https://doi.org/10.1016/j.wocn.2019.01.002

Influence of Within-Category Tonal Information in the Recognition of Mandarin-Chinese Words by Native and Non-Native Listeners: An Eye-Tracking Study

Zhen Qin[a,b*], Annie Tremblay[c], & Jie Zhang[c]

*Corresponding author: zhen-quentin.qin@polyu.edu.hk

[a] Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
, Hung Hom, Hong Kong
[b] Research Center for Language, Cognition, and Neuroscience, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
[c] Department of Linguistics, University of Kansas, Lawrence, 66045 Kansas, USA

Abstract

This study investigates how within-category tonal information influences native and non-native Mandarin listeners' spoken word recognition. Previous eye-tracking research has shown that the within-category phonetic details of consonants and vowels constrain lexical activation. However, given the highly dynamic and variable nature of lexical tones, it is unclear whether the within-category phonetic details of lexical tones would similarly modulate lexical activation. Native Mandarin listeners and proficient adult English-speaking Mandarin learners were tested in a visual-world eye-tracking experiment. The target word contained a level tone and the competitor word contained a high-rising tone, or vice versa. The auditory stimuli were manipulated such that the target tone was either canonical (Standard condition), phonetically more distant from the competitor (Distant condition), or phonetically closer to the competitor (Close condition). Growth curve analyses on fixations suggest that, compared to the Standard condition, Mandarin listeners' target-over-competitor word activation was enhanced in the Distant condition and inhibited in the Close condition, whereas English listeners' target-over-competitor word activation was inhibited in both the Distant and Close conditions. These results suggest that within-category tonal information influences both native and non-native Mandarin listeners' word recognition, but does so differently for the two groups.

Keywords: Mandarin tones; spoken word recognition; within-category tonal information; eye tracking.

1.0 Introduction

Research on spoken word recognition that uses continuous measures of lexical activation such as the visual-world eye-tracking paradigm has shown that within-category phonetic details in consonants and vowels modulate lexical access (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; McMurray, Clayards, Tanenhaus, & Aslin, 2008; McMurray, Tanenhaus, & Aslin, 2002, 2009; Salverda, Dahan, & McQueen, 2003). For example, McMurray et al. (2002) investigated whether native English listeners would show sensitivity to within-category variability in voice onset time (VOT) during the course of spoken word recognition. English listeners heard minimal pairs that began with /b/ or /p/ (e.g., *bear* vs. *pear*) in a nine-step VOT continuum. Listeners' eye fixations showed gradient effects of VOT, such that proportions of fixations to competitors increased linearly as the VOT approached the category boundary. Importantly, a gradient effect of VOT was found even if the analyses did not include the VOT steps that were at the end points of the continuum or near the category boundary. These results suggest that the within-category information available in the speech signal affects lexical activation in a gradient way (i.e., its effect is not limited to differences between exemplars that are near vs. far from the category boundary). These findings have important implications for models of spoken word recognition: They suggest that within-category phonetic information must be stored as part of listeners' lexical representations in order to modulate lexical activation (e.g., McMurray et al., 2002). Exemplar-based models can account for these findings by stipulating that listeners store the fine-grained phonetic details of the individual words they hear in memory, and in turn incorporate these details into the sound categorization process (e.g., Goldinger, 1991, 1996, 1998; Goldinger et al., 1999; Johnson, 1997, 2007).

These findings, however, also raise the question of whether within-category tonal information would similarly modulate lexical activation in languages that have lexical tones. In Mandarin Chinese (henceforth Mandarin), tonal information is extremely important for distinguishing words, as segmentally identical words that contain different lexical tones differ in their meaning. Mandarin has four lexical tones: Tone (T) 1 (e.g., /pā/ 'eight'), T2 (e.g., /pá/ 'to pull'), T3 (e.g., /pǎ/ 'to hold'), and T4 (e.g., /pà/ 'father') (Chao, 1968; Li & Thompson, 1989). The four tones have different tone shapes: T1 is flat (or level), T2 has a rising pitch contour, T3 has a dipping then rising pitch contour, and T4 has a falling pitch contour. Lexical tones are thus highly dynamic, requiring Mandarin listeners to evaluate the pitch they hear in the signal against the pitch range of the talker and continuously update this evaluation as more of the pitch contour is heard over time (e.g., Moore & Jongman, 1997). Importantly, a given Mandarin talker will show variability in the production of tonal categories, thus also requiring listeners to evaluate the pitch they hear against the talker's different realizations of the same tonal categories and of different tonal categories (e.g., Wang, Spence, Jongman, & Sereno, 1999; Wang, Jongman, & Sereno, 2003). Thus, although within-category tonal information may be very informative to the listener for distinguishing among competing words, the dynamic and variable nature of lexical tones may also make it difficult for listeners to use this information in spoken word recognition.

Existing research on the perception of lexical tones suggests that native Mandarin listeners perceive Mandarin tones categorically, with listeners relying primarily on pitch contour differences to distinguish the tones (Burnham & Mattock, 2007; Hallé, Chang, & Best, 2004; Sun & Huang, 2012; Wang, 1976). For instance, using identification and discrimination tasks, Wang (1976) found that Mandarin listeners showed the typical pattern of categorical perception when perceiving a dynamic tonal continuum varying from a level tone (T1) to a rising tone (T2);

that is, listeners showed steep slopes at the category boundary in their categorization function and marked peaks at the category boundary in their discrimination function. Later studies replicated Wang's (1976) findings using both speech and non-speech tones, with Mandarin listeners showing categorical perception of tonal contour differences and little sensitivity to within-category pitch variations (e.g., Hallé et al., 2004; Peng et al., 2010; Sun & Huang, 2012; Xu, Gandour, & Francis, 2006). In other words, Mandarin listeners' perception of lexical tones was not found to be proportional to the size of the acoustic changes heard in the speech signal. These results were interpreted as suggesting that Mandarin listeners established a clear linguistic boundary between level and rising tones, and discarded the phonetic details within each of the tonal categories. This research, however, leaves open the possibility that within-category tonal information would nonetheless modulate Mandarin listeners' lexical access in continuous measures of lexical activation such as the visual-world eye-tracking paradigm (e.g., Dahan et al., 2001; McMurray et al., 2002, 2009; Salverda et al., 2003). Although recent eye-tracking studies have examined Mandarin listeners' use of tonal information in spoken word recognition (e.g., Malins & Joanisse, 2010; Shen, Deutsch, & Rayner, 2013; Wiener & Ito, 2015), these studies shed little light on the influence of within-category tonal information as the speech signal unfolds.

The present study uses the visual-world eye-tracking paradigm to investigate whether within-category tonal information modulates native Mandarin listeners' lexical activation in spoken word recognition. If the within-category phonetic details of tones constrain lexical access, Mandarin listeners should show *more* competition from tonal competitors when the pitch of the target word (heard in the signal) is acoustically *closer* to that of the tonal competitor word (in listeners' lexical representation) than when it is standard (i.e., prototypical), and they should show *less* competition when the pitch of the target word is acoustically *more distant* from that of the tonal competitor word than when it is standard. Investigating whether within-category tonal information modulates Mandarin listeners' lexical access will have important implications for models of spoken word recognition, indicating whether the within-category phonetic details of lexical tones should be stored as part of listeners' lexical representations in order to be incorporated into the sound categorization process.

