantage in the tonal condition for L2 learners as compared to native listeners was not confirmed. Although the results seem to be trending in that direction, with higher proportion correct responses for L2 learners than native listeners in the early gates, the effect did not reach significance. By contrast, the prediction of a lack of difference between the two groups on the segmental condition was confirmed. Second, comparing tones and segments, the predictions were again confirmed in that, for both groups, segments had higher accuracy rates overall and faster improvement rates compared to tones. Each of these effects will now be discussed in more detail.

4.1.6.1. Chinese vs. English listeners' categorization of tones

The results of the GCA shown in Table 4 revealed that group did not significantly improve the model. A visual inspection of Figure 7 indicated that L2 learners did show a numerical trend towards increased target responses in Gates 2-6 as compared to native listeners; however, this effect was not significant. This suggests that while L2 learners of Chinese are more sensitive to pitch height than native listeners, the present work was not able to confirm that this ability extends to the categorization of tones (Gandour, 1983; Gandour & Harshman, 1978; Kaan et al., 2007; Xu et al., 2006). The figures clearly showed that the tonal condition was subject to more variation than in the segmental condition. It is possible that the number of participants in this study was not sufficiently large to overcome the variability in the tonal data to reveal an early effect of condition on the proportion of target responses. From the present results, we can only conclude that English-speaking L2 learners of Chinese did not make greater use of early pitch height than native Chinese listeners to categorize the tones of Chinese.

While the proportions of target responses began at above-chance levels, for both groups the highest accuracy rate seen for the tonal condition was about 86% at Gate 12, or about 200 ms into the rime. This means that having heard half of the tone and rime, listeners were approximately 86% accurate at identifying which tone they were hearing. Given the drastically different onsets and slopes of these tone pairs, it is perhaps surprising that the accuracy rates were not higher at that point. However, these results are consistent with other work on the categorization of tones that showed difficulty for native listeners when making explicit tone judgments (e.g., Cutler & Chen, 1997; Taft & Chen, 1992).

4.1.6.2. Chinese vs. English listeners' categorization of segments

In terms of the segmental contrasts, the results of the GCA shown in Table 4 revealed no differences between native Chinese listeners and English-speaking L2 learners of Chinese in the proportions of target responses. This was confirmed by the almost identical response curves seen in Figure 6. These results were as predicted, since the segmental contrasts chosen for this task existed in both Chinese and English.

4.1.6.3. Categorization of tones vs. segments

When comparing the performance on the tonal and segmental conditions across language groups, two main conclusions can be drawn. First, the results presented in Table 4 revealed that across all gates, there were more correct responses in the segmental condition than in the tonal condition, as indicated by the significant effect of condition for native Chinese listeners and English-speaking L2 learners of Chinese. The lower accuracy for tones than for segments with the same duration of acoustic input is consistent with other non-gated studies on the topic. Recall that Taft and Chen (1992) found more errors for native Chinese and Cantonese listeners to judge homophony of word pairs that contained tonal changes compared to segmental changes. Likewise, Cutler and Chen (1997) found that native Cantonese listeners made more errors when the prime heard mismatched the target word in tone than when the mismatch was in segments. These results support the conclusions made by previous work in showing lower accuracy for the identification of tonal contrast compared to segmental contrasts.

Additionally, in terms of time course, we see that the improvement rate for the categorization of segments is greater than that for tones. This means that, as the duration of acoustic information increases in equal increments, the number of target responses increases more for segments compared to tones. While this is consistent with previous studies showing an advantage for segments over tones, the present results are novel in that it shows a direct relationship between the segment advantage in terms of accuracy (Cutler & Chen, 1997; Taft & Chen, 1992) and the information accumulation notion discussed in other gated studies with non-naturally produced stimuli (Ritsma et al., 1965; Robinson & Patterson, 1995). Recall that it has been proposed that pitch information requires a longer duration to perceive than vowel information, since the acoustic information needs to accumulate in order to apply it. Therefore, these results show a relationship between the duration of information needed to perceive pitch information, and the duration of acoustic input needed to categorize pitch into linguistic categories. Not only does the categorization of segments have an accuracy advantage on the whole, but also, with each incoming unit of information, listeners are able to make greater use of the segmental information than of an equal amount of tonal information. These results support the results of Ritsma et al. (1965) by showing that tones need an extended amount of information to be explicitly identified. Additionally, recall that Robinson and Patterson (1995) found that participants could identify the vowel with less acoustic input than they could identify the musical note on the same vowel. The present results extend this conclusion into the linguistic realm by showing the same results with naturally produced linguistic stimuli and linguistic pitch categories as opposed to pitch in terms of musical notes.

From these findings, one important question that remains is, why do tones and segments differ in this way? That is to say, why can segments be more reliably identified than tones with the same duration of acoustic information? As discussed previously, these results may be attributable to the number of cues that signal segmental versus tonal contrasts. Vowel perception (relevant here since the rimes differed in vowels) is arguably most dependent on the relationship between F1 and F2; however, other cues such as center of gravity, skewness, kurtosis, and variance for the preceding consonant, as well as features of the actual vowel such as F3 and, specifically for diphthongs, the information in the formant transitions is also relevant. By contrast, in the present study, the F0 contour was the only cue to tonal identity, since both duration and intensity were normalized. This means that vowels may have had many more cues to their identity than tones. This leaves tones requiring more of the acoustic signal to reliably identify, since the reliability of the cue is less than what can be taken from a rime portion of equal length.

4.1.7. CONCLUSIONS

The gated category-identification task presented in this chapter revealed two significant results. First, unlike what was predicted, English-speaking L2 learners of Chinese did not have an advantage in categorizing the tones of Chinese over native Chinese listeners at early points in the tone. Second, as predicted, segments had accuracy and improvement rate advantages over tones.

The present study showed that English listeners did not differ from Chinese listeners in their performance on either tonal or segmental contrasts. However, this task was non-lexical and offline, leaving open the question of what occurs during lexical and online tasks: Do English L2 learners of Chinese still perform similarly to native Chinese listeners on the use of tonal and segmental contrast in online spoken word recognition? Do segments have an advantage over tones for Chinese and English listeners when processed online? Are tones and segments processed on a different time scale (as in Experiment 1) or on the same time scale (as documented in previous research, (e.g., Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011). Before moving on to a second experiment that aims to answer these and other questions, we move on to a discussion of the timing of use of tones and segments in L1- and L2-Chinese lexical access.

Part II

CHAPTER 5: THE TIMING OF USE OF TONES AND SEGMENTS IN LEXICAL ACCESS

5.1. INTRODUCTION

The previous chapter focused on Chinese and English listeners' ability to categorize tones and segments. Both the timing of accurate categorization and the rate at which target responses improved with gates were compared; however, the forced-choice gating task used to investigate these issues was offline and inherently non-lexical. Part II of this dissertation aimed to investigate Chinese and English listeners' use of tones and segments using an online lexical measure: visual-world eye-tracking. Recent research targeting the online use of tones and segments in lexical access (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011) has yielded different results from those of earlier offline studies (e.g., Cutler & Chen, 1997; Taft & Chen, 1992), showing no disadvantage for tones compared to segments. By comparing the results of Experiment 1 to those of Experiment 2 (presented in Chapter 6) using a task that targets the online use of tonal and segmental information in lexical access with comparable materials and with the same participants, it can be determined if tones are in fact disadvantaged in general, or if this disadvantage is specific to certain types of tasks. Additionally, using an online lexical measure can investigate issues specific to native and non-native Chinese listeners' use of tonal information in L2 lexical processing.

The literature on native listeners' use of tones and segments has suggested that tones and segments begin constraining the lexical search simultaneously (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011), (as opposed to a 2-stage process proposed by Lee, 2007). However, only one of these studies produced results that speak to the rate at which these two types of information are used as the speech signal unfolds over time (henceforth referred to as speed of use) (Malins & Joanisse, 2010), with this study showing no speed-of-use difference between the tonal and segmental conditions.

Within the research on L2 learners' use of suprasegmental features in lexical processing, lexical stress has received the most attention (e.g., Cooper et al., 2002; Cutler, Wales, Cooper, & Janssen, 2007), with suprasegmental categories such as lexical tone left largely unstudied. To date, only one study (to the best of my knowledge) has investigated the online processing of tones in an L2 (Sun, 2012). Using priming, Sun (2012) showed that while English-speaking L2 learners processed Chinese words more slowly than native Chinese listeners, they displayed the same pattern of results as native speakers (as seen in Experiment 1 of the present work as well), in that they showed no difference in the timing with which they used tones and segments in online word recognition. In the realm of lexical processing, these results are surprising, in that learners typically show non-native like use of features/cues not present in their L1 (e.g., Dupoux, Pallier, Sebastian, & Mehler, 1997; Dupoux, Peperkamp, & Sebastián-Gallés, 2001; Dupoux, Sebastian-Galles, Navarrete, & Peperkamp, 2008; Tremblay, 2009). Second language learners of Chinese must first form the tonal categories before they can use them in online spoken word recognition. In other words, one might expect that L2 learners might struggle with the use of tonal information in lexical access.

The experiment presented in the following chapter (Experiment 2) aimed to re-investigate this conclusion using a more time-sensitive measure: eye-tracking. Eye-tracking is more informative than priming in that it provides detailed time course data about the activation and competition between differing lexical items and may reveal differences in how English-speaking L2 learners of Chinese use tones and segments in Chinese spoken word recognition. Before turning to Experiment 2, the present chapter reviews the tonal and segmental processing literature for native and L2 Chinese listeners.

5.2. Lexical tone processing in native tone language listeners

In order to understand L2 learners' use of tones in lexical access, it is important to review the existing literature on native listeners' use of tones in lexical access (for comparison). The literature on native listeners' lexical tone processing reveals a debate about when tone is used in relation to segments in lexical access, and the relative weighting of each type of information (where weighting refers to the degree of importance placed on each type of information). Although the present study focuses on the relative timing of use of tonal and segmental information, the issue of relative weighting, or which of tonal or segmental information has more influence on lexical access, inevitably becomes relevant. Of the studies that have investigated the use of tonal and segmental information in word recognition, some report that tone is used later than segments (Chen & Cutler, 1997; Lee, 2007; Taft & Chen, 1992), whereas others have claimed that tone and segments are used concurrently (e.g., [Cantonese] Cutler & Chen, 1995; [Mandarin Chinese] Malins & Joanisse, 2010). Ultimately, however, several of these studies present results that address the issue of the weighting of tones and segments rather than the issue of the timing.

For example, Cutler and Chen (1995), who conducted a priming experiment in Cantonese, showed that a prime-target mismatch in tone (e.g., *ji6liu4* 'treatment' – *ji6liu5* 'feed') and a prime-target mismatch in segments (e.g., *to4fa1* 'peach flower' – *to4foo1* 'butcher') created similar inhibition effects in response times to the target. The authors claim that this supports the view that tones and segments are used in the same way to constrain the word search. However, as discussed in Part I of the dissertation, Cutler and Chen (1997) found that native Cantonese listeners made more errors in a priming task when the prime word mismatched the target word in tone (e.g., *dzi1-gam1 – dzi1-gam2*) and in an AX discrimination task when the two words contained the same segments but a different tone (AX discrimination; *ma3 – ma2*) than when they contained the same tone but a different segment (e.g., *dzi1-gam1 – dzi1-ham1*; *ma3 – na3*).[14] From these results, and counter to their previous claim, the authors argued that Chinese listeners may have more difficulty using tonal information compared to segmental information.

Notice, however, that reaction times from priming experiments with a single ISI and AX discrimination responses to complete words do not shed direct light on the *timing* of the use of tonal and segmental information. Decreases in overall response times and accuracy rates following a tone mismatch as compared to a segmental mismatch might instead reflect the overall *weight* of tonal and segmental cues. Hence, the results of Cutler and Chen (1995, 1997) may be more relevant to a discussion of weighting of tonal and segmental information rather than one of timing (see also Sereno & Lee, 2015).

One way to test for the timing of use of tonal and segmental information in word recognition is to conduct a priming experiment in which the ISI between the prime and target is varied. By varying the ISI, it is possible to test whether specific cues constrain lexical access similarly at the same word recognition stage. For example, Andruski, Blumstein, and Burton (1994) showed that sub-phonetic variation causes gradience in priming effects at a 50-ms ISI; however, no gradience was seen at a 250-ms ISI. The authors interpreted these results as suggesting that sub-phonetic variation impacts lexical access only in early word-recognition

[14] English glosses were not provided in the original paper.

stages. Using different ISIs can thus give insight into whether specific cues to word recognition impact lexical access *on the same timescale.*

This is precisely the method used by Lee (2007), who conducted several priming experiments that examined how tonal match (e.g., *lóu* 'hall'-*lóu* 'hall') and tonal mismatch (e.g., *lǒu* 'hug'-*lóu* 'hall') affected the recognition of the target word. Word pairs were presented auditorily one after another, and participants were asked to indicate whether or not the second word in the pair was a real Chinese word or not. Both experiments, with ISIs of 50 ms and 250 ms, showed only repetition (e.g., *lóu* 'hall'-*lóu* 'hall') priming effects; segmentally matched pairs that differed in tones (e.g., *lǒu* 'hug'-*lóu* 'hall') showed no priming effects. These results were later replicated with different stimuli by Sun (2012).

To ascertain whether the lack of priming effects for the segmental pairs was due to the mismatching tone inhibiting lexical access, a similar experiment was conducted using mediated priming in which the author instead used a target that was semantically related to the prime. The new sets included primes with a direct semantic relationship with the target (e.g., *lóu* 'hall' – *jiànzhu* 'building'), called the semantically related condition, and primes that were segmentally identical to the primes in the previous condition, but differed in tones (e.g., *lǒu* 'hug' – *jiànzhu* 'building'), called the non-semantically related condition (this tone difference causes the prime to have no semantic relationship to the target in this second condition). Priming for the first pair is expected, since there is a direct semantic relationship between the prime and the target. Priming is expected in the second pair only if the tonal information in *lǒu* 'hug' is not sufficient to inhibit lexical competition from the semantically related prime *lóu* 'hall'.

The results showed that, at a 250-ms ISI, only the primes in the semantically related condition yielded significant priming effects. By contrast, and contrary to his previous experiment, at an ISI of 50 ms, there was a significant priming effect for both the semantically related and the non-semantically related pairs, indicating that the segmentally identical words that differed in tone were also activated. On the basis of these results, the author claims that at early stages of lexical access, tone is not used to constrain the word candidate list, and thus there is priming between *lǒu* 'hug' and *jiànzhu* 'building,' since all words with the segmental structure of *lou* have been activated. At later stages (as shown by the 250-ms ISI results), however, tone is used to select among the segmental candidates already active. This is shown by the significant priming effect for only the semantically related pairs, with *lǒu* 'hug' no longer priming *jiànzhu* 'building'. The author concluded that "[the] auditory presentation of a Mandarin word activates its minimal tone pairs in the early phase of lexical activation, but tonal information is used soon afterwards to resolve segmental ambiguity, ruling out candidates that are mismatching in tone" (Lee, 2007, p. 188).

Lee's (2007) results suggest that segmental and tonal information may be used on a different time course in lexical access, with segmental information being used immediately to constrain the lexical search and with tonal information being used later. However, there are concerns with the design and controls used in Lee (2007). First, even with the varied ISIs, the priming technique does not shed direct light on the time course of use of segmental and tonal information in lexical access.

Second, as noted by Sereno and Lee (2015), the materials were not carefully controlled, and many confusable tone pairs were used (e.g., Tone 2 – Tone 3) in an unbalanced way, which could have contributed to the pattern of results. Sereno and Lee (2015) conducted an auditory-auditory lexical decision priming task similar to that of Lee (2007); however, in this experiment the number of each tone being presented as primes and targets were balanced. The task included targets such as ru4 paired with identity primes (ru4 – ru4), segmental primes (ru3 – ru4), tonal primes (sha4 – ru4) and an unrelated condition (qin1 – ru4). Targets and primes were represented with a 50 ms ISI only. Timing was not investigated. The results showed that when the prime and target match in segments and tones, there was significant facilitation compared to the unrelated condition. This facilitation was also seen when only segments overlapped, but to a lesser extent than when both segments and tones matched. When overlap was only in the tones, there was significantly inhibition compared to the unrelated condition.

These results cast doubt on the conclusion of Lee (2007), since neither the segmental only priming nor the tonal only inhibition was found by Lee (2007). Sereno and Lee (2015) claim that this is due to the unbalanced selection of tones and tones pairs used in Lee (2007). Since the results of Lee (2007) appear to be inconclusive, and Sereno and Lee (2015) did not test the timing of use of tones and segments, the timing of use of tones and segments was still an open question. Additionally, priming may not be sufficiently time sensitive to capture the differences between the timing of use of tones and segments.

An arguably more informative measure of the time course of use of lexical processing is visual-world eye-tracking. Eye-tracking has been well established as a method to examine the online time course of lexical access (e.g., Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001; Marian & Spivey, 2003; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1996) and provides detailed time course information about the competition between multiple lexical items.

Recently, Malins and Joanisse (2010) conducted a visual-world eye-tracking study to investigate the time course of the integration of tonal and segmental information in native Chinese listeners' lexical access. Participants heard monosyllabic words in isolation while looking at an array of four pictures that corresponded to a target word, a competitor word, and two distracter words. The experiment included a condition where the target and competitor contained the same segments but differed in tone (tonal condition), and a condition where the target and competitor contained the same tone but differed in (rime) segments (segmental condition).[15] For example, for the target *chuáng* [t͡ʃuaŋ2] 'bed', the competitor in the tonal condition was *chuāng* [t͡ʃuaŋ1] 'window' and the competitor in the segmental condition was *chuán* [t͡ʃuan2] 'ship'.

The stimuli in Malins and Joanisse (2010) were naturally produced Chinese words. The authors report that across all tokens and items, there were no timing differences between the tonal and segmental conditions in terms of when the relevant acoustic information became reliable in the signal. In other words, across all of their stimuli, on average the target and competitor words in the tonal condition disambiguated in the acoustics at the same point in time as the target and competitor words in the segmental condition.

The authors argued that if tonal and segmental information are available on the same timescale (i.e., if the tonal competitor *chuāng* [t͡ʃuaŋ1] 'window' competes with the target *chuáng* [t͡ʃuaŋ2] 'bed' on the same timescale as the competitor *chuān [t͡ʃuan2]* 'ship' competes with the same target *chuáng* [t͡ʃuaŋ2] 'bed', then eye fixations in the tonal condition should pattern similarly to eye fixations in the segmental condition. The authors conducted growth curve analyses on the fixations to target items. The results showed no significant difference between the two conditions in listeners' target fixations. The authors took these results as evidence that tonal information is processed on the same timescale as segmental information.

[15] Malins and Joanisse (2010) refer to the segmental condition as "cohort" condition. Their experiment included additional conditions, not mentioned here since they are not relevant to the present discussion.

Similarly, Schirmer et al. (2005) investigated the time course of tonal and segmental information in Cantonese using event related potentials (ERP). The authors created sentences with a critical word (e.g*., beng6* [peŋ6] 'illness') and then manipulated either the tone (e.g., *beng2* [peŋ2] 'biscuit') or the segments (*e.g., bou6* [pou6] 'step') to create two conditions of semantically incongruous words. When measuring the timing and amplitude of the N400 ERP component, a negative going waveform that has been shown to reflect the detection of semantic violations, the authors found similar latencies and amplitudes for both types of violations. The authors concluded from these results that tones and segments play similar roles in online spoken word processing in Cantonese in terms of their timing and degree of influence.

These results of comparable tonal and segmental timing have since been replicated in Mandarin, also using ERPs. Zhao et al. (2011) developed a new ERP task called the "picture/spoken-word/picture task" to investigate the timing of competition between pairs of words. In this task, participants saw two pictures one after the other, and their task was to decide if the objects represented in the two pictures were from the same semantic category (e.g., animals, body parts, etc.). Between the presentations of the two pictures, participants heard a spoken word. They were instructed not to do anything with that word, but simply to listen to it and make a judgment about whether the second picture matched the semantic category of the first picture. The spoken word either matched the word represented in the second picture in all aspects (match condition*,* e.g., *bi2 – bi2)* or mismatched it in terms of onset (onset mismatch condition, e.g., *bi2 – li2*)*,* rime, (rime mismatch condition, e.g., *bi2 – bo2*)*,* tone (tone condition, e.g., *bi2 – bi3*), or syllable (syllable condition, e.g., *bi2 – ge1*)*.* With this task, the authors hoped to avoid having participants make explicit judgments during spoken word processing, but to still capture effects due to lexical competition between the spoken word and the second picture.

The authors investigated the latency and amplitude of the N400 ERP component, which has also been shown to be modulated by rime matches, showing an earlier N400, and rime mismatches, showing a later N400, in response to spoken Chinese (Desroches, Newman, & Joanisse, 2009). The authors predicted that the various violations between the expectation set up by the spoken word and the violation of that expectation upon seeing the visual picture would influence the N400 latency and peak, thereby shedding light on the nature and linguistic units of Chinese lexical competition. Of their findings, the most critical to the present discussion was the lack of difference between the tone condition (e.g., *bi2 – bi3)* and the rime condition (e.g., *bi2 – bo2)*: Both conditions elicited comparably larger N400s compared to the match condition (e.g., *bi2 – bi2)*, due to the failed expectation set up by the spoken word. These results indicate no timing difference in the use of tonal and segmental information in lexical competition, contrary to the claims of Lee (2007).

Malins and Joanisse (2012) also conducted an ERP experiment to investigate the timing of tonal and segmental information in Chinese. The authors argue that while Zhao et al. (2011) found no difference between tonal and segmental mismatches, the task did not require lexical access and thus cannot be said with certainty to reflect lexical processing. Malins and Joanisse (2012) instead used a picture-word matching task. Participants saw a picture presented on the screen and heard an auditory word. They would then press one of two buttons to indicate whether the picture matched the spoken word. The pictures and auditory words were manipulated to exhibit different match/mismatch relationships (match, e.g., *hua1* [xua1] 'flower' – *hua1* [xua1] 'flower'; tonal mismatch, e.g., *hua1* [xua1] 'flower' – hua4 [xua4] 'painting'; rime mismatch, e.g., *hua1* [xua1] 'flower' – *hui1 [*xui1] 'gray'; onset mismatch, e.g*., hua1* [xua1] 'flower' – *gua1* [kua1] 'melon'; syllable mismatch, e.g., *hua1* [xua1] 'flower' – *jing1* [tɕiŋ1] ‘whale’; unrelated, e.g., *hua1* [xua1] ‘flower’ – *lang2* [laŋ2]‘wolf’; note: the tonal and rime mismatches are the conditions of interest to the present discussion).[16] In that study, the ERP component of interest was the phonological mapping negativity (PMN), which is an early negative deflection believed to be related to pre-lexical processing. This component has been shown to be modulated by word-initial phoneme mismatches between expected and observed words (e.g., Connolly & Phillips, 1994; Connolly, Stewart, & Phillips, 1990). The results showed that the two conditions of interest (tonal mismatch, e.g., *hua1* ‘flower’ – *hua4* ‘painting’; and rime mismatch, e.g., *hua1* ‘flower’ – *hui1* ‘gray’) showed similar timings of PMN responses. The authors concluded that tones and segments are used at the same time and are used as early as the information becomes available.

These studies presented here, which used different materials, tasks, and measures, have all come to the conclusion that tones and segments are used at the same time in lexical processing. When discussing learners, we can be sure of the assertion that a native-like timing pattern would be for tones and segments to be used concurrently in lexical access.

5.3. Lexical processing in second language listeners

5.3.1.1. Lexical stress

Although not much work has been done on the L2 processing of tones, there is a substantial literature on how learners use stress in L2 lexical access. Stress in English is signalled by segmental cues such as vowel reduction (e.g., Gay, 1978; Lindblom, 1963) and suprasegmental cues such as duration, intensity, and pitch (e.g., Beckman, 1986; Lieberman, 1960). Looking at what is known about the use of stress in L2 lexical access will inform the predictions about the

[16] These are referred to as the “match,” “segmental,” “cohort,” “rime,” “tonal,” and “unrelated” conditions in Sun (2012). The labels were changed here for consistency with the present study.

use of tonal information in L2 lexical access. Hence, the literature on L2 learners' processing of lexical stress is discussed first to lay the foundation.

Beginning with how native English listeners use stress, Cutler (1986) originally suggested that since most stress minimal pairs in English are differentiated with segmental cues such as vowel reduction, English listeners would not use suprasegmental cues to stress in the presence of segmental cues to disambiguate minimal pairs. This prediction was tested using cross-modal priming with minimal pairs of words that differed only in the suprasegmental cues to stress, such as *FORbear – forBEAR* (with the capital letters indicating stress placement), where no vowel reduction is present in the unstressed syllables. Participants heard sentences ending in the prime words and made lexical decisions to a visually presented semantically related target word. The targets were either semantically related (e.g., *FORbear – ancestor*) or unrelated (e.g., *FORbear – tolerate*) to the prime. It is crucial to note that in the semantically unrelated pairs, the target was semantically related to the prime word's stress minimal pair (e.g., *forBEAR*). It was predicted that if English listeners could use the stress information to constrain the word search, then a semantically unrelated target (e.g., *tolerate*) should not be primed by the semantically unrelated prime word (e.g., *FORbear*), since stress would have ruled it out: If *forBEAR* were not initially activated, it would not activate its semantically related word (e.g., *tolerate*).

The results showed equal priming for semantically related and semantically unrelated items, suggesting that English listeners did not use the suprasegmental information when accessing the lexicon, since *FORbear* was treated as a homophone of *forBEAR.* It was explained that in English, suprasegmental cues to stress are insufficient to constrain the lexical search.

There are pairs that are disambiguated solely by suprasegmental cues, but that they are so infrequent that English speakers do not need to use this information to access the lexicon.

In a later study, however, Cooper et al. (2002) questioned the power of the results of Cutler (1986), which had used only stress minimal pairs and thus had very few test items. Cooper et al. (2002) hypothesized that stress may play a role in earlier stages of word recognition, before the whole word has been heard. For this reason, they retested native English listeners' ability to use suprasegmental cues to stress in lexical access and additionally included Dutch-speaking L2 learners of English. The stimuli included were word pairs whose first or first two syllables were identical segmentally but differed in stress, such as *ADmiral – admiRAtion*. Participants heard a sentence ending with the truncated critical word (e.g., *admi*-). At this point, they were presented with a visual target that matched or mismatched the truncated word in stress (e.g., "admiral" or "admiration") and made a lexical decision to the word. If participants were able to use the suprasegmental cues to stress in the initial syllables, upon hearing the first two syllables of *admiral,* there should be no facilitative priming effects when the target word was *admiration,* since *admiration* does not match *admiral* in its stress. Over the course of several experiments, they found that English listeners were able to use the suprasegmental cue to stress to constrain lexical access, contrary to the results of Cutler (1986). It was argued that the earlier results were conducted on too small a set of words, since English has few pairs of that differ suprasegmentally but not segmentally, and as a result failed to reveal the effect. The authors also suggest that identity priming may be more sensitive to the subtle effects of suprasegmental cues than associative priming, as was used in Cutler (1986).

In addition to establishing that English listeners use stress in online lexical access, this study included a group of Dutch-speaking L2 learners of English. Dutch is very similar to

English in terms of stress, with the exception that an unstressed vowel does not reduce to the extent that it would in English. This means that in Dutch, the suprasegmental cues to stress are weighted more heavily in identifying stress than in English. The results showed that the Dutch listeners were able to use the stress information to correctly constrain the lexical search in their L2, and in some cases, they even made greater use of the suprasegmental cues to stress than native English listeners. This greater use of suprasegmental cues to stress than native English listeners was also found by Cutler et al. (2007).

Cooper et al. (2002) showed that when both the L1 and L2 use stress to signal lexical identity, L2 learners have little difficulty in using stress to constrain the L2 lexical search and access L2 words. However, other studies have shown that in cases where the L1 and L2 differ in their use of stress, L2 learners have great difficulty using stress in L2 lexical access. French differs from English or Dutch in that it does not have stress: Prominence is not lexically contrastive and is instead used to mark word boundaries, falling on the final syllable of phrase-final words (according to Dell & Vergnaud, as cited in Dupoux, Pallier, Sebastian, & Mehler, 1997; see also Jun & Fougeron, 2000; Jun & Fougeron, 2002). Studies on French speakers learning lexical-stress languages have reliably shown a stress "deafness," that is, an inability to encode stress in tasks such as AXB, sequence recall, and speeded lexical decision (Dupoux et al., 1997; Dupoux et al., 2001; Dupoux, Sebastian-Galles, et al., 2008; Tremblay, 2009). It is argued that this stress "deafness" is not so much a lack of ability to perceive the stress differences, but rather an inability to encode stress in phonological representations (Dupoux, Sebastián-Gallés, Navarrete, & Peperkamp, 2008). The studies by Dupoux and colleagues were the basis for the Stress Parameter Model (SPM; Peperkamp & Dupoux, 2002), which argues that if stress is not encoded in the L1 lexical representations, listeners will not be able to encode stress in L2 lexical representations and use it in online word recognition.

Along this line, Tremblay (2008) investigated the use of stress in L2 lexical access in relation to proficiency and knowledge of stress placement by Canadian French-speaking L2 learners of English. By correlating the results of a cross-modal priming task with the results of a stress production task, the author showed two major effects relevant for the current discussion. First, knowledge of stress placement was essential, but not always enough, for L2 learners to use stress in processing. This means that all learners who were able to use stress in processing had good knowledge of stress placement, but not all participants who had good knowledge of stress placement were able to use it in processing. While accuracy for some learners was high (up to 80%), many learners struggled to use this information online, with L2 learners' RTs remaining higher than native English participants. These results were interpreted as suggesting that the difficulty for French speakers comes from lexical access itself, and not necessarily from an inability to encode the stress in the lexicon, given the high stress production accuracy for some learners. The second important effect was the lack of effect of proficiency: The results showed that learners struggled to use stress information to access the lexicon in general, and that increased proficiency did not increase the likelihood that the learners would be able to use stress to access the lexicon. Years of experience with English, however, did significantly predict stress use in L2 lexical access.

The results for French speakers learning stress languages collectively show that when an L1 does not have lexical stress, learners struggle to use stress in L2 lexical access. Additionally, this effect does not seem to be merely an effect of general proficiency in the L2, with participants who have reached a high level of proficiency still failing to use stress online (Tremblay, 2008).

Lin, Wang, Idsardi, and Xu (2014) conducted a study to test the predictions of the SPM with Chinese- and Korean-speaking L2 learners of English by contrasting the presence and absence of segmental cues to stress (i.e., vowel reduction). According to the SPM, listeners must encode stress in their L1 lexical representations to use it in an L2. Based on the SPM, the authors predicted that Chinese listeners, who encode stress in the L1, would succeed at using stress online, whereas Korean listeners, who do not have lexical stress in the L1, would fail to do so. In both sequence recall and lexical decision tasks, the English and Chinese listeners outperformed the Korean listeners, both in the presence and in the absence of vowel reduction. This was taken as support for the SPM, with Chinese listeners being able to use lexical stress in English but with Korean speakers being unable to use stress in English.

However, a study on L2 learners' processing of stress that is framed within the cue-weighting theory of speech perception (Francis, Ciocca, Ma, & Fenn, 2008; Francis & Nusbaum, 2002; Holt & Lotto, 2006; Ingvalson, Holt, & McClelland, 2012; Zhang & Francis, 2010) has shown that what is more important to stress processing is the specific cues used to process stress, not the presence or absence of stress in the language (Qin, Chien, & Tremblay, 2017). Specifically, Qin et al. (2017) tested L2 learners of English in their ability to use duration as a cue to stress in English based on the properties of their L1. In Standard Mandarin Chinese, duration serves a lexical stress function, in that disyllabic words contrast based on a stressed-stressed or stressed-unstressed pattern (e.g., Duanmu, 2007), where the second syllable in a stressed-unstressed word has a shorter duration that the second syllable in a stressed-stressed sequence (Chen & Xu, 2006; Lin & Yan, 1980, as cited in Qin et al., 2017). Taiwan Mandarin, on the other hand, does not have this contrast, and so duration does not play a significant role in signaling lexical identity (Huang, 2012).

The authors tested proficiency-matched native Standard Mandarin and Standard Taiwan Mandarin L2 learners of English as well as a control group of native English listeners with a sequence recall task to test their use of durational cues in identifying stress in English non-words. Stimuli were manipulated to contain only durational cues, only F0 cues, both cues, or conflicting cues (i.e., duration indicating stress, but F0 indicating unstressed and vice versa). The results showed that when only durational cues to stress were present, native Standard Mandarin listeners outperformed native Taiwan mandarin listeners. These results were attributed to Standard Mandarin listeners' use of durational cues in their L1, and the lack of reliance on these cues in Taiwan Mandarin. Additionally, when only F0 was a cue to stress, Taiwan Mandarin listeners did not differ from Standard Mandarin listeners. Recall that Taiwan Mandarin does not have lexical stress, and so this shows that these Taiwan Mandarin learners of English were able to transfer their use of pitch from their L1 tones to process their L2 English stress. Finally, when the cues conflicted, the results showed that both Taiwan and Standard Mandarin listeners relied more on F0 than they did on duration. This is true of the Standard Mandarin listeners as well, who could have relied on duration, which is a cue to stress in Standard Mandarin as well. These results again suggest that it is the cues to lexical identify that are important, not the presence or absence of stress, since Taiwan Mandarin does not have lexical stress, and yet they were able to use pitch alone as a cue to English stress as well as Standard Mandarin listeners, who do have stress in their L1.

Thus, moving to English speakers learning Chinese lexical tones, while the SPM would have predicted that L1-English L2-Chinese listeners would have failed to use lexical tones online since they do not have tones in the L1, the cue weighting theory suggests that L1-English L2-

Chinese listeners should have the ability to encode lexical tones, in that they should be able to transfer their use of pitch from lexical stress to L2 lexical tones.

5.3.1.2. LEXICAL TONE

The previous section discussed how stress is processed in an L2. This work predicts that L1-English-L2 Chinese listeners should be able to encode tones, given their ability to transfer their use of pitch as a cue to stress to being a cue to tones. While it is predicted that they will be able to encode the tones, it remains unclear if they will be able to use those tones in online lexical processing. As previously mentioned, only one study has directly investigated L2 learners' tone processing in Chinese. In his dissertation, Sun (2012) examined native and L2-Chinese listeners' processing of tones. More specifically, using two priming experiments, it investigated tone neighborhood density effects and the timing of use of segmental and tonal information. Only the experiment on the timing of use of the two types of information will be discussed here.

Experiment 2 was a form-priming task based largely on Lee (2007). Participants heard two words and were asked to judge whether or not the second word in the pair was a real word of Chinese. Prime-target pairs included an identity condition (e.g., *mā* [ma1] 'mother'-*mā* [ma1] 'mother'), a tonal condition where the pairs shared segments but differed in tones (e.g., *mǎ* [ma3] 'horse' – *mā* [ma1] 'mother'),[17] a segmental condition where the pairs shared the tone but differed in all segments (e.g., *tū* [tʰu1] 'bald' – mā [ma1] 'mother') [18], and an unrelated condition where the pairs shared nothing (e.g., *fó* [fo2] 'Buddha' – *mā* [ma1] 'mother'). Unlike Lee (2007), the author controlled for tonal neighborhood density by selecting only words that

[17] This was referred to as the "segmental" condition in. Sun (2012). The labels were changed here for consistency with the present study.
[18] This was referred to as the "tonal" condition in the paper. Sun (2012). The labels were changed here for consistency with the present study.

could appear as real words with all four tones, but it is unclear whether it also controlled for segmental neighborhood density or word/token frequency. Two ISIs were used to investigate the timing of use of tonal and segmental information (50 ms and 250 ms), as was used in Lee (2007). The results showed that for native listeners and L2 learners, in both ISI conditions, only identity pairs produced facilitation compared to the unrelated pairs, and no inhibition effects were found for either segmental or tonal pairs. Overall, L2 learners' reaction times were slower and their accuracy rates lower than those of native listeners. The author claims that these L2 results show that L2 learners process Chinese similarly to native listeners, but with their processing being slower.

The author argues that one explanation for the L1 effect (that L2 learners were in general slower to process the words) was because tones are too similar to English stress. The author discusses perceptual models such as the Speech Learning Model (SLM; Flege, 1995) and the Perceptual Assimilation Model (PAM; Best, 1995). While these models differ slightly, their main argument is the same: Both claim that when an L2 sound is similar enough to an L1 sound, this L2 sound will be "assimilated" to the L1 category. If the L2 sound is inappropriately assimilated to an L1 category, learners will not be able to separate the new L2 sound from the existing L1 sound. This will cause learners not to perceive the difference between the two; consequently, acquiring the new sound in perception and production will be extremely difficult. The author cites these models to explain that since both tone and stress are similar in their acoustic properties (signaled in part by pitch), English-speaking L2 learners of Chinese may equate lexical tone to stress. Because they assimilate the tones to stress, they will have difficulty perceiving the differences between the tones, and therefore this will affect their lexical access; however, this explanation was not directly tested.