The current study also examines whether English-speaking second-language (L2) learners of Mandarin differ from native Mandarin listeners in their use of within-category tonal information. In contrast to Mandarin, languages like English do not have lexical tones; this means that tonal information at the syllable level does not contribute to lexical identity in English. Native speakers of English who learn Mandarin after the offset of the so-called 'critical period' for language acquisition (e.g., Lenneberg, 1967) thus find it difficult to use lexical tones in the recognition of spoken Mandarin words (e.g., Braun, Galts, & Kabak, 2014; Sun, 2012). Previous studies on the learning of Mandarin tones have focused on whether L2 learners have difficulty discriminating and/or identifying Mandarin tones if the native language (L1) does not have lexical tones (e.g., Chandrasekaran, Sampath, & Wong, 2010; Gandour, 1983; Hallé et al., 2004) and on the types of training that are most effective for enhancing the learning of tones (e.g., Wang et al., 1999; Wayland & Li, 2008). Hence, little is currently known about the effect of within-category tonal information on L2 learners' lexical activation.

Whether within-category tonal information constrains English listeners' recognition of Mandarin words is likely to depend on the degree to which English listeners' tone representations are phonetically detailed. In order for English-speaking L2 learners of Mandarin to use within-category tonal information in a target-like manner, they should both tune in to the fine-grained phonetic details of lexical tones and relate these phonetic details to prototypical tonal categories. Categorical perception studies have shown that native English listeners are more sensitive to subtle within-category pitch changes compared to native Mandarin listeners, whose perception of lexical tones is more categorical (e.g., Leather, 1987; Peng et al., 2010; Shen & Froud, 2016; Stagray & Downs, 1993; Xu et al., 2006). These findings suggest that English-speaking L2 learners of Mandarin should be able to tune in to the fine-grained phonetic details of lexical tones in spoken word recognition; less clear is whether they can also relate these phonetic details to prototypical tonal categories. L2 tonal training studies have found that English listeners often have difficulty dealing with the tonal variability caused by different speakers, tonal contexts, and speech rates (e.g., Chang & Bowles, 2015; Liu & Zhang, 2016; Wang et al., 1999, 2003). This difficulty has been attributed to L2 learners' limited exposure to tonal variability, which may cause L2 learners' lexical representations of tones to be coarser or less phonetically detailed, making it more difficult for them to interpret within-category tonal information in relation to prototypical tonal categories in lexical access (e.g., Díaz, Mitterer, Broersma, & Sebastián-Gallés, 2012; Shen & Froud, 2018).

The present study sheds further light on these issues. If English-speaking L2 learners of Mandarin are sensitive to the fine-grained phonetic details of lexical tones but have difficulty relating these details to prototypical tonal categories, we might expect them to show *more* competition from tonal competitors when the pitch of the target word (heard in the signal) is *different* from the standard (i.e., prototypical) tone, whether or not the pitch heard is acoustically closer to or more distant from that of the tonal competitor word (in L2 learners' lexical representation). Investigating the effect of within-category tonal information on English listeners' recognition of Mandarin words will provide insights into whether within-category tonal information modulates native and non-native Mandarin listeners' lexical activation in a *qualitatively* different way, which may in turn help explain L2 learners' difficulties in the processing of suprasegmental information.

The present study uses the visual-world eye-tracking paradigm, a method which has proven to be effective for testing listeners' continuous use of tonal information in spoken word recognition (e.g., Malins & Joanisse, 2010; Shen et al., 2013; Wiener & Ito, 2015). We report the results of an eye-tracking experiment with T1 (i.e., the level tone) and T2 (i.e., the rising tone) as the critical tone pair, because these two tones differ in their early pitch information and can be distinguished from the onset of the tones, thus making it possible to determine whether the within-category tonal information heard from the onset of the tone modulates lexical activation.

2.0 Method

2.1 Participants

Thirty-six native Mandarin listeners (mean age: 23.3, standard deviation (SD): 1.7, 25 females) and twenty-six native English listeners (mean age: 22.3, SD: 2.9, 9 females) who learned Mandarin after the age of 12 and considered themselves proficient in Mandarin participated in this study. The testing took place at the Center for Brain and Cognitive Sciences at Peking University, China. All participants were college students. They reported normal hearing and no history of speech or language disorders. In compensation for their time, the Mandarin listeners each received the equivalent of 20 US dollars, and the English listeners, who also completed Mandarin proficiency tasks, each received the equivalent of 30 US dollars.

All Mandarin-speaking participants identified Beijing Mandarin (i.e., Putonghua) to be their L1, alone or together with another Mandarin variety (Northern Mandarin dialect = 9, Southwestern Mandarin dialect = 8, Jianghuai Mandarin dialect = 2). The Mandarin varieties the participants knew are similar to Beijing Mandarin in their prosodic systems (Norman, 1988), and these participants reported using Beijing Mandarin as their dominant language. None of the Mandarin-speaking participants knew other Chinese languages (e.g., Shanghainese and Cantonese).

All L2 learners reported that English was their L1, that both their parents were native English speakers, that English was the only language spoken in their household during childhood, and that English was the primary language of their K-12 education. Additionally, they reported having learned Mandarin after the age of 12 and not having been exposed to tone languages other than Mandarin. The L2 learners' proficiency in Mandarin was tested with a Mandarin lexical decision task adapted from LexTALE (Lemhöfer & Broersma, 2012), as well as a Mandarin cloze (i.e., fill-in-the-blank) test (Yuan, 2009). The Mandarin LexTALE included a total of 120 items, 80 of which were words. The cloze test included a total of 40 missing words. The L2 learners' Mandarin learning experience and proficiency information (mean scores converted into percentages) is summarized in Table 1.

Table 1 ***L2 learners' Mandarin Learning Experience and Proficiency Information***

	AOE (year)	Years of Mandarin Instruction	LOR (month)	Lexical Decision Task (%)	Cloze Test (%)
English (n = 26)	17.6 (3.4)	4.1 (2.3)	13.0 (14.3)	66.5 (8.0)	82.1 (10.0)

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

2.2 Materials

In each trial of the eye-tracking experiment, participants heard a target word and saw four images in a 2 x 2 visual display, one of which corresponded to the target word, one of which corresponded to a competitor word, and two of which corresponded to distracter words, as illustrated in Figure 1. All words were imageable monosyllabic nouns.[1]

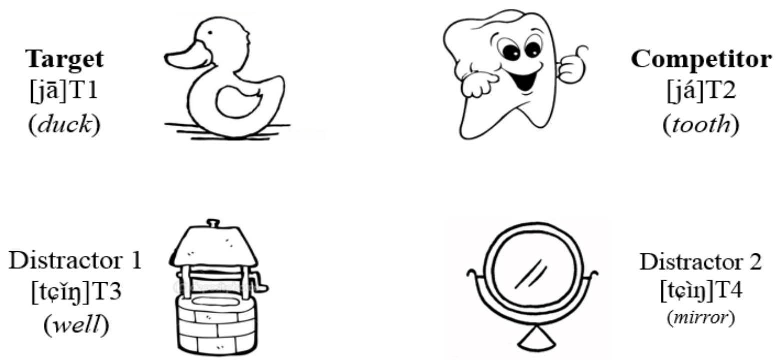


Figure 1. A visual display of experimental trial (the orthographic transcriptions were not presented in the actual experiment)

In the experimental trials, six T1-T2 word pairs (e.g., T1, a level tone: /jā/ 'duck'; T2, a rising tone: /já/ 'tooth') were used as target-competitor words. The two words in each pair shared the same segments, contrasted in tones, and were not semantically related. The target and competitor words in these trials began with an approximant (i.e., /j/, /w/, or /ɥ/, e.g., /jā/ 'duck') to reduce the likelihood of a tone-bias effect.[2] The T1-T2 words were matched for log morphemic frequency based on the SUBTLEX-CH database (Cai & Brysbaert, 2010; T1: 2.8 [SD: 1.2], T2: 3.2 [SD: 0.4]), and they were matched for tonal neighborhood (defined as the

---

[1] The nouns were represented with images rather than orthographically because little is currently known about how Chinese characters are processed in the context of a visual-world eye-tracking experiment.

[2] The analysis of SUBTLEX-CH database (Cai & Brybaert, 2010) showed that syllables beginning with an obstruent are less evenly distributed across all four of Mandarin tones compared to syllables beginning with an approximant. Thus, approximant-initial words were selected so that the results would not be affected by tone bias effects, reported in Wiener and Ito (2015).

number of legal tones that can be associated with a given syllable; Sun, 2012; Yip, 2002, p. 140; T1: 4 [SD: 0], T2: 4 [SD: 0]). For each of the T1-T2 word pair, a T3-T4 word pair (e.g., T3, /tɕǐŋ/ 'well'; T4 /tɕìŋ/ 'mirror') was selected to be used as distracter words. The two words in each pair were phonologically and semantically unrelated to the target and competitor words, but like the target and competitor words, as a pair they shared the same segments and differed only in tones. Their similar phonological overlap thus prevented possible baseline effects in the results (i.e., the participants did not know ahead of time which pair would be the target and competitor). The T3-T4 words were also matched for log morphemic frequency (T3: 3.5 [SD: 0.6], T4: 3.3 [SD: 0.6]) and tonal neighborhood (T3: 3.7 [SD: 0.5], T4: 3.7 [SD: 0.5]). The T1 and T2 words were each heard three times as targets, once in each experimental condition (see Section 2.3). All words from these trials can be found in Appendix A.

The experiment also included two types of filler trials. One type of filler trials was similar to the experimental trials, but with the T1-T2 target-competitor words beginning with a fricative or an affricate (e.g., T1, /tʂʰā/ 'fork'; T2 / tʂʰá/, 'tea') so that the approximant-initial T1-T2 words in the experimental trials would not stand out. Six T1-T2 fricative-/affricate-initial word pairs were selected for these trials; for each of these word pairs, a T3-T4 word pair (e.g., T3, /ɕɥě/ 'snow'; T4 /ɕjàu/ 'smile') was selected for the distracter words.[3] These T1 and T2 words were also each heard three times as targets (see Section 2.3) throughout the experiment. All words from these trials can also be found in Appendix A. The second type of filler trials had all distracter words from the experimental trials and from the previously described filler trials as

[3] The T3 and T4 words in the filler trials were not always perfectly matched segmentally because of the lack of T3-T4 minimal pairs.

targets so that participants would not be biased to expect a word containing T1 or T2, with each distracter word also being heard three times as targets (see Section 2.3).

Black-and-white images were selected to represent each word in the experiment. To ensure that the images were representative of the words, twelve native Mandarin speakers rated the goodness of each image for representing the corresponding word using a 1-6 point scale (with 1 meaning very bad and 6 meaning very good). The images representing T1 and T2 words were all highly rated (T1: 5.5 [SD: 0.8], T2: 5.0 [SD: 1.3]) as well as T3 and T4 words (T3: 5.4 [SD: 1.1], T4: 5.5 [SD: 0.8]).

The experiment included a total of 36 experimental trials (6 word pairs × 2 tones × 3 tokens) and 108 filler trials (18 word pairs × 2 tones × 3 tokens). Across experimental and filler trials, the four tones were each heard 36 times, and the 48 images corresponding to the T1-T2 and T3-T4 word pairs were each seen 12 times. The location of target, competitor, and distracter images in the display was counterbalanced throughout the experiment. The relative location of targets and competitors and of the different tones was also counterbalanced, and experimental and filler trials were pseudorandomized.

2.3 Stimulus Preparation

One male native speaker of Standard Mandarin was recorded producing all the stimuli in a quiet room. The speaker read a randomized list of words in isolation three times at a normal speech rate. One token was chosen for each word based on the clarity of the spoken words. The intensity of all selected tokens was normalized to 70 dB. Then, both the T1-T2 and T3-T4 words were resynthesized.

Given the different duration of the naturally produced tones, the duration of all targets stimuli was normalized at 445 ms (the duration mean of T1, T2, and T4 natural tokens) using the "To Manipulation" function in Praat (Boersma & Weenink, 2015).[4,5] Next, the T1-T2 words had their pitch height resynthesized such that their pitch contour would be either more similar to or more different from that of the competitor word. Three levels of a tonal continuum were created for T1 (T1-Standard, T1-Distant, and T1-Close) and for T2 (T2-Standard, T2-Distant, and T2-Close). T1-Standard and T2-Standard are standard exemplars of T1 and T2 natural tokens. T1-Distant is acoustically more distant from T2 than T1-Standard is, but T1-Close is acoustically closer to T2 than the T1-Standard is. Likewise, T2-Distant is acoustically more distant from T1 than T2-Standard is, but T2-Close is acoustically closer to T1 than the T2-Standard is. The standard (i.e., prototypical) T1 and T2 tokens were created by using the average time-normalized pitch values of the natural T1 tokens and the average time-normalized pitch values of the natural T2 tokens. The acoustically closer and more distant tokens were created by using the average pitch values of the natural tokens but raising or lowering the starting point of the contour by 10 Hz. A manipulation value of 10 Hz was chosen, because it was small enough that the acoustically closer tone would not be close to the boundary (e.g., McMurray et al., 2002), but the difference between the acoustically closer and more distant tones and the canonical tones was larger than the Just Noticeable Difference (JND) of tone contour discrimination for both

---

[4] The duration of T3 words was not considered when establishing the normalized duration that all the auditory stimuli would have, because participants completed an additional experiment (not reported here) with T1, T2, and T4 target words that required duration to be normalized based on these three tones. The same duration normalization was used in both experiments.

[5] The mean duration for each tone was as follows: 454 ms for T1, 524 ms for T2, and 387 ms for T4. Thus, T1 words had very similar normalized and natural durations (445 ms and 453 ms, respectively), whereas T2 words had a shorter normalized duration compared to its natural duration (443 ms vs. 524 ms). One might thus hypothesize that if listeners used duration as a cue to recognize the words, they would show more lexical competition from T1 words in T2 trials than from T2 words in T1 trials. Crucially, the results suggest that this is not the case (see the statistical analyses in Appendix D).

Mandarin and English listeners (Liu, 2013; Jongman, Qin, Zhang & Sereno, 2017).[6] The acoustically closer and more distant contours were obtained via incremental interpolation between the new starting points and the natural endpoint in 10 measurements. The resynthesized pitch contours, illustrated in Figure 2, were superimposed on the duration-normalized stimuli.