Additionally, it remains unclear why no priming was found for the tonal condition, where all the segments overlapped and only the tone mismatched. An examination of Sun's (2012) segmental items reveals that the distribution of the tones in the target words was not balanced, with seven T1 targets, two T2 targets, seven T3 targets, and four T4 targets. The distribution of the tones in the prime words varied just as much. Sereno and Lee (2015) showed that when the number of appearances of each tone are controlled, segmental-only overlap causes facilitative priming, though not as much as the identity condition. This effect was not seen in Sun (2012) and could be due to a lack of controls in the materials. The tone pairs in the tonal condition were also not evenly distributed, with 10 out of 20 pairs being Tone 1 – Tone 3 pairs. While this unbalancing cannot directly cause the effects seen, the sheer variation in the tones and tone pairs is more than enough to cast doubt on the results. It is also unclear whether the frequency of the words and L2 learners' familiarity with these words was controlled. Especially for L2 learners, word familiarity is an important measure to ensure participants were equally familiar with the words in all conditions.

In sum, there has been only one study that investigated L2-Chinese learners' use of tonal information in lexical access. That study used priming and varied the ISI to investigate the timing of use of tones and segments, and concluded that learners use tones and segments at the same time, but process slower overall compared to native speakers.

In terms of processing segmental information in the L2, a substantial body of literature indicates that when the L2 has sounds that correspond to a single category in the L1, learners struggle to use those segments in online lexical access. Broersma and Cutler (2008) showed that Dutch-speaking L2 learners of English activate English near-words due to their misperception or incorrect identification of vowels. Dutch listeners often confuse the English vowels /æ/ and /ɛ/, since they do not contrast in the L1 (Schouten, 1975). In order to investigate if this difficulty created additional lexical competition for L2 learners of English as compared to native English listeners, Broersma and Cutler (2008) conducted two experiments. The first was an auditory lexical decision task looking to see if Dutch L2 learners of English would incorrectly accept near-words such as [lɛmp] by mis-categorizing the vowel [ɛ] as [æ], which would create the real English word [lamp]. The results showed the Dutch L2 learners of English incorrectly accepted the near-words significantly more than native English listeners, showing that the L1 segmental inventory does interfere with correct lexical access.

The first experiment showed that difficulty in learning vowels in the L2 causes learners to incorrectly accept nonce words as words of English. In a second experiment, Broersma and Cutler (2008) investigated whether this lack of a vowel contrast in the L2 could create additional lexical competition for learners by resulting in embedded words that would not be present in native listeners, who can correctly identify the vowels. They investigated if Dutch listeners would activate the word *deaf* upon hearing the first syllable of the word *daffodil* due to the misperception of the first syllable as [dɛf-] instead of the correct [dæf-]. The authors conducted a cross-modal priming experiment using word pairs such as *definite* and *daffodil*. The initial syllable was extracted from each word (e.g., [dɛf-] and [dæf-]) and served as the auditory prime in two conditions: either the same-word condition (prime: [dɛf-], visual target: *deaf*) or in a near-word condition (prime: [dæf-], visual target: *deaf*). For English listeners, less facilitation (or no facilitation) should be found in the near-word condition compared to the same-word condition, since the vowel will reduce the likelihood that the target word would compete. For Dutch L2 learners of English, however, if the incorrect representation of the vowel leads to unwanted lexical competition, then we would expect to see facilitation in both cases.

Native listeners' results showed facilitation only in the same-word condition, with no facilitation in the near-word condition. For Dutch L2 learners of English, however, facilitation of the target word was seen in both the same-word and near-word conditions. The authors concluded that the mis-identification of the L2 sounds created additional lexical competition when processing the L2: not distinguishing between these two sounds resulted in the activation of not only the intended word (e.g., *daffodil*), but also words that did not share the same vowel (e.g., *deaf, definite*).

This pattern of results of incorrect representations impacting L2 lexical processing with increased competition has been found not only for Dutch L2 learners of English (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011), but also for Japanese L2 learners of English due to the difficult [r]/[l] distinction (Cutler & Otake, 2004; Cutler, Weber, & Otake, 2006). These findings may further extend to the L2 learners' processing of Chinese tones. The four tones of Chinese do not map straightforwardly onto any English category (cf. Sun, 2012). Given that English listeners' perception of tones is ultimately non-native (Wang et al., 1999), we might predict that English-speaking L2 learners of Chinese would also experience more lexical competition when having to distinguish segmentally identical target and competitor words that differ only in tones in lexical access. Furthermore, one would also predict that L2 learners would experience more lexical competition with word pairs that differ only in tones than with word pairs that differ only in segments (when the L2 segments map straightforwardly onto L1 segments), a result not found by previous priming research with English L2 learners of Chinese (Sun, 2012).

The issue of L2 learners' use of Chinese tones and segments in online lexical processing warrants a reinvestigation with a more time-sensitive measure. Eye-tracking has the ability to investigate the timing and speed of use of various types of information, and thus it is a very useful measure of processing. In an online visual-world eye-tracking task comparing native and non-native listeners' use of tonal information in lexical access, two main outcomes could be predicted. If the previous priming work conducted by Sun (2012) is an adequate representation of L2 tonal processing, then we should find that English L2 learners of Chinese use tones at the same time as segments, but that they will use both more slowly than native Chinese listeners. Alternatively, if the research on L2 learners' difficulty in using segmental contrasts that do not exist in the L1 is a more suitable representation of how L2 learners use tones in lexical access (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011; Cutler & Otake, 2004; Cutler et al., 2006), then a difference between the L2 learners' processing of segments and tones should be found, with tones being disadvantaged compared to segments (when the L2 segments map straightforwardly onto L1 segments).

5.4. Timing vs. speed in lexical processing

Experiment 1 investigated both the *timing* with which tones and segments can be categorized (i.e., the gate at which responses rose above chance) and the rate of improvement of categorization responses as the speech signal unfolded (i.e., with increasing gate). We know that the literature on lexical processing has focused on the timing of use of tones and segments, but perhaps tones and segments could also be compared on the lexical processing equivalent of improvement rate from the gating task: the *speed* of use of tones and segments in lexical access.

It is relevant here to elaborate on a distinction in processing alluded to by Sun (2012) between timing (as in early or late) and speed (as in fast or slow). In the use of tonal information in lexical access, we can look at timing of use, as is common in the current literature. Timing is defined here as the point in time at which information begins to be used. With respect to tones, we can look at the point in time when tonal information begins to constrain the word search: How much of the tone is needed before the tonal information begins to have an effect on the lexical search? Importantly, is that timing the same between tonal and segmental information? Recent research suggests that the timing of use of tonal and segmental information in native Chinese listeners' lexical access is the same (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011).

On the other hand, speed of use is how fast information can be used once it is available. Whereas tones and segments may be available to constrain the word search at the same time, perhaps one constrains the word search faster, or has a more rapid effect on the lexical search, than the other.[19] This was seen in Experiment 1, where the words in the tonal and segmental conditions received accurate responses above chance at the same time but differed in their improvement rates. Malins and Joanisse (2010) reported no difference in native Chinese listeners' eye fixations between their tonal and segmental conditions. However, variability in the materials and lack of controls for initial consonants and duration (both relevant to the issue of timing) warrant a second look into the timing and speed of use of tonal and segmental information in Chinese listeners' lexical access. Additionally, there were no controls for the timing of the arrival of the information between tones and segments in perception, though the timing of arrival in the acoustics was controlled.

[19] It would be difficult to tease apart speed from weighting, since both a speed effect and a weighting effect would incur similar results. Therefore, it may not be useful or possible to tease them apart, especially in the present work.

5.5. The present experiment

The following chapter aimed to investigate the time course of use of tones and segments in lexical access for both native Chinese listeners and English-speaking L2 learners of Chinese. To do this, a visual-world eye-tracking experiment was conducted contrasting the processing of tonal and segmental information, and comparing Chinese and English listeners.

For native listeners, it is predicted that tones and segments will begin to influence the lexical search at the same time (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011). In terms of speed of use, the predictions are unclear. No prior study suggests a speed difference between the processing of tones and segments; however, of all of the studies showing no timing of use difference between tones and segments, only Malins and Joanisse (2010) used an appropriate method to investigate the speed of use, and the amount of variability in the materials could be masking speed effects, if present. Based on the current literature, it would be predicted that tones and segments would be used with similar speeds. If they differ in their speed of use, the results of Experiment 1 and other research on the relation of tones and segments (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011) would suggest that tones would be used more slowly than segments.

For L2 learners, previous priming results suggest that tones and segments would be used at the same time in relation to each other (Sun, 2012). Research from the segmental domain, however, would predict difficulty in using tones in online word recognition compared to native listeners and compared to using L2 segments that map straightforwardly onto L1 segments (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011; Cutler & Otake, 2004; Cutler et al., 2006). This difficulty could manifest itself as a delay of use of tonal information in relation to segmental information.

With respect to L2 learners' speed of use of tones and segments, three outcomes can be predicted. This first predicted outcome follows from the results of Experiment 1, which showed that L2 learners performed identically to native listeners on both tonal and segmental contrasts, not only in terms of timing but also in terms of rate of improvement. From the results of Experiment 1, one would predict that in an online task, L2 learners and native listeners might perform similarly in terms of speed of use of tones and segments, with segments being used faster than tones. The second predicted outcome follows the results of previous priming research (Sun, 2012): The findings of this research instead predict that English L2 learners of Chinese will exhibit equally slower use of tones and segments as compared to native listeners. The third predicted outcome, following prior studies on L2 lexical processing, is that L2 learners of Chinese will exhibit slower use of tones than segments, but with segmental contrasts showing similar speed compared to native listeners. This would result from L2 learners' difficulty in using new cues to lexical identity in L2 lexical processing (e.g., Tremblay, 2008). The present study aimed to tease apart these possible outcomes to better understand the nature of the use of tones and segments in native and L2 lexical processing.

Part II of this dissertation reports on a visual-world eye-tracking study that used similar materials as those used in Experiment 1. This task is ideal for investigating native and L2 lexical processing, as the main measure (eye movements) is unconscious. The task itself is simply to click on the picture that matches the spoken word; as such, it is not particularly taxing on L2 learners. Eye movement data reveals the direct lexical competition between two items. This makes it more sensitive to fine-grained differences that may be missed in other tasks. Additionally, eye-tracking allows for the investigation of both timing and speed effects in lexical access, which will be crucial to the present investigation of tones and segments.

CHAPTER 6: THE TIMING OF USE OF TONES AND SEGMENTS IN LEXICAL ACCESS: EXPERIMENT 2

6.1. Introduction

This chapter reports on a visual-world eye-tracking experiment designed to investigate the timing of use of tonal and segmental information in lexical access. Previous priming work has shown that English-speaking L2 learners of Chinese use tones and segments in similar ways, although they use the information more slowly compared to native Chinese listeners (Sun, 2012). The absence of a difference between the use of tones and segments in Sun's (2012) results may seem counterintuitive, since it would likely be expected that English-speaking L2 learners of Chinese would struggle to use tones in online spoken word recognition based on the difficulty that L2 learners have when using new segments in L2 lexical access (Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011). To further investigate this issue, the processing of the words corresponding to the tonal and segmental minimal pairs in Experiment 1 was investigated using eye-tracking to obtain a highly precise measure of the time course of lexical access.

6.2. Methods

This experiment used the visual-world eye-tracking paradigm. Eye-tracking has been well established as a method that examines the online time course of lexical access (e.g., Allopenna et al., 1998; Tanenhaus et al., 1995). In this method, participants see four objects on a computer screen arranged in a grid. They hear a spoken word and are instructed to do something with the visual object corresponding to the word (e.g., use a mouse to click on the object that depicts the word they heard). The participants' eye movements are recorded as they perform the task, and these eye movements are analysed in relation to when the spoken word unfolds in time.

6.2.1. Participants

The same participants who participated in Experiment 1, described in Section 4.1.1, also participated in Experiment 2, with the exception of two native Chinese listeners, who were excluded from this experiment for complete lack of recorded fixations in the pre-set interests areas around the pictures. Lack of recorded fixations can result from participants using their peripheral vision to complete the task and not making eye movements to any of the objects in the display despite instructions to look at these objects. The analyses on fixations were therefore run on 22 native Chinese listeners (13 females; mean age: 25.4; standard deviation (SD) 3.6) and all 22 English-speaking L2 learners of Chinese described in Section 4.1.1.

6.2.2. Materials and design

The same auditory materials used in Experiment 1, described in Section 4.1.2 were used in Experiment 2, with the exception that the auditory stimuli were *not* gated and the complete word was heard. All other manipulations (duration and intensity) remained. As a reminder, there were 16 items in the tonal condition and 16 items in the segmental condition. An additional 16 tonal and 16 segmental pairs identical to the critical trials in all respects (except for controls in frequency) were selected as target and competitor words for the filler trials. Each of the four tones was played as the target an equal number of times across the experiment. Each tone was visually represented in the display an equal number of times.

The frequency of all items was controlled. The lack of significant differences between conditions was confirmed by paired samples *t*-tests on the log-10-transformed frequency per million words, as listed in SUBTLEX-CH Corpus (Cai & Brysbaert, 2010). For the target items, there were no significant differences for the frequencies between the target and competitor pairs for the tonal condition ($p > 0.1$) or segmental condition ($p > 0.1$). Additionally, there were no differences between the target or competitor items between the two conditions (p's> 0.1).

All words were represented with black and white line drawings. In every trial, participants saw four pictures in the display: two pictures representing one tonal pair and two pictures representing one segmental pair. In the tonal condition, the target and competitor words were a tonal pair and the two distracter words were a segmental pair; the reverse was true in the segmental condition. The words that appeared as target and competitor words in the filler trials served as distracter words in the critical trials, and the words that appeared as target and competitor words in the critical trials served as distracter words in the filler trials. The filler trials were therefore identical to the critical trials in all respects, with one tonal pair and one segmental pair present in each display. All displays were thus repeated twice, once in each condition.[20] For example, if the first time a display was seen, the tonal pair was targeted, the second time it appeared, the segmental pair was targeted. In order to reduce the bias that once an item or a pair has been targeted, it will not be heard again, some filler trials repeated the same target word as other filler trials.

Figure 8 below illustrates how the filler trials were balanced, with the red circle indicating which item in the two possible pairs (tonal or segmental) was the target for each presentation. Filler Type 1 demonstrates that for half of all of the filler trials, the exact same target was repeated in the display's second presentation. Filler Type 2 demonstrates that for the other half of the filler trials, the opposite item within the same pair was targeted in the display's second presentation. An equal number of Type 1 and Type 2 fillers are present for each block. Half of both Type 1 and Type 2 filler trials were segmental trials and the other half were tonal

[20] In order to reduce the number of word-picture associations that participants memorized, each display was repeated twice. As a result, participants needed to remember half the number of word-picture associations. This helped accommodate the L2 learners, who were not as familiar with some of the words.

trials. As previously mentioned, the filler items were designed in this way to ensure that when a display was repeated, participants could not predict which item would or would not be the target.

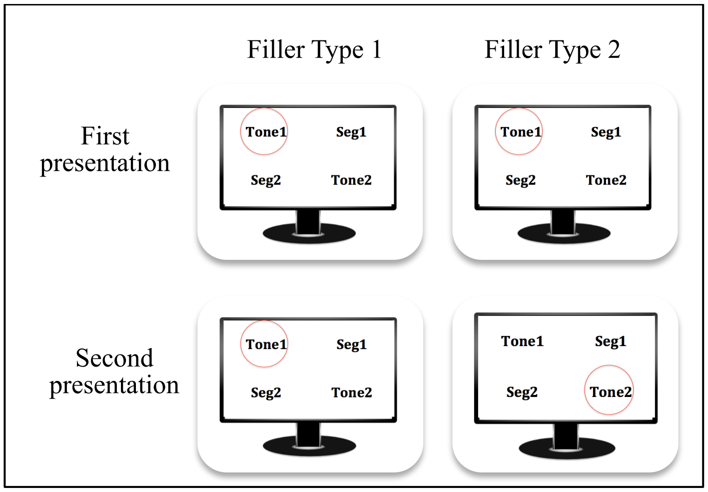

Figure 8: Filler trial balancing example

Two lists were created such that on the second list, the target-competitor relationship was reversed (i.e., the target word on List 1 was the competitor word on List 2, and vice versa). Participants were randomly assigned to a list. The experiment was broken into four blocks to offer breaks for the participants if needed, and also to allow for recalibration of the eye tracker as necessary. Each block contained 16 items: four items from each condition (tonal and segmental) and eight filler trials (half tonal and half segmental). Items were randomized within blocks.

In addition to filler items, eight stimulus sets were selected to serve as practice items before the experiment began. In addition to familiarizing the participants with the task, these practice items set up the expectation of the repeated displays. Practice items were identical in form to the critical trials, with one tonal and one segmental pair presented. These eight trials consisted of four trials that mimicked the critical trials, of which two had a tonal target and two had a segmental target. On the second presentation of the display, the opposite pair on the screen was targeted. The other four trials mimicked the filler items and included two Type I-like and two Type II-like filler items.

6.2.3. Procedures

6.2.3.1. Word familiarity task

Participants first completed a word familiarity task in which they were presented with the written character for each word and heard the corresponding auditory stimulus played through headphones. All 128 critical and filler items were included in the word familiarity task. Participants completed a word familiarity task by giving a rating for how familiar they were with that word based on the rating scale provided in Table 5. Participants entered the number that corresponded to their response. This task was implemented with Psychopy (Peirce, 2007) on a MacBook Pro.

Table 5: Word familiarity rating scale

Rating	Rating description
0	I have never seen/heard this word.
1	I have occasionally seen/heard this word, but I don't know what it means.
2	I have occasionally seen/heard this word and I know what it means in context, but I could not provide a definition for it.
3	I have frequently seen/heard this word and I know what it means in context, but I could not provide a definition for it.
4	I have frequently seen/heard this word, I know what it means, and I can provide a definition for it.

This familiarity task was completed after Experiment 1 but before the training phase for Experiment 2 (described below). This was done so as not to influence the results of Experiment 1 with the words in the familiarity judgments. If participants had taken the word familiarity task before Experiment 1, they may have activated the lexical item in Experiment 1 despite hearing only the first half of the word. Since Experiment 1 was designed to by-pass lexical access, this was an undesired effect. The familiarity rating could also not be conducted at the end of Experiment 2, since L2 learners' experience with the words used in Experiment 2 (on which they were trained prior to Experiment 2) could have influenced their ratings. Therefore, the word familiarity task was conducted after Experiment 1 but before the training for Experiment 2.

The mean word familiarity ratings for target and competitor words in the tonal and segmental conditions are reported in Table 6. As can be seen from Table 6, the word familiarity ratings for each of the conditions were very similar. The same is true of targets and competitors within and between conditions. Thus, any effect of condition in the eye-tracking experiment cannot be due to differences in the L2 listeners' familiarity with the words. It is also worthwhile noting that familiarity was generally very high for both the tonal and segmental conditions. Paired samples t-tests did not yield a significant difference between the targets or competitors between the two conditions (targets $t(351)=-0.45$, $p > 0.1$, competitors $t(351)=-0.23$, $p > 0.1$).

Table 6: Average word familiarity ratings for target and competitor words by condition

	Tonal		Segmental	
	Target	Competitor	Target	Competitor
Mean	3.15	3.11	3.11	3.09
SD	1.30	1.36	1.38	1.31
Condition Mean	3.10		3.13	
Condition SD	1.34		1.33	

6.2.3.2. Word-picture association training

After the word familiarity task, participants completed a training phase in which they learned the 128 word-picture associations for the eye-tracking experiment. This training served two purposes: (i) to familiarize participants with the words and word-picture parings; and (ii) to familiarize participants with the talker's voice and pitch range, and to encourage them to tune in to pitch cues, as duration had been normalized. In this way, participants were not expecting duration as a cue to tone identity in the main experiment.

The training included a familiarization phase and a test phase. The familiarization phase was completed over the course of two days. In the first familiarization portion, participants saw the pictures and heard the words associated with them, one by one. Participants simply pressed the space bar to move on to the next word and were instructed to try to remember the associations. All items were presented twice, after which participants were asked if they would like to see the items again. Participants could look at the pictures as many times as they wished.

On the second day of testing (at least 5 days after the first day), participants completed the familiarization again. Participants first repeated the familiarization from day 1. Once they were confident they knew the word-picture correspondences, they were tested on these correspondences. There were four blocks to the test phase. Each block showed the same 32 pictures in each trial, with each picture labeled with a letter (a-z) or number (1-6). The pictures corresponding to the critical word pairs (tonal or segmental) appeared together in the same display of 32 pictures to ensure that the participants could distinguish the target from the competitor items. In a given trial, one word was heard, and participants were instructed to enter the letter or number that identified the corresponding picture.

Feedback was given after every trial of the test phase. If the response was correct, participants saw "CORRECT 对" written on the screen above the correct picture. The word was also repeated auditorily. If the response was incorrect, the participant saw "Incorrect, the correct picture was… 不对， 对的是…" above the correct picture, and the correct word was heard. Items in the test phase were repeated only if an incorrect response was given in that block. This means that if the participant gave an incorrect response, the entire block was repeated, up to a maximum of three times. Once participants were able to reliably select the correct picture for every word, or the maximum number of block repetitions had been reached, the eye-tracking portion of the experiment began. All tasks were implemented using the software PsycoPy (Peirce, 2007) on a MacBook Pro.

6.2.3.3. Eye-tracking experiment

The experiment was compiled using Experiment Builder software (SR Research). Participants' eye-movements were recorded with a desktop EyeLink 1000 Eye Tracker (Beijing, China), a tower mounted EyeLink 1000 Plus Eye Tracker (College Park, MD), and a head-mounted EyeLink II (Lawrence, KS) (SR Research), each recording at 250 Hz (1 eye movement recorded every 4 milliseconds). The experiment began with a calibration of the participants' left pupil and corneal reflection. This calibration was followed by the practice session. After the practice session, participants were encouraged to ask any questions they had. Written instructions were presented in English and Simplified Chinese. The main experimenter, a native English speaker and L2 speaker of Chinese, gave verbal instructions and answered questions in the language preferred by the participant. After any questions were answered, the experiment began.

Figure 9 illustrates the procedures for a single trial of the experiment. A trial began with four images appearing on the screen in a non-displayed 2x2 grid. The images remained on the screen for 2,000 ms (preview time). This time allowed participants to pre-activate the names for each of the pictures and to familiarize themselves with their locations. No auditory stimulus was heard during this presentation. After the 2,000 ms preview, the images disappeared, and a fixation cross appeared in the middle of the screen for 500 ms. As the fixation cross disappeared, the images reappeared on the screen in the same locations as during the preview, and an auditory stimulus was heard though headphones. This auditory stimulus was the target word for that trial, heard in isolation. Participants were instructed to click on the picture that matched the spoken word as quickly as possible. Once the participant clicked, a blank screen appeared for 700 ms, after which the next trial began. Both eye-movements (recorded from the target-word onset in the auditory stimulus) and selection accuracy were recorded.

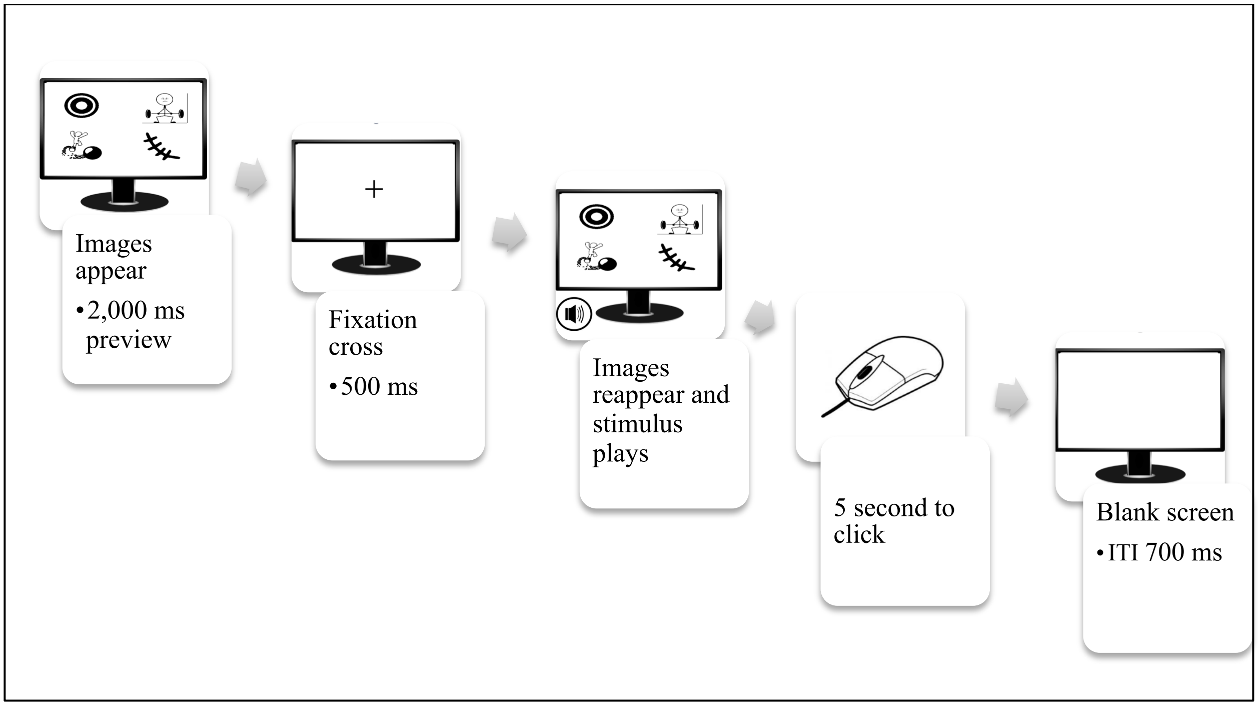

Figure 9: Experiment 2 trial procedure

6.2.4. Data analysis

Both eye movements as well as accuracy rates in selecting the target object were recorded. Accuracy was 100% in both conditions for native Chinese listeners; as a result, no data were excluded from native listeners' results. For L2 learners, accuracy was 93.6% in the tonal condition and 97.8% in the segmental condition (when L2 learners were inaccurate, they clicked on the competitor or distracters, or they did not click on any object).

Trials for which the target object had not been correctly identified or trials which generated no response from the participants were excluded from the eye movement analyses. Excluding the trials for which the target object had not been correctly identified or trials which generated no response resulted in a loss of 4.3% of the L2 data, 3.2% was from the tonal condition and 1.1% was from the segmental condition. Of the remaining trials, any trials for which the learner did not reliably identify the target or competitor picture in the training phase

(<50% accuracy) were excluded, resulting in an additional exclusion of 18.6% of the L2 data, 8.4% was from the tonal condition and 10.2% from the segmental condition. For the remaining trials, fixations in a pre-determined interest area around each picture were included and analyzed.

GCA was used to analyze the eye-tracking fixations. GCA has recently been suggested as a method for analyzing time-course data, making it ideal for eye-tracking research (Mirman, 2014; Mirman et al., 2008). One important concern with using methods such as ANOVA is that by including time in the analysis, the assumption of independence of observations is violated: Given that the proportion of fixations to target at one time point is related to the proportion of fixations to target at the next time point, analyses such as ANOVAs violate the assumption of independence of the observations. As a result, GCA has been used in several recent eye-tracking studies (e.g., Magnuson et al., 2007; Malins & Joanisse, 2010; Tremblay et al., 2016).

The important effects for these analyses include the effect of condition, equivalent to a simple effect, and interactions between condition and time. A significant effect of condition and no interaction with time would indicate a timing difference between the tonal and segmental conditions,[21] such that the *shapes* of the two fixation curves are the same, but due to a horizontal shift, one condition has on average higher proportions of target fixations than the other (i.e., if the fixation curves in the two conditions have the same shape but one is shifted horizontally in time, this will result in higher proportions of fixations in the condition where the proportions of target fixations rise earlier). This would indicate that the information being used to identify the target in the condition where proportions of target fixations rose first would be used *earlier* than in the other condition. Alternatively, a non-significant effect of condition together with an interaction between condition and any of the time terms would suggest that while the two conditions have on average equal proportions of target fixations, the fixation lines differ in

[21] It is also possible that this could be due to baseline effects that persist into the critical word.

shape. This could indicate that although the two types of information are used at the same stage, they are used in different ways, and possibly with different speeds, depending on the time term with which condition interacts. Finally, a simple effect of condition and an interaction between condition and time could be seen, but in such a case additionally analyses may need to be conducted.

If a simple effect of condition and interactions with time occur, it is unclear if the effect of condition is due to the interactions with time. Imagine two lines with drastically different slopes with the same intercept. These lines would show both an effect of condition (since overall one line has higher values than the other) and interactions with time (one line has a more positive slope than the other). In this case, the effect of condition, while significant, does not indicate that the lines are shifted in time, and is simply reflecting the larger difference in values over the whole time window due to the differences in the slope. For this reason, if the results pattern with an effect of condition and interactions with time, a follow up linear mixed-effects model using various time windows would be needed to investigate timing independent of speed.

The data onset was the onset of the word, excluding a 200-ms baseline for the time it takes to plan and launch a saccade (Hallett, 1986). From there, the GCA was conducted over a time period of 750 ms. Each stimulus had a total duration of exactly 524 ms; hence, this window includes the whole word and 226 ms after the word offset. This window was selected to include the time from the onset of the word to where the fixation lines plateaued at their highest level. Graphs of the full time window including the baseline to 1,000 ms are provided in Appendix G.

In the event of an interaction with condition and time, an additional analysis would be needed to more carefully determine the timing of use of tones and segments. An interaction with time could cause an additional effect of condition. This effect of condition would be interpreted as a timing effect by itself; however, given the interaction with time, this timing effect would be unreliable. In this case, a secondary analysis would be conducted using time windows to ascertain the initial point where the target and competitor fixations diverge. This analysis would not be influenced by the speed effect, and would more precisely determine the timing of use of tones and segments if the results of the GCA are inconclusive.

6.2.5. Results

The proportions of target fixations were analyzed with GCA (Mirman, 2014) using the *lme4* (Bates et al., 2015) and *LMERConvenienceFunctions* (Tremblay & Ransijn, 2015) packages in R (R Core Team, 2013). First, in order to investigate each group's use of tones and segments, separate models were run on L1- and L2-Chinese listeners. These models included the effects of condition, time, and all interactions. The segmental condition served as the baseline to which all comparisons were made. The model also included random effects of participant on all time terms. The most complex model was evaluated using the backward fit function bfFixfLMER_F.fnc of the *LMERConvinienceFunctions* package (Tremblay & Ransijn, 2015), which uses log-likelihood ratio to test improvements to the model. Only the results of the model with the best fit are presented, with *p* values being calculated using the *lmerTest* package (Kuznetsova et al., 2016).

6.2.5.1. Native Chinese listeners

Figure 10 shows Chinese listeners' proportion of fixations to target, competitor, and distracters over the time window of the analysis. The shaded area represents one standard error above and below the mean.

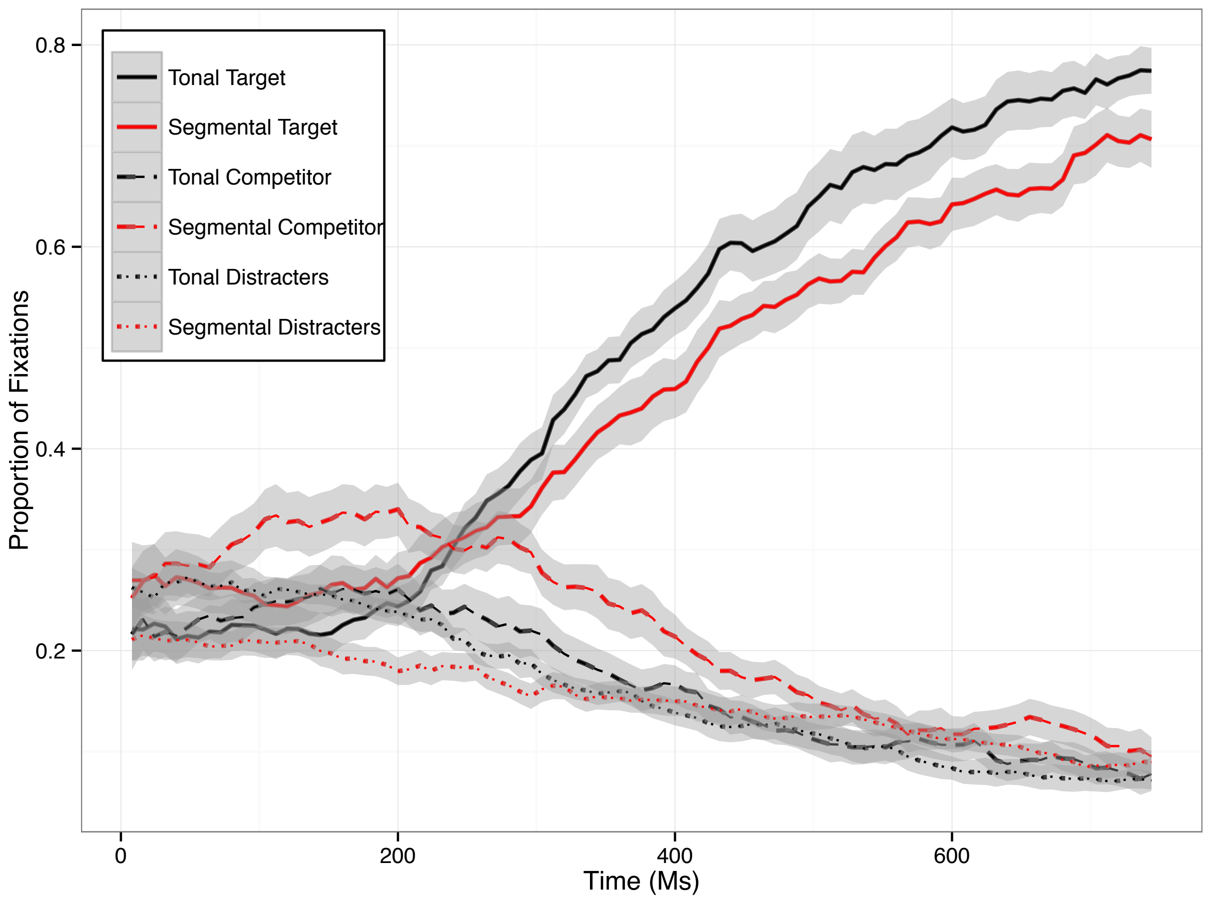

Figure 10: Native Chinese listeners' proportion of target, competitor, and distracter fixations, with fixations in the tonal condition in black and fixations in the segmental condition in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

Figure 11 shows Chinese listeners' actual and predicted proportions of target fixations in the segmental and tonal conditions. The solid lines represent the actual fixations and the dashed lines represent the fixations predicted by the model listed in Table 7 (presented next). The shaded area represents one standard error above and below the mean.