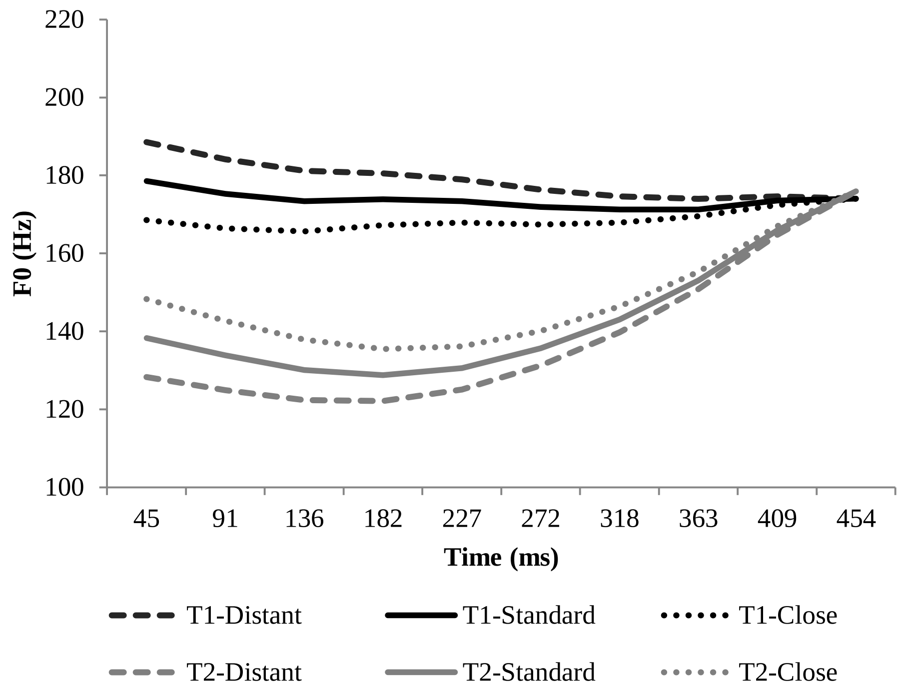


Figure 2. Tonal continua of Tone 1 and Tone 2 used in the stimuli

To ensure that all three levels of the two tonal continua would be within-category variations of the same tonal category and be far enough from the tonal boundary of T1-T2, six native Mandarin listeners judged the naturalness of the resynthesized T1-T2 stimuli on a 6-point scale (1-6, 1 meaning "The word sounds really bad" and 6 meaning "The word sounds really good") and categorized them into different tones. The stimuli were rated as natural (mean: 5.1;

---

[6] A tone continuum was not used because doing so would have resulted in either tokens that approach the tone boundary or tokens whose distance from each other is smaller than the JND.

SD: 1.1) and had their target tonal category correctly identified (mean: 99.5%; SD: 6.8%), and the conditions did not differ significantly in their naturalness rating ($t < |1|$) or in their identification accuracy ($t < |1|$): Standard (T1, mean accuracy: 99%, SD: 1.2%; mean rating: 5.3, SD: 1.2; T2, mean accuracy: 100%; mean rating: 5.1, SD: 1.1); Distant (T1, mean accuracy: 100%, SD: 1.2%; mean rating: 5.3, SD: 1.1; T2, mean accuracy: 100%; mean rating: 5.0, SD: 1.3); Close (T1, mean accuracy: 100%; mean rating: 5.3, SD: 1.1; T2, mean accuracy: 100%; mean rating: 4.9, SD: 1.2).

To ensure that the resynthesized T1-T2 targets would not stand out, the T3-T4 targets in the filler trials were resynthesized following a method similar to that described for the T1-T2 words. The T3 and T4 words also had their duration normalized to 454 ms so that none of the words in the experiment would differ in their duration, which would help direct participants' attention to pitch information rather than durational cues. The standard T3 and T4 tokens were created by using the average time-normalized pitch values of the natural T3 tokens and the average time-normalized pitch values of the natural T4 tokens, and the acoustically closer and more distant tokens were created by using the average pitch values of the natural tokens but raising or lowering the starting point of the contour by 10 Hz and interpolating between the new starting points and the natural endpoint in 10 measurements (see Appendix B). Like the T1-T2 stimuli, the T3-T4 stimuli were rated as natural (mean: 5.0; SD: 1.2) and had their target tonal category correctly identified (mean: 92.8%; SD: 2.6%), and the conditions did not differ significantly in their naturalness rating ($t < |1|$) or in their identification accuracy ($t < |1|$).

2.4 Procedures

Prior to completing the eye-tracking experiment, participants first completed a word familiarity task in which they rated their familiarity with the words corresponding to all the images in the experiment on a scale from 0 to 4 (0 = "I have never seen/heard this word"; 4 = "I have frequently seen/heard this word, I know what it means, and I can provide a definition for it"). In this task, a spoken stimulus was played, its printed name was presented in Pinyin and (both simplified and traditional) Chinese characters, and participants rated their familiarity with the word. English listeners' familiarity ratings were not significantly different ($t < |1|$) between the T1 trials (Mean: 2.2; SD: 1.7) and the T2 trials (Mean: 2.4; SD: 1.6).

Next, participants completed a word-picture association training: First, they went through a look-and-listen phase in which they heard a spoken word with its natural pitch contour (after duration normalization) and saw the corresponding picture on the screen; second, they completed a picture selection test in which they heard a spoken word with its natural pitch contour (after duration normalization) and selected the picture corresponding to the word from a large number of candidate pictures. The target and competitor words from the same trial were displayed in the same set of pictures on the screen to make sure that the participants could distinguish the tonal contrast before the eye-tracking experiment. Participants received feedback on their responses, and the task ended only when they correctly identified all the pictures. This training was essential for four reasons: First, although all words were imageable, it was difficult to perfectly control the imageability of the words; second, the L2 learners were not as familiar with the words as the native listeners; third, and crucially, listeners' exposure to the auditory words allowed them to familiarize themselves with the pitch range of the speaker, which was crucial for L2 learners to store some form of prototypical tone for the speaker; fourth, the training helped direct participants' attention to pitch information, since the training stimuli also had their duration and intensity normalized.

Following this training session, the visual-world eye-tracking experiment was administered. Eye movements were recorded using a desktop-mounted Eyelink 1000 (sampling rate: 1000 Hz), and the experiment was delivered using the software Experiment Builder (www.sr-research.com). In the task, participants were instructed to click on the picture corresponding to the monosyllabic Mandarin word they heard through headphones. The visual display contained four black-and-white images (200 x 200 pixels) in a non-displayed 2 x 2 grid, as illustrated in Figure 1. In each trial of the experiment, participants first saw these four pictures for 2,000 ms (preview phase). The pictures then disappeared and a fixation cross centered on the screen appeared and stayed on the screen for 500 ms. The goal of this fixation cross was to bring participants' fixations back to the middle of the display. The fixation cross then disappeared and the four pictures reappeared on the screen (in the same location) and the auditory stimulus was synchronously heard (through headphones). The target word was heard in isolation to eliminate possible effects of tonal context.[7] Participants were instructed to click on the correct picture immediately after hearing the auditory stimulus, and their eye movements were recorded from the onset of the auditory stimulus in each of the four regions of interest (300 x 300 pixels surrounding the images). Trials were split into four different blocks, with each block containing only one token of each target word. The number of times each word and each

[7] The recognition of T1 and T2 targets may be differentially affected by the tone in the preceding word, even if the T1 and T2 targets were recorded in isolation and thus not influenced by co-articulatory cues. For example, T1 targets, which begin with a high pitch, may be recognized faster following a tone that ends in a low pitch (e.g., T4) than following a tone that ends with a higher pitch (e.g., T2) because of the pitch difference between the end of the first tone and the beginning of the second tone; conversely, T2 targets, which begin with a low pitch, may be recognized faster following a tone that ends in a high pitch (e.g., T1) than following a tone that ends in a low pitch (e.g., T4) for the same reason.

tone was heard was the same within and across blocks. The task began with 8 practice trials in which the four words on the screen differed segmentally and suprasegmentally (e.g., /ɕjā/ 虾 ‘shrimp, ’/tʂ^{h}wán/ 船 ‘boat’, /ts^{h}ǎu /草 ‘grass’, /ɥuɤ̀/ 月 ‘moon’), followed by the main experiment.

To reduce participants’ (especially L2 learners’) memorization burden, the word familiarity task, the word-picture association training, and the eye-tracking experiment were conducted over three sessions, with one third of all the words being used in each session and with at least two days between sessions. Since the experiment included a total of 48 words (and 144 trials), the word familiarity task and the word-picture association training contained 16 words in each session, and the eye-tracking experiment contained the same 16 words (and 48 trials). Only words that had been trained on the same day would appear in the eye-tracking experiment. Per session, the word familiarity task took approximately 5 minutes to complete, the word-picture association training approximately 10-15 minutes for native Mandarin listeners and 20-30 minutes for L2 learners, and the eye-tracking experiment approximately 15 minutes. The order of the different sessions was counterbalanced across participants.

2.5 Data Analysis

Only trials in which participants clicked on the target word were analyzed, resulting in the exclusion of 1% of the data for Mandarin listeners and 21.5% of the data for English listeners. Proportions of fixations to the target, competitor, and distracter words were extracted in 20-ms time windows from the onset of the target word to 800 ms post-target-word onset using a Python script. The dependent variable for the statistical analyses was the difference between the empirical log-transformed proportions of target and competitor fixations (i.e., the transformed proportion of competitor fixations was subtracted from the transformed proportion of target fixations) from the target word onset to the word offset (i.e., 454 ms), both with a delay of 200-ms (i.e., the window of analysis was 200-654 ms) to account for the fact that eye fixations take approximately 200 ms to reflect speech processing (Hallet, 1986). This dependent variable is ideal in that it simultaneously takes into consideration listeners' activation of both the target and competitor words (for a similar analysis of visual-world eye-tracking data, see Creel, 2014 and Connell et al., 2018).

Listeners' transformed fixation differences were modeled using growth curve analysis (GCA; Mirman, 2014), a type of curvilinear regression that can model the linear (i.e., capturing the overall angle of a curve), quadratic (i.e., capturing a curve with a single inflection), and cubic (i.e., capturing a curve with two inflections) shapes of the differential fixation lines. GCAs are similar to mixed-effects models (Bates, Maechler, & Walker, 2015) but include time polynomials, thus making it possible to model the shape of participants' overall eye fixation curve rather than their average eye fixations at arbitrary points in time. Since this experiment focused on the participants' sensitivity to the fine-grained phonetic details of tonal contours as the speech signal unfolds, GCAs are appropriate for analyzing participants' fixations, because they can model subtle changes in the curvilinear patterns of eye fixations over time and capture differences in the slope and curvature of the differential fixation lines.

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

To be able to conclude that our tonal manipulation had an effect on participants’ fixations, the GCAs must show *interactions* between the effect of condition (Standard vs. Distant or Close) and the linear and/or quadratic time polynomials, and possibly the cubic time polynomial.[8] These interactions would indicate that, as the speech signal unfolds over time, the *shape* of participants’ differential fixation line changes differently across the different tonal conditions. Importantly, an effect of Condition without a significant interaction with the linear, quadratic, or cubic time polynomials indicates that fixation proportions are either higher or lower in one tonal condition than in another, but the *shape* of participants’ differential fixation lines are similar among the different conditions; thus, on its own, such an effect cannot be attributed to the tonal manipulation in the speech signal, and may instead be understood as a baseline difference between the two conditions (i.e., a difference that exists independently of the auditory stimuli; Barr, Gann, & Pierce, 2011).

---

[8] The cubic time polynomial can capture an asymptote effect in the tail of the fixation line. Since the asymptote effect does not have a meaningful cognitive interpretation in many cases (Mirman, 2014), the interaction between condition and the cubic time polynomial is less relevant than the interaction between the linear and/or quadratic time polynomials.

The GCAs were conducted with the *lme4* package in R (Bates et al., 2015). For the sake of clarity, we first present the analysis of the individual language group's results. These analyses included the three time polynomials (linear, quadratic, and cubic), condition (Standard, Distant, and Close; baseline: Standard), the tone of the target word (T1 vs. T2; Baseline: T1), and all interactions as fixed effects.[9] A back-fitting function from the package *LMERConvenienceFunctions* in R (Tremblay & Ransijin, 2015) was used to identify the model that accounted for significantly more of the variance than simpler models, as determined by log-likelihood ratio tests; only the results of the model with the best fit are presented, with *p* values calculated using the *lmerTest* package in R (Kuznetsova, Brockhoff, & Christensen, 2016). Analyses yielding significant interactions between condition and tone were followed up by subsequent GCAs conducted separately on trials with T1 targets and trials with T2 targets, with the alpha level adjusted to .025. All analyses included participant as random intercept and the orthogonal time polynomials as random slopes for the participant variable, which allowed the analysis to model a line of a different shape for each participant.[10] To determine whether the English-speaking L2 learners differed from native Mandarin listeners in the use of within-category tonal information, we also conducted analyses that tested three-way interactions between the effects of condition, L1 (Mandarin vs. English, baseline: Mandarin), and time separately for the trials with T1 targets and trials with T2 targets.

---

[9] Tone was included in the analyses because L2 learners of Chinese have been shown to have less difficulty recognizing T1 compared to T2 (e.g., Hao, 2012; Wang et al., 1999); it is thus possible that they will show greater sensitivity to within-category tonal information with T1 compared to T2.

[10] We followed Mirman's (2014) recommendation and did not include item as a random intercept. Data from a single visual-world paradigm trial consist of a sequence of categorical fixations rather than a smooth fixation curve. Even at a lower-sampling rate (such as that used in the present study), the fixations in a single trial do not yield a smooth curve and thus would be difficult to model. Since participants are more likely to differ in the rate of activation of lexical representations than items, participant was the chosen random intercept for the analyses.

If Mandarin listeners show sensitivity to within-category tonal information, the GCAs should yield one or more of the following effects: (i) significant interactions between the effect of condition and the linear time polynomial, with the differential fixation line being more ascending in the Distant condition than in the Standard condition and less ascending in the Close condition than in the Standard condition; (ii) potential significant interactions between the effect of condition and the quadratic time polynomial, with the differential fixation line being less U-shaped in the Distant condition than in the Standard condition and more U-shaped in the Close condition than in the Standard condition; (iii) a potential significant interaction between the effect of condition and the cubic time polynomial, with the differential fixation line having a sharper 's' shape in the Distant condition than in the Standard condition. A more ascending and/or less U-shaped differential fixation line would be indicative of faster target word recognition due to increased target activation and decreased competitor activation, whereas a less ascending and/or more U-shaped differential fixation line would be indicative of slower target word recognition due to decreased target activation and increased competitor activation. A differential fixation line with a sharper 's' shape would be due to fixations reaching an asymptote towards the end of the trial as a result of a fast word recognition (i.e., listeners have recognized the word, so they start looking elsewhere; see Footnote 2).