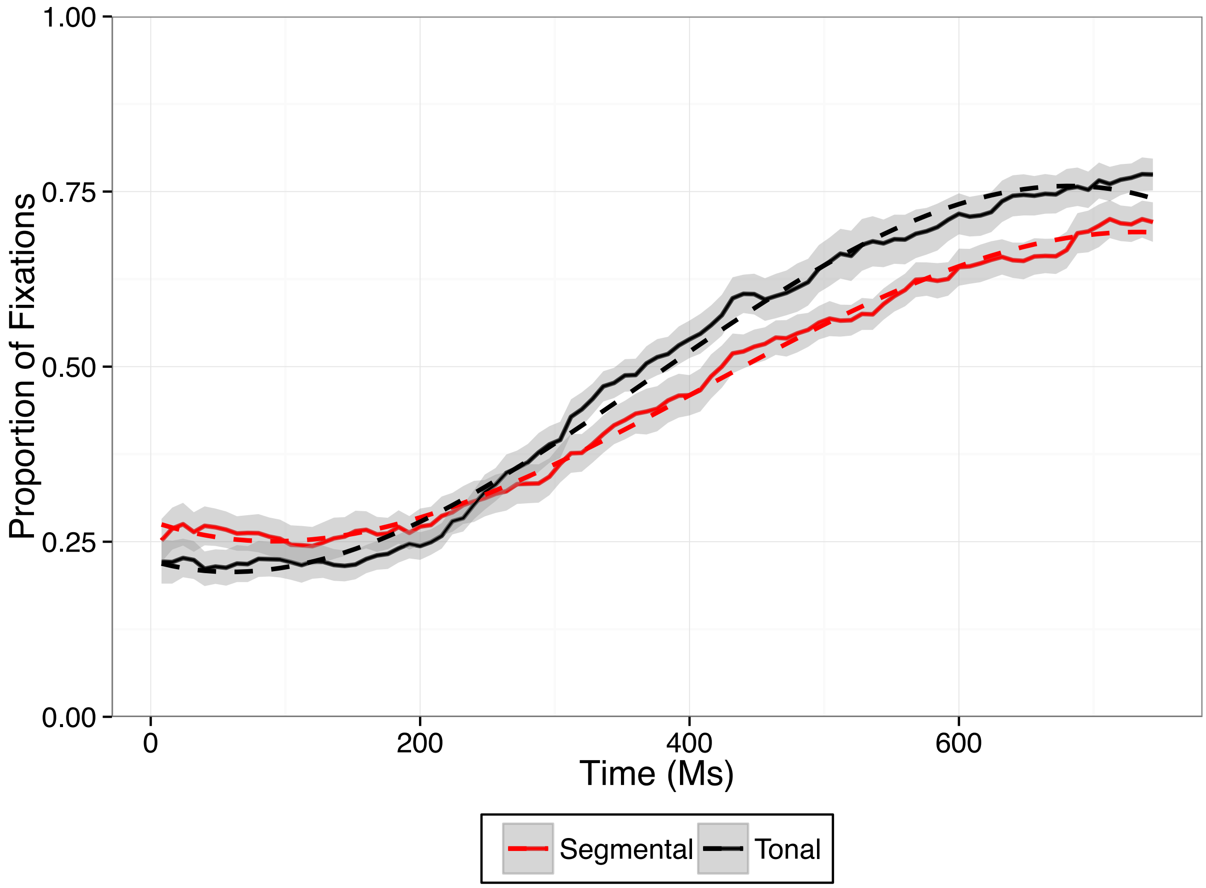

Figure 11: Native Chinese listeners' predicted proportion of target fixations, with fixations in the tonal condition in black and fixations in the segmental condition in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

The GCA on Chinese listeners' proportions of target fixations with the best fit included the fixed effect of condition and its interaction with the linear, quadratic, and cubic time terms. As a reminder, the segmental condition served as the baseline. The results of this GCA are provided in Table 7.

Table 7: Results of GCA on native Chinese listeners' proportion fixations to target

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.45	0.01	33.46	<.001
Linear	1.54	0.09	17.15	<.001
Quadratic	0.14	0.06	2.38	.026
Cubic	–0.25	0.05	–4.99	<.001
Condition	0.04	0.00	11.90	<.001
Linear : Condition	0.43	0.03	15.04	<.001
Quadratic : Condition	–0.17	0.03	–6.07	<.001
Cubic : Condition	–0.09	0.03	–3.28	.001

Significance codes ("***" < .001, "**" < .01, "*"<.05)

As shown in Table 7, the significant effect of the linear term with a positive estimate indicates that native listeners' proportion of target fixations in the segmental condition had a positive linear trend. The significant effect of the quadratic term with a positive estimate indicates that the proportion of target fixations in the segmental condition had a convex (i.e., U) shape. The significant effect of the cubic term with a negative estimate indicates that the proportion of target fixations in the segmental condition had a canonical S-shape. The effect of condition was significant with a positive estimate, indicating that collapsing across time, the proportion of target fixations was higher in the tonal compared to segmental condition. The interaction between condition and the linear term was significant with a positive estimate, indicating that the tonal condition had a more positive slope than the segmental condition. This means that the rate at which participants increased their proportion of target fixations was faster in the tonal condition compared to the segmental condition. The interaction between condition and the quadratic term with a negative estimate indicates that the fixation curve was less U-shaped in the tonal condition than the segmental condition. The interaction between condition and the cubic term with a negative estimate indicates that the fixation curve had a more exaggerated canonical S-shape in the tonal condition than in the segmental condition.

The effect of condition suggests that native Chinese listeners use tones before they use segments; however, in the presence of interactions with time, this conclusion is unstable and will be more appropriately investigate in 6.2.7 with a separate analysis. The results also support the conclusion that native Chinese listeners' speed of use of tonal information was faster than their speed of use of segmental information.

6.2.5.2. SECOND LANGUAGE CHINESE LISTENERS

Figure 12 shows the proportions of fixations to target, competitor, and distracters over the time window of the analysis for English-speaking L2 learners of Chinese. The shaded area represents one standard error above and below the mean.

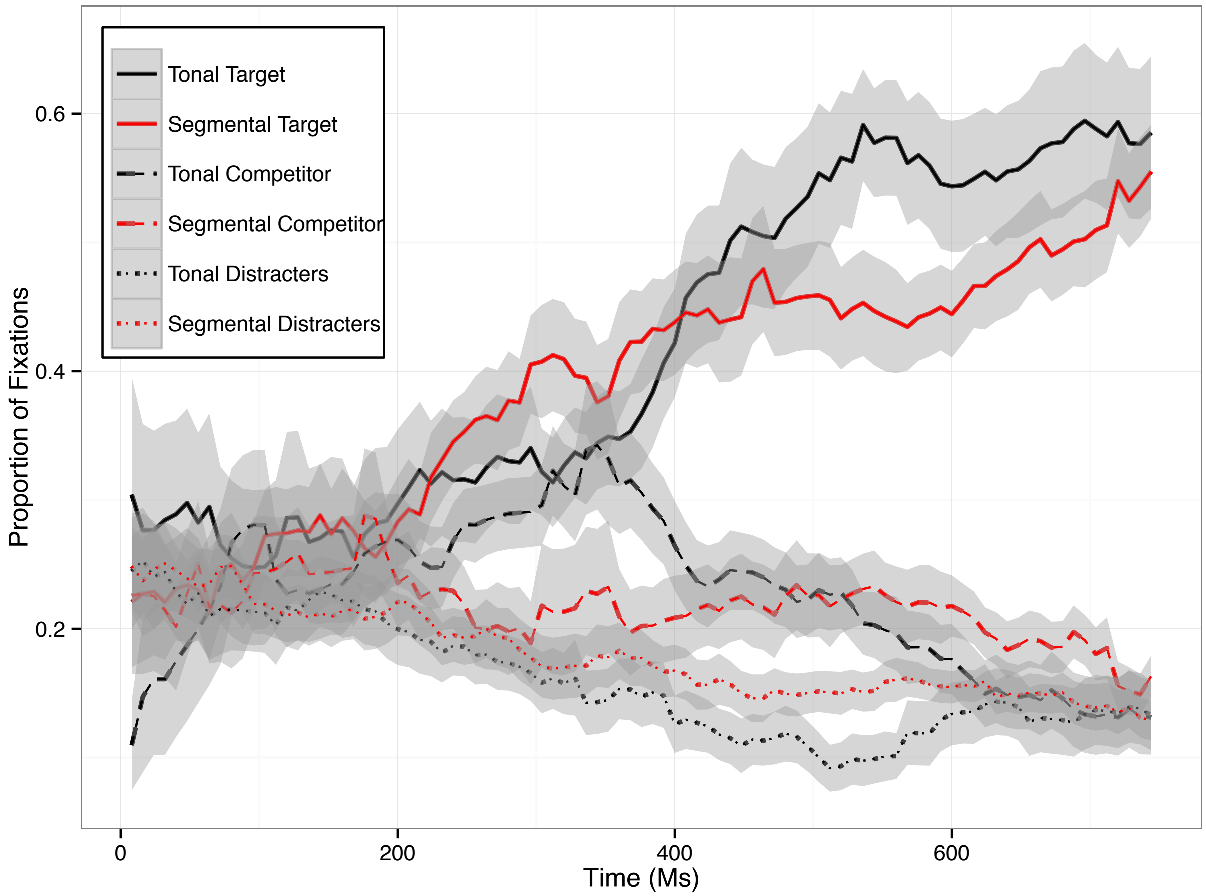

Figure 12: Second language Chinese listeners' proportion of target, competitor, and distracter fixations, with fixations in the tonal condition in black and fixations in the segmental condition in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

Figure 13 shows L2 learners' actual and predicted proportions of target fixation in the segmental and tonal conditions. The solid lines represent the actual fixations and the dashed lines represent the fixations predicted by the model listed in Table 8 (presented next). The shaded area represents one standard error above and below the mean.

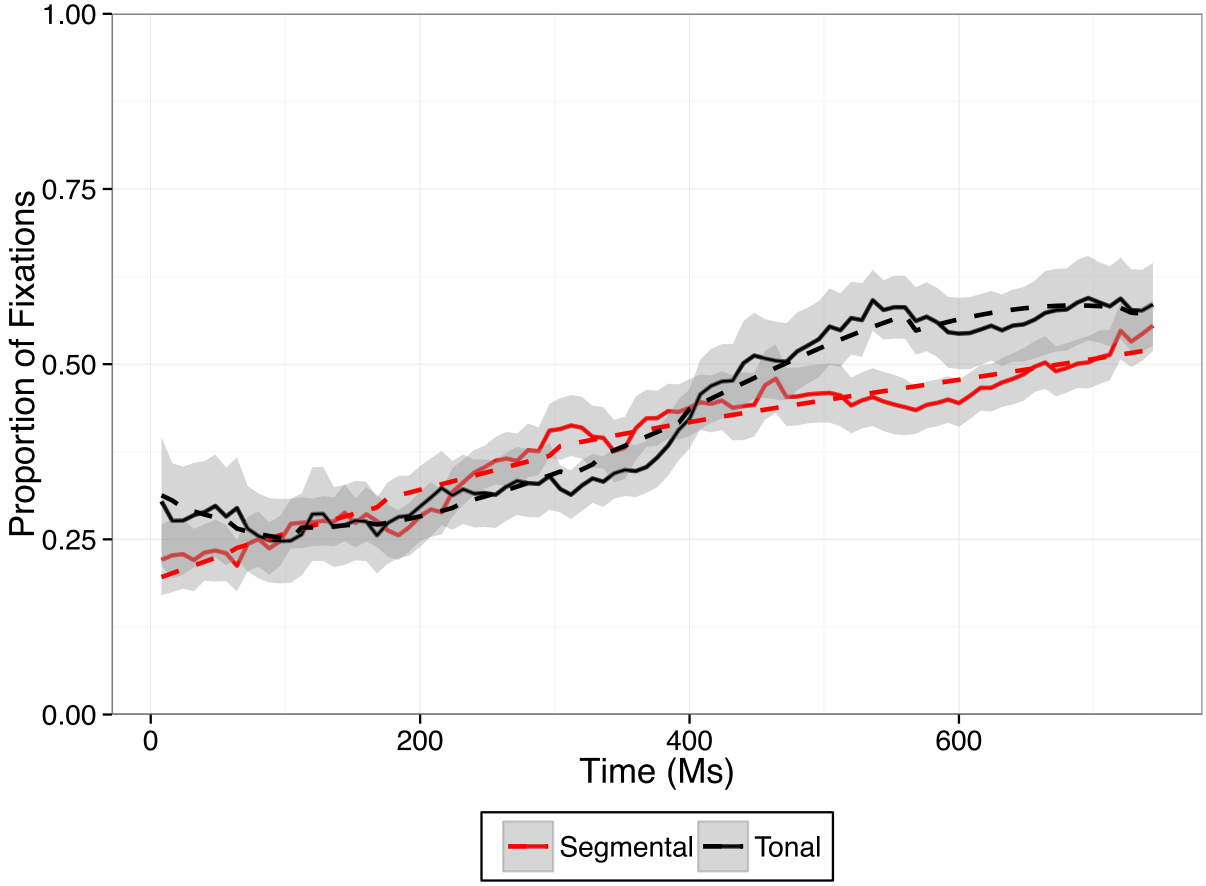

Figure 13: Second language Chinese listeners' predicted proportion of target fixations, with fixations in the tonal condition in black and fixations in the segmental condition in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

The model with the best fit included the fixed effect of condition and its interaction with the linear, quadratic, and cubic time terms. Recall that the segmental condition served as the baseline to which all comparisons were made. The results of the GCA are provided in Table 8.

Table 8: Results of GCA on second language Chinese listeners' proportion fixations to target

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.38	0.03	14.64	<.001
Linear	0.69	0.19	3.57	.002
Quadratic	–0.12	0.12	–1.01	.324
Cubic	0.06	0.12	0.50	.619
Condition	0.03	0.00	6.06	<.001
Linear : Condition	0.43	0.05	9.37	<.001
Quadratic : Condition	0.26	0.05	5.76	<.001
Cubic : Condition	–0.29	0.05	–6.34	<.001

Table 8 shows that there was a significant effect of the linear term with a positive estimate, indicating that L2 learners' proportion of target fixations in the segmental condition had a positive linear trend. The effects of the quadratic and cubic terms were not significant, which indicates that L2 learners' fixation line in the segmental condition did not have U or S shape. The effect of condition was significant with a positive estimate, indicating that collapsing across time, the proportion of target fixations was higher in the tonal condition compared to the segmental condition. The interaction between condition and the linear term was significant with a positive estimate, indicating that the fixation line in the tonal condition had a more positive slope than that in the segmental condition. This means that the rate at which participants increased their proportion of target fixations was faster in the tonal condition than in the segmental condition. The interaction between condition and the quadratic term with a negative estimate indicates that the fixation curve in the tonal condition was less U-shaped than that in the segmental condition. The interaction between condition and the cubic term with a negative estimate indicates that the fixation curve in the tonal condition had a more exaggerated canonical S-shape than in the segmental condition.

The effect of condition suggests that second language Chinese listeners use tones before they use segments; however, in the presence of interactions with time, this conclusion is again unstable, and does not seem likely given the graphs. This timing effect will be more appropriately investigated in 6.2.7. The results support the conclusion that, like native listeners, second language Chinese listeners' speed of use of tonal information was faster than their speed of use of segmental information.

6.2.6. Native vs. second language learner comparisons

In order to investigate the difference in the use of tones and segments between L1- and L2-Chinese listeners, a large model was conducted that included the effects of condition, group, time, and all interactions. Given that three-way interactions with time are difficult to interpret, this analysis serves the purpose of looking for a three-way interaction that would justify running GCAs separately for each condition. The segmental condition for native listeners served as the baseline to which all comparisons were made. The model also included random effects of participant on all time terms. The most complex model was evaluated using the backward fit function bfFixfLMER_F.fnc of the *LMERConvinienceFunctions* package (Tremblay & Ransijn, 2015), which uses log-likelihood ratio to test improvements to the model. Only the results of the model with the best fit are presented, with p values being calculated using the *lmerTest* package (Kuznetsova et al., 2016).

The model with the best fit included the fixed effects of condition and group, their individual interactions with all time terms, and a three-way interaction between condition, group, and the quadratic time term. The results of this GCA are provided in Table 9.

Table 9: Results of GCA on native and second language Chinese listeners' proportion fixations to target

	Estimate	Std. Error	*t* value	P(>F)
(Intercept)	0.44	0.02	24.87	<.001
Linear	1.41	0.10	13.53	<.001
Quadratic	0.17	0.08	2.09	.043
Cubic	–0.14	0.05	–3.00	.004
Condition	0.04	0.00	8.71	<.001
Group	–0.06	0.03	–2.27	.029
Linear : Condition	0.45	0.03	14.16	<.001
Quadratic : Condition	–0.18	0.04	–4.55	<.001
Cubic : Condition	–0.13	0.03	–4.28	<.001
Linear : Group	–0.60	0.14	–4.45	<.001
Quadratic : Group	–0.26	0.12	–2.27	.028
Condition : Group	–0.02	0.01	–3.81	<.001
Quadratic : Condition : Group	0.32	0.06	4.98	<.001

The significant interaction between condition, group, and the quadratic time term suggests that the two groups differed in the effect of condition they showed in relation to the time terms. For this reason, GCAs will be conducted separately for the two groups to investigate how Chinese and English listeners use tones and segments.

Figure 14 shows the proportions of fixations to target, competitor, and distracters for native Chinese listeners and English-speaking L2 learners of Chinese, with the segmental condition in the left panel and the tonal condition in the right panel. The shaded area represents one standard error above and below the mean.

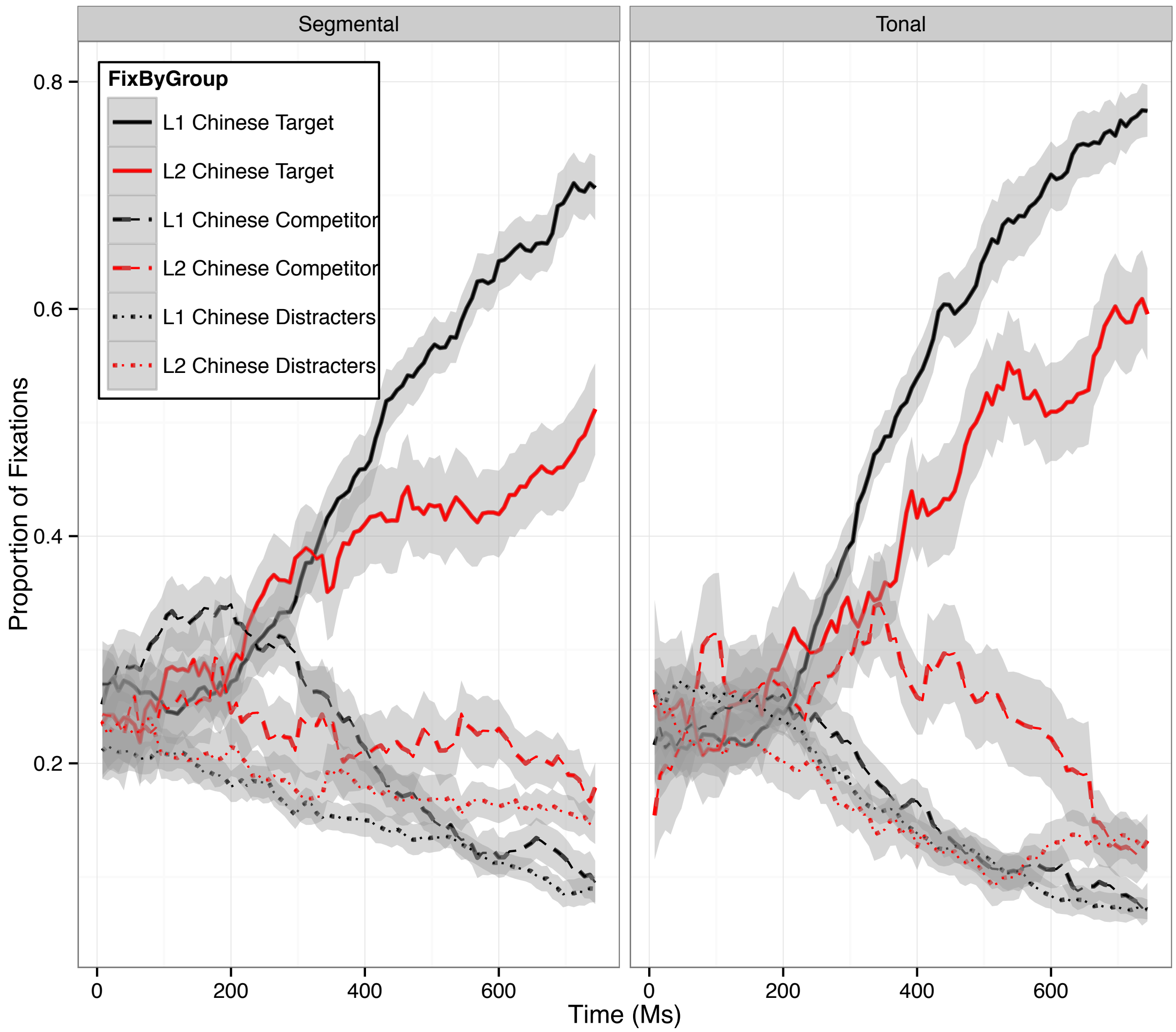

Figure 14: Native and second language Chinese listeners' proportion target, competitor, and distracter fixations with fixations of native listeners in black and the second language listeners in red. The segmental condition is represented in the left panel and the tonal condition is represented in the right panel. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

In order to compare Chinese and English listeners on their use of tones and segments in word recognition, two follow-up analyses were conducted looking at the effect of group separately for the tonal condition and segmental condition. For these models, native listeners were the baseline to which all comparisons were made. The segmental model will be discussed first, followed by the tonal model.

Figure 15 shows the actual and predicted proportions of target fixations for native Chinese listeners and English L2 learners of Chinese in the segmental condition. The solid lines represent the actual fixations and the dashed lines represent the fixations predicted by the model in Table 10 (presented next). The shaded area represents one standard error above and below the mean.

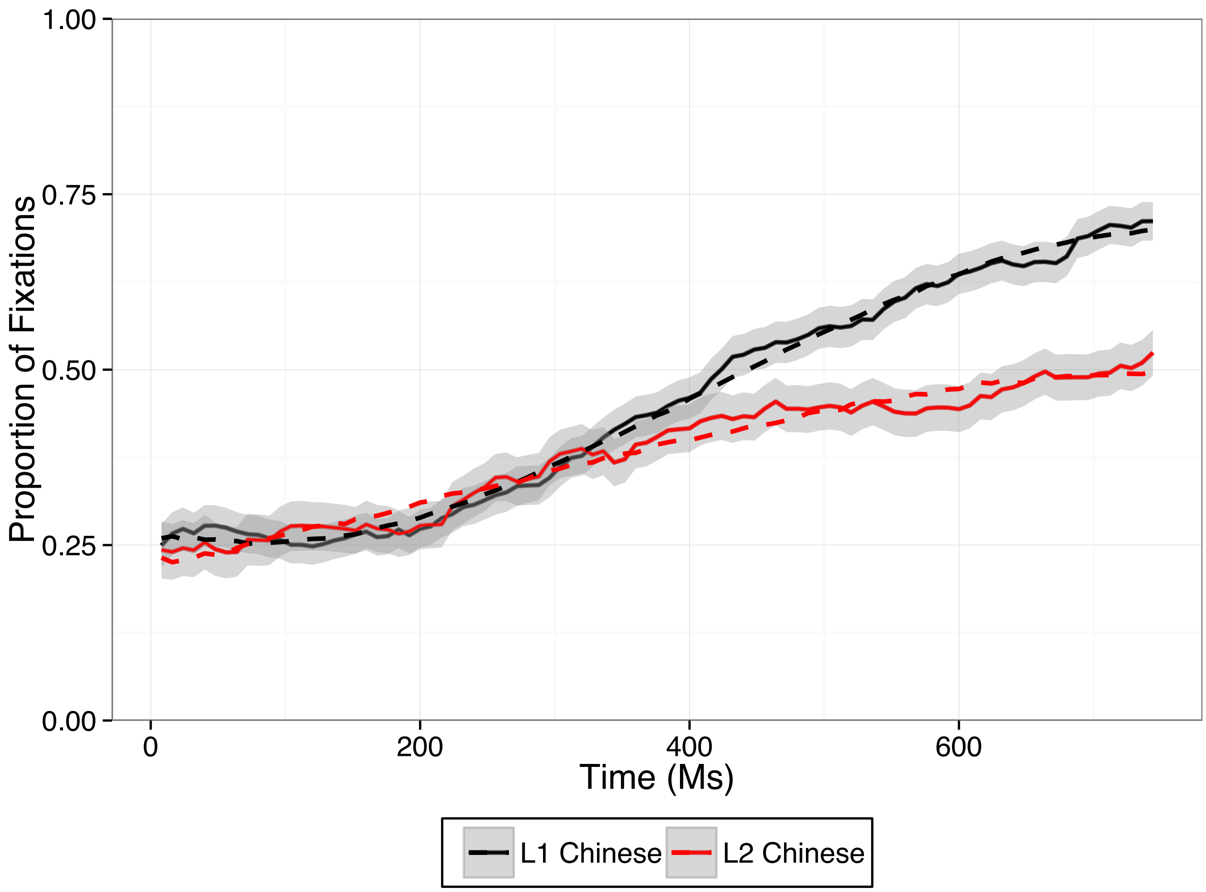

Figure 15: Native and second language Chinese listeners' predicted proportion of target fixations in the segmental condition, with fixations of native listeners in black and fixations of second language learners in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

The model with the best fit included the fixed effect of group and its interaction with the linear time term. As a reminder, the native Chinese listeners served as the baseline to which all comparisons were made. The results of this GCA are provided in Table 10.

Table 10: Results of GCA on native and second language Chinese listeners in the segmental condition

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.45	0.02	21.48	<.001
Linear	1.33	0.11	12.21	<.001
Group	–0.09	0.03	–2.86	0.007
Linear : Group	–0.62	0.16	–3.90	<.001

Table 10 shows that there was a significant effect of the linear term with a positive estimate, indicating that Chinese listeners' proportion of target fixations had a positive linear trend. Neither the quadratic term nor the cubic term improved the fit of the model, which indicates that Chinese listeners' fixation line did not have a U or S shape. The effect of group was significant with a negative estimate, indicating that collapsing across time, the proportion of target fixations was lower for L2 learners compared native listeners. The interaction between condition and the linear term was significant with a negative estimate, indicating that the proportion of target fixations had a less positive slope for L2 learners than for native listeners. This means that the rate at which participants increased their proportion of target fixations in the segmental condition was slower for L2 learners compared to native listeners.

Figure 16 shows the actual and predicted proportions of target fixations for native Chinese listeners and English L2 learners of Chinese in the tonal condition. The solid lines represent the actual fixations and the dashed lines represent the fixations predicted by the model listed in Table 11 (presented next). The shaded area represents one standard error above and below the mean.

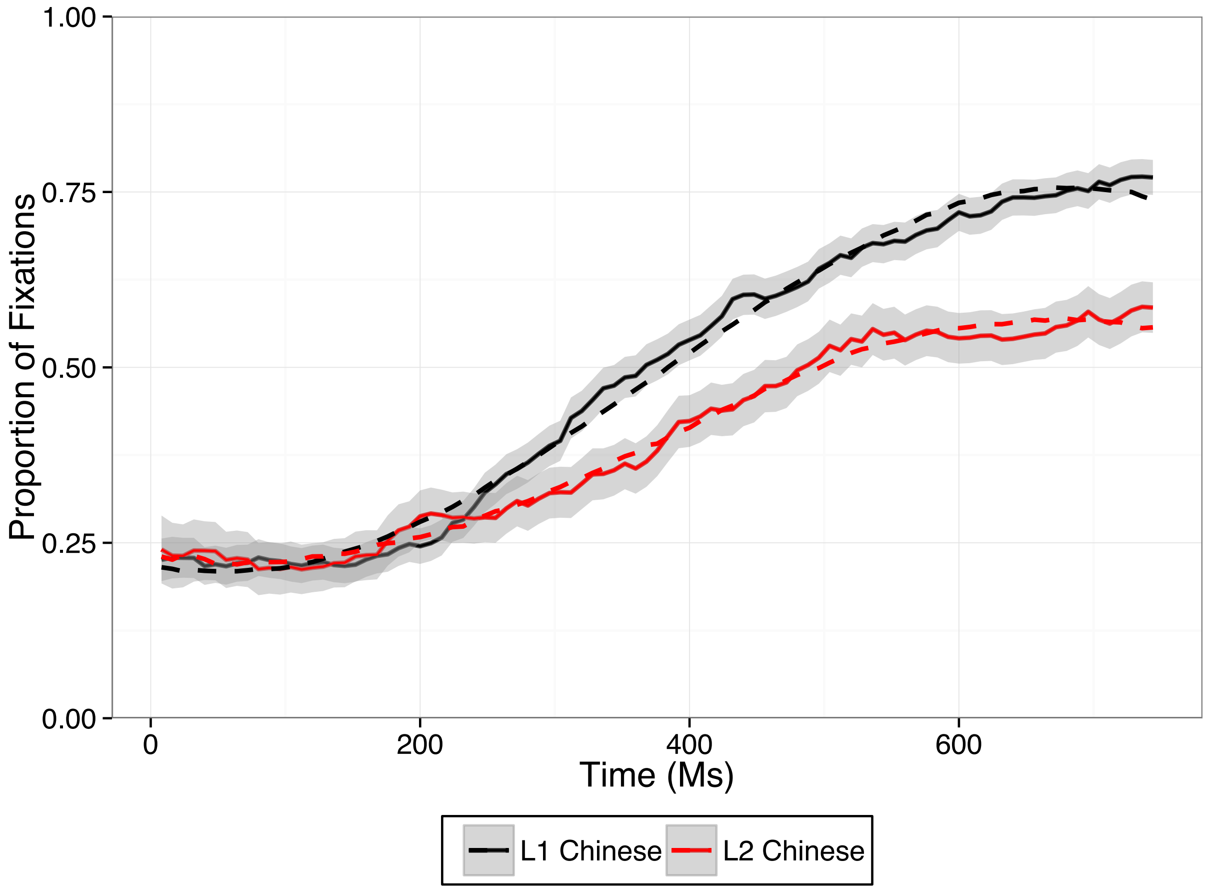

Figure 16: Native and second language Chinese listeners' predicted proportion of target fixations in the tonal condition, with fixations of native listeners in black and fixations of second language learners in red. Time in milliseconds is presented on the *x*-axis and proportion of fixations is presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

The model with the best fit included the fixed effect of group and its interaction with the linear time term. As a reminder, the native Chinese listeners served as the baseline to which all comparisons were made. The results of this GCA are provided in Table 11.

Table 11: Results of GCA on native and second language Chinese listeners' in the tonal condition

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.48	0.02	24.30	<.001
Linear	1.89	0.15	12.63	<.001
Cubic	–0.28	0.06	–4.53	<.001
Group	–0.08	0.03	–2.77	0.008
Linear : Group	–0.64	0.20	–3.14	0.003

Table 11 shows that there was a significant effect of the linear term with a positive estimate, indicating that Chinese listeners' proportion of target fixations had a positive linear trend; there was also a significant effect of the cubic term with a negative estimate, indicating that Chinese listeners' proportion of target fixations had a canonical S shape. The effect of the quadratic term did not improve the fit of the model. The effect of group was significant with a negative estimate, indicating that collapsing across time, the proportion of target fixations was lower for the L2 learners compared to native listeners. The interaction between condition and the linear term was significant with a negative estimate, indicating that the proportion of target fixations had a more negative slope for L2 learners than the native listeners. This means that the rate at which participants increased their proportion of target fixations in the tonal condition was slower for L2 learners than for native listeners.

6.2.7. Target-competitor divergence point

In order to investigate the timing of use of tones and segments in lexical access more directly, a second analysis was conducted to determine the target-competitor divergence point, or the point in time where proportions of fixations to the target and to the competitor diverge from each other and never cross again. This can be taken as the moment in time where participants first have enough of the acoustic information to look more at the target than at the competitor. This point will give the ability to directly compare the timing of use of tones and segments between native Chinese listeners and English L2 learners of Chinese.

This analysis was conducted in several steps. First, differential proportions of fixations were calculated for use in the analysis by subtracting the proportions of competitor fixations from the proportions of target fixations for every line of data. Chinese and English listeners' differential proportions of fixations are presented in Figure 17. In this figure, data points below 0 reflect that participants were looking at the competitor more than the target; points at 0 reflect equal proportion fixations to target and competitor; and points above 0 reflect that participants were looking more at the target than the competitor. This way of analysing the data is ideal, in that the goal of this analysis is to find the point in time where participants begin to look at the target more than the competitor.

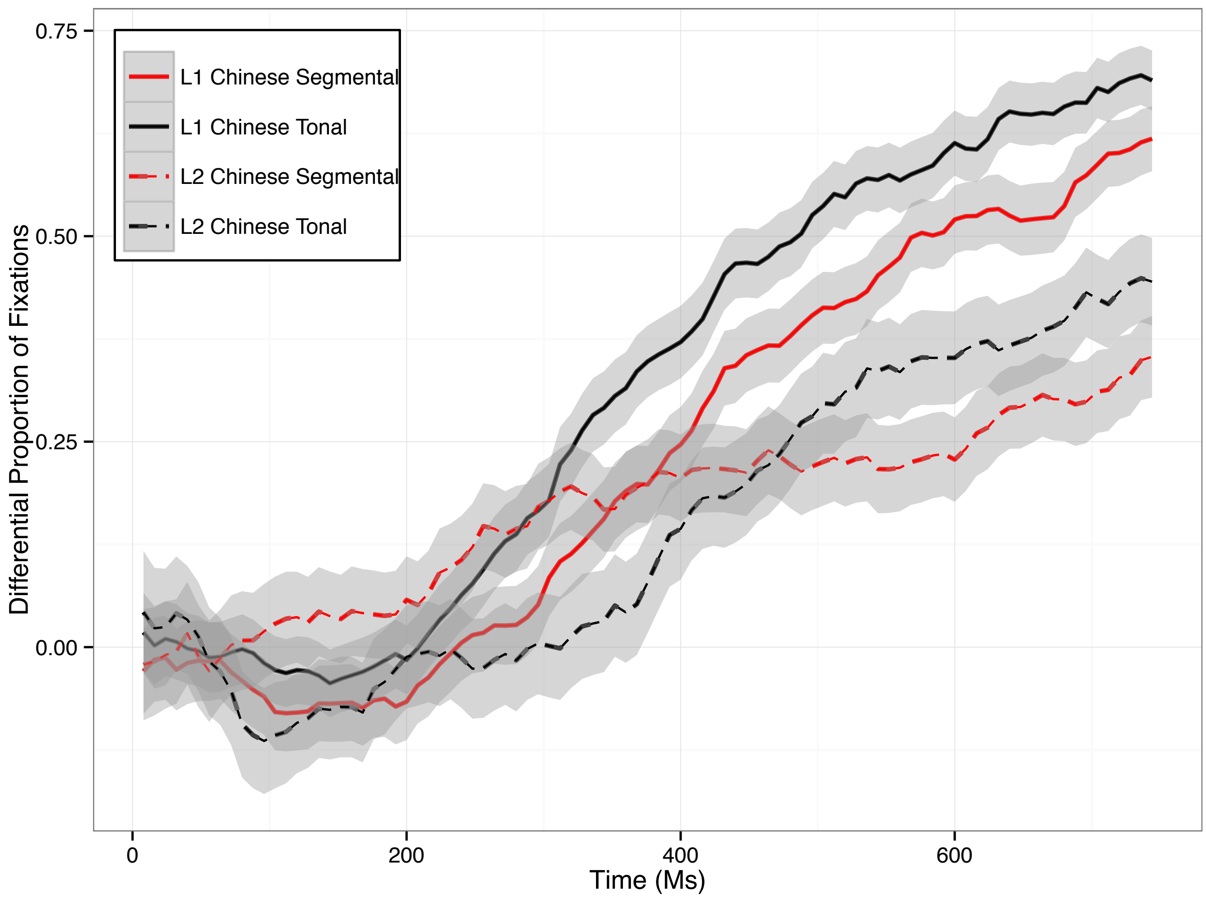

Figure 17: Differential proportion fixations for native Chinese listeners (solid lines) and English L2 learners of Chinese (dashed lines) with fixations in the segmental condition represented in red and the tonal condition represented in black. Time in milliseconds is presented on the *x*-axis and differential proportions of fixations are presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

In order to find the target-competitor divergence point, the time variable was divided into 31 bins of 24 ms each. This bin size includes 6 points of data, since the data processing script extracted a data point with a resolution of a measurement every 4 ms. Beginning with the first time bin, linear mixed-effects models compared participants' differential proportions of fixations to 0 separately for the tonal and segmental conditions using the lme4 package (Bates et al., 2015). These models included group (native listeners vs. L2 learners) as the fixed effect and item and participant as random effects. The target-competitor divergence point was then defined as the first of a continuous set of bins where the differential proportion of fixations were significantly *above* zero. Although group is entered as a variable in the models, the models actually output results separately for each group. These models made it possible to identify the point at which participants were looking at the target significantly more than the competitor. Table 12 reports the corresponding time in ms of the bin (ms values correspond to the starting time point of the bin). Figure 18 displays the same results as in Figure 17, but with lines added for each group and condition at the target-competitor divergence point for visual comparison.

Table 12: Target-competitor divergence point in milliseconds by group and condition

	L1	L2
Tonal	336	504
Segmental	384	408

Note: Time is presented in ms.