If L2 learners are sensitive to the fine-grained phonetic details of lexical tones but have difficulty relating these details to prototypical tonal categories, the GCAs should yield one or more of the following effects: (i) significant interactions between the effect of condition and the linear time polynomial, with the differential fixation lines being less ascending in both the Distant and Close conditions than in the Standard condition; (ii) potential significant interactions between the effect of condition and the quadratic time polynomial, with the differential fixation lines being more U-shaped in both the Distant and Close conditions than in the Standard condition. We do not expect a significant interaction between the effect of Condition and the cubic time polynomial, because L2 learners are expected to show a slower recognition of the words in the Distant and Close conditions compared to those of the Standard condition.

If Mandarin or English listeners show different sensitivities within-category tonal information, the GCAs should yield a three-way interaction between the effect of L1, the effect of condition (particularly for the Distant condition), and one or more of the time polynomials. Such an interaction would indicate that the conditions had a different effect on the differential fixation lines of the two language groups.

3.0 Results

While Mandarin listeners performed at ceiling in identifying the target word (mean accuracy: 99%; SD: 9.8%), English listeners were less accurate (mean accuracy: 78.5%; SD: 44.7%). Mandarin listeners clicked on the target picture at an average of 1,521 ms (SD: 445 ms), whereas English listeners did so at an average of 2,380 ms (SD: 959 ms). These mouse-click results are consistent with previous studies showing that Mandarin listeners had a higher accuracy and shorter RTs than English listeners when recognizing Mandarin words (e.g., Sun, 2012).

We now turn to the analysis of participants' eye fixations, reported separately for Mandarin and English listeners.

3.1 Native Mandarin Listeners' Fixations

Figure 3 shows Mandarin listeners' differential proportions of fixations in the Standard, Close, and Distant conditions for trials with T1 targets and trials with T2 targets in the first 800 ms post target-word onset (Mandarin listeners' proportions of target, competitors, and distracter fixations are shown in Figure C1 of Appendix C). Differential proportions of fixations above 0 mean that participants looked more at the target than at the competitor.

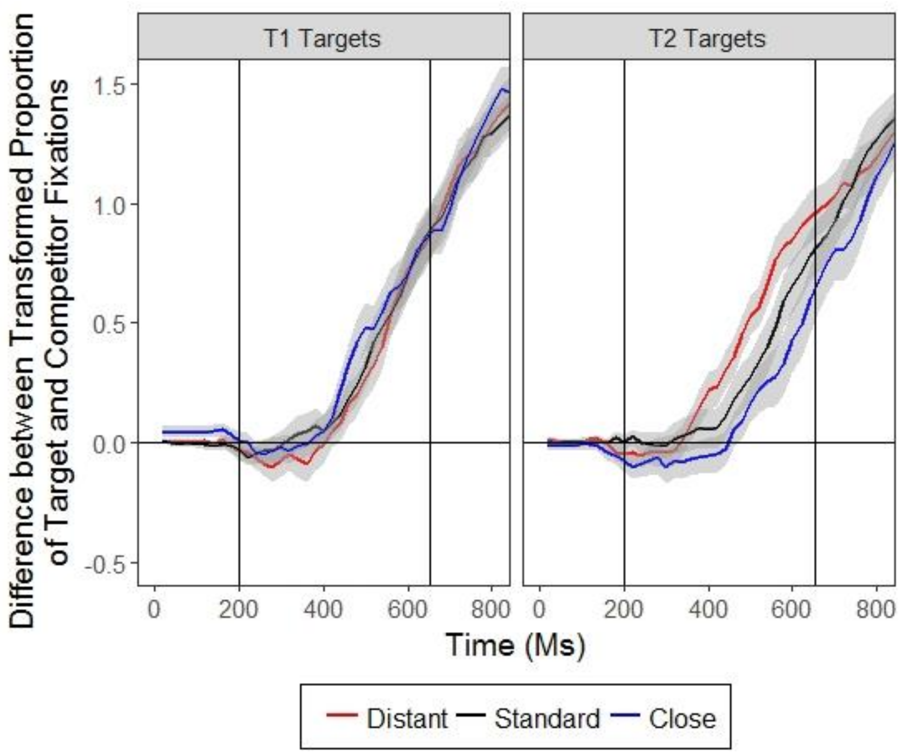


Figure 3. Mandarin listeners' differential proportions of fixations in the Distant (red), Standard (black), Close (blue) conditions for trials with T1 targets and trials with T2 targets in the first 800 ms; the shaded area represents one standard error above and below the participant mean; the vertical lines represent the onset and offset of the target word (and tone) with a 200-ms delay (time window of analysis)

Recall that GCAs were performed on the differential proportions of fixations corresponding to the tonal portion (duration of 454 ms) of the target word (from the tonal onset to the tonal offset with a 200-ms delay, represented with vertical lines in Figure 3). The GCA with the best fit on Mandarin listeners' data included the linear and quadratic time polynomials, condition, tone, and their interactions. The results of this GCA are provided in Table D1 of Appendix D. This GCA yielded significant three-way interactions between the linear time polynomial, condition, and tone for both the Close and Distant conditions. Subsequent GCAs were therefore performed on the differential proportions of fixations separately for each target tone. Table 2 presents the results of the follow-up GCAs with the best fit. For trials with T1 targets, the GCA with the best fit included the simple effects of the two time polynomials, and condition; for trials with T2 targets, the GCA with the best fit included all simple effects and the interaction between the linear time polynomial and condition.

Table 2 ***Growth Curve Analyses on the Difference between Mandarin Listeners' Transformed Proportions of Target and Competitor Fixations Separately for Trials with T1 Targets and Trials with T2 Targets***

Tone	**Effect**	**Estimate**	***t***	***p***
T1	(Intercept)	0.223	6.705	**< .001**
	Time			
	Linear	1.228	9.936	**< .001**
	Quadratic	0.474	7.201	**< .001**
	Condition (Distant)	–0.043	–2.348	**< .019**
	Condition (Close)	0.034	1.829	.07
T2	(Intercept)	0.214	5.160	**< .001**
	Time			
	Linear	1.076	6.594	**< .001**
	Quadratic	0.421	5.499	**< .001**
	Condition (Distant)	0.109	5.839	**< .001**
	Condition (Close)	–0.133	–7.084	**< .001**
	Time × Condition (Distant)			
	Linear	0.517	5.766	**< .001**
	Time × Condition (Close)			
	Linear	–0.201	–2.238	**.025**

Note. α = .025; significant results are in bold; T1: *n* = 2484 observations; T2: *n* = 2484 observations

The effects that speak about the influence of within-category tonal information on lexical activation for Mandarin listeners can be summarized as follows. For trials with T1 targets, the negative estimate for the significant effect of condition (Distant) means that Mandarin listeners had a *lower* differential proportion of fixations in the Distant condition than the Standard condition (contrary to predictions). Crucially, the interaction between the time polynomials and condition (both Distant and Close) did not improve the model, indicating that the differential fixation lines in the Distant and Standard conditions did not differ in their slope or U-shape and suggesting that the effect of condition with Distant tokens may not be attributable to the speech signal. These results suggest that, for trials with T1 targets, within-category tonal information did not enhance target-over-competitor word activation in the Distant condition, and it did not inhibit it in the Close condition.