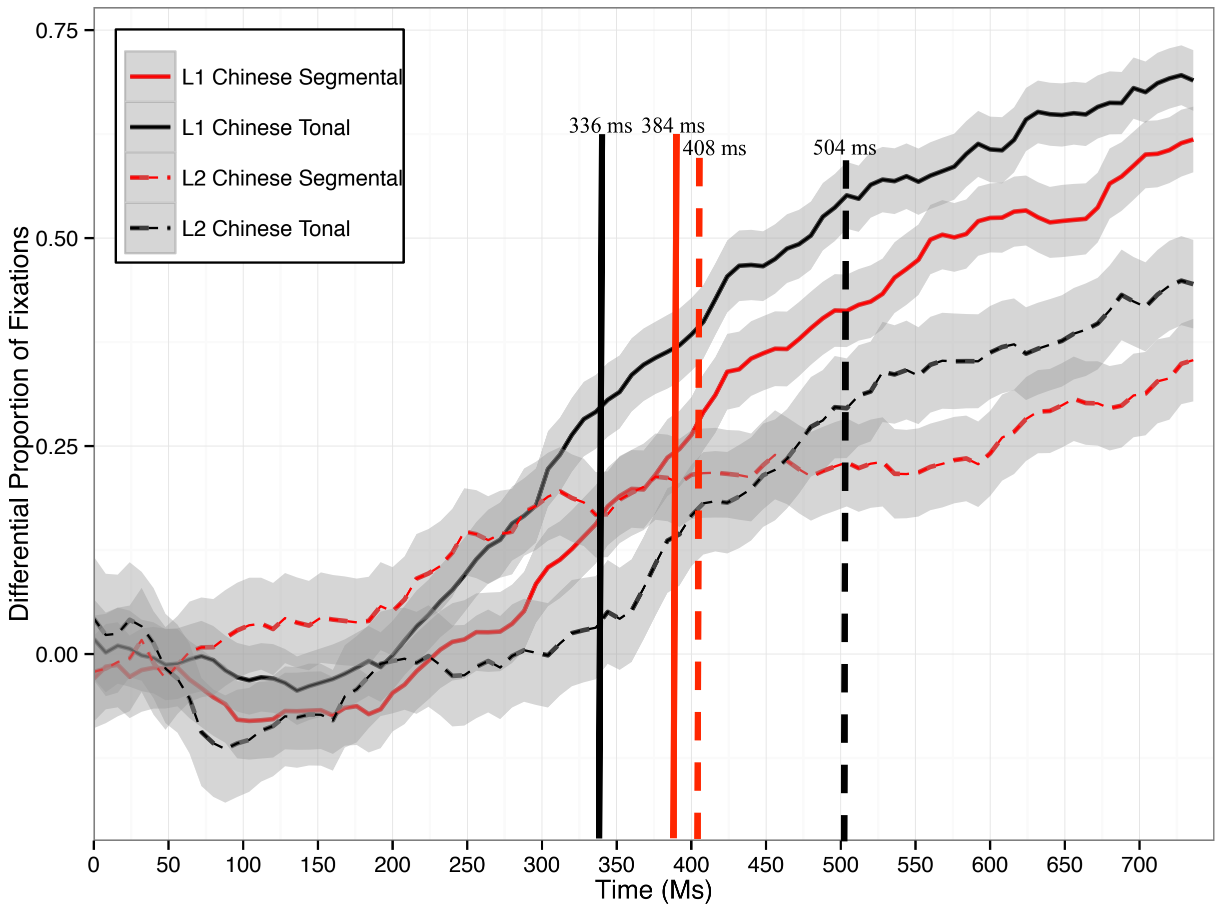

Figure 18: Differential proportion fixations for native Chinese listeners (solid lines) and English L2 learners of Chinese (dashed lines) with the segmental condition represented in red and the tonal condition represented in black. Vertical lines represent the target-competitor divergence, with for L1-Chinese listeners (red lines) and L2-Chinese listeners (black lines) with the tonal condition represented in solid lines and the segmental condition represented in dashed lines Time in milliseconds is presented on the *x*-axis and differential proportions of fixations are presented on the *y*-axis. The shaded region represents ±1 standard error of the mean.

These results reveal several interesting observations. Beginning with native listeners (displayed in the red and black solid lines in Figure 18), we see a difference in the timing with which target and competitor fixations diverge, with the divergence point being slightly earlier for tonal items than for segmental items (by 48 ms). This indicates that tonal information may be processed slightly earlier that segmental information, contrary to recent results (c.f., Malins & Joanisse, 2010, 2012; Zhao et al., 2011). The effect of condition could indicate that across time, these two conditions are different in terms of proportion looks to the target. If there had been no other effects, this effect of condition would have been interpreted as a timing effect; however, since there were interactions with time, this effect of condition is insufficient to make the claim that there is a timing-of-use difference between tones and segments. Alternatively, this difference may instead be attributable to the speed difference found in the GCA analysis of native listeners' proportions of target fixations. Since the proportion fixations to target rise faster in the tonal compared to segmental conditions, this could cause the differential proportion fixations to target to become significantly above 0 earlier than for the segmental condition. For this reason, no timing difference between tones and segments is claimed for native listeners.

Looking at L2 learners' data (displayed in the red and black dashed lines in Figure 18), the opposite effect is seen, with target and competitor fixations diverging approximately 100 ms later in the tonal condition than in the segmental condition. This indicates that although L2 learners' target fixations accrue at a faster rate in the tonal condition than in the segmental condition (as shown in the GCAs), L2 learners appear to begin using tonal information later than segmental information. In other words, L2 learners' proportion fixations to the target and competitor diverge later for the tonal condition than for the segmental condition, but once it diverges, the speed with which these proportions of fixations increase is greater for the tonal condition than for the segmental condition. This means that L2 learners show a difference in both the timing with which they use tonal and segmental information (i.e., tonal information begins constraining lexical access later than segmental information) and in the speed with which they use these two types of information (once tonal information begins constraining lexical access, it does so more rapidly than segmental information). Comparing the Chinese and English groups, the results in the segmental condition show very little difference in the timing by which the groups' target and competitor fixations diverge, with only a 24-ms delay for L2 learners. The greatest difference between the groups comes in the tonal condition, with L2 learners using tones 168 ms later than native listeners do. These results indicate that the timing by which L2 learners process segments is native-like, but the speed at which they process tones differs from that of native listeners from native listeners.

6.2.8. DISCUSSION

This chapter reported on a visual-world eye-tracking study comparing the use of tonal and segmental information in Chinese online spoken-word recognition. Accuracy in selecting the target over the competitor word and eye-tracking fixation data were collected, and GCA and LME models were conducted on the fixation data to investigate how native Chinese listeners and English-speaking L2 learners of Chinese use tonal and segmental information in lexical access. The results presented in this section reveal several important effects. First, contrary to recent work, differences between the processing of tonal and segmental information *were* found for native Chinese listeners, with a definite speed advantage for the processing of words that differ in tones compared to the processing of words that differ in segments. Second, L2 learners showed later use of tones compared to segments, but nonetheless showed a speed advantage for tones as compared to segments. Finally, comparing native listeners and L2 learners, several interesting effects were found: the two groups showed comparable timing of use of segments, but L2 learners showed a delayed use of tones compared to their own use of segments and compared to native listeners' use of tones. We now turn to a discussion of each of these findings.

6.2.8.1. Native Chinese listeners

The results of the present study for native Chinese listeners shed new light on the use of tonal and segmental information in lexical access, and are not in line with the findings of previous studies on this topic (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011). The present study revealed two significant differences between tones and segments. Beginning with the issue of timing, the results of the GCA model reported in Table 7 revealed a significant effect of condition. This effect of condition could indicate that across time, these two conditions are different in terms of proportion fixations to the target. If there had been no other effects, this effect of condition would have been interpreted as a timing effect (in the absence of baseline effects), with tones being used earlier than segments for both groups; however, since there were interactions with time, this effect of condition is insufficient to make the claim that there is a timing-of-use difference between tones and segments. Therefore, a second analysis was conducted to directly investigate the timing of use of tones and segments in more detail.

This follow-up analysis investigated the timing by which target fixations diverged from competitor fixations using time bins and comparing the differential proportion of fixations to 0. This point can be taken as the first time bin where the proportion of target fixations increased over the proportion of competitor fixations. Using this measure, the timing of initial use of tones and segments was identified and compared. These results showed that tones were used approximately 48 ms *earlier* than segments. This effect is in opposition to the conclusions of recent works that found no timing difference between the use of tones and segments using both eye-tracking and EEG measures (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011); however, from the analysis conducted in this study, it is not possible to determine if the observed timing difference between the use of tonal and segmental information is reliable.

The difference in timing between the tonal and segmental conditions is numerically small, and no statistical analyses were possible to determine if this numerically small difference is significant; therefore, no solid claims can be made. [22] As such, it is not claimed that these results counter the claims of studies finding no timing difference. Instead, it is more prudent to claim that, like previous work, that tones are not used after segments in online spoken word recognition measures (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011).

In terms of speed, the results of the current study are also not in line with those of previous research. For native listeners, the results presented in Table 7 are in direct opposition to the prediction that tones and segments would be used with the same speed. The GCA model revealed that the proportions of target fixations had a steeper slope in the tonal condition compared to the segmental condition. The increased slope of the fixation curve in the tonal condition indicates that native listeners were able to use the tonal information to constrain the lexical search more rapidly than they were able to use the segmental information. This means that although tones and segments begin to influence the lexical search at similar times, the tonal information constrains the lexical search more rapidly than the segmental information.

The discrepancy between the results of the present study and those of previous eye-tracking research (Malins & Joanisse, 2010) is likely due to a difference in the degree of controls. The current study more tightly controlled the timing of the arrival of information by selecting tonal and segmental pairs that disambiguated with the same duration of acoustic input. Recall from the norming study that the syllable pairings selected for this study did not differ in the timing by which native Chinese and naïve English listeners were able to discriminate the pairs. The segmental items used in Malins and Joanisse (2010) included pairs from the four hypothesized segmental disambiguation timings and the two hypothesized tonal disambiguation

[22] For learners however, this difference is numerically very large, and is more likely to be reliable.

timings used in the norming study. The present work also controlled the onsets of the target word so as to only have non-sonorant-initial words. This controlled the timing of the arrival of the tonal information, since onsets like [m] would carry tonal information, whereas onsets like [p] would not. The stimuli in Malins and Joanisse (2010) included both sonorant-initial and non-sonorant-initial words, which makes the arrival of the tonal information variable (i.e., earlier for sonorant-initial words and later for non-sonorant-initial words). It is possible that this variation in the materials in Malins and Joanisse (2010) resulted in the speed advantage of tones being masked in their study.

One possible explanation for the underlying cause of this tonal advantage can immediately be ruled out. Since the word frequencies between targets and competitors were controlled within and between conditions, the greater accrual of proportions of target fixations in the tonal condition than in the segmental condition cannot be due to the relative frequency of the words in the tonal and segmental conditions.

A second possibility is that this difference may be related to the number of competitors removed from the lexical search when processing tonal information as compared to when processing the rime portion of a word. There are only four lexical tones in Chinese, and as many as 36 rimes (Chen et al., 2004). Recall that the norming study investigated the duration of acoustic input needed to discriminate pairs of tones and rimes, and the stimuli for the eye-tracking experiment were selected so that the tonal discriminations could be made with the same amount of acoustic information as the segmental pairs. When thinking about the incremental process of eliminating lexical competitors, tones may have an advantage over segments. Beginning with segments, recall that the segmental pairs all took the form of the items sharing an initial consonant and onglide vowel (either [i] or [u]), and the difference between the target and competitor came in the vowel directly following the onglide. In all of these cases, however, at the point where the processor has accumulated enough information to be certain of the initial onglide being heard, the rime could still continue in 3 to 10 different ways depending on the vowel. For example, the vowel [i] could continue as [ia], [ian], [iaŋ], [iau], [iɛ], [in], [iŋ], [ioŋ], [iuo] or simply continue as [i]. Even if we assume some co-articulation between the vowels, for example, determining from the onset of [i] that the sequence of onglide and following vowel is [ia], the rime could still continue as [ian], [iaŋ], [iau], or simply as [ia]. All of the words that correspond to all of these rime continuations may all compete for selection with the target word being heard. This would increase lexical competition and slow down word recognition.

For tones, on the other hand, the amount of competition is quite different. Remember that all of the tone pairs disambiguated at their onsets. This means that upon perceiving the initial consonant and first portion of the vowel and pitch height information, two tones can already be excluded from the lexical search. If the early pitch is high, the tone cannot be Tone 2 or Tone 3, both of which begin with a low pitch; the only options compatible with the signal are Tone 1 and Tone 4. This allows the processor to exclude half of the lexical items based on initial pitch height alone. For segmental contrasts on the other hand, upon perceiving the initial consonant and first portion of the vowel certainty of the onglide does not mean that only words with that rime will be activated, since there are anywhere from three to ten rimes that will have that share that initial onglide Therefore, upon being certain of the consonant and initial portion of the vowel, all rimes with that onglide, and all words with those rimes, will compete with the target for selection. This drastic removal of lexical candidates for tones, and more gradual removal of lexical candidates for rimes, may be what allow have robust representations s the proportions of target fixations to rise more rapidly in the tonal condition than in the segmental condition.

There are two assumptions crucial to this argument that need further discussion, both dealing with the dynamics of competition due to lexical items not present in the eye-tracking display. The first assumption is that lexical competition effects can be found even when no competitor is present on the screen (e.g., when the target is presented with three distractor items): Magnuson et al. (2007) showed that words from denser lexical neighborhoods were recognized more slowly than words from sparser lexical neighborhoods, even when the neighbors (which compete for lexical activation) were not displayed. This indicates that while competition between items presented on the screen is arguably stronger, lexical items not presented in the visual display can also compete for lexical access. Therefore, it is reasonable to assume that the overall dynamics of the lexicon can come into play to explain the current speed advantage for tones (Assumption 1).

The second assumption is that the items that are not in the display do not compete to an equal extent. The explanations provided above suggested that in the tonal condition, only words that have the same segments but differ in tones compete for lexical selection, whereas in the segmental condition, only words that have the same tones but differ in segments compete for selection. In other words, this explanation assumes that the lexical competitor present in the display further narrows the lexical search to only words that overlap in segments (in the tonal condition) or in tone (in the segmental condition) (Assumption 2).

Table 13 illustrates the dynamics of lexical competition from words not included in the display. This table shows an (intentionally unrealistically small) example cohort of words all beginning with an initial consonant [p], with five different rimes listed in the rows, and the four tones for each listed in the columns. Each lexical entry is provided in Pinyin on the left and in IPA on the right. For example, for the target *ba1* ([pa1]) with a high tone and the competitor *ba2* ([pa2]) with a rising tone (top shaded area of Table 13), only the words in Row 1 are assumed to compete with each other, but when the processor identifies that the early pitch of the target is high, words with Tone 3 and Tone 4 become excluded from the lexical search. In theory, the non-displayed segmental competitors of the target *ba1* ([pa1]) and competitor *ba2* ([pa2]) (bottom un-shaded area of Table 13) could also compete with the target, and so could words with other rimes and tones (bottom shaded area of Table 13). Their not competing as much thus relies on the aforementioned Assumption 2.

Table 13: Example of on-screen/off-screen lexical competition dynamics

	Tone 1		Tone 2		Tone 3		Tone 4	
Rime 1	ba	[pa]	ba	[pa]	ba	[pa]	ba	[pa]
Rime 2	bai	[pai]	bai	[pai]	bai	[pai]	bai	[pai]
Rime 3	bao	[pau]	bao	[pau]	bao	[pau]	bao	[pau]
Rime 4	ban	[pan]	ban	[pan]	ban	[pan]	ban	[pan]
Rime 5	bang	[paŋ]	bang	[paŋ]	bang	[paŋ]	bang	[paŋ]

When presented with a display of a tonal pair and a segmental pair, the participant cannot *a priori* know which pair will be targeted. However, upon hearing the initial consonant, the targeted pair becomes clear. After hearing the initial consonant [p], the processor knows that the tonal pair of a target *ba1* ([pa1]) and a competitor *ba2* ([pa2]) are the relevant items, and that the decision to be made is a tonal one, not a segmental one. Therefore, Assumption 2 is that the target word pairs (which can be identified from the onset of the consonant) can alter the degree to which non-displayed lexical items compete, such that off-screen tonal competitors with the same segments (top un-shaded area of Table 13) will compete to a greater degree than either off-screen segmental competitors with the same tones (bottom shaded area of Table 13) or off-screen competitors with the rime-tone combinations (bottom un-shaded area of Table 13).

If Assumption 2 holds, then when the initial high pitch of Tone 1 is perceived on the target word *ba1* ([pa1]), then the activations of words with Tone 2 (the on-screen competitor) and Tone 3 will be inhibited, and the activations of the target and a non-displayed Tone 4 competitor will rise, since both Tone 1 and Tone 4 have high initial pitch. Importantly, since only Tone 1 and Tone 2 items are present in the display, the Tone 1 item will be correctly selected and recognised. The off-screen segmental competitors will also compete, but they will not compete to the same extent as the off-screen Tone 4 word that contains the same segments. The initial consonant informed the processor that the decision to be made was about tones, not segments, and so processor will give higher weight to tones in the lexicon.

This explanation of the results would be stronger, however, if it could be shown that the target words used in the segmental and tonal conditions were balanced for the number of possible rimes they can take. If the target words in the segmental and tonal conditions did not differ in the number of possible rimes they can take, then the results found in this study could not be attributed to this particular characteristic of the test items.

The SUBTLEX-CH Corpus (Cai & Brysbaert, 2010) was used to identify all the possible unique syllables of Chinese. From there, the number of rime+tone combinations for each onset was calculated (e.g., the initial [p] occurs with a total of 54 rime+tone combinations). Then, each stimulus used in the experiment was coded as the number of possible rime+tone combinations based on the word onset (e.g., all items with the initial [p] were coded as 54). These codings were then summed separately for each condition. The results are presented in Table 14. A paired samples *t*-test did not yield a significant difference between the two conditions ($t(15)=–0.36$, $p>0.1$).

Table 14: Summed totals of the number of rime+tone combinations by condition

Condition	Sum
Segmental	705
Tonal	689

This analysis shows that the number of possible rime+tone combinations was not significantly different between the segmental and tonal conditions. This means that the speed effect seen in Experiment 2 is likely not due to the target words in the segmental condition having more possible rime continuations than those in the tonal condition. Although this does not provide direct evidence in support of Assumption 2, it rules out that the observed speed difference between the tonal and segmental conditions is due to a lack of control in possible rime continuations of the target word. We thus propose that the information in the visual display helps narrow down the lexical search outside of that display, but this will have to be tested in further research.

One remaining effect that needs to be discussed is the early boost to competitor fixations, as seen in Figure 10. From about 0 to approximately 200 ms (not including the baseline), the segmental competitor fixation line was well above the fixation lines for all other objects presented on the screen. This numerical increase in proportions of competitor fixations would have to mean that participants were more likely to look at the competitor item than the target item in the segmental condition, but only for the first 200 ms of the word. This early increased proportions of competitor fixations in the segmental condition could be considered problematic, in that it may have contributed to the speed difference observed between the tonal and segmental conditions.[23] However, we argue that this early difference is not what caused the observed speed difference between tones and segments. First, the same speed effect is seen for L2 learners, who

[23] Special thanks to Dr. James Magnuson for a helpful discussion of this effect

do not show an early boost in segmental competitor fixations. Second, further inspection revealed that only two individual participants are responsible for this early advantage for competitor fixations: These participants began looking, and continued looking, at the competitor item, with proportions of competitor fixations ranging from 0.5 to 0.75. This bias to look at the competitor items visually differed from other participants' proportions of competitor fixations, which were at chance (0.25). A secondary analysis excluding these two participants confirms that this competitor boost was driven by these two participants (updated data visualizations reveal no early increase in proportions of competitor fixations in the segmental condition), and the interaction between condition and the linear time term remains significant even when these two participants are excluded (see Appendix G for graphs and GCA results), suggesting that this early competitor boost is not responsible for the observed speed difference between the tonal and segmental conditions.

The present results make a very concrete prediction that could be tested in other languages: It predicts that the number of tones compared to segments (possible rimes) in any given language should mediate the slope difference in eye-fixation data in a predictable way. For example, a language with more tones but the same number of rimes as Chinese should show a more comparable slope between tones and segments; however, a language with a similar number of tones but more rime possibilities should show a more exaggerated tone speed advantage. Future work should investigate these issues with other tonal languages to see if this speed difference of tones is indeed modulated by the ratio of tones to rimes, which would help to confirm this possible explanation. If this speed effect cannot be attributed to a differing number of competitors being removed from the lexical search, then more exploratory studies combined with computational modelling would likely be the most informative method for discovering the underlying cause of the speed difference found by the current work.

The discrepancy between the results of the present study and those of previous eye-tracking research (Malins & Joanisse, 2010) may again be due to the tighter control of the stimuli in the present study. The greater variability in Malins & Joanisse's (2010) stimuli may have masked the speed-of-use difference observed between the tonal and segmental conditions in this study.

6.2.8.2. Second language Chinese listeners

The results presented in this chapter also included a group of L2 learners of Chinese. Previous priming research with L2 learners would have predicted that L2 learners would show equal timing of use between tones and segments and equal speed of use (Sun, 2012). By contrast, previous work with L2 learners' use of segments that exist in the L2 but not in the L1 predicted the L2 learners would have difficulty using tones in online processing (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011). In terms of comparing tones and segments within the L2 group, the results showed that tonal information is used later than segmental information, but once used, it constrains the lexical search faster than do segments.

Beginning with the accuracy with which participants selected the target, the results showed that learners were 93.6% accurate in the tonal condition and 97.8% in the segmental condition. While learners' accuracy rates were numerically worse in the tonal condition than in the segmental condition, these accuracy rates were nonetheless very high. Sun (2012) reported tonal accuracy of around 50% and segmental accuracy of about 74% in his priming task). There are several reasons for the higher accuracy reported in the present study. First, the L2 learners were more proficient in Chinese, with L2 learners reporting an average of 4.5 years of instruction and with a majority of them living/studying abroad in China at the time of testing (with Sun (2012) reported that his L2 learners had approximately 3 years of instruction). Additionally, and perhaps more importantly, because the L2 learners in the present study were trained on the word-picture associations, they had prior exposure to the stimuli used in the experiment. This may have increased their accuracy across the board and reduced the difference between their performances in the segmental and tonal conditions.

In terms of the eye-tracking fixation data, effects of both timing and speed were found. Beginning with timing, it was predicted that L2 listeners would either show native-like timing of use of tonal information or they would show a delayed use of tones compared to segments. Like with native listeners, for the L2 learners of Chinese, the words in the tonal condition generated higher proportions of target fixations than the words in the segmental conditions, indicating that tones were used later than segments in lexical access, however, this effect is unreliable given the interactions with time. This effect of condition in the absence of interactions with time would have indicated that tones were used earlier than segments for second language learners. The results of the target-competitor divergence point analysis presented in Table 12 and seen in Figure 18 revealed that in fact L2 listeners made later use of tonal information than of segmental information (by approx. 100 ms), which supports the second predicted possible outcome. This means that tones are beginning to constrain the lexical search much later than segments are beginning to be used. Like research on L2 learners' use of new segments (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011), these results suggest that L2 learners are having difficulty making efficient use of tones in online word recognition; consequently, tonal information constrains the lexical search much later than does segmental information.

These results are attributed to the fact that the L1 of the L2 learners is not a tone language. The L2 segments in this study were selected so that they would map straightforwardly onto L1 segments. By contrast, English does not have lexical tones. Thus, learners must acquire four new tonal categories and use them in online spoken word recognition. Whether L2 learners adapt their L1 stress system or use some other mechanism to form tonal categories and process lexical tones, this seems to have the consequence of delaying their use of this new information in lexical access. The segments used in this study, which are similar between Chinese and English, were used on a native-like timescale, and it is only the use of lexical tones that was delayed. For this reason, it is argued that the timing of use of information is L1-dependent.

Although L2 learners used tones later than segments, this does not appear to be a two-stage process like that described by Lee (2007). Recall that Lee (2007) suggested, based on priming research with native listeners, that segments were used online to constrain the word search and tones were used at a post-lexical selection stage. While this has been shown not to be the case for native Chinese listeners' use of tones, a two-stage process could have been found for L2 listeners, with the new tonal categories being processed at a post-lexical decision stage. In the present study, L2 learners used tones at about 500 ms into the word (excluding the 200-ms baseline). Since the words in this experiment had a duration of 524 ms, this indicates that the L2 learners were using the tonal information at around the end of the word. Although this timing is close to the end of the word, it is still within the word itself, which would indicate that this is likely not a post-lexical decision process;

rather, the use of tonal information appears to be part of L2 learners' spoken word recognition process.

These results differ from the results of the other L2 lexical tone processing research in several key respects. Recall that Sun (2012) concluded that English-speaking L2 learners of Chinese used tones and segments on the same time scale. The difference between Sun's (2012) results and those of the present study is likely to stem from the different methodologies used in the two studies. In Sun (2012), timing of use was established using two ISIs, one of 50 ms and one 250 ms. The effects observed in the present research reflect direct, simultaneous competition between two words, and these effects happen over the course of the word. It is possible that the effects reported by Sun (2012) reflected the state of the processing system after lexical access had been completed. In order to understand the processes that underlie spoken word recognition, time-sensitive measures are essential, and the temporal quality of eye-tracking is better suited to make timing claims than is priming.

With regards to the speed of use of tones and segments, perhaps surprisingly, the same pattern that was observed with native listeners was also present for L2 learners: The results in Table 8 show a steeper slope in the tonal condition than in the segmental condition in the proportion fixations to target. The cause of this tonal speed advantage is believed to be the same as for native listeners. This does not necessarily entail that the native listeners' and L2 learners' lexicons are identical, however; it merely suggests that the use of tonal information more rapidly reduces the set of word candidates that compete for lexical access.

The conditions were controlled for frequency, but this frequency measure is more likely to reflect the structure of native listeners' lexicon than that of L2 learners. It would be naïve to assume that these frequencies hold for L2 learners as well, since their experience with Chinese is very different from that of native listeners. For this reason, word familiarity data were collected to gain an understanding of the 'frequency' of these words in the L2 lexicon. If L2 learners were more familiar with the words in the tonal condition than with those in the segmental condition, it could explain the speed effect found; however, L2 learners' word familiarity ratings in the tonal and segmental conditions did not differ significantly. L2 learners' word familiarity ratings were also not significantly different between targets and competitors, and between conditions, as shown in Table 6. Hence, the speed-of-use difference between the tonal and segmental conditions cannot be due to L2 learners' familiarity with the words used in the experiment. This speed difference can also not be due to a difference in L2 learners' ability to associate the pictures to the spoken words in the segmental condition as compared to the tonal condition, since the words that learners failed to reliably associate with the corresponding pictures in the training were excluded from the data analysis. Finally, the speed difference cannot be attributed to the greater accuracy in one condition over another, since trials where participants did not click on the target were excluded from the analysis.

It is believed that this speed of use difference of tones and segments for native listeners is due to an inherent property of the lexicon, namely the ratio of tones to segments in the language. Given the high proficiency of the L2 learners, it is plausible that native listeners' and L2 learners' lexicons have similar ratios of tones to segments. In other words, L2 learners may have had enough exposure to Chinese for their lexicon to be similarly structured as the lexicon of native Chinese listeners. We thus conclude that the ratio of tones to segments in Chinese is what caused L2 learners to also show a speed advantage for the processing of tones as compared to the processing of segments.

Once again, these L2 results differ from the results of the other lexical processing research with L2 learners of Chinese. Recall that Sun (2012) concluded that while L2 learners of Chinese were slower than native listeners to process tonal and segmental information, there was no difference in the reaction times to tonal and segmental items. The present work has shown that tones are used later, and once they begin to constrain the word search, they constrain it much more rapidly than segments. Again, this discrepancy between the two studies may be due in part to the greater sensitivity of eye-tracking to the time course of lexical processing, to the different stimuli used in the two studies, and to the different proficiency and lexical knowledge of the L2 learners in the two studies.

6.2.8.3. Native vs. second language Chinese comparison

When comparing native Chinese listeners and English L2 learners of Chinese, the results reveal a mixture of the predicted outcomes. From the results of Sun (2012) it was predicted that L2 learners would use tones and segments at the same time, but would process them much more slowly than would native listeners. By contrast, from the research on the processing of segmental contrasts that do not exist in the L1 (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011), it was predicted that L2 learners would have difficulty using tones compared to native listeners and compared to segmental information that can be mapped onto L1 categories, in timing, speed, or both.

With regards to timing, the results of the present study are not in line with the results of Sun (2012). The results presented in Table 12 and Figure 18 showed that L2 learners used segmental information at almost at the same time as native listeners, with only a 24 ms delay. This is in line with the results of Experiment 1, which showed comparable segmental performance across the groups. The segments chosen for this experiment existed in both languages; hence, it is not surprising that L2 learners were able to use this information with a native-like timing. L2 learners' timing of use of tones, however, was not native-like, with L2 learners beginning to use tones 172 ms after native listeners. This is a substantial delay that would not have been predicted based on Sun (2012), who showed that L2 learners' processing of *both* the tonal and segmental contrasts was delayed in comparison to native listeners.

As previously discussed, L2 learners' delay in the use of tonal information compared to native listeners' use of the same information is likely a consequence of L2 learners having to create and use new tonal categories for online lexical processing. This could cause difficulty for learners in processing the new lexical tones. Previous work on L2 learners' use of new segmental categories (e.g., Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011) suggested that when L2 learners have difficulty distinguishing between L2 phonetic categories, they experience difficulty using this information to recognize the words, which leads to increased lexical competition. The present results extend this conclusion to new categories in the suprasegmental domain, and suggest that when L2 learners form new suprasegmental categories these tonal categories may not have robust representations, and consequently they may not be mapped efficiently from the speech signal to L2 learners' lexical representations. As a result, the mapping between speech signal and L2 lexical representations may not be as efficient as for native listeners. As such, L2 learners may experience more difficultly in word recognition compared to L2 segmental categories that can be mapped onto L1 categories; this difficulty results in the delayed use of tonal information in constraining the lexical search.

As for the speed of use, the predicted outcome for L2 learners was slower overall processing of both tonal and segmental information compared to that of native listeners. This effect was indeed found and can be seen most clearly in the differential proportion of target fixations in Figure 17, where the slopes of both the tonal and segmental conditions were much shallower for L2 learners than for native listeners.

The present results also solidified the distinction between timing and speed, in that the timing of use does not predict the speed of use. That is to say, earlier does not necessarily mean faster. While L2 learners are slower in using tonal information compared to native listeners (likely due to unstable tonal representations), their lexicons are argued to be similar in structure, and so the ratio of competitor removed from the search between tones and segments would be similar for L1- and L2- Chinese listeners. This similarity would therefore cause tones to constrain the lexical search faster than segments for both L1- and L2-Chinese listeners, even though L2-listeners are slower than native listeners overall. This distinction is the clearest for the L2 group. L2 learners processed segments earlier than tones, but once tones were used, they more rapidly constrained the lexical search. Additionally, native listeners and L2 learners showed comparable timing in the use of segmental information; these results are believed to be due to the shared segments between the languages used in the task.

It is interesting to note, then, that whereas L2 learners began to use segmental information at the same time as did native listeners, their speed of use of the same segments was slower than that of native listeners. This suggests that, whereas the timing of use of information in lexical access appears to be L1 dependent, the speed of use of the same information appears to be L1 independent, with both segments and tones being slower to be used in L2 processing than in native processing. This slower processing extends to categories that exist in the L1 (segments) and categories that do not exist in the L1 (tones). Segments are also slower regardless of the native-like *timing* of use: Whereas L2 learners begin to use segments to constrain the word search at the same time as native listeners, they are unable to use this information to constrain the lexical search as rapidly as native listeners. In other words, although L2 learners are using segmental information to begin constraining the lexical search at the same time (with the same duration of acoustic in as native listeners, once this information is used, it does not constrain the lexica search as rapidly as for native listeners.

This slower L2 processing is most likely due to increased lexical competition, as is commonly seen in both L2 (new) segmental and suprasegmental processing (Broersma, 2002, 2005; Broersma & Cutler, 2008, 2011; Cutler & Otake, 1999; Cutler et al., 2006; Sebastián-Gallés, Echeverría, & Bosch, 2005). This increased competition is likely due to weak representations. However, in the case of English speakers learning Chinese, it may come as a surprise that while they used segmental information with the same timing as native listeners, they were still slower to use this information to constrain the lexical search compared to native listeners. Even though L2 segments that could be mapped onto L1 segments were used in the current work, it is possible that the mapping may not have been perfect (i.e., the L2 sounds may differ phonetically, at least to some degree, from the corresponding L1 sounds). This may be the reason why L2 learners showed slower use of segments (shallower slope of fixations to target) than for native listeners. In other words, the mapping was sufficiently stable to result in a native-like timing, but not perfect, and thus it may have resulted in some uncertainty on the part of the learners, who then ended up showing more competition as compared to native listeners. In order to understand if the slower speed of use is due to imperfect mappings between L1- and L2 sounds, or is simply due to timing being L1-independent, more detailed cross-linguistic acoustic analysis of the sounds involved would be needed to see if the degree of difference between L1- and L2 sounds predicts L2 learners' speed of use.

6.2.9. Conclusions

The results of the present eye-tracking study revealed several key findings. First, for both native listeners and L2 learners, it was found that tones are used to constrain the lexical search more rapidly than segments. Future work will be needed to determine the exact cause of this tonal advantage, but whatever the cause, L2 listeners are sensitive to it as well. In terms of timing, it was found that native listeners use tones slightly earlier than segments. By contrast, L2 learners showed native-like timing of use of segments, but a substantial delay in the timing by which they used tones compared to segments and compared to native listeners' use of tones. This suggests that the timing of use of information is dependent on the use of that cue in first language, while timing is slowed in the second language regardless of whether or not the cue is used in the listeners' L1 or not.

CHAPTER 7: GENERAL DISCUSSION AND CONCLUDING REMARKS

7.1. Introduction

The studies presented in the previous chapters investigated the categorization of tones and segments and use of tones and segments in lexical access by native Chinese listeners and English-speaking L2 learners of Chinese with a norming study to select materials, and two main experiments.

Before the time-course of the categorization or use in lexical access of tones and segments could be investigated, a norming study was conducted to select tone and rime pairs that were optimally matched in their disambiguation timing. This study tested native Chinese listeners and naïve English listeners using a gated-AX discrimination task to identify the duration of acoustic input needed for listeners to discriminate the tonal and segmental pairs. Five tonal sets were included, including four early disambiguating sets and one late disambiguating set for comparison. The segmental pairs fell into four hypothesized disambiguation timings form early to late in order to best match a segmental set with the early tonal sets in terms of disambiguation timing.

The results showed that for native Chinese listeners and naïve English listeners, the early tonal pairs best matched with the post-onglide segmental timing. The materials for Experiment 1 were selected so as to fit the early tonal timing and the post-onglide segmental timing so as to control as much as possible the low-level psycho-acoustic differences between tone and vowel perception.

Experiment 1 focused on the time course of categorization of tones and segments (rimes) and tested L1- and L2-Chinese listeners. This study aimed to investigate if L2 learners of Chinese had an initial advantage in categorizing tones that differ in onset pitch and to investigate if tones were once again disadvantaged compared to segments. In order to investigate these two issues, a gated forced-choice identification task was used. The stimuli were selected so as to fit the early tonal timing and the post-onglide segmental timing determined from the norming study.

It was predicted that L2 learners would have an initial advantage over native listeners given their superior use of pitch height in the perception of tones. The results showed that while L2 learners appeared to show higher accuracy rates than native listeners in early gates, this effect was not significant. As predicted, for segmental contrasts, L2 learners of Chinese performed identically to native listeners, and both groups had higher accuracy on segmental contrasts than on tonal contrasts.

Experiment 2 followed up on the results of Experiment 1 by investigating the use of tones and segments in L1- and L2-Chinese listeners' lexical access. Experiment 1 revealed no differences between L1- and L2-Chinese listeners' categorization of tones and segments; however, this task was offline, non-lexically focused, and meta-linguistic in nature, leaving open the question of whether L2-Chinese listeners would perform natively in online lexical processing. Experiment 2 therefore investigated the use of tones and segments in an online, lexically focused, and unconscious measure to determine if L2-Chinese listeners were still able to perform similarly to native speakers, and if not, in what ways did their use of tones and segments differ from native listeners. By using visual-world eye-tracking, both the timing of use and the speed of use of tones and segments were investigated for L1- and L2-Chinese listeners.

The results of Experiment 2 showed that for native listeners, tones were used at about the same time as segments. However, tones showed a distinct advantage in terms of the speed of use, constraining the lexical search more rapidly than did segments. For L2 learners, the same speed-of-use effect was seen, with tonal information influencing the lexical search more rapidly than segmental information. In terms of timing, however, L2 learners showed a significant delay in the use of tones compared to the use of segments and compared to native listeners' use of tones.