For trials with T2 targets, the positive estimate for the significant effect of condition (Distant) means that Mandarin listeners had a *higher* differential proportion of fixations in the Distant condition than the Standard condition (as predicted), and the negative estimate for the significant effect of condition (Close) indicates that Mandarin listeners had a *lower* differential proportion of fixations in the Close condition than the Standard condition (as predicted). Importantly, the positive estimate for the significant interaction between the linear time polynomial and condition (Distant) suggests that Mandarin listeners had a more ascending differential fixation line in the Distant condition than in the Standard condition. The significant interaction with the linear time polynomial stems from the increasingly higher proportions of target fixations and lower proportions of competitor fixations over time in the Distant condition compared to the Standard condition (see Figure C1 in Appendix C). The negative estimate for the significant interaction between the linear time polynomial and condition (Close) indicates that Mandarin listeners' differential fixation line was less ascending in the Close condition than in the Standard condition. This interaction stems from the increasingly lower proportions of target fixations and higher proportions of competitor fixations in the Close condition compared to the Standard condition. Thus, for trials with T2 targets, the acoustically greater tonal distance between the target and competitor enhanced target-over-competitor word activation in the Distant condition compared to the Standard condition; conversely, the acoustically smaller tonal distance between the target and competitor inhibited target-over-competitor word activation in the Close condition compared to the Standard condition. These results suggest that the within-category phonetic details of T2, as manipulated in the present study, constrained Mandarin listeners' lexical access.

3.2 English-Speaking L2 Learners of Mandarin

Figure 4 shows English listeners' differential proportions of fixations in the Standard, Close, and Distant conditions for trials with T1 targets and trials with T2 targets in the first 800 ms post target-word onset (English listeners' proportions of target, competitors, and distracter fixations are shown in Figure C2 of Appendix C).

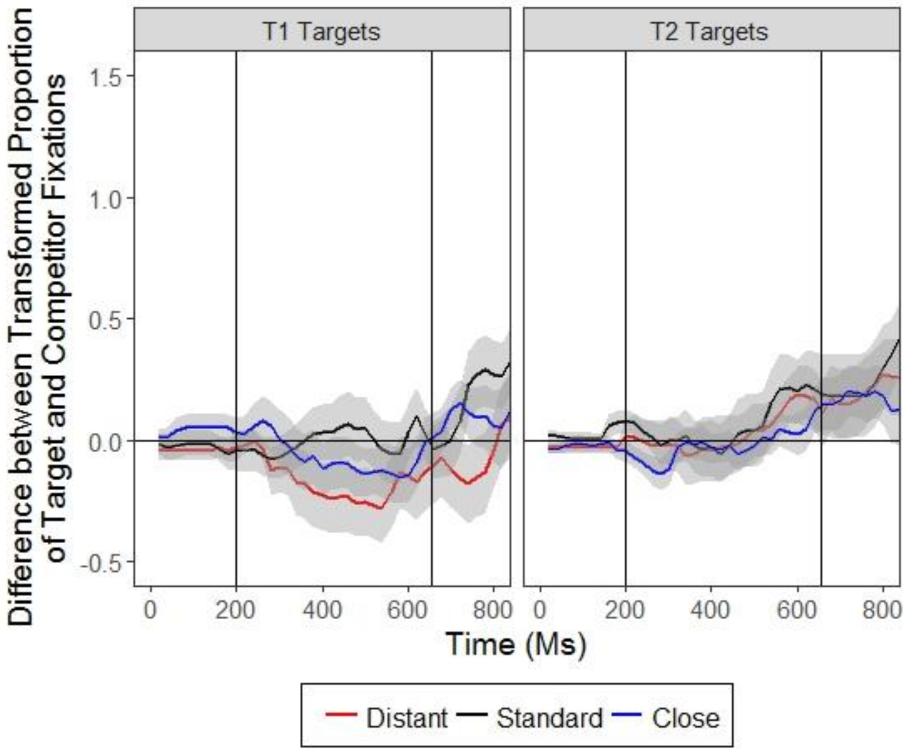


Figure 4. English listeners' differential proportions of fixations in the Distant (red), Standard (black), Close (blue) conditions for trials with T1 targets and trials with T2 targets in the first 800 ms; the shaded area represents one standard error above and below the participant mean; the vertical lines represent the onset and offset of the target word (and tone), with a 200-ms delay (time window of analysis)

The GCA with the best fit on English-speaking L2 learners' data included the linear time polynomial, condition, tone as well as their interactions. The results of this GCA are provided in Table D2 of Appendix D. This GCA yielded significant interactions between the linear time polynomial and tone, as well as a significant interaction between condition and tone for the Distant condition. Subsequent GCAs were therefore performed on the differential proportions of fixations separately for each target tone. Table 3 presents the results of the follow-up GCAs with the best fit. For trials with T1 targets, the GCA with the best fit included the linear time polynomial, condition, and their interactions; for trials with T2 targets, the GCA with the best fit only included the simple effects of condition.

Table 3 ***Growth Curve Analyses on the Difference between English Listeners' Transformed Proportions of Target and Competitor fixations Separately for Trials with T1 Targets and Trials with T2 Targets***

Tone	Effect	Estimate	*t*	*p*
T1	(Intercept)	0.050	1.297	.20
	Time			
	Linear	0.133	0.788	.43
	Condition (Distant)	–0.159	–6.042	**< .001**
	Condition (Close)	–0.059	–2.225	**.025**
	Time × Cond (Distant)			
	Linear	–0.370	–2.930	**< .01**
	Time × Cond (Close)			
	Linear	–0.391	–3.097	**< .01**
T2	(Intercept)	0.056	1.969	.05
	Condition (Distant)	–0.031	–1.198	.23
	Condition (Close)	–0.083	–3.165	**< .01**

Note. α = .025; significant results are in bold; T1: *n* = 4368 observations; T2: *n* = 4368 observations

The effects that speak about the influence of within-category tonal information on lexical activation for English listeners can be summarized as follows. The negative estimates for the significant effects of Distant and Close conditions mean that English listeners had *lower* differential proportions of fixations in the Distant and Close conditions compared with the Standard condition, as predicted. Crucially, for both the Distant and Close conditions, the negative estimates for the significant interactions between condition and the linear time polynomial indicate that English listeners had a less ascending differential fixation line in the Distant and Close conditions than in the Standard condition. The significant interactions with the linear time polynomial stem from the increasingly lower proportions of target fixations and higher proportions of competitor fixations over time in the Distant and Close conditions compared to the Standard condition (see Figure C2 in Appendix C). These results suggest that, for trials with T1 targets, within-category tonal information inhibited English listeners' target-over-competitor word activation in both the Distant and Close conditions compared to the

Standard condition. In other words, the tones that were acoustically *different* from the prototypical tone (i.e., the Standard condition) disrupted English listeners' word recognition.[11]

For trials with T2 targets, the negative estimate for the significant effect of Close condition suggest that English listeners had a *lower* differential proportion of fixations in the Close condition compared with the Standard condition. However, the interaction between these effects and the time polynomials did not improve the model, indicating that the differential fixation line shapes in the Distant and Close conditions did not differ from that in the Standard condition. These results suggest that, for trials with T2 targets, within-category tonal information did not clearly inhibit English listeners' word recognition in either the Distant or Close condition.[12]

3.3 Native Listeners vs. L2 Learners

To determine whether the English-speaking L2 learners differed from native Mandarin listeners in the use of within-category tonal information, analyses were conducted that tested

---

[11] Exploratory analyses with centered mean proficiency score (the average of the last two columns in Table 1) and its interactions with condition and the time polynomials revealed significant interactions between proficiency and condition, with the effect of condition being more negative (i.e., more inhibiting) for more proficient L2 learners than for less proficient L2 learners in both the Distant and Close conditions (Distant: $\beta = -0.019$, SE = 0.004, $t = -5.149$, $p < .001$; Close: $\beta = -0.018$, SE = 0.004, $t = -4.729$, $p < .001$). These results suggest that the inhibiting effect of the Distant and Close conditions became larger as proficiency in Mandarin increased.