Taken together, the results of these two experiments leave some open questions. First, these studies aimed to resolve the contradiction between the tonal disadvantage observed in earlier offline work (e.g., Cutler & Chen, 1997; Taft & Chen, 1992) and the analogous use of tones and segments see in more recent online work (Malins & Joanisse, 2010, 2012; Schirmer et al., 2005; Zhao et al., 2011). Experiment 1 was an offline categorization task, and the tonal disadvantage was found. Experiment 2 used similar materials as Experiment 1 in an online eye-tracking task, and showed that tones are used at the same time as segments, and are actually used faster than segments. These results indicate that tones are not globally disadvantaged, and suggest that the type of task used (i.e., offline, less lexically focused, meta-linguistic vs. online, lexically focused, unconscious) may determine if tones are disadvantaged. This leaves an open question: *Why* is it that, in the present offline categorization task, segments have an advantage, whereas in the present online eye-tracking task, tones seem to have a (speed-of-use) advantage for L1 listeners? What causes this reversal of effects from offline to online tasks? Second, for L2 learners, the effects between the two tasks do not reverse as they do for native listeners, in that L2 learners take more time to distinguish target and competitor words based on the tone they hear than based on the segments they hear in both the offline categorization task and the online eye-tracking task, except in terms of online speed of use, where they constrain the lexical search more rapidly. Why do L2 learners not show the same reversal of effects seen with native listeners? Third, L2 learners showed identical performance to native listeners in the offline categorization task; however, L2 learners differed from native listeners in online processing, showing later and slower processing of tones compared to native listeners. Why is it that L2 learners do not show the same pattern as native listeners in online processing?

This chapter will discuss each of these findings in more detail, as well as discuss some methodological implications of the current research and the impact it may have on the computational modelling of lexical access.

7.2. Native Chinese listeners

For native listeners, there was a striking difference in the pattern of effects from Experiment 1 to Experiment 2. In Experiment 1, an advantage was seen for segments over tones in terms of the overall accuracy and the rate of improvement in correct responses across the gates. In Experiment 2, a similar timing of use of tones and segments and a speed advantage for tones over segments was observed. While these experiments used different tasks, it would at first seem odd that the pattern of effects would change between a gating task and an eye-tracking task; however, the explanation is relatively simple. In Experiment 1, participants were simply asked to identify the tone or the rime of the fragment they were hearing. When identifying the tone, participants could only rely on F0 as a cue to tonal identity. For segments, on the other hand, rime identity was signaled by a host of cues, including, but not limited to, F1, F2, and F3. With more cues to rely on, participants could be more certain of their segmental responses than they could be of their tonal choices. With the same amount of acoustic input, segmental information possessed more cues, and therefore allowed for greater identification accuracy. An opposite prediction is also possible in relation to the number of cues. It could also be the case that with more cues, the processor needs more time to integrate them. If this were the case, we would in fact expect the pattern of results to be the opposite, with segments having lower accuracy and slower rates of improvement compared to tones. Since this is not the case, and tones were shown to be disadvantaged compared to segments, it is argued that, in this case, more cues lead to greater accuracy instead of processing difficulty in the integration of multiple cues.

This advantage does not transfer to the eye-tracking task in Experiment 2 for one important reason: While participants' identification of segments may be more reliable even in online processing, the larger number of lexical competitors that overlap in onset (and first vowel) makes it so that the processing of segmental contrasts is slower compared to that of tonal contrasts. As previously discussed, identifying a vowel in a rime still leaves several possible continuations for that rime, with the words ending with all possible rimes competing to some extent and slowing processing. In the forced-choice gating task of Experiment 1, this lexical competition does not come into play. Participants were asked to identify a tone or a rime, not a word, and the full words were never heard. For this reason, it is not surprising that the pattern of results of Experiment 1 differs from that of Experiment 2, with segments showing an advantage over tone in the offline gating task, as in previous offline tasks (e.g., Cutler & Chen, 1997; Taft & Chen, 1992), and with tones showing no difference compared to segments (in terms of timing) or an advantage (in terms of speed) in the eye-tracking task.

7.3. Second language Chinese listeners

When comparing native Chinese listeners and English-speaking L2 learners of Chinese over Experiments 1 and 2, one lingering question remains to be answered: Why is it that L2 learners performed natively in the gating task but non-natively in the eye-tracking task 2? Experiment 1 showed that L2 learners patterned like native listeners for segments *and* tones; however, Experiment 2 showed that L2 learners did not begin to use the tonal information to constrain the lexical search until almost the end of the tone. This means that they began to use the tonal information well after native listeners. Additionally, L2 learners were slower than native listeners to use both tonal and segmental information.

The answer is twofold. On the one hand, the gating task may have stacked the cards in favor of L2 learners showing native-like performance. Not only is this a relatively easy task when not under time pressure, but it is also a meta-linguistic task, which could have aided learners in their identifications of the tones. Additionally, given the gated nature of the stimuli, the task was focused on the onset of words and highlighted the early pitch differences of the tones in question. For segments, the task was simply to identify the rime heard. For both conditions, the gated forced-choice task is a relatively easy task when not under pressure to comprehend or to link the sounds to a word. The fact that lexical effects were not at play in the gating task may have also made the two groups more similar in their performance. Hence, it is not surprising that learners were able to perform as accurately as native listeners.

On the other hand, with respect to the eye-tracking experiment, whole words were presented; therefore, participants had to process the words, not just the tones. Processing the whole word in an eye-tracking experiment introduces lexical competition, which, as seen by the slower overall processing of tones and segments, is greater for the L2 learners than for native listeners for a variety of reasons, likely caused by mapping onto unstable tonal representations. While there were only two segmentally matching items on the screen, it is likely that words with all four of the tones were competing to some extent in the eye-tracking experiment, even if not present on the screen. With the added burdens and complications of processing whole words and attempting to map onto unstable tonal representation, it is not surprising that learners are unable to show native-like performance in online lexical processing. Additionally, unlike Experiment 1, which was more meta-linguistic in nature, Experiment 2 used an unconscious measure. Such an unconscious task may be more likely to reveal non-native-likeness in L2 learners, because they don't explicitly control where they look at on the screen. Therefore, while L2 learners were native-like in the categorization of tones in offline tasks, this effect did not extend to online processing of spoken words.

7.4. Methodological implications

The results of this dissertation also have an important methodological implication. This study has shown, once again, that eye-tracking is an effective method for studying the online process of lexical access. Eye-tracking has long been used to study the time course of lexical processing (e.g., Allopenna et al., 1998), and the present work has shown that this method can be used to simultaneously compare the timing and speed of use of various types of information present in the speech signal. Specifically, with L2 learners, it was seen that while segments were processed earlier than tones, tones were used to constrain the lexical search faster than segments. By using a combination of statistical analyses, these two effects were successfully teased apart, and the results revealed novel information about how tones and segments are used to constrain the lexical search online. To my knowledge, eye-tracking is the only methodology capable of making this timing vs. speed distinction; even EEG only focuses on the timing of effects and cannot reveal the speed with which information constrains the lexical search once used.[24]

---

[24] An amplitude difference in EEG would also give rise to a change in the slope of the line between conditions; however, that would not be interpreted as a speed effect since the shape of the EEG waveform is not directly linked to the speed of processing as it is in eye-tracking. The link between the interaction with the linear term and the speed effect discussed here is specific to eye-tracking data, and does not extend to other methodologies, even if GCA is used.

7.5. Computational modeling implications

In addition to methodological implications, the results of the present work also have implications for lexical access theory and computational models. Most current models of lexical access deal exclusively with segmental information (e.g., (TRACE) McClelland & Elman, 1986; (Shortlist B) Norris & McQueen, 2008; (Merge) Norris, McQueen, & Cutler, 2000), with an attempt at accounting for suprasegmental information by modeling the segmental cue of vowel reduction in English stress (Norris & McQueen, 2008). In terms of modeling tones specifically, two models have been proposed, the first being a theoretical model of how tones would be incorporated into lexical access models by introducing the notion of a 'toneme', analogous to a phoneme (Ye & Connine, 1999), and the second being a computational models of tones and segments in lexical access.

Shuai and Malins (2016a) introduced a computational model built off of the TRACE theory (McClelland & Elman, 1986) that was implemented in the jTRACE computation model (Strauss, Harris, & Magnuson, 2007). The authors call this new implementation TRACE-t. This model takes the existing jTRACE architecture and modifies it to include tones. The underlying theoretical model TRACE is a connectionist model of spoken word recognition made up of three layers of units: featural, phonemic, and lexical. Each phoneme is activated by a matrix of feature nodes, which are activated by the input. Activation feeds forward from the activated features to their connected phonemes, which then spreads to the connected words. Nodes on the same level are linked by inhibitory connections, whereas connections between levels are excitatory. The model is interactive, in that there are feedback connections from the lexical layer to the phoneme layer.

The original jTRACE implemented the TRACE model by using a feature matrix of seven features with nine values each to encode the phonemes of the language. These phonemes were used to make up lexical entries, and were also used as the input to the model. As such, the input to the model was a linear ordering of the phonemes of the word, and each phoneme was ramped on and off, with each phoneme overlapping by several cycles to simulate coarticulation. jTRACE is a user-friendly program to run TRACE simulations and has been shown to account for over 12 phenomena (e.g., lexical effects, frequency effects, etc.) in spoken word recognition (Strauss et al., 2007). As such, each phoneme is made up of a feature matrix of seven features with nine possible values each, and input to the model consists of a linear ordering of these phonemes (e.g., *^br^pt – 'abrupt'*).

The modification done by Shuai and Malins (2016a) included taking the original feature coding matrix of seven features used to encode consonants and vowels (that had been rigorously validated) and recoding three of them for consonants (encoding voicing, manner, and place), three for vowels (encoding height, frontness, and roundness), and one for tone height (encoding height (1-5) and direction (rising, falling, level)). By doing so, the authors created segmental phonemes as well as tonal 'phonemes'.[25] In order to encode the simultaneous use of tones and segments, the authors made each lexical entry ten segments long, with alternating segmental and tonal units. Only mono-morphemic words were used in this simulation. Thus, for a word like *má*, the encoding would be something similar to $mT^{2L}mT^{2L}aT^{3R}aT^{3R}aT^{4L}$, with the *T* representing the tonal units and the superscript representing the pitch height (1-5) and direction (L=level, R=rising, F= falling) of that tonal unit. In this way, the authors felt that they had captured the simultaneous nature of the use of tones and segments, and were somewhat able to model the eye-

[25] The word 'phoneme' is used here instead of toneme to indicate that the tones were represented in the model in the same way that segmental phonemes are. They created phonemes in the model that represent the tones instead of the segments.

tracking results of Malins and Joanisse (2010). The model predicted the same pattern of results, but the timing and size of effects was very different between the real and simulated fixations.

While this is a worthwhile attempt at incorporating tones into lexical access models, it has several theoretical and practical flaws. First, alternating tonal and segmental units is not *simultaneous* use; it is *alternating* use of tones and segments. Figure 19 below illustrates how jTRACE handles coarticulatory information. Each phoneme ramps on and off over the course of 11 cycle and adjacent phoneme overlap for 6 cycles, indicated by the green spaces where the segments overlap in Figure 19. By having both phonemes activated together this is able to simulate coarticulatory effects between adjacent segments.

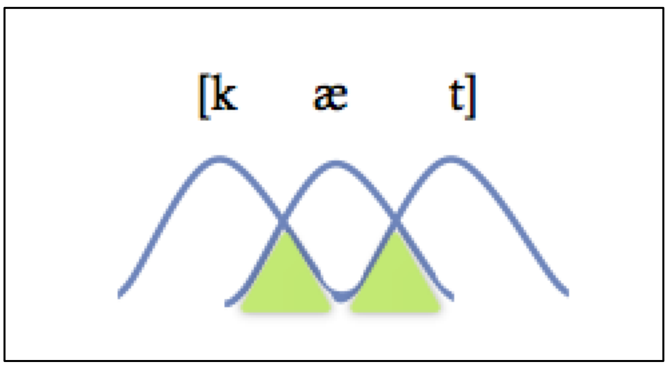

Figure 19: Illustration of jTRACE coarticulatory effects

TRACE-t, on the other hand, has intervening tonal units, as illustrated in Figure 20 below. This intervening tonal unit makes it so that adjacent segments would overlap for at most 1 cycle, as opposed to 6 cycles in the original model. This alternation of tones and segments therefore substantially reduces the model's ability to account for segmental coarticulatory effects.

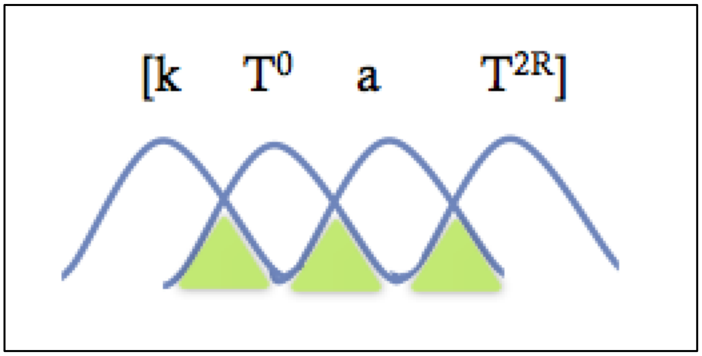

Figure 20: Illustration of TRACE-t lack of ability to capture coarticulatory effects.

Additionally, the encoding of the tones themselves is based off of what appears to be a classic phonological description of the tones, using the 5-point pitch height description (Chao, 1930) and including tonal direction. Shuai and Malins have mentioned that the model works significantly better by including pitch direction than when direction is removed (Shuai & Malins, 2016b). However, while these pitch points are commonly used to describe the tones, there is no evidence that native listeners divide the pitch space into these five heights to perceive the tones. Furthermore, while it is clear that native listeners rely on pitch direction to process the tones of Chinese (e.g., Gandour & Harshman, 1978), encoding direction in the tone 'phonemes' implies that in a single time unit of speech information, direction information is not only available, but also encoded in the speech signal. It makes more sense to assume that direction is computed from moment-by-moment changes in pitch values perceived by the listener. Thus, using the arbitrary 5-point pitch height and pitch direction for encoding tones does not provide a psychologically realistic model of the process going on in the use of tones in lexical access.

Shuai and Malins' (2016a) model did not show a speed-of-use difference between tones and segments, which is likely due to two properties of their model. First, by encoding both tones and segments as phonemes and alternating them, it is impossible to weight tonal or segmental information to differing degrees. The current explanation of the speed difference relies on the assumption (Assumption 1, Section 6.2.5.1) that the display allows the processor to weight information to varying degrees so that off-screen items do not compete to the same extent. With tones and segments encoded as alternating tonal and segmental units, modeling the effect is impossible. Secondly, as is common in eye-tracking simulations, the simulation only included the items in the eye-tracking study of Malins and Joanisse (2010). The speed effect is also claimed to be a property of the lexicon as a whole, where being certain of a tone will remove more competitors than a vowel will. This relies on the ratio of tones to rimes in the lexicon as a whole, and will likely not surface with such a restricted lexicon.

As such, there are several important limitations of the TRACE-t model that will need to be addressed in the future to accurately model lexical processing as is in the minds of native listeners, including mainly its inability to utilize coarticulatory effects and its treatment of tones as equivalent to segments. Additionally, by encoding pitch in five pitch heights and encoding direction, this system is only able to account for the tones of Chinese, and is not extendable to any other tone languages, let alone other types of suprasegmental information. Ideally, a model of lexical access should model not just one single type of suprasegmental information, but also all types in the same model. Otherwise, a separate model will be needed for every language, and that is neither efficient nor productive to the study of spoken word recognition.

7.6. Concluding remarks and future directions

To conclude, the results of the experiments presented here revealed several key effects. Beginning with native Chinese listeners, this work has shown that in terms of the time course of categorization, segments have an advantage over tones. In online lexical access, on the other hand, this work confirmed once again that tones and segments are used at around the same time in lexical access. A novel finding of the current eye-tracking work was the result that tones are used to constrain the word search faster than segments. In terms of second language learners, the present work revealed several novel effects. First, L2 learners of Chinese perform in a native like pattern in categorizing the tones of Chinese in a gating task; however, this native-like performance does not extend to online lexical access, with learners displaying significantly delayed use of tones in online spoken word recognition. Learners' processing of segments was native-like in the offline gating task, and in terms of the timing of use in online lexical access. With respect to the speed of use, however, L2 learners showed a faster use of tones compared to segments, the same pattern as seen with native listeners. Although tones were used faster than segments for learners, both tonal and segmental information had proportion of fixations to target slopes that were equally shallower than native listeners in constraining the word search. From these results it is claimed that the timing of use of information is dependent on how/if that information is used in the L1. The speed of use, however, is L1 independent, with L2 processing proceeding more slowly than native processing, regardless of similarity to the L1.

This work could continue in several directions. If taken down a computational modeling track, this work could be extended to use current computational models to examine the speed of use difference between tones and segments. If the proposed explanation (that it is the ratio of tones to rimes in the language) is indeed the underlying cause, then this effect should fall out from properties of the lexicon. If the lexicon of the simulation were sufficiently large and captured the general trends of a native lexicon, then this speed effect would be predicted to be seen with no adjustment to the model. If this is not the case, computational modeling, in conjunction with experimental research, could be used to further explore why tones were used to constrain lexical access more rapidly than segments.

Additionally, the speed of use difference could be investigated by investigating another tone language. If the ratio of tones to segments in the lexicon is the cause of the speed difference, then varying this ratio should modulate this effect in predictable ways. A language with the same number of rimes, but more tones should show a smaller speed effect, or an advantage for segments, whereas a language with the same number of tones and more rime should show a more exaggerated tone speed effect.

In terms of the L2 processing of tones, this work could be continued by investigating different L1 groups based on their use of pitch as a lexical cue. Based on the results of Connell et al. (2016), it would be predicted that speakers of a language with no lexical pitch distinctions (e.g., French or Korean) would fail to use the tones of Chinese in online lexical access. English speakers, who use pitch in lexical stress, were able to use tones in early word recognition in the present work. It could be predicted that the more a listener relies on pitch distinctions in the L1, the better their use of the L2 pitch contrasts would be in terms of timing. The present results suggest that timing is dependent on the cue's use in the L1, whereas speed (slower use of tones and segments compared to L1 listeners) is L1-independent, meaning that learners will be slower to use any type of information in the L2. If this is the case, we could predict Dutch speakers would outperform English listeners. Since stress is signaled solely by suprasegmental cues in Dutch, they may be able to utilize the tonal information earlier than English listeners. Even more than Dutch speakers, it could be predicted that Japanese speakers, who use a pitch-accent distinction, would outperform Dutch speakers, in that they differentiate high and low accents as opposed to the presence or absence of increased pitch, as in Dutch. In this case, it could be predicted that they would use tonal information even earlier than Dutch listeners.

One step further would be to investigate native tone language speakers learning Mandarin, such as Thai or Vietnamese speakers. Two predictions can be made of L1 tone language speakers' use of L2 tones in a similar experiment based on the current studies. First, it could be predicted that, barring interferences from the L1 tonal system, L1 tone speakers would process L2 Mandarin tones natively in terms of timing. This would be on account of tones *not* being a new feature that the L2 system must accommodate. Second, it would be predicted that the processing of tonal and segmental contrasts would *still* be slower than native Chinese listeners, since the speed of use of information in the L2 would be expected to be L1 independent, based on the current results.

REFERENCES

Abramson, A. (1979). The noncategorical perception of tone categories in Thai. In S. O. B. Lindblom (Ed.), *Frontiers of speech communication* (pp. 127–134).

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language, 38*(4), 419-439.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition, 52*(3), 163-187.

Bates, D., Maechler, B., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1-48).

Beckman, M. (1986). *Stress and non-stress accents*. Dordrecht: Foris.

Best, C. T. (1995). A direct realistic perspective on cross-language speech perception. In W.Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 229-273). Timonium, MD: York Press.

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

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. i. (1999). Training Japanese listeners to identify English/r/and/l: Long-term retention of learning in perception and production. *Attention, Perception, & Psychophysics, 61*(5), 977-985.

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. i. (1997). Training Japanese listeners to identify English /r/and/l: IV. Some effects of perceptual learning on speech production. *J Acoust Soc Am, 101*(4), 2299-2310.

Broersma, M. (2002). *Comprehension of non-native speech: inaccurate phoneme processing and activation of lexical competitors.* Paper presented at the Interspeech.

Broersma, M. (2005). *Phonetic and lexical processing in a second language (MPI Series in Psycholinguistics No. 34).* Doctoral dissertation, Radboud University, Nijmegen, The Netherlands.

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System, 36*(1), 22-34. doi:10.1016/j.system.2007.11.003

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Q J Exp Psychol (Hove), 64*(1), 74-95. doi:10.1080/17470218.2010.499174

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

Chan, S. W., Chuang, C. K., & Wang, W. S. Y. (1975). Cross-linguistic study of categorical perception for lexical tone. *J Acoust Soc Am, 58*(1), 119.

Chang, Y. C., Halle, P., Best, C. T., & Abramson, A. (2008). *Do non-native language listeners perceive Mandarin tone continua categorically.* Paper presented at the 8th Phonetic Conference of China and the International Symposium on Phonetic Frontiers, Beijing.

Chao, Y. (1930). A system of tone letters. *Le Maitre Phonetique, 45*, 24-27.

Chen, H. C., & Cutler, A. (1997). Auditory priming in spoken and printed word recognition. *The cognitive processing of Chinese and related Asian languages*, 77-81.

Chen, X., Anderson, R. C., Li, W., Hao, M., Wu, X., & Shu, H. (2004). Phonological awareness of bilingual and monolingual Chinese children. *Journal of Educational Psychology, 96*(1), 142-151.

Chen, Y., & Xu, Y. (2006). Production of weak elements in speech–Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica, 63*(1), 47-75.

Connell, K., Hüls, S., Martínez-García, M. T., Qin, Z., Shin, S., Yan, Y., & Tremblay, A. (2016). *Examining the Role of English Stress in Lexical Access: Evidence from the Visual-World Paradigm.* Paper presented at the Acoustical Society of America, Honolulu, HI.

Connolly, J. F., & Phillips, N. A. (1994). Event-related potential components reflect phonological and semantic processing of the terminal word of spoken sentences. *Journal of Cognitive Neuroscience, 6*(3), 256-266.

Connolly, J. F., Stewart, S., & Phillips, N. (1990). The effects of processing requirements on neurophysiological responses to spoken sentences. *Brain and language, 39*(2), 302-318.

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*(3), 207-228.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*(3), 201-220.

Cutler, A., & Chen, H. C. (1995). *Phonological similarity effects in Cantonese word recognition*. Paper presented at the The Thirteenth International Congress of Phonetic Sciences.

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

Cutler, A., & Otake, T. (1999). Pitch accent in spoken-word recognition in Japanese. *J Acoust Soc Am, 105*(3), 1877-1888.

Cutler, A., & Otake, T. (2004). Pseudo-homophony in non-native listening. *J Acoust Soc Am, 115*(5), 2392-2392.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). *Dutch listeners' use of suprasegmental cues to English stress.* Paper presented at the 16th International Congress of Phonetic Sciences (ICPhS 2007).

Cutler, A., Weber, A., & Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics, 34*(2), 269-284.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cogn Psychol, 42*(4), 317-367. doi:10.1006/cogp.2001.0750

Dell, F., & Vergnaud, J.-R. (1984). Les développements récents en phonologie: quelques idées centrales. *Forme sonore du langage*, 1-42.

Desroches, A. S., Newman, R. L., & Joanisse, M. F. (2009). Investigating the time course of spoken word recognition: Electrophysiological evidence for the influences of phonological similarity. *Journal of Cognitive Neuroscience, 21*(10), 1893-1906.

Duanmu, S. (2007). *The phonology of standard Chinese*: Oxford University Press.

Dupoux, E., Pallier, C., Sebastian, N., & Mehler, J. (1997). A destressing “deafness” in French? *Journal of Memory and Language, 36*(3), 406-421.

Dupoux, E., Peperkamp, S., & Sebastián-Gallés, N. (2001). A robust method to study stress “deafness”. *J Acoust Soc Am, 110*(3), 1606-1618.

Dupoux, E., Sebastian-Galles, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress 'deafness': the case of French learners of Spanish. *Cognition, 106*(2), 682-706. doi:10.1016/j.cognition.2007.04.001

Dupoux, E., Sebastián-Gallés, N., Navarrete, E., & Peperkamp, S. (2008). Persistent stress ‘deafness’: The case of French learners of Spanish. *Cognition, 106*(2), 682-706.

Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research, 92*, 233-277.

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics, 36*(2), 268-294. doi:10.1016/j.wocn.2007.06.005

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*(2), 349.

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5*(4), 171-189.

Fujisaki, H., & Kawashima, T. (1968, August). *The influence of various factors on the identification and discrimination of synthetic speech sounds*. Paper presented at the 6th International Congress on Acoustics, Tokyo, Japan.

Fujisaki, H., & Kawashima, T. (1969). On the modes and mechanisms of speech perception. . *Annual Report of the Engineering Research Institute, 28*, 67-73.

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

Gandour, J., & Harshman, R. A. (1978). Crosslanguage differences in tone perception: A multidimensional scaling investigation. *Language and Speech, 21*(1), 1-33.

Gay, T. (1978). Physiological and acoustic correlates of perceived stress. *Language and Speech, 21*, 347-353.

Gottfried, T. L., & Suiter, T. L. (1997). Effect of linguistic experience on the identification of Mandarin Chinese vowels and tones. *Journal of Phonetics, 25*(2), 207-231.

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*(3), 395-421. doi:10.1016/s0095-4470(03)00016-0

Hallett, P. (1986). Eye movements. *Handbook of perception and human performance, 1*, 10-11.

Holt, L. L., & Lotto, A. J. (2006). Cue weighting in auditory categorization: Implications for first and second language acquisition a. *J Acoust Soc Am, 119*(5), 3059-3071.

Huang, K. (2012). *A study of neutral-tone syllables in Taiwan Mandarin* (unpublished doctoral dissertation) University of Hawaii at Manoa, Honolulu.

Huang, T., & Johnson, K. (2010). Language specificity in speech perception: Perception of Mandarin tones by native and nonnative listeners. *Phonetica, 67*(4), 243-267. doi:10.1159/000327392

Ingvalson, E. M., Holt, L. L., & McClelland, J. L. (2012). Can native Japanese listeners learn to differentiate/r–l/on the basis of F3 onset frequency? *Bilingualism: Language and Cognition, 15*(02), 255-274.

Jun, S.-A., & Fougeron, C. (2000). A phonological model of French intonation *Intonation* (pp. 209-242): Springer.

Jun, S.-A., & Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus, 14*(1), 147-172.

Kaan, E., Wayland, R., Bao, M., & Barkley, C. M. (2007). Effects of native language and training on lexical tone perception: An event-related potential study. *Brain Res, 1148*, 113-122.

Kukona, A., Fang, S.-Y., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition, 119*(1), 23-42.

Kuznetsova, A., Brockhoff, B., & Christensen, H. (2016). Tests in linear mixed effects models. Version 2.0.32.

Lai, Y., & Zhang, J. (2008). Mandarin Lexical Tone Recognition: The Gating Paradigm. *Kansas Working Papers in Linguistics, 30*, 183-194.

Lee, C.-Y. (2007). Does Horse Activate Mother? Processing Lexical Tone in Form Priming. *Language and Speech, 50*(1), 101-123.

Lee, C.-Y., Tao, L., & Bond, Z. (2008). Identification of acoustically modified Mandarin tones by native listeners. *Journal of Phonetics, 36*(4), 537-563.

Lemhofer, K., & Broersma, M. (2012). Introducing LexTALE: a quick and valid Lexical Test for Advanced Learners of English. *Behav Res Methods, 44*(2), 325-343. doi:10.3758/s13428-011-0146-0

Li, A. (2002). *Chinese prosody and prosodic labeling of spontaneous speech.* Paper presented at the Speech Prosody 2002, International Conference.

Li, Q., & Chen, Y. (2016). An acoustic study of contextual tonal variation in Tianjin Mandarin. *Journal of Phonetics, 54*, 123-150.

Li, Y. L. (2016). *Effects of high variability phontetic training on monosyllabic and disyllabic Mandarin Chinee tones for L2 Chinese learners.* University of Kansas.

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

Lin, C. Y., Wang, M., Idsardi, W. J., & Xu, Y. (2014). Stress processing in Mandarin and Korean second language learners of English. *Bilingualism: Language and Cognition, 17*(02), 316-346.

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi [Acoustic characteristics of neutral tone in Beijing Mandarin]. *Fangyan [Dialects], 3*, 166-178.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America, 35*, 1173-1181.

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. *J Acoust Soc Am, 94*(3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y. i., & Yamada, T. (1994). Training Japanese listeners to identify English /r/and/l/. III. Long-term retention of new phonetic categories. *J Acoust Soc Am, 96*(4), 2076-2087.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/and/l: A first report. *J Acoust Soc Am, 89*(2), 874-886.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cogn Sci, 31*(1), 133-156.

Malins, J. G., & Joanisse, M. F. (2010). The roles of tonal and segmental information in Mandarin spoken word recognition: An eyetracking study. *Journal of Memory and Language, 62*(4), 407-420. doi:10.1016/j.jml.2010.02.004

Malins, J. G., & Joanisse, M. F. (2012). Setting the tone: an ERP investigation of the influences of phonological similarity on spoken word recognition in Mandarin Chinese. *Neuropsychologia, 50*(8), 2032-2043. doi:10.1016/j.neuropsychologia.2012.05.002

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within-and between-language competition. *Bilingualism: Language and Cognition, 6*(02), 97-115.

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

Mirman, D. (2014). *Growth curve analysis and visualization using R*: CRC Press.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *J Mem Lang, 59*(4), 475-494. doi:10.1016/j.jml.2007.11.006

Norris, D., & McQueen, J. M. (2008). Shortlist B: a Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357-395. doi:10.1037/0033-295X.115.2.357

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences., 23*(3), 299-325.

Peirce, J. W. (2007). PsychoPy - Psychophysics software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. doi:10.3389

Peperkamp, S., & Dupoux, E. (2002). A typological study of stress 'deafness'. *Laboratory phonology, 7*, 203-240.

Pisoni, D. B. (1971). *On the Nature of Categorical Perception of Speech Sounds.* . Michigan University, Ann Arbor.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Percept Psychophys, 13*(2), 253-260. doi:10.3758/BF03214136

Qin, Z., Chien, Y.-F., & 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*(3), 318-338.

R Core Team. (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/.

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: lexical stress drives eye movements immediately. *The Quarterly Journal of Experimental Psychology, 63*(4), 772-783. doi:10.1080/17470210903104412

Ritsma, R. J., Cardozo, B. L., Domburg, G., & Neelen, J. J. M. (1965). The buildup of the pitch percept. *IPO Annual Progress Report*(1), 12-15.

Roberts, T. P. L., Flagg, E. J., & Gage, N. M. (2004). Vowel categorization induces departure of M100 latency from acoustic prediction. *NeuroReport, 15*(10), 1679-1682. doi:10.1097/01.wnr.0000134928.96937.10

Robinson, K., & Patterson, R. D. (1995). The stimulus duration required to identify vowels, their octave, and their pitch chroma. *Journal of the Acoustical Society of America, 98*(4), 1858-1865.

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

Schouten, M. E. H. (1975). *Native-language interference in the perception of second-language vowels: An investigation of certain aspects of the acquisition of a second language*: Utrecht.

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language, 52*(2), 240-255.

Sereno, J. A., & Lee, H. (2015). The contribution of segmental and tonal information in Mandarin spoken word processing. *Language and Speech, 58*(2), 131-151.

Shuai, L., & Malins, J. G. (2016a). Encoding lexical tones in jTRACE: a simulation of monosyllabic spoken word recognition in Mandarin Chinese. *Behav Res Methods*. doi:10.3758/s13428-015-0690-0

Shuai, L., & Malins, J. G. (2016b). *TRACE-T with and without tonal directions*. Paper presented at the Aspects of Language, Buffalo, NY.

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

Strauss, T. J., Harris, H. D., & Magnuson, J. S. (2007). jTRACE : A reimplementation and extension of the TRACE model of speech perception and spoken word recognition. *Behavior Research Methods, 39*, 19-30.

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

Sun, K. C., & Huang, T. (2012). A cross-linguistic study of Taiwanese tone perception by Taiwanese and English listeners. *Journal of East Asian Linguistics, 21*(3), 305-327. doi:10.1007/s10831-012-9092-9

Taft, M., & Chen, H. C. (1992). Judging homophony in Chinese: The influence of tones. *Advances in psychology, 90*, 151-172.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*(5217), 1632-1634.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1996). Using eye movements to study spoken language comprehension: Evidence for visually mediated incremental interpretation.

Tremblay, A. (2008). Is second language lexical access prosodically constrained? Processing of word stress by French Canadian second language learners of English. *Applied Psycholinguistics, 29*(04). doi:10.1017/s0142716408080247

Tremblay, A. (2009). Phonetic variability and the variable perception of L2 word stress by French Canadian listeners. *International Journal of Bilingualism, 13*(1), 35-62.

Tremblay, A., Broersma, M., Coughlin, C. E., & Choi, J. (2016). Effects of the Native Language on the Learning of Fundamental Frequency in Second-Language Speech Segmentation. *Frontiers in Psychology, 7*.

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, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *J Acoust Soc Am, 113*(2), 1033. doi:10.1121/1.1531176

Wang, Y., Spence, M. M., Jongman, A., & Sereno, J. A. (1999). Training American listeners to perceive Mandarin tones. *J Acoust Soc Am, 106*(6), 3649-3658.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*(1), 35-44. doi:10.3758/bf03207136

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *J Acoust Soc Am, 120*(2), 1063-1074.

Ye, Y., & Connine, C. M. (1999). Processing Spoken Chinese: The Role of Tone Information. *Language and Cognitive Processes, 14*(5-6), 609-630. doi:10.1080/016909699386202

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 Representational Deficits in SLA: In Honour of Roger Hawkins*, 69-103.

Zhang, J., & Meng, Y. (2016). Structure-dependent tone sandhi in real and nonce disyllables in Shanghai Wu. *Journal of Phonetics, 54*, 169-201.

Zhang, Y., & Francis, A. (2010). The weighting of vowel quality in native and non-native listeners' perception of English lexical stress. *Journal of Phonetics, 38*(2), 260-271.

Zhao, J., Guo, J., Zhou, F., & Shu, H. (2011). Time course of Chinese monosyllabic spoken word recognition: evidence from ERP analyses. *Neuropsychologia, 49*(7), 1761-1770. doi:10.1016/j.neuropsychologia.2011.02.054

Appendix A: Norming study materials

Table A.1. Norming Study Words

Condition	Set	Word	Pinyin	IPA	English Gloss	Word	Pinyin	IPA	English Gloss
Tonal	Tone 1 – Tone 3	疤	ba1	pa1	scar	靶	ba3	pa3	target/mark
		班	ban1	pan1	class/squad	板	ban3	pan3	board
		刀	dao1	tɑo1	knife	岛	dao3	tɑo3	island
		兜	dou1	tou1	pocket	抖	dou3	tou3	tremble
	Tone 1 – Tone 2	漆	qi1	t^{h}çi1	paint	棋	qi2	t^{h}çi2	Chess
		花	hua1	xua1	flower/blossom	滑	hua2	xua2	to slip
		敲	qiao1	t^{h}çiɑu1	knock	桥	qiao2	t^{h}çiɑu2	bridge
		秋	qiu1	t^{h}çiu1	autumn/fall	球	qiu2	t^{h}çiu2	ball
	Tone 4 – Tone 3	病	bing4	piŋ4	ailment/sickness	饼	bing3	piŋ3	round flat cake
		醉	zui4	tsuei4	intoxicated	嘴	zui3	tsuei3	mouth
		豹	bao4	pɑo4	leopard	宝	bao3	pɑo3	a jewel
		救	jiu4	tçiou4	to save	酒	jiu3	tçiou3	wine/liquor
	Tone 4 – Tone 2	菜	cai4	t^{h}sai4	vegetables	财	cai2	t^{h}sai2	money/wealth
		饭	fan4	fan4	food	烦	fan2	fan2	to feel vexed
		嚏	ti4	t^{h}i4	sneeze	啼	ti2	t^{h}i2	to cry
		递	di4	ti4	to hand over	笛	di2	ti2	flute
	Tone 4 – Tone 1	抱	bao4	pɑu1	to hold	包	bao1	pɑu3	to wrap
		悖	bei4	pan1	rebellious	碑	bei1	pan3	a monument
		渍	zi4	tɑo1	saturate/soak	资	zi1	tɑo3	capital
		烫	tang4	tou1	to scald	汤	tang1	tou3	soup

Segmental	Vowel	沙	sha1	ʃa1	granule	梳	shu1	ʃu1	comb
		闭	bi4	pi4	to close	爸	ba4	pa4	father
		河	he2	xə2	river	湖	hu2	xu2	lake
		土	tu3	tʰu3	earth/dust	塔	ta3	tʰa3	pagoda
	Allophonic	盗	dao4	tɑu4	steal	袋	dai4	tai4	a pouch/bag
		债	zhai4	tʂɑo4	debt	罩	zhao4	tʂai4	cover
		逃	tao2	tʰɑu2	to escape	抬	tai2	tʰai2	to lift
		柴	chai2	tʰʂai2	firewood	潮	chao2	tʰʂɑu2	tide/current
	Offglide	贴	tie1	tʰie1	to stick/to paste	挑	tiao1	tʰiau1	carry on a pole
		龟	gui1	kuei1	tortoise/turtle	锅	guo1	kuo1	pot
		敲	qiao1	tʰçiau1	knock	切	qie1	tʰçie1	to cut/to slice
		角	jia3	tçia3	angle	姐	jie3	tçie3	older sister
	Nasal Coda	抱	bao4	pau4	to hold	棒	bang4	paŋ4	a stick/club
		泡	pao4	pʰau4	puffed/swollen	胖	pang4	pʰaŋ4	fat/plump
		搡	sang3	saŋ3	push back	扫	sao3	sau3	to sweep
		逃	tao2	tʰɑu2	to escape	糖	tang2	taŋ2	sweets/candy

Appendix B: Norming study acoustic analyses

Figure B.1. F0 measurements of Tone 1 – Tone 2 tonal condition pairs

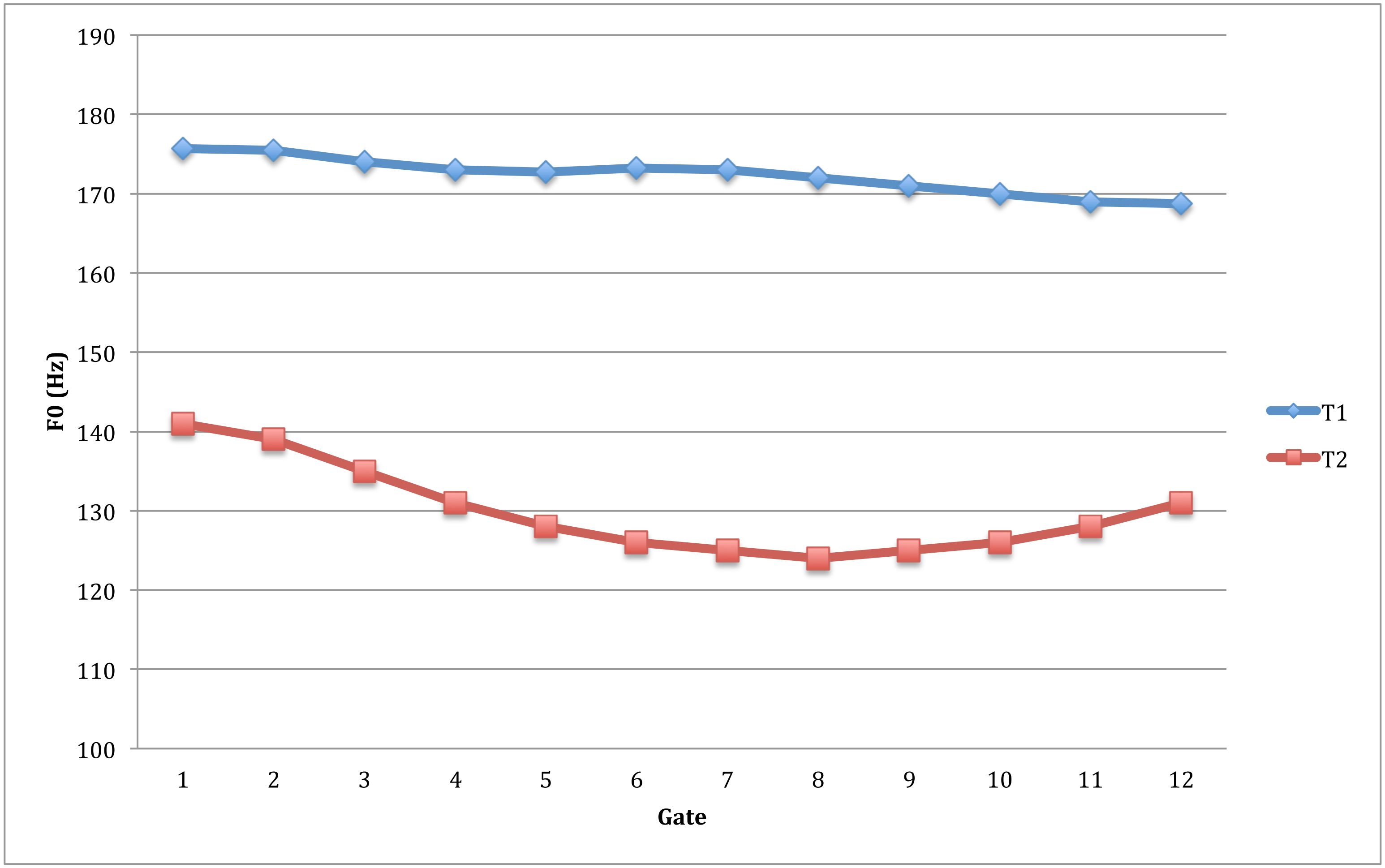

Figure B.2. F0 measurements of Tone 1 – Tone 3 tonal condition pairs

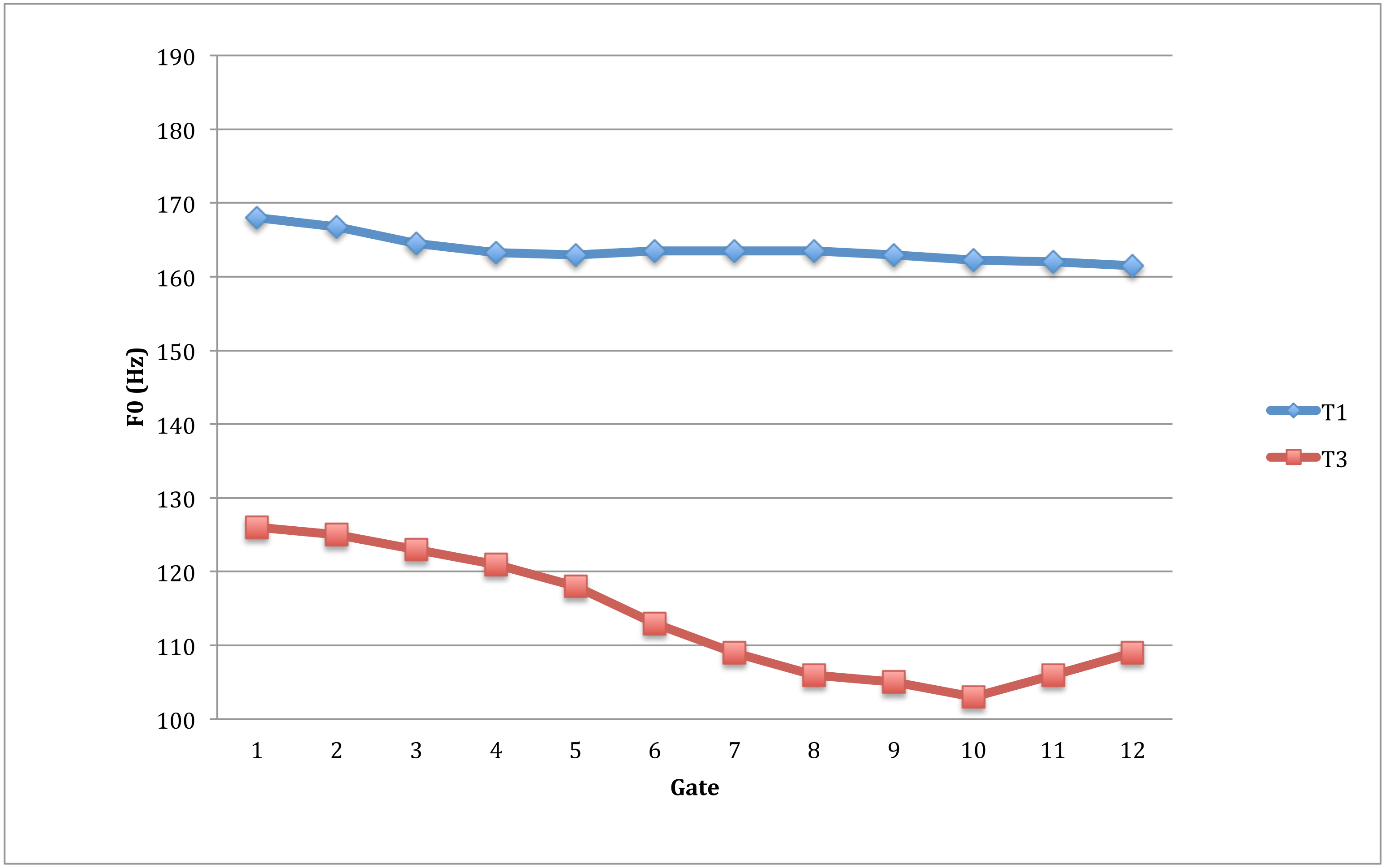

Figure B.5. F0 measurements of Tone 4 – Tone 2 tonal condition pairs

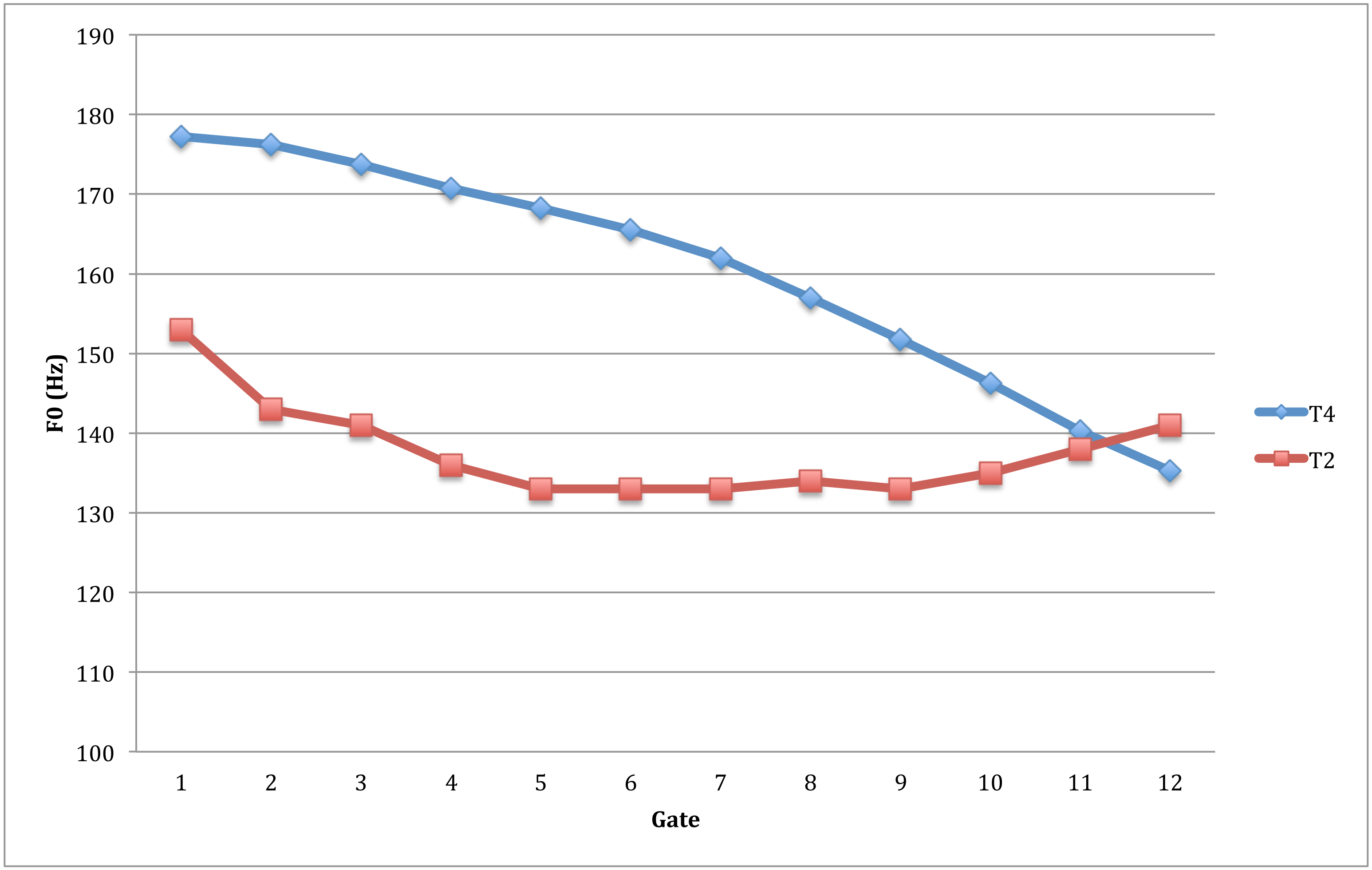

Figure B.6. F0 measurements of Tone 1 – Tone 4 tonal condition pairs

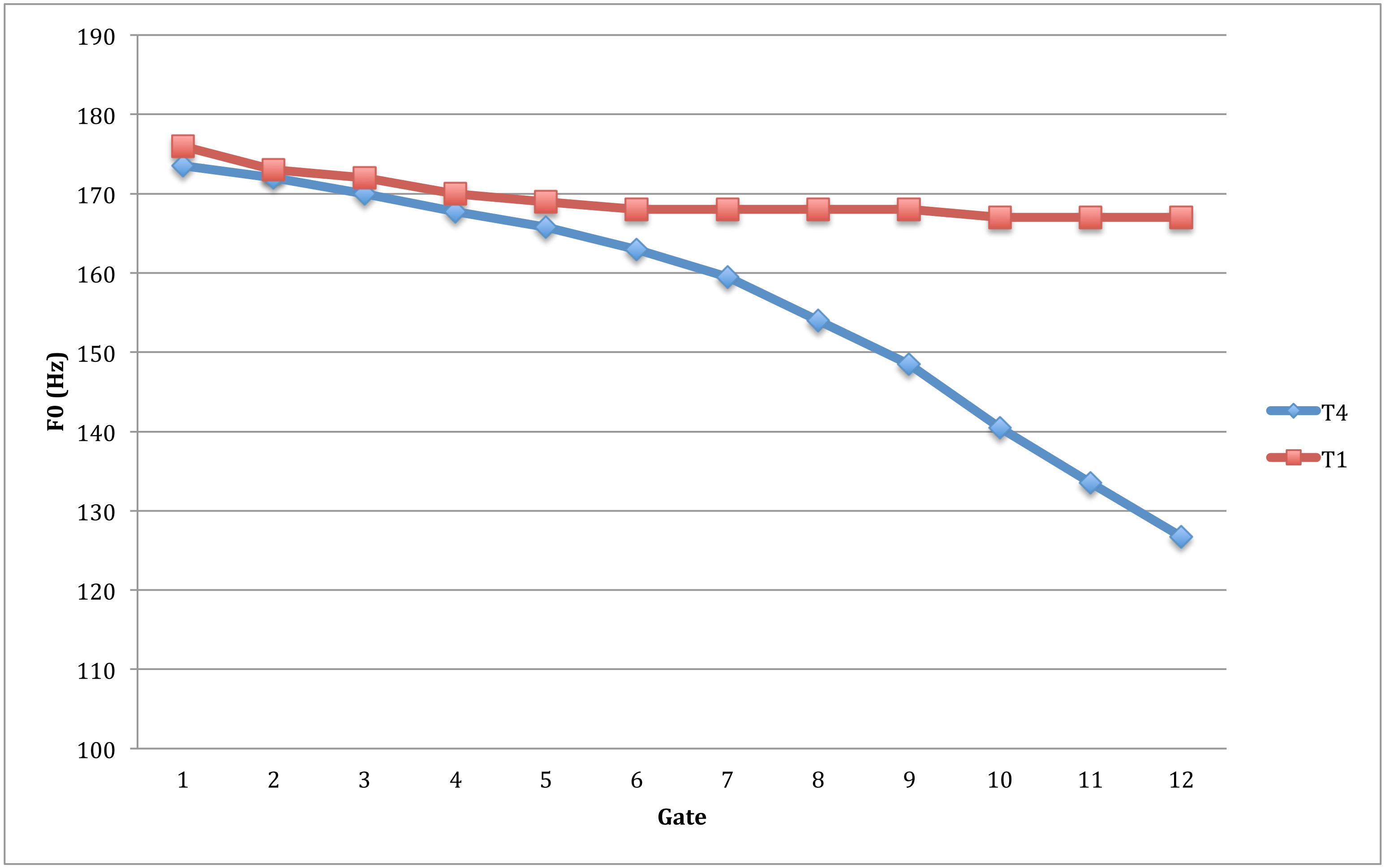

Figure B.7. F0 measurements of Tone 3 – Tone 4 tonal condition pairs

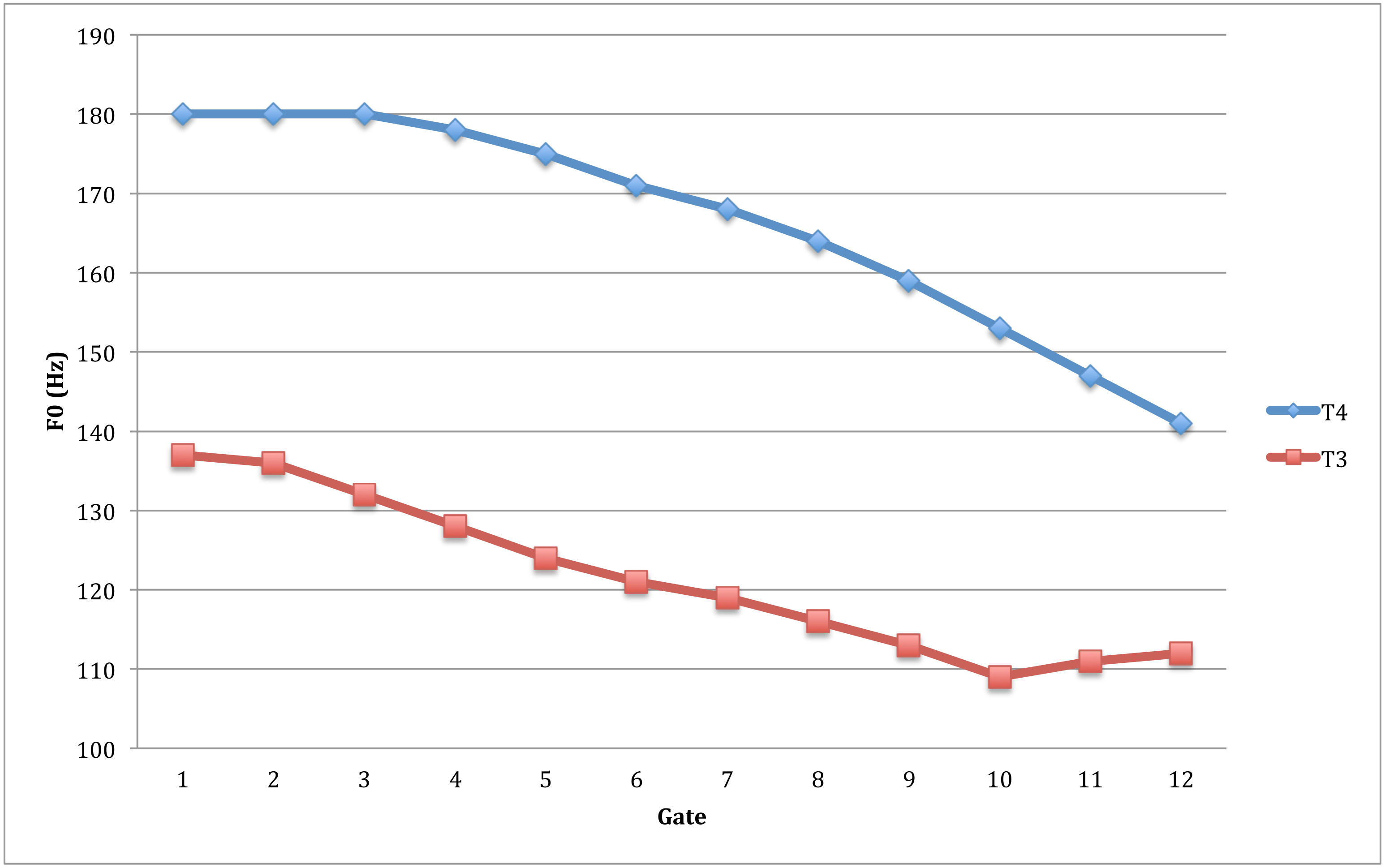

Table B.2. F1 measurements of the segmental items by Set

Gate	1	2	3	4	5	6	7	8	9	10	11	12
Vowel	1185	653	641	648	657	686	676	678	687	716	681	709
	1129	660	373	363	354	354	349	348	356	356	345	331
Allophonic	1276	671	717	734	746	744	737	731	721	712	711	701
	1209	704	714	718	728	714	709	705	704	689	673	653
Offglide	1061	369	387	431	481	529	565	594	596	590	583	574
	1081	314	351	381	352	342	344	349	360	373	385	405
Nasal Coda	1224	698	757	776	785	769	769	791	758	815	829	839
	1090	736	720	729	733	740	731	727	734	741	717	705

Table B.3. F2 measurements of the segmental items by Set

Gate	1	2	3	4	5	6	7	8	9	10	11	12
Vowel	1909	1387	1288	1291	1268	1279	1266	1255	1255	1232	1226	1235
	1800	1604	1350	1298	1288	1272	1242	1241	1208	1142	1127	1175
Allophonic	2059	1602	1579	1569	1597	1624	1658	1701	1750	1789	1830	1866
	1858	1301	1254	1206	1177	1144	1115	1087	1063	941	913	895
Offglide	2130	1931	1911	1876	1815	1695	1585	1513	1495	1513	1494	1455
	2181	2006	1896	2003	2014	2006	1994	1966	1946	1913	1862	1831
Nasal Coda	1881	1327	1315	1312	1309	1275	1267	1231	1210	1252	1250	1303
	1883	1327	1183	1148	1114	1105	1083	1091	1029	1050	1073	1019

Table B.4. F1 measurements for individual segmental pairs by gate

Item	Pinyin	IPA	1	2	3	4	5	6	7	8	9	10	11	12
1	sha1	ʃa1	1655	607	684	740	777	788	787	790	807	803	796	824
	shu1	ʃu1	1779	1444	385	386	389	389	386	386	392	403	372	347
2	bi4	pi4	1031	346	351	341	329	313	294	272	269	272	271	277
	ba4	pa4	819	708	732	744	753	763	763	755	753	758	753	754
3	he2	xə2	906	494	390	388	386	387	391	399	415	433	455	474
	hu2	xu2	613	375	385	384	377	390	379	389	404	381	366	368
4	tu3	tʰu3	1092	473	369	341	320	323	336	344	357	368	369	330
	ta3	tʰa3	1360	801	756	720	713	805	762	769	774	871	718	782
5	dao4	tɑu4	1103	687	718	720	725	723	716	709	699	685	700	690
	dai4	tai4	1353	654	729	749	768	772	763	751	747	749	763	765
6	zhai4	tʂai4	1564	579	625	679	726	751	763	768	774	766	755	742
	zhao4	tʂɑo4	1347	637	676	723	753	750	742	727	718	706	691	688
7	tao2	tʰɑu2	1169	696	721	704	708	689	665	667	677	647	590	559
	tai2	tʰai2	1270	755	761	748	747	727	700	690	658	649	653	648
8	chai2	tʰʂai2	917	694	752	760	741	725	721	714	705	684	674	650
	chao2	tʰʂɑu2	1217	794	740	724	726	695	711	715	722	719	711	675
9	tie1	tʰie1	1131	281	328	332	330	331	333	334	335	339	347	358
	tiao1	tʰiau1	1015	322	367	398	464	537	589	620	636	619	589	557
10	gui1	kuei1	717	473	373	379	378	369	362	354	347	342	339	336
	guo1	kuo1	561	416	438	540	415	362	358	360	364	368	374	394
11	qiao1	tʰçiau1	1137	322	386	445	506	561	607	641	627	608	608	609
	qie1	tʰçie1	1221	297	339	344	342	343	345	351	357	367	382	406
12	jia3	tʰçia3	1376	359	422	503	576	647	701	759	775	790	796	794
	jie3	tʰçie3	1411	260	299	309	321	333	339	352	384	418	437	462
13	bao4	pau4	722	707	722	737	740	729	715	707	700	688	680	692
	bang4	paŋ4	1216	727	761	768	765	774	782	783	780	752	770	773
14	pao4	pʰau4	1027	890	737	734	740	743	747	745	736	731	738	737
	pang4	pʰaŋ4	1147	656	750	768	771	759	762	615	579	760	767	769

15	sang3	saŋ3	1282	602	722	758	795	754	749	993	914	980	1050	1142
	sao3	sau3	1415	629	733	749	755	779	766	769	842	903	821	784
16	tao2	tau2	1194	716	686	697	698	707	697	688	657	643	630	607
	tang2	taŋ2	1250	805	794	808	808	790	783	774	758	768	727	670
17	ba1	pa1	985	746	767	782	791	791	792	787	792	781	792	788
	ba3	pa3	1321	726	768	768	779	781	784	797	846	1084	1005	1111
18	ban1	pan1	1045	691	751	766	776	790	795	789	771	742	601	430
	ban3	pan3	1206	739	787	813	809	837	845	807	781	823	792	731
19	dao3	tɑo3	1340	713	756	738	750	741	747	785	727	839	826	772
	dao1	tɑo1	1350	665	708	697	678	674	663	660	653	667	631	616
20	dou1	tou1	1148	522	518	506	480	458	444	430	417	403	395	386
	dou3	tou3	1050	614	624	606	585	561	552	506	484	449	417	401
21	qi2	t^hçi2	1383	300	296	289	279	274	272	273	275	277	275	276
	qi1	t^hçi1	1375	280	314	318	318	319	319	321	323	324	323	322
22	hua2	xua2	620	403	467	505	576	667	719	742	760	759	777	783
	hua1	xua1	602	414	469	497	542	622	682	734	756	773	784	800
23	qiao1	t^hçiao1	1610	340	396	449	509	564	624	665	671	636	616	613
	qiao2	t^hçiao2	1182	390	444	519	620	690	677	625	656	632	670	685
24	qiu2	t^hçiu2	1125	340	353	360	369	378	377	375	379	381	381	385
	qiu1	t^hçiu1	1475	264	342	354	357	360	363	361	360	359	357	353
25	bing3	piŋ3	1403	340	352	339	329	306	317	314	314	354	351	310
	bing4	piŋ4	722	347	364	360	357	351	345	338	335	333	320	299
26	zui3	tsuei3	1320	369	396	392	390	384	393	437	469	433	409	348
	zui4	tsuei4	1103	407	419	421	422	421	423	439	449	438	422	407
27	bao4	pɑo4	1193	679	703	729	729	728	715	703	700	680	662	646
	bao3	pɑo3	893	716	724	712	702	699	717	737	741	702	721	743
28	jiu3	t^hçiou3	1251	269	285	292	314	332	351	389	436	452	442	397
	jiu4	t^hçiou4	1380	281	335	334	331	329	327	324	325	331	345	366
29	cai4	t^hsai4	1103	679	759	804	813	800	781	796	796	799	796	794
	cai2	t^hsai2	941	748	766	741	719	713	700	687	673	659	633	588

30	fan4	fan4	1198	642	732	790	803	794	769	745	741	781	787	780
	fan2	fan2	1403	654	737	788	829	826	840	825	807	783	721	615
31	ti2	tʰi2	1405	296	311	304	298	295	292	293	290	288	286	283
	ti4	tʰi4	959	312	323	319	305	299	292	289	290	291	282	279
32	di2	ti2	1092	312	306	295	286	279	275	273	273	274	274	273
	di4	ti4	935	349	340	325	311	289	276	276	280	281	279	278
33	bao1	pɑo1	734	659	695	717	723	719	718	712	708	684	651	628
	bao4	pɑo4	1251	693	739	745	745	731	717	697	689	705	692	681
34	bei1	pei1	631	558	552	503	451	406	377	366	357	345	340	339
	bei4	pei4	1103	573	587	578	553	529	507	490	480	461	445	428
35	zi1	tsɨ1	1441	338	358	361	359	358	357	360	358	358	361	362
	zi4	tsɨ4	1339	328	354	352	342	338	332	333	333	338	338	338
36	tang1	tʰaŋ1	1156	653	697	701	740	752	761	764	760	742	745	369
	tang4	tʰaŋ4	1199	701	753	791	790	756	727	721	724	729	762	744

Table B.5. F2 Measurements for Individual Segmental Pairs by Gate

Item	Pinyin	IPA	1	2	3	4	5	6	7	8	9	10	11	12
1	sha1	ʃa1	2562	1547	1497	1478	1474	1474	1463	1452	1456	1444	1433	1432
	shu1	ʃu1	2281	2337	1472	1257	1216	1102	974	954	868	709	692	711
2	bi4	pi4	1868	2357	2464	2533	2580	2627	2653	2670	2644	2625	2594	2660
	ba4	pa4	1917	1221	1281	1291	1300	1314	1323	1319	1319	1305	1280	1273
3	he2	xə2	1358	1406	976	1009	992	981	979	974	913	888	927	935
	hu2	xu2	1249	667	720	717	716	717	713	720	727	721	721	740
4	tu3	t^hu3	1802	1054	745	684	638	642	629	619	594	511	502	588
	ta3	t^ha3	1799	1374	1396	1384	1305	1348	1300	1275	1333	1291	1264	1300
5	dao4	tɑu4	2202	1365	1278	1217	1217	1197	1177	1156	1143	1109	1043	988
	dai4	tai4	1922	1549	1534	1496	1518	1550	1569	1598	1625	1649	1659	1682
6	zhai4	tʂai4	2584	1669	1627	1596	1613	1609	1599	1622	1648	1685	1712	1743
	zhao4	tʂɑo4	1989	1530	1430	1356	1271	1261	1242	1202	1182	1155	1126	1099
7	tao2	t^hɑu2	1502	1110	1122	1087	1070	1009	958	947	938	742	697	721
	tai2	t^hai2	1800	1550	1568	1593	1641	1684	1888	1929	1917	1967	2031	2068
8	chai2	t^hʂai2	1929	1639	1585	1590	1616	1654	1576	1654	1811	1853	1919	1971
	chao2	t^hʂɑu2	1740	1199	1184	1163	1148	1107	1083	1043	990	758	787	771
9	tie1	t^hie1	2565	2495	2530	2515	2529	2509	2506	2477	2457	2406	2338	2291
	tiao1	t^hiau1	2466	2302	2306	2265	2161	1959	1736	1502	1333	1168	1028	936
10	gui1	kuei1	1407	856	857	903	996	1086	1294	1526	1879	2249	2387	2402
	guo1	kuo1	1413	895	655	627	648	665	668	674	675	677	682	712
11	qiao1	t^hçiau1	2320	2328	2257	2141	2003	1770	1502	1326	1159	1096	1083	1037
	qie1	t^hçie1	2391	2282	2005	2428	2423	2413	2396	2371	2353	2316	2240	2187
12	jia3	t^hçia3	2326	2236	2224	2195	2100	1966	1808	1698	1610	1537	1478	1445
	jie3	t^hçie3	2355	2352	2394	2441	2454	2436	2405	2340	2299	2253	2186	2133
13	bao4	pau4	1776	1063	1061	1067	1059	1023	1033	1068	1041	1014	995	952
	bang4	paŋ4	2033	1142	1212	1235	1250	1256	1260	1244	1225	1196	1190	1173
14	pao4	p^hau4	1807	1662	1198	1182	1170	1164	1133	1115	1104	1075	1063	993
	pang4	p^haŋ4	1548	1248	1261	1242	1216	1214	1218	1111	1059	1185	1177	1184

15	sang3	saŋ3	2168	1566	1475	1470	1484	1398	1372	1346	1354	1426	1475	1735
	sao3	sau3	2145	1415	1360	1258	1180	1189	1180	1243	1111	1314	1455	1325
16	tao2	tau2	1803	1168	1113	1085	1047	1045	986	939	860	796	777	804
	tang2	taŋ2	1775	1353	1310	1300	1285	1231	1218	1221	1203	1202	1158	1120
17	ba1	pa1	2083	1252	1290	1306	1319	1325	1326	1330	1322	1312	1330	1328
	ba3	pa3	2049	1230	1234	1251	1307	1291	1290	1237	1263	1235	1276	1376
18	ban1	pan1	1874	1251	1356	1384	1443	1488	1521	1591	1654	1672	1670	1659
	ban3	pan3	2183	1346	1360	1379	1379	1481	1485	1559	1544	1552	1604	1605
19	dao3	tɑo3	2233	1502	1340	1252	1137	1009	1014	991	861	990	1318	1267
	dao1	tɑo1	2360	1425	1267	1181	1137	1102	1052	1021	989	963	899	847
20	dou1	tou1	2154	1407	1091	876	803	797	793	793	791	780	763	733
	dou3	tou3	2169	1387	1122	977	882	821	770	701	669	640	610	626
21	qi2	t^{h}çi2	2404	2507	2527	2520	2531	2503	2450	2509	2504	2506	2484	2461
	qi1	t^{h}çi1	2469	2481	2454	2407	2430	2402	2427	2455	2574	2596	2574	2590
22	hua2	xua2	1923	644	721	796	858	956	1069	1118	1186	1239	1229	1224
	hua1	xua1	2123	671	775	803	826	871	997	1090	1138	1173	1197	1235
23	qiao1	t^{h}çiao1	2453	2258	2225	2135	2013	1734	1492	1319	1150	1070	1009	963
	qiao2	t^{h}çiao2	2376	2321	2164	1967	1590	1433	1285	1098	1076	972	988	822
24	qiu2	t^{h}çiu2	2352	2300	2225	2098	1909	1576	1336	1106	930	782	745	741
	qiu1	t^{h}çiu1	2640	2309	2284	2246	2191	2078	1869	1590	1366	1198	1015	908
25	bing3	piŋ3	2446	2404	2490	2503	2525	2540	2566	2510	2477	2382	2196	2023
	bing4	piŋ4	1753	2414	2469	2522	2553	2592	2622	2692	2810	2678	2038	1606
26	zui3	tsuei3	2433	1380	1228	1181	1248	1506	1718	1955	2124	2147	2221	2267
	zui4	tsuei4	2185	1424	1264	1247	1285	1385	1477	1629	1863	2064	2186	2246
27	bao4	pɑo4	2026	1080	1108	1137	1139	1145	1126	1101	1083	1071	1033	953
	bao3	pɑo3	1947	1095	1081	1062	1041	1008	988	1021	962	930	850	1286
28	jiu3	t^{h}çiou3	2326	2427	2446	2443	2391	2171	1833	1494	1197	993	829	759
	jiu4	t^{h}çiou4	2329	2416	2447	2448	2454	2463	2464	2413	2350	2232	2071	1945
29	cai4	t^{h}sai4	1760	1434	1416	1434	1445	1440	1436	1459	1487	1506	1557	1609
	jiu4	t^{h}çiou4	2329	2416	2447	2448	2454	2463	2464	2413	2350	2232	2071	1945

30	fan4	fan4	1920	1187	1190	1220	1283	1300	1282	1272	1277	1337	1377	1408
	fan2	fan2	2021	1287	1392	1423	1461	1478	1531	1547	1580	1607	1364	1052
31	ti2	tʰi2	2574	2623	2577	2564	2543	2585	2570	2561	2545	2555	2528	2527
	ti4	tʰi4	2292	2434	2544	2579	2621	2620	2631	2620	2609	2630	2603	2612
32	di2	ti2	2296	2515	2568	2598	2602	2599	2614	2624	2616	2612	2624	2608
	di4	ti4	2031	2416	2528	2617	2654	2712	2687	2688	2692	2697	2673	2679
33	bao1	pɑo1	1602	999	1104	1136	1133	1116	1100	1052	1008	937	878	853
	bao4	pɑo4	2109	1096	1141	1188	1190	1176	1158	1126	1082	1056	977	972
34	bei1	pei1	1928	1831	1945	2091	2202	2249	2299	2388	2434	2436	2440	2513
	bei4	pei4	2169	1869	1919	1982	2044	2108	2157	2191	2215	2256	2257	2278
35	zi1	tsɨ1	2253	1425	1323	1329	1315	1309	1309	1324	1299	1310	1306	1307
	zi4	tsɨ4	2173	1600	1432	1445	1416	1470	1439	1440	1431	1440	1420	1428
36	tang1	tʰaŋ1	1552	1209	1197	1181	1189	1184	1180	1175	1173	1162	1087	931
	tang4	tʰaŋ4	1520	1351	1303	1287	1270	1228	1208	1199	1183	1184	1189	1157

Table B.6. COG Measurements for Consonants of Segmental Condition

Item	Pinyin	IPA	COG Gate 1	COG last 10 ms
1	sha1	ʃa1	4932	2398
	shu1	ʃu1	5997	6381
2	bi4	pi4	556	502
	ba4	pa4	645	674
3	he2	xə2	2204	2604
	hu2	xu2	928	555
4	tu3	tʰu3	1967	1077
	ta3	tʰa3	2855	1726
5	dao4	tɑu4	516	509
	dai4	tai4	3101	2356
6	zhai4	tʂai4	5366	1848
	zhao4	tʂɑo4	5076	2764
7	tao2	tʰɑu2	3526	2655
	tai2	tʰai2	2996	2264
8	chai2	tʰʂai2	4987	2053
	chao2	tʰʂɑu2	4703	2042
9	tie1	tʰie1	4118	4532
	tiao1	tʰiau1	3911	3729
10	gui1	kuei1	800	723
	guo1	kuo1	601	565
11	qiao1	tʰçiau1	6404	3663
	qie1	tʰçie1	7189	3817
12	jia3	tʰçia3	6883	3693
	jie3	tʰçie3	6988	4628
13	bao4	pau4	766	746
	bang4	paŋ4	850	804
14	pao4	pʰau4	3000	2114
	pang4	pʰaŋ4	3128	2129
15	sang3	saŋ3	7409	6787
	sao3	sau3	7735	7225
16	tao2	tau2	3106	2412
	tang2	taŋ2	3152	1950

Section B.1. Follow-up analysis of stop vs. non-stop initial items

The following is an analysis conducted to determine whether the modification of the stop vs. non-stop initial items influenced the results of the norming task. Figures A.7 and A.8 below show the proportion different responses for each set with stop vs. non-stop initial trials indicated by the solid or dashed lines.

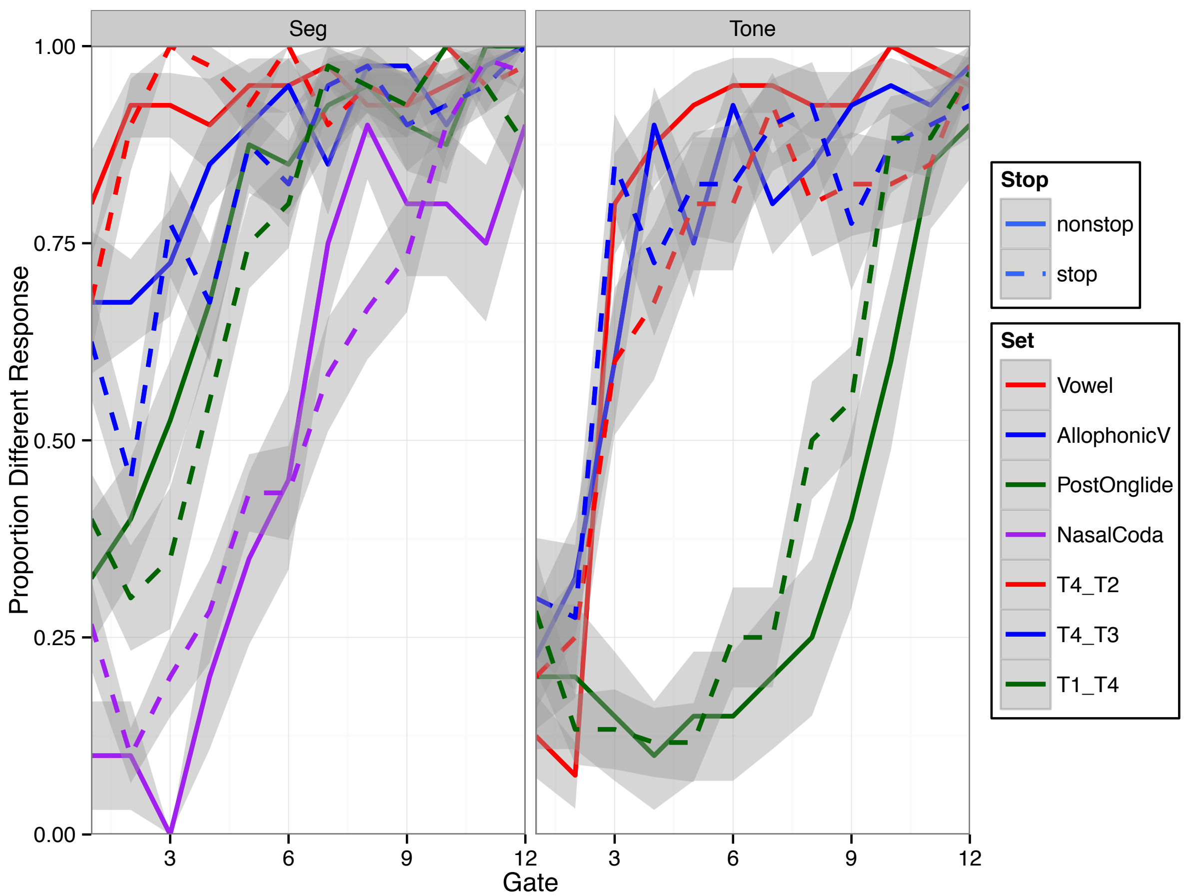

Figure A.8. Native Chinese listeners' different responses with the segmental condition in the panel on the left and the tonal condition in the panel on the right. Solid lines represent non-stop initial items and dashed lines represent stop initial items. Gate is presented on the *x*-axis and proportion different responses presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

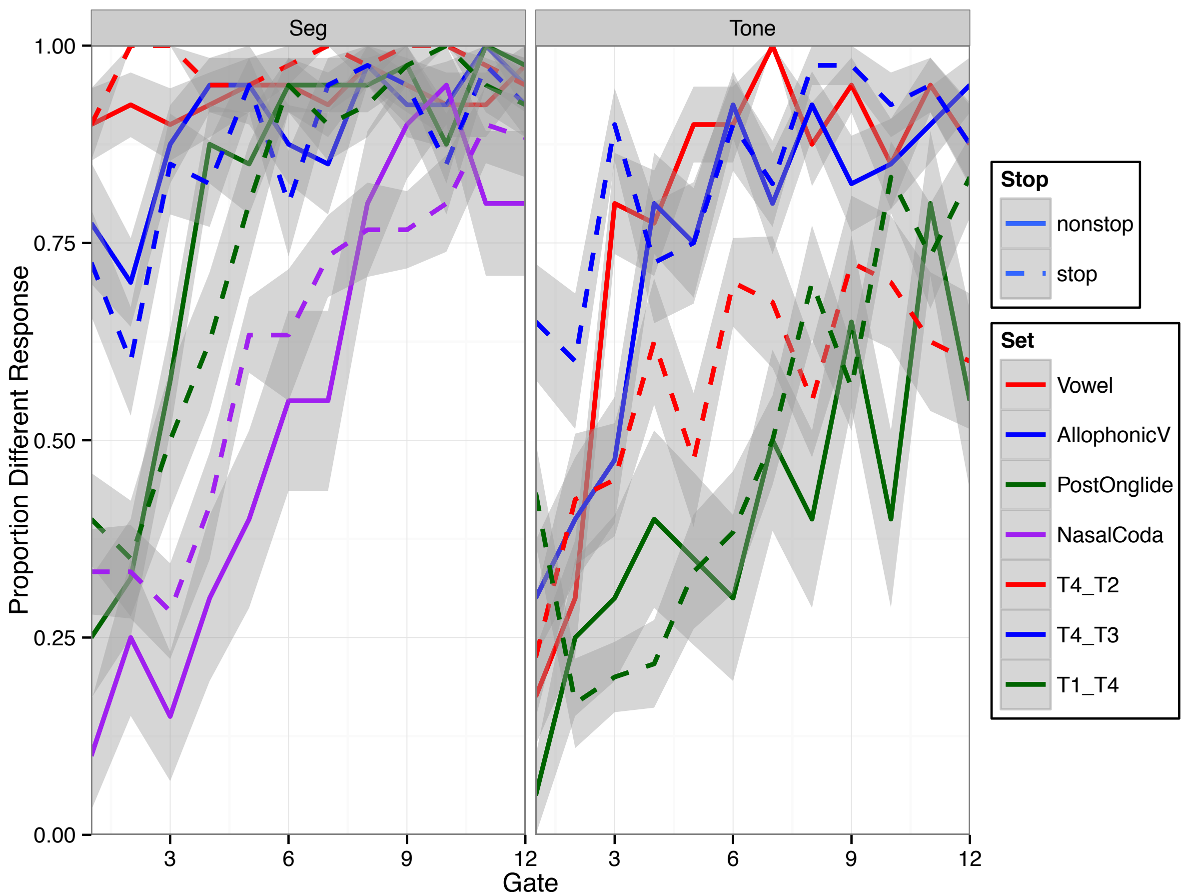

Figure A.9. Naïve English listeners' different responses with the segmental condition in the panel on the left and the tonal condition in the panel on the right. Solid lines represent non-stop initial items and dashed lines represent stop initial items. Gate is presented on the *x*-axis and proportion different responses presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

All items were coded as having a stop or non-stop (affricates and fricatives) initial. Then, for each set and each group separately, LME models compared the correct responses between the stop and non-stop initial items. There were two tonal sets for which this was not possible, since the items were either all stops or all non-stops in the set. Non-stops were set as the baseline.

Table B.7. Results of LME on native responses in the vowel set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.93	0.01	54.0	77.28	< .001
Stop	–0.002	0.02	459.0	–0.13	0.900

Table B.8. Results of LME on native responses in the allophonic vowel set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.87	0.02	39.2	40.86	< .001
Stop	–0.04	0.02	459.0	–1.85	.06

Table B.9. Results of LME on native responses in the post–onglide set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.78	0.02	43.4	29.20	< .001
Stop	–0.04	0.03	459.0	–1.28	.203

Table B.10. Results of LME on native responses in the nasal coda set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.05	0.03	34.0	13.18	< .001
Stop	0.04	0.03	459.0	0.96	.336

Results of the segmental sets reveal no difference between items with stop vs. non-stop initials for native listeners. This indicates that the modification of the initial stops did not influence the results of the segmental results for native listeners. The following set of tables presents the results for the tonal sets. The set of T1 – T2 is not included since all initials were stops and the set of T1 – T3 was not included since all initials were non-stops.

Table B.11. Results of LME on native responses in the set T4 – T2 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.80	0.03	23.9	25.45	< .001
Stop	–0.08	0.03	699.0	–2.59	.009

Table B.12. Results of LME on native responses in the set T4 – T3 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.76	0.03	23.5	23.37	< .001
Stop	–0.004	0.03	699.0	–0.14	.893

Table B.13. Results of LME on native responses in the set T1 – T4 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.35	0.04	40.8	10.0	< .001
Stop	0.08	0.04	459.0	2.0	.052

Results of the tonal sets revealed that the difference between stop and non-stop initials was only significant in the T4 – T2 set. This indicates that the modification of the initial stops influenced the results of the tonal results for native listeners, with the stop initial items having fewer different responses. Since this is only 1 of the five tonal sets included, this result does not undermine the results of the norming study, and overall it appears the effect of manipulating the consonant durations was minimal. The following set of tables presents the results for naïve English listeners.

Table B.14. Results of LME on English listener responses in the vowel set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.94	0.01	39.7	81.63	< .001
Stop	0.04	0.01	459.0	2.93	.004

Table B.15. Results of LME on English listener responses in the allophonic vowel set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.90	0.02	35.2	43.53	< .001
Stop	–0.03	0.02	459.0	–1.45	.142

Table B.9. Results of LME on English listener responses in the post-onglide set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.78	0.02	52.1	33.24	< .001
Stop	–0.02	0.03	459.0	–0.69	.493

Table B.10. Results of LME on English listener responses in the nasal coda set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.55	0.04	37.2	15.54	< .001
Stop	0.08	0.04	459.0	2.06	0.040

Results of the segmental sets reveal differences between items with stop vs. non-stop initials for English listeners in the vowel and nasal coda sets. This indicates that the modification of the initial stops not influence the results of the segmental results for English listeners in these two conditions, with stop initial items having more different responses than non-stop items. This is argued to not be problematic to the results since there were no differences found for the selected segmental set; the post-onglide set.

The following set of tables presents the results for the tonal sets. Again, the set of T1 – T2 is not included since all initials were stops and the set of T1 – T3 was not included since all initials were non-stops.

Table B.11. Results of LME on English listener responses in the set T4 – T2 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.78	0.03	26.4	29.26	< .001
Stop	–0.21	0.03	699.0	–6.90	< .001

Table B.12. Results of LME on English listener responses in the set T4 – T3 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.74	0.03	23.9	24.53	< .001
Stop	0.10	0.03	699.0	3.21	.001

Table B.13. Results of LME on English listener responses in the set T1 – T4 set

	Estimate	Std. Error	*df*	*t* value	P(>F)
Intercept	0.41	0.04	30.8	10.30	< .001
Stop	0.08	0.04	459.0	2.13	.034

Results of the tonal sets revealed that the difference between stop and non-stop initials was only significant in all three tone sets analyzed. This indicates that the modification of the initial stops influenced the results of the tonal results for native listeners, with the stop initial items having more different responses.

Appendix C: Experiments 1 and 2 words

Table C.1. Experiments 1 and 2 stimuli words and frequency

Condition	Trial Type	Word 1	Pinyin	IPA	Frequency	English Gloss	Word 2	Pinyin	IPA	Frequency	English Gloss
Tonal	Critical	班	ban1	pan1	3.35	class	板	ban3	pan3	3.08	board
		刀	dao1	tɑu2	3.44	knife	岛	dao3	tɑu3	3.47	island
		兜	dou1	tou1	2.37	pocket	抖	dou3	tou3	2.37	tremble
		夸	kua1	k^{h}ua1	2.39	to boast	垮	kua3	k^{h}ua3	2.58	collapse
		棋	qi2	t^{h}çi2	2.13	Chess	漆	qi1	t^{h}çi1	2.42	paint
		结	jie2	tçie2	3.21	knot	街	jie1	tçie1	3.51	street
		夺	duo2	tuo2	3.11	to rob	多	duo1	tuo1	4.72	many
		搏	bo2	pɔ2	2.35	fight	波	bo1	pɔ1	3.08	wave
		饼	bing3	piŋ3	3.19	round flat cake	病	bing4	biŋ4	3.58	sickness
		嘴	zui3	tsuei3	3.93	mouth	醉	zui4	tsuei4	3.40	intoxicated
		鼠	shu3	ʃu3	2.74	rat/mouse	树	shu4	ʃu4	3.33	tree
		水	shui3	ʃuei3	3.88	water	睡	shui4	ʃuei4	3.90	to sleep
		菜	cai4	t^{h}sai4	3.29	vegetables	财	cai2	t^{h}sai2	2.12	wealth
		饭	fan4	fan4	3.26	food	烦	fan2	fan2	3.26	to feel vexed
		闭	bi4	pi4	3.90	shut	鼻	bi2	pi2	2.59	nose
		递	di4	ti4	2.33	to hand over	笛	di2	ti2	2.07	flute

Note: Frequency listed in log10 words/million

Condition	Trial Type	Word 1	Pinyin	IPA	Frequency	English Gloss	Word 2	Pinyin	IPA	Frequency	English Gloss
Tonal	Filler	揪	jiu1	tçijou1	2.22	to clutch	九	jiu3	tçijou3	3.09	nine
		规	gui1	kuei1	1.98	a rule	轨	gui3	kuei3	1.95	rail
		疤	ba1	pa1	2.42	scar	靶	ba3	pa3	1.95	target
		杯	bei1	pei1	3.88	cup	北	bei3	pei3	3.00	north
		碟	die2	tie2	2.59	dish/plate	爹	die1	tie1	2.42	dad
		球	qiu2	tʰçiou2	3.85	ball/sphere	秋	qiu1	tʰçiou1	2.13	autumn/fall
		袭	xi2	ɕi2	2.39	to attack	膝	xi1	ɕi1	2.23	knee
		十	shi2	ʃɨ2	3.56	ten/10	尸	shi1	ʃɨ1	3.17	corpse
		拐	guai3	kuai3	2.40	to turn	怪	guai4	kuai4	3.67	bewildering/odd
		宝	bao3	pɑu3	2.89	a treasure	豹	bao4	pɑu4	2.25	leopard
		腿	tui3	tʰuei3	3.57	leg	退	tui4	tʰuei4	3.20	retreat
		酒	jiu3	tçijou3	3.80	wine/liquor	救	jiu4	tçijou4	3.91	to save/to rescue
		嚏	ti4	tʰi4	1.04	sneeze	啼	ti2	tʰi2	1.04	to crow/to hoot
		拜	bai4	pai4	3.07	worship	白	bai2	pai2	3.43	white
		帝	di4	ti4	2.37	emperor	笛	di2	ti2	2.41	enemy
		碎	sui4	suei4	3.07	four/4	碎	sui4	suei4	3.30	plain

Condition	Trial Type	Word 1	Pinyin	IPA	Frequency	Gloss	Word 2	Pinyin	IPA	Frequenc y	Gloss
Segmental	Critical	桥	qiao2	tʰçiɑu2	3.03	bridge	求	qiu2	tʰçiou2	4.04	to seek
		匣	xia2	ɕia2	1.30	box	鞋	xie2	ɕie2	3.37	shoe
		滑	hua2	xua2	3.09	slippery	回	hui2	xuei2	4.48	to go back
		踝	huai2	xuai2	1.59	ankle bone	活	huo2	xuo2	4.06	to live
		笑	xiao4	ɕiɑu4	3.82	smile	谢	xie4	ɕie4	3.46	to thank
		罪	zui4	tsuei4	3.44	crime	坐	zuo4	tsuo4	4.15	to sit
		队	dui4	tuei4	3.92	team	剁	duo4	tuo4	2.05	to cut
		挂	gua4	kua4	3.54	to hang	贵	gui4	kuei4	2.98	expensive
		推	tui1	tʰuei1	3.35	to push	脱	tuo1	tʰuo1	3.49	to take off
		胶	jiao1	tçiɑu1	2.54	glue	揭	jie1	tçie1	2.03	lift off
		消	xiao1	ɕiɑu1	2.50	subside	休	xiu1	ɕiou1	2.85	to rest
		龟	gui1	kuei1	2.18	turtle	锅	guo1	kuo1	2.54	pot/pan
		假	jia3	tçia3	3.52	fake	姐	jie3	tçie3	2.46	older sister
		鬼	gui3	kuei3	3.62	ghost	裹	guo3	kuo3	2.40	to wrap
		耍	shua3	ʃua3	3.26	to juggle	甩	shuai3	ʃuai3	3.08	to throw
		挑	tiao3	tʰiɑu3	3.23	incite	铁	tie3	tʰie3	2.97	iron Fe

Condition	Trial Type	Word 1	Pinyin	IPA	Frequency	Gloss	Word 2	Pinyin	IPA	Frequency	Gloss
Segmental	Filler	逃	tao2	t^{h}ɑu2	3.44	to escape	抬	tai2	t^{h}ai2	3.02	to lift
		潮	chao2	tʂ^{h}ɑu2	2.22	tide/current	柴	chai2	tʂ^{h}ai2	2.06	firewood
		河	he2	xɤ2	3.06	river	湖	hu2	xu2	2.92	lake
		茶	cha2	tʂ^{h}a2	3.10	tea	池	chi2	tʂ^{h}ɨ2	2.37	pool/pond
		撬	qiao4	t^{h}çiau4	2.41	to pry open	妾	qie4	t^{h}çie4	1.63	concubine
		叫	jiao4	tçiɑu4	4.54	to call/to yell	借	jie4	tçie4	3.50	to borrow
		盗	dao4	tɑu4	2.41	steal/robber	袋	dai4	tai4	3.04	a pouch
		闭	bi4	pi4	2.68	shut	爸	ba4	pa4	2.07	father
		叼	diao1	tiɑu1	1.98	to hold in the mouth	跌	die1	tie1	2.37	to fall
		交	jiao1	tçiɑu1	3.56	to deliver	接	jie1	tçie1	4.04	to answer
		瓜	gua1	kua1	2.48	melon	归	gui1	kuei1	3.14	to return
		花	hua1	xua1	3.95	flower	挥	hui1	xuei1	2.71	to wave
		毁	hui3	xuei3	3.58	to destroy	火	huo3	xuo3	3.55	fire
		土	tu3	t^{h}u3	2.89	earth/dust	塔	ta3	t^{h}a3	3.15	pagoda/tower
		窄	zhai3	tʂai3	2.03	narrow	沼	zhao3	tʂɑu3	1.95	swamp
		洒	sa3	sa3	2.58	to sprinkle	伞	san3	san3	2.36	umbrella

Appendix D: Experiments 1 and 2 stimuli acoustic analyses

Figure D.1. F0 measurements of Tone 1 – Tone 2 tonal condition pairs

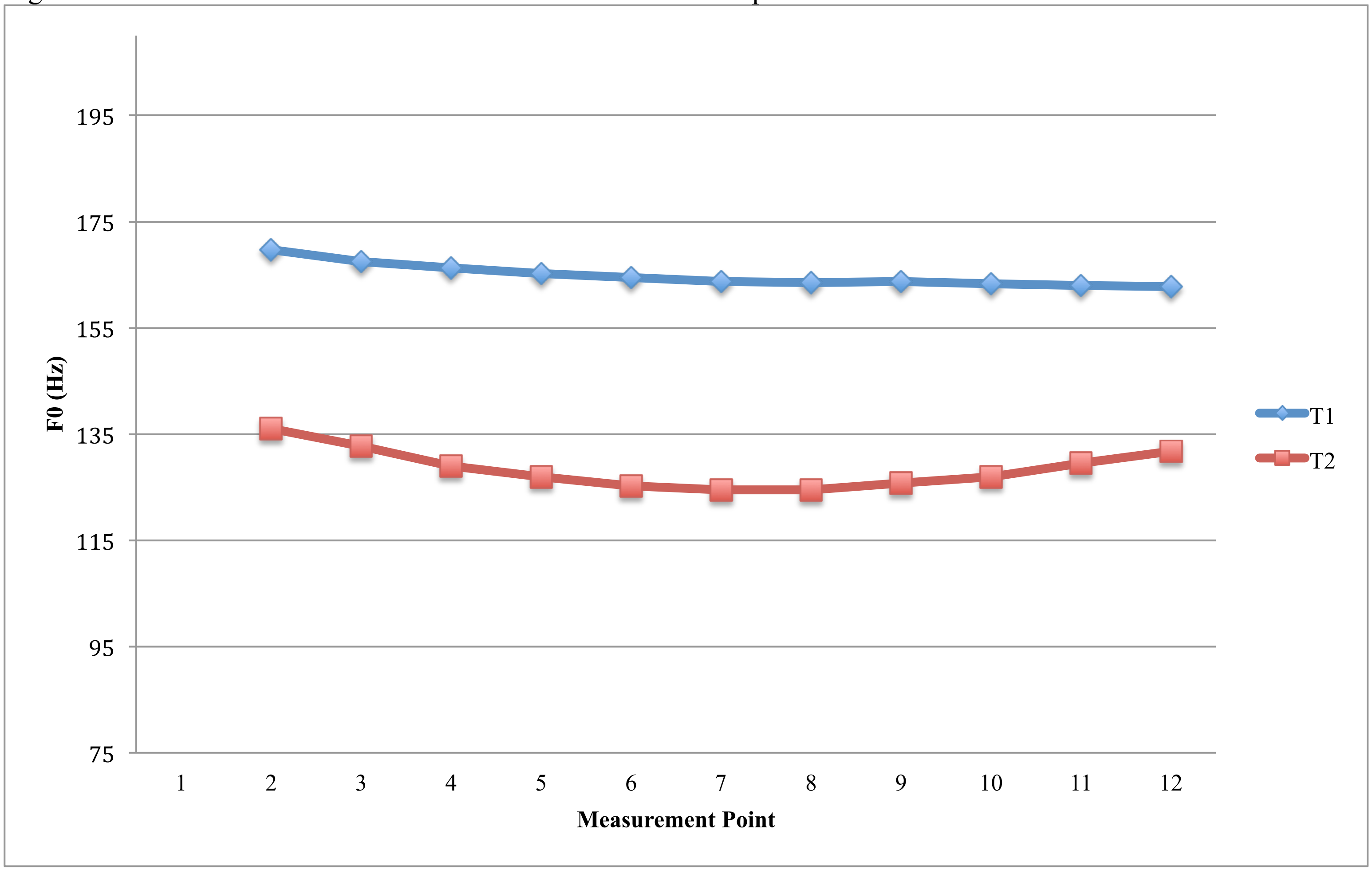

Figure D.2. F0 measurements of Tone 1 – Tone 3 tonal condition pairs

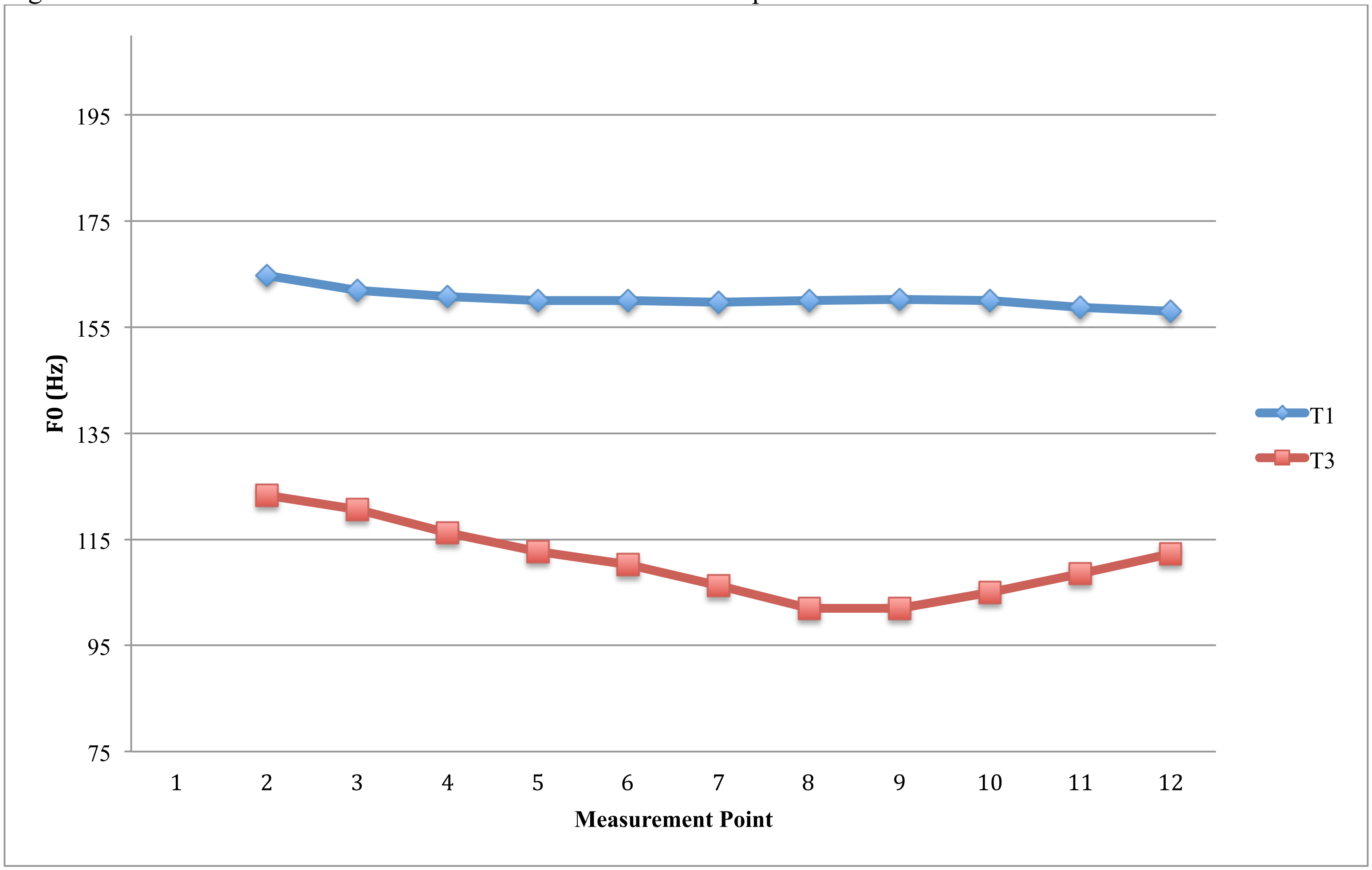

Figure D.3. F0 measurements of Tone 4 – Tone 3 tonal condition pairs

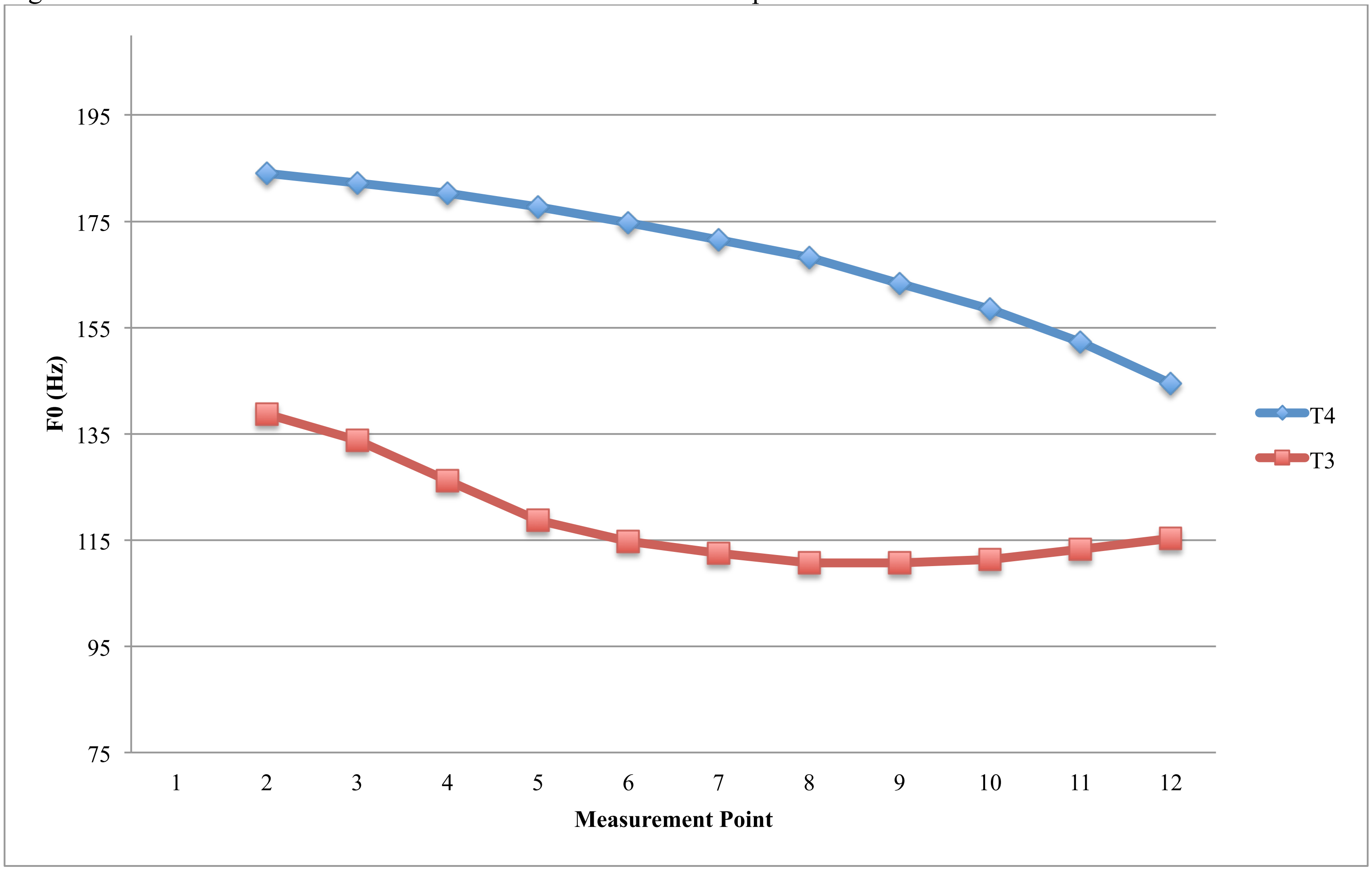

Figure D.4. F0 measurements of Tone 4 – Tone 2 tonal condition pairs

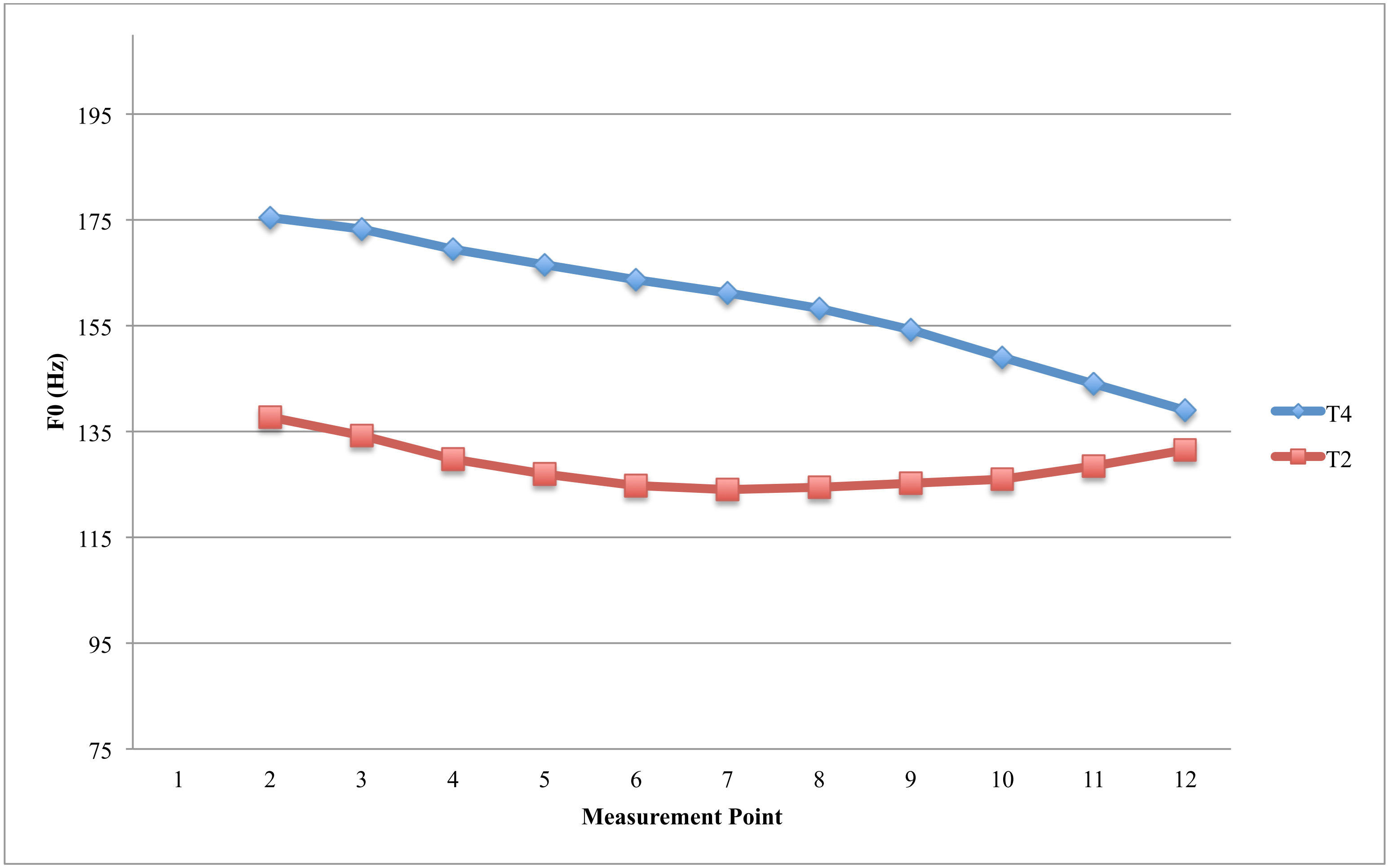

Table D.3. F1 measurements for individual segmental pairs by gate

Item	Pinyin	IPA	1	2	3	4	5	6	7	8	9	10	11	12
	hui2	xuei2	677	549	390	390	390	388	384	379	373	370	368	363
1	hua2	xua2	662	435	512	581	683	746	780	792	800	818	830	825
	dui4	tuei4	1446	379	383	385	383	387	393	403	427	433	424	410
2	duo4	tuo4	1049	371	382	428	477	423	392	386	363	363	376	384
	xiao1	ɕiɑu1	1213	256	320	327	332	336	339	342	345	347	346	341
3	xiu1	ɕijou1	1346	310	338	353	392	458	539	590	629	630	618	601
	shua3	ʃua3	1736	800	427	438	489	592	684	720	732	802	771	703
4	shuai3	ʃuai3	1718	742	449	482	524	598	671	742	835	839	833	849
	huai2	xuai2	559	281	378	372	378	369	372	373	383	397	415	443
5	huo2	xuo2	595	474	478	532	606	645	671	690	693	697	682	674
	gua4	kua4	917	263	372	383	388	390	399	403	402	404	404	395
6	gui4	kuei4	509	400	419	463	505	571	650	726	756	777	786	814
	gui1	kuei1	831	319	345	355	363	367	369	356	344	337	333	328
7	guo1	kuo1	740	331	349	368	364	355	365	395	406	418	419	428
	tiao3	t^{h}iɑu3	1158	300	325	322	325	338	350	356	389	395	422	441
8	tie3	t^{h}ie3	1418	365	416	479	611	689	754	763	775	710	662	679
	qiao2	t^{h}çiɑu2	1679	325	361	367	370	371	378	382	382	379	382	385
9	qiu2	t^{h}çiou2	1365	432	521	595	644	675	688	698	706	710	710	722
	xiao4	ɕiɑu4	1300	261	339	349	352	347	339	331	337	360	389	403
10	xie4	ɕie4	1337	768	311	316	314	315	329	371	438	504	611	690
	tui1	t^{h}uei1	1184	531	372	375	371	364	356	351	346	342	339	335
11	tuo1	t^{h}uo1	1026	423	375	382	389	405	402	396	398	410	415	429
	jia3	tçia3	1128	295	334	352	354	366	377	396	399	416	436	463
12	jie3	tçie3	1316	308	349	379	448	591	738	809	918	956	888	856
	xie2	ɕie2	1296	942	291	304	319	333	344	351	358	366	376	389
13	xia2	ɕia2	1222	272	372	435	537	654	721	752	759	756	772	784
14	zuo4	tsuo4	1346	716	396	391	394	390	387	391	400	411	405	389

	zui4	tsuei4	1321	327	395	397	406	430	484	481	426	405	400	394
	jiao1	tçiɑu1	1378	287	313	307	307	307	311	317	324	331	347	363
15	jie1	tçie1	1556	337	352	372	426	481	526	554	570	575	563	555
	gui3	kuei3	885	344	374	379	395	424	432	430	384	328	281	309
16	guo3	kuo3	1062	380	396	389	390	422	449	448	443	481	498	490

Table D.4. F2 measurements for individual segmental pairs by gate

Item	Word	IPA	1	2	3	4	5	6	7	8	9	10	11	12
1	hui2	xuei2	1752	1254	746	834	959	1221	1540	1775	2004	2290	2363	2384
	hua2	xua2	1887	1107	743	809	893	1021	1124	1197	1256	1289	1308	1313
2	dui4	tuei4	2722	1473	1383	1282	1277	1307	1400	1519	1707	1864	2005	2161
	duo4	tuo4	2077	1248	945	928	1203	887	737	674	632	607	618	642
3	xiao1	ɕiɑu1	2442	2411	2382	2376	2371	2250	1986	1733	1512	1292	1059	897
	xiu1	ɕijou1	2422	2280	2325	2274	2195	2083	1878	1671	1512	1372	1193	1058
4	shua3	ʃua3	2484	1918	1179	1008	944	1071	1237	1343	1460	1584	1692	1745
	shuai3	ʃuai3	2558	1680	1275	1102	969	885	918	1076	1272	1305	1371	1298
5	huai2	xuai2	1434	575	610	599	592	607	621	639	666	688	716	747
	huo2	xuo2	1632	1422	764	855	976	1215	1401	1487	1563	1675	1746	1796
6	gua4	kua4	1858	824	823	858	990	1143	1306	1496	1654	1827	1969	2092
	gui4	kuei4	1411	725	770	803	855	888	957	1024	1102	1138	1186	1223
7	gui1	kuei1	1559	883	827	889	1066	1263	1471	1713	1953	2229	2449	2481
	guo1	kuo1	1558	790	687	666	650	640	634	624	635	653	671	697
8	tiao3	t^{h}iɑu3	2489	2544	2724	2687	2683	2579	2474	2437	2367	2260	2215	2180
	tie3	t^{h}ie3	2700	2481	2338	2055	1783	1550	1389	1212	1123	1040	971	951
9	qiao2	t^{h}çiɑu2	2499	2350	2353	2105	1774	1452	1244	1012	903	854	774	732
	qiu2	t^{h}çiou2	2434	2160	1840	1603	1485	1382	1275	1208	1156	1092	1010	938
10	xiao4	ɕiɑu4	2175	2328	2595	2646	2541	2516	2544	2665	2705	2710	2688	2599
	xie4	ɕie4	2127	2629	2548	2552	2545	2576	2314	2407	2253	1930	1766	1623
11	tui1	t^{h}uei1	1683	1522	1005	982	1176	1420	1595	1815	2172	2404	2465	2452
	tuo1	t^{h}uo1	1718	1007	689	690	687	647	660	685	689	699	720	750
12	jia3	tçia3	2384	2330	2384	2457	2438	2420	2424	2430	2339	2318	2267	2187
	jie3	tçie3	2354	2533	2584	2550	2365	1998	1700	1539	1472	1485	1399	1338
13	xie2	ɕie2	2369	2715	2280	2276	2428	2395	2375	2367	2357	2311	2256	2213
	xia2	ɕia2	2244	2291	2390	2265	2062	1844	1691	1569	1490	1444	1442	1414
14	zuo4	tsuo4	2328	1850	1399	1231	1193	1214	1273	1401	1539	1728	1948	2224
	zui4	tsuei4	2451	1374	1187	973	827	729	586	551	665	703	696	692

15	jiao1	tçiau1	2190	2490	2502	2488	2516	2508	2498	2483	2432	2416	2384	2376
	jie1	tçie1	2683	2347	2321	2207	2028	1856	1627	1468	1322	1186	1091	1012
16	gui3	kuei3	1713	773	793	839	988	1241	1590	1933	2199	2306	2349	2372
	guo3	kuo3	1939	808	741	738	726	734	747	766	740	742	800	816

Table D.4. Consonant acoustics for individual segmental pairs

Item	Pinyin	IPA	COG	Variance	Skewness	Kurtosis
1	hui2	xuei2	375	434	7	56
	hua2	xua2	754	402	2	13
2	dui4	tuei4	464	546	4	22
	duo4	tuo4	354	225	10	178
3	xiao1	ɕiɑu1	543	1118	5	25
	xiu1	ɕijou1	1011	1473	3	9
4	shua3	ʃua3	2630	2737	1	–2
	shuai3	ʃuai3	1845	2295	1	0
5	huai2	xuai2	484	349	6	58
	huo2	xuo2	583	534	3	20
6	gua4	kua4	439	469	5	33
	gui4	kuei4	660	347	4	47
7	gui1	kuei1	406	554	5	24
	guo1	kuo1	382	218	8	154
8	tiao3	tʰiɑu3	913	1254	2	6
	tie3	tʰie3	647	820	4	21
9	qiao2	tʰçiɑu2	718	1363	4	13
	qiu2	tʰçiou2	1061	1551	3	8
10	xiao4	ɕiɑu4	756	1661	4	14
	xie4	ɕie4	1160	2261	3	5
11	tui1	tʰuei1	333	344	9	107
	tuo1	tʰuo1	405	261	6	77
12	jia3	tçia3	1430	2120	2	3
	jie3	tçie3	1457	2216	2	4
13	xie2	ɕie2	1248	2231	2	4
	xia2	ɕia2	1231	1567	3	11
14	zuo4	tsuo4	558	942	5	26
	zui4	tsuei4	451	767	7	57
15	jiao1	tçiɑu1	653	1357	4	17
	jie1	tçie1	645	908	4	23
16	gui3	kuei3	593	881	3	9
	guo3	kuo3	591	424	5	38

Appendix E: Language background survey

语言背景调查问卷 - Language Background Questionnaire

请回答以下关于您语言学习经历的问题。回答完毕请点击“提交”。

Please answer the following questions regarding your language learning experiences.
Click "Submit" at the bottom when you are done.

1.

2. **您的年龄？**
How old are you?

3. **性别**
Sex:
Mark only one oval.
- 女 Female
- 男 Male

4. **您的专业/研究领域？**
What is your major/ field of study?

5. **您的年级？**
What university year are you?

6. **您的身份？**
What university level?
Mark only one oval.
- 本科生 Undergraduate
- 研究生 Graduate
- 教授/讲师 Professor/Lecturer
- 其他 Other

7. **您本人的母语是什么？**

What is YOUR native language?

8. **您母亲的母语是什么？**

What is your MOTHER's native language?

9. **您父亲的母语是什么？**

What is your FATHER's native language?

10. **您的身份？ (请勾选所有符合您的选项)**

Have you ever had
Check all that apply.

- [ ] 视觉问题 Vision problems
- [ ] 听力损伤 Hearing impairment
- [ ] 语言失能 Language disability
- [ ] 学习障碍 Learning disability

11. **如果您符合上述任何一项，请详述之。 例如，您的视觉问题可能是“佩戴眼镜”**

If you checked any of the boxes above, please provide details. For example, "Glasses" if you checked vision problems

12. **从您出生直至五周岁，您在家里使用何种语言？**

What languages were used in your home between birth and 5 years old?

13. **13 从六周岁直至十一周岁，您在家使用何种语言？**

What languages were used in your home between 6 years old and 11 years old?

14. **岁直至十七周岁，您在家使用何种语言？**

What languages were used in your home between 12 years old and 17 years old?

15. **您在幼年时期，在哪些国家生活过？**

In what country/countries did you live during childhood?

16. **您在少年时期，在哪些国家生活过？**
    In what country/countries did you live as a teenager?

17. **您在成年后，在哪些国家生活过？**
    In what country/countries did you live as an adult?

18. **除去专门的外语类课程，您在小学期间，课堂（如数学、历史等）里使用何种教学语言？**
    EXCLUDING language classes, in what language were your classes (e.g., math, history) taught in elementary school?

19. **除去专门的外语类课程，您在初中期间，课堂（如数学、历史等）里使用何种教学语言？**
    EXCLUDING language classes, in what language were your classes (e.g., math, history) taught in middle school?

20. **除去专门的外语类课程，您在高中期间，课堂（如数学、历史等）里使用何种教学语言？**
    EXCLUDING language classes, in what language were your classes (e.g., math, history) taught in high school?

21. **请将您所能使用的语言按使用频率降序排列. 点击回车键换行以使每行仅输入一种语言**
    Please list all languages you know in order of DOMINANCE. Hit Enter to have each language on its own line

22. **请将您学会的语言按学习时间先后顺序排列（从您的母语开始） 点击回车键换行以使每行仅输入一种语言**

Please list all languages you know in order of ACQUISITION (starting with native language). Hit Enter to have each language on its own line

....

....

....

....

....

23. **请选择您使用英语的总体熟练程度**

Please give your global proficiency of English
Mark only one oval.

- ◯ 母语使用者 Native speaker
- ◯ 接近于母语使用者 Near-native
- ◯ 高级 Advanced
- ◯ 中等 Intermediate
- ◯ 初学 Beginner

24. **请选择您使用汉语普通话的总体熟练程度**

Please give your global proficiency of Mandarin Chinese
Mark only one oval.

- ◯ 母语使用者 Native speaker
- ◯ 接近于母语使用者 Near-native
- ◯ 高级 Advanced
- ◯ 中等 Intermediate
- ◯ 初学 Beginner

25. **请列出您在日常中使用各种语言的比率（各种语言的比率之和应为100%） 请列出每种语言并使用回车键换行以使每行仅输入一种语言**

Please give the percentage of time you currently use each language (your percentages should add to 100%). Please name each language and hit Enter to have each language on its own line

....

....

....

....

....

26. **如果同一段文字被译成您所前举的各种语言，您会选择分别花多少比率的时间来阅读每种语言的文字？（各种语言的比率之和应为100%). 请列出每种语言并使用回车键换行以使每行仅输入一种语言**

    If a text were available in all your languages, what percentage of the time would you choose to read it in each language? (your percentages should add to 100%). Please name each language and hit Enter to have each language on its own line

27. **27 当和另外一个对前举各种语言与您有同等熟练程度的人交谈时，您会选择分别花多少比率的时间来用每种语言交谈？（各种语言的比率之和应为100%） 请列出每种语言并使用回车键换行以使每行仅输入一种语言**

    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? (your percentages should add to 100%). Please hit Enter to have each language on its own line

Second language learners of Chinese only

Native Chinese and monolingual English speakers do not need to continue

28. **How many years of Chinese instruction have you received?**

29. **Were the majority of your Chinese instructors native speakers of Chinese?**
    Mark only one oval.

    - Yes
    - No

30. **At what age did you begin learning Chinese at school?**

31. **At what age did you begin listening to Chinese?**
    This may or may not be the same age as when you began Chinese classes

32. **At what age did you begin interacting with native Chinese speakers?**

Experiences in Chinese speaking environments

Please provide information about any time you have spent in a Chinese speaking environment. If you have not spent time in a Chinese speaking country, you may skip this section.

33. **Study abroad #1: Country you visited**

34. **Study abroad #1: Your age at time of visit**

35. **Study abroad #1: Length, in months, of your visit**

36. **Study abroad #1: Context of your visit**
    E.g. Study abroad, family vacation, etc.

37. **Study abroad #2: Country you visited**

38. **Study abroad #2: Your age at time of visit**

39. **Study abroad #2: Length, in months, of your visit**

40. **Study abroad #2: Context of your visit**
    E.g. Study abroad, family vacation, etc.

41. **Study abroad #3: Country you visited**

42. **Study abroad #3: Your age at time of visit**

43. **Study abroad #3: Length, in months, of your visit**

44. **Study abroad #3: Context of your visit**
E.g. Study abroad, family vacation, etc.

45. **Study abroad #4: Country you visited**

46. **Study abroad #4: Your age at time of visit**

47. **Study abroad #4: Length, in months, of your visit**

48. **Study abroad #4: Context of your visit**
E.g. Study abroad, family vacation, etc.

49. **How would you rate your proficiency for READING in Chinese?**
Mark only one oval.

- Beginner
- Intermediate
- Advanced
- Near-Native

41. **Study abroad #3: Country you visited**

42. **Study abroad #3: Your age at time of visit**

43. **Study abroad #3: Length, in months, of your visit**

44. **Study abroad #3: Context of your visit**
    E.g. Study abroad, family vacation, etc.

45. **Study abroad #4: Country you visited**

46. **Study abroad #4: Your age at time of visit**

47. **Study abroad #4: Length, in months, of your visit**

48. **Study abroad #4: Context of your visit**
    E.g. Study abroad, family vacation, etc.

49. **How would you rate your proficiency for READING in Chinese?**
    Mark only one oval.

    - Beginner
    - Intermediate
    - Advanced
    - Near-Native

50. **How would you rate your proficiency for WRITING in Chinese?**
Mark only one oval.

- Beginner
- Intermediate
- Advanced
- Near-Native

51. **How would you rate your proficiency for LISTENING in Chinese?**
Mark only one oval.

- Beginner
- Intermediate
- Advanced
- Near-Native

52. **How would you rate your proficiency for SPEAKING in Chinese?**
Mark only one oval.

- Beginner
- Intermediate
- Advanced
- Near-Native

Please provide information about the activities in which you use Chinese, and the frequency you do so

E.g., "Talking with Friends" (Sometimes)
"Watching movies" (Frequently)

53. **Activity 1**

54. **Activity 1 frequency**
Mark only one oval.

- Rarely
- Sometimes
- Often
- Daily

55. **Activity 2**

56. **Activity 2 frequency**
Mark only one oval.

- Rarely
- Sometimes
- Often
- Daily

57. **Activity 3**

58. **Activity 3 frequency**
Mark only one oval.

- Rarely
- Sometimes
- Often
- Daily

59. **Activity 4**

60. **Activity 4 frequency**
Mark only one oval.

- Rarely
- Sometimes
- Often
- Daily

61. **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)
Mark only one oval.

1	2	3	4	5	6	7	8	9	10

Powered by Google Forms

Appendix F: Proficiency measure materials

Table F.2. Lexical decision proficiency test materials

Difficult		Easy		Nonce	
Word	Log W/Million	Word	Log W/Million	Word	Log W/Million
省长	0.70	石头	2.86	果春	n/a
浑浊	1.00	唱歌	2.86	往过	n/a
奸商	1.04	箱子	2.80	查听	n/a
惊恐	1.97	孤独	2.94	美秀	n/a
校庆	0.70	好奇	2.95	电巫	n/a
师弟	0.30	相处	3.01	血纸	n/a
问诊	0.90	效果	2.98	火反	n/a
晨练	0.90	蓝色	2.91	采术	n/a
宝剑	1.57	阳光	2.87	平决	n/a
书柜	1.04	教育	2.89	适海	n/a
品茶	0.78	变化	2.99	保棍	n/a
念经	0.95	环境	2.96	自宝	n/a
姑父	0.85	大约	2.98	兴重	n/a
豪放	1.00	前面	3.17	方圈	n/a
学历	1.76	后悔	3.01	电牌	n/a
小鹿	1.00	旅馆	2.93	极面	n/a
猛增	0.95	小鸟	2.32	女击	n/a
老年	1.91	神奇	2.94	映镜	n/a
卷尺	0.95	许多	3.11	批会	n/a
近景	0.95	聊天	2.91	天该	n/a
采煤	0.60	干净	3.18	幻手	n/a
宿敌	1.04	巨大	2.98	师行	n/a
国语	0.48	健康	3.13	舰船	n/a
票务	0.90	年级	2.88	结开	n/a
药厂	0.48	魔鬼	2.79	宝定	n/a
上山	1.71	女生	2.89	掉校	n/a
知音	1.04	暑假	2.25	座质	n/a
划定	1.00	太阳	2.89	下态	n/a
水车	0.78	蛋糕	2.92	亏警	n/a
新型	1.97	新鲜	2.95	圆完	n/a
改写	1.84	空气	2.97	服包	n/a
孤苦	1.04	好看	2.98	错娘	n/a
吟诗	1.04	政治	2.85	数事	n/a
谅解	1.96	爷爷	2.74	衣动	n/a

重负	1.00	姐妹	2.87	害鬼	n/a
校花	0.95	惩罚	2.96	引光	n/a
远近	0.85	学会	2.95	怪后	n/a
电器	1.90	护士	2.84	随共	n/a
雨林	1.45	放心	2.93	欢敢	n/a

Survey Materials F.2. Cloze Test

1. 有一____人在路上遇到（meet）一个神仙（fairy），这个神仙以前是他____朋友。他告神仙，现在他的情况越____越不如从____，生活很困____。神仙听完____的话，用手一____路旁的一块小石头，那块石头立刻变____了金子，神仙把这块金子____了他。这个____得到金子，还不满意。神____又用手一指，把一块大石头____变成了金子，又给了____。这个人____是不满意。神仙____他："怎么样你____满意呢？"这个人回____说："我想……我要你的____。"

2. 有一个老人和他的儿子赶（drive）着一头驴（donkey）____ 集市（market）上去卖，没走____远，遇到一群人。其____一个姑娘说："看！____两个人真傻（stupid），有驴不骑（ride），倒要走路。"老人____到这些话，就让儿子骑到驴背上。过了一会儿，遇到一位____大爷，老大爷说："这个年轻人，____老人太不尊敬（respect），老人走路，却骑驴！" 于是，父亲____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____真狠心（merciless），自己骑驴，却让这个可怜____孩子跟在驴后面走。"老人只好也_____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："____个人骑一头驴，它能____得了吗？"这下老人可为难（puzzled）____，他只好把驴腿捆起来（tie up），____儿子一起抬（carry）着驴。驴可____干了，他们过一座的时候，这头驴挣脱（get rid of）了绳子，父子俩____这头驴都掉到河____去了。

Appendix G: Additional eye-tracking analyses and graphs

Figure G.1 Native Chinese listeners' proportion target, competitor and distracter fixations over 1,500 ms time window (including 200 ms baseline) with the tonal condition target, competitor and distracters in black and the segmental condition target, competitor and distracters in red. Time in ms is presented on the *x*-axis and proportion fixations to target presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

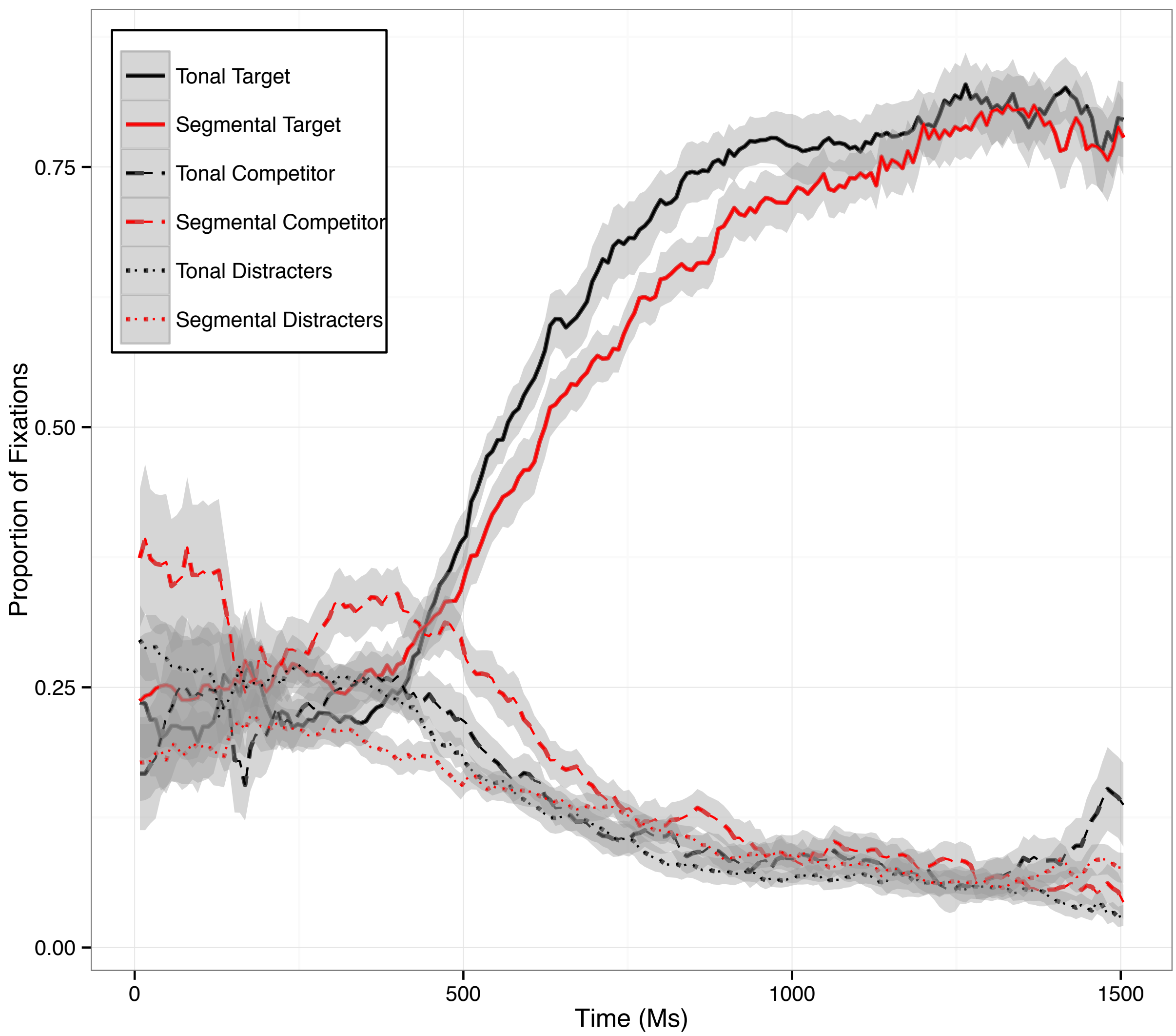

Figure G.2 Second language Chinese listeners' proportion target, competitor and distracter fixations over 1,500 ms time window (including 200 ms baseline) with the tonal condition target, competitor and distracters in black and the segmental condition target, competitor and distracters in red. Time in ms is presented on the *x*-axis and proportion fixations to target presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean

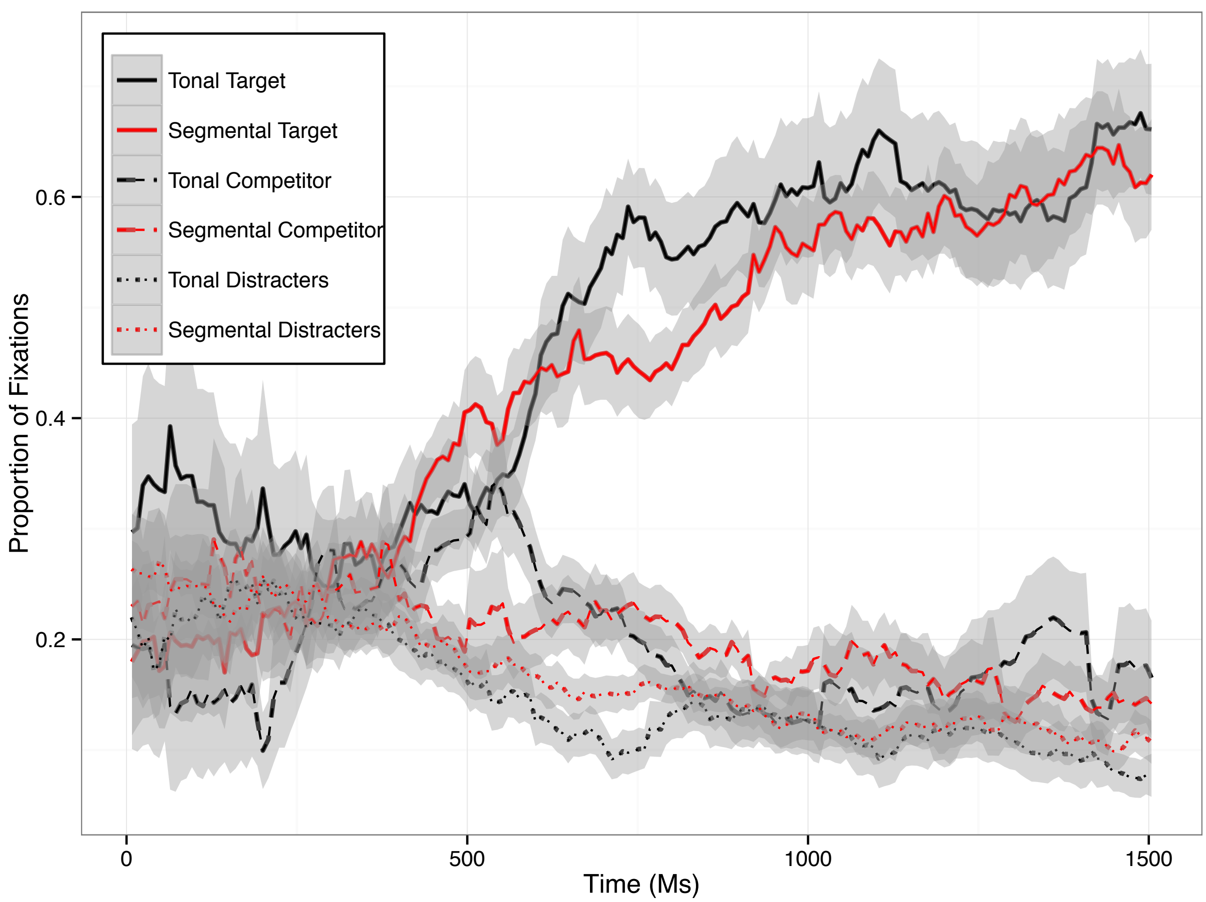

Figure G.3. Native Chinese listeners' proportion target, competitor and distracter fixations excluding two outliers with high initial segmental competitor fixations with the tonal condition target, competitor and distracters in black and the segmental condition target, competitor and distracters in red. Time in ms is presented on the *x*-axis and proportion fixations to target presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

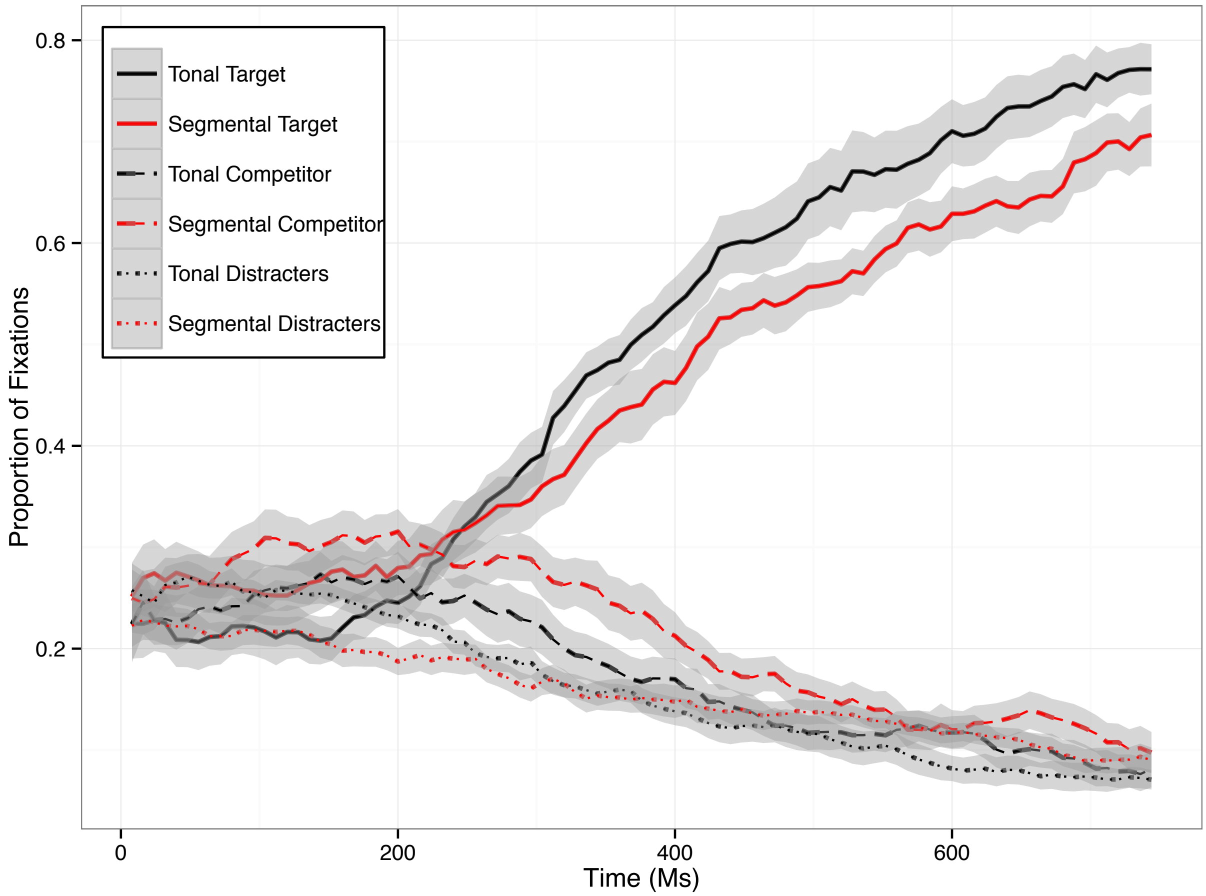

Table G.1. Results of GCA on native language Chinese listeners' proportion fixations to target excluding two outliers with high initial segmental competitor fixations

	Estimate	Std. Error	*t* value	P(>F)
Intercept	0.45	0.01	30.20	<.001
Linear	1.44	0.10	14.99	<.001
Quadratic	0.14	0.06	2.20	0.038
Cubic	–0.28	0.05	–5.97	<.001
Condition	0.03	0.00	7.60	<.001
Linear : Condition	0.53	0.04	13.20	<.001
Quadratic : Condition	–0.18	0.04	–4.52	<.001

Figure G.4. Native Chinese listeners' predicted proportion target, competitor and distracter fixations excluding two outliers with high initial segmental competitor fixations with the tonal condition target, competitor and distracters in black and the segmental condition target, competitor and distracters in red. Time in ms is presented on the *x*-axis and proportion fixations to target presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.

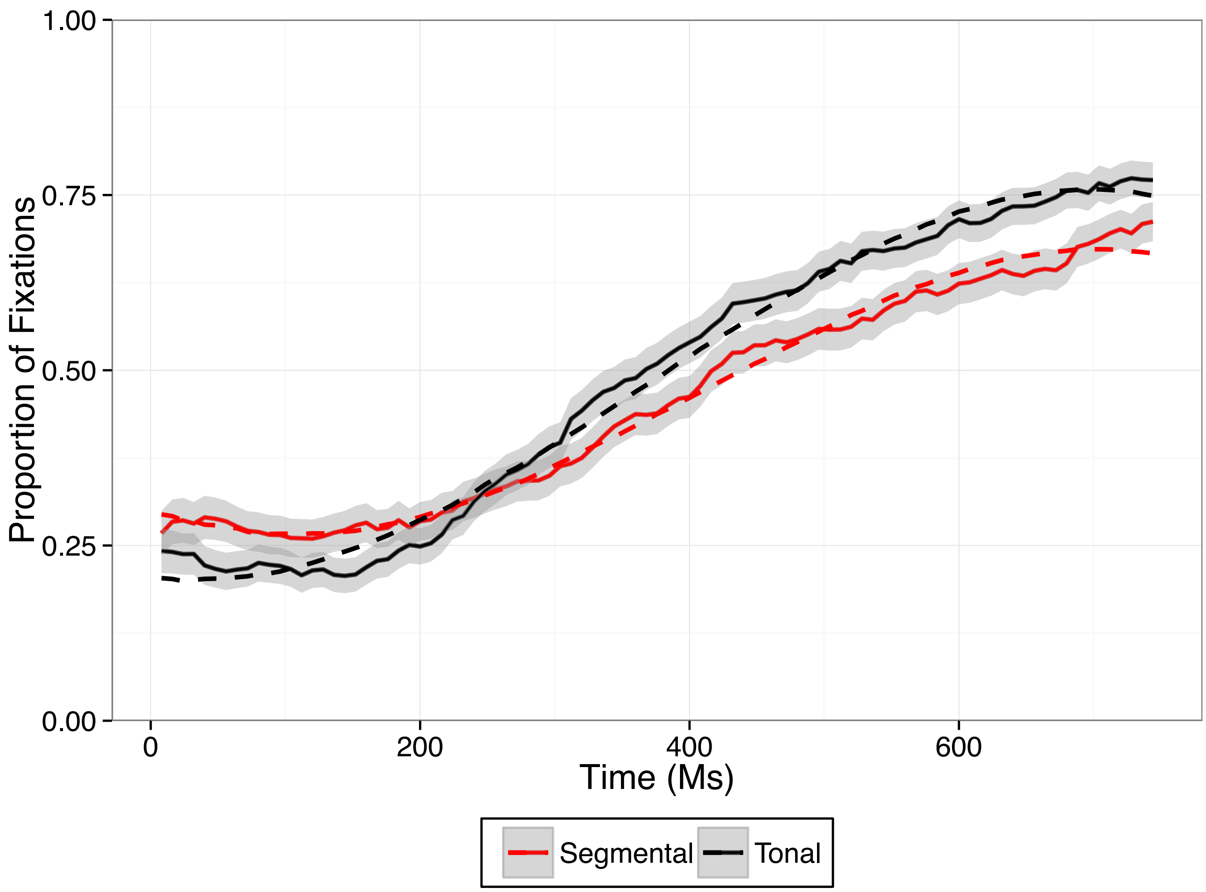

Figure G.5 Native Chinese listeners' proportion target, competitor and distracter fixations over 1,500 ms time window (including 200 ms baseline) excluding two outliers with high initial segmental competitor fixations with the tonal condition target, competitor and distracters in black and the segmental condition target, competitor and distracters in red. Time in ms is presented on the *x*-axis and proportion fixations to target presented on the *y*-axis. The shaded regions represent ±1 standard error of the mean.