[12] However, exploratory analyses with mean proficiency score (the average of the last two columns in Table 1) and its interactions with condition and the time polynomials revealed significant interactions between proficiency and condition, with the effect of condition being more negative (i.e., more inhibiting) for more proficient L2 learners than for less proficient L2 learners in both the Distant and Close conditions (Distant: $\beta = -0.020$, SE = 0.004, $t = -5.297$, $p < .001$; Close: $\beta = -0.012$, SE = 0.004, $t = -3.351$, $p = .001$). This suggests that, like in T1 trials, the tokens heard in the Distant and Close conditions of T2 trials were more likely to inhibit spoken word recognition as L2 learners' proficiency in Mandarin increased.

three-way interactions between the effects of condition, L1, and the time polynomials separately for the trials with T1 targets and trials with T2 targets.

For trials with T1 targets, the GCA with the best fit included linear and quadratic time polynomials, condition, L1, and all interactions. The results of this GCA are reported in Table D3 of Appendix D. Importantly, this GCA revealed significant three-way interactions between the linear time polynomial, condition, and L1 for both the Distant and Close conditions. These results confirm that, in trials with T1 targets, within-category tonal information differentially affected Mandarin and English listeners' fixation lines in both the Distant and Close conditions.

For trials with T2 targets, the GCA with the best fit included the linear and quadratic time polynomials, condition, L1, and their interactions. The results of this GCA can be found in Table D4 of Appendix D. Crucially, this GCA yielded significant a three-way interaction between the linear time polynomial, condition, and L1 for the Distant condition. These results indicate that, in trials with T2 targets, within-category tonal information had a different effect on Mandarin and English listeners' fixation lines in the Distant condition.

4.0 Discussion

This study investigated whether within-category tonal information would modulate native Mandarin listeners' lexical access, and whether it would differentially affect lexical activation in native Mandarin listeners and English-speaking L2 learners of Mandarin. A visual-world eye-tracking experiment was used to examine Mandarin and English listeners' fixation patterns when the tonal information heard in the target word was acoustically closer to or more distant from that of the competitor word (in listeners' lexical representations) compared to when the tonal information was standard (thus, more prototypical). T1 (a level tone) and T2 (a rising tone) were used as critical tone pair to examine listeners' use of within-category tonal information because these two tones differ in their early pitch and can thus be distinguished from the onset of the tones. It was hypothesized that native Mandarin listeners' target-over-competitor word activation would be enhanced in the Distant condition and inhibited in the Close condition (compared to the Standard condition), whereas English listeners' target-over-competitor word activation would be inhibited in both the Distant and Close conditions (compared to the Standard condition). The results provided partial support for the formulated hypotheses, with Mandarin listeners clearly showing the predicted effects of within-category tonal information only when the target word contained T2 and with English listeners showing a clear inhibiting effect for both the Distant and Close conditions only when the target word contained T1 (when the target word contained T2, the Distant and Close conditions were more likely to inhibit word recognition as English listeners' proficiency in Mandarin increased; see Footnote 12).

Mandarin listeners' results in trials with T2 targets are in line with those of previous research showing that the within-category phonetic details of consonants and vowels influences lexical activation (e.g., Dahan et al., 2001; McMurray et al., 2002, 2008, 2009). These results provide further evidence that the speech processing system does not discard within-category phonetic details, even when those details are not near the category boundary (e.g., McMurray et al., 2002). These findings also add to the literature on the perception of Mandarin tones. Previous research that used identification and discrimination tasks showed that native Mandarin listeners' perception of lexical tones was not proportional to the size of the acoustic changes heard in the speech signal (e.g., Hallé et al., 2004; Peng et al., 2010; Sun & Huang, 2012; Xu et al., 2006). These results were interpreted as suggesting that Mandarin listeners established a clear linguistic boundary between level and rising tones, and discarded the phonetic variability within each of the tonal categories. The present study, which used an experimental paradigm that is highly sensitive to the time course of lexical activation and can reveal differences elicited by fine-grained acoustic details, provided evidence that Mandarin listeners effectively do not discard the within-category phonetic details of T2, with small changes in the early pitch of the contour modulating listeners' target-over-competitor word activation such that lexical representations are increasingly or decreasingly activated as the target tone becomes acoustically closer to or more distant from that of the competitor word. An important theoretical implication of the present results is that models of spoken recognition should incorporate the within-category phonetic details of tonal information in spoken word recognition. One type of model that can account for the use of this information is exemplar-based models, which posit that listeners store the fine-grained phonetic details of the individual words they hear in memory, and in turn incorporate these details into the sound categorization process (for such a proposal with lexical tones, see Liu & Zhang, 2016; Zhang et al., 2016; Zhang & Chen, 2016; Zhang, 2018).

An important question that the current results raise, however, is *why* native Mandarin listeners showed sensitivity to within-category tonal information when it was heard in T2 targets but not when it was heard in T1 targets. There are a few possible explanations. On the one hand, the category boundary between T1 and T2 has been shown to be closer to T2 than to T1 (Peng et al., 2010; Xu et al., 2006). Hence, T2 may have less room for within-category phonetic variability than T1, and thus an equal increase/decrease in Hz at the onset of the tone may be perceived as greater for T2 than for T1. An alternative (but related) possibility is that because the pitch onset of T2 is lower than that of T1, an equal decrease/increase in Hz may be perceived as greater for T2 than T1. In other words, it is possible that the effect of within-category tonal information would have emerged for both tones if the early pitch differences had been held constant on a perceptual scale (e.g., semitone) rather than on an acoustic scale (Hz). A third possibility (also related to the first one) is that an equal increase/decrease in Hz at the onset of the tone may have a greater impact on the dynamic changes of pitch over time for rising tones (e.g., T2) than for the level tone (T1). The present study was not designed to compare the use of within-category information in T1 vs. T2 (the two tones were heard as targets for the purpose of counterbalancing the tones in the experiment). Further research should investigate whether a greater acoustic difference in the onset of T1 would modulate lexical activation in Mandarin listeners.

Importantly, the results also showed that within-category tonal information *inhibited* English listeners' target-over-competitor word activation in trials with T1 targets and was more likely to inhibit English listeners' word recognition with increased Mandarin proficiency for both T1 and T2 targets (see footnotes 11-12). English-speaking L2 learners of Mandarin thus differed *qualitatively* from native Mandarin listeners in that within-category tonal information disrupted English listeners' word recognition independently of the phonetic distance between the target and competitor tones. These results indicate that English listeners are sensitive to the fine-grained phonetic details of lexical tones but have difficulty relating these details to prototypical tonal categories. These findings are consistent with those of recent studies showing that English-speaking L2 learners of Mandarin showed heightened sensitivity to fine-grained phonetic differences of Mandarin tones, but had difficulty categorizing tones at early, pre-attentional levels of processing (as indexed by MMN), or at later, attentional-modulated levels of processing (as indexed by P300) (Shen & Froud, 2016, 2018). These difficulties may be due to L2 learners having coarser or less phonetically detailed representations of lexical tones as a result of their limited exposure to tonal variability in Mandarin. Specifically, English listeners—even those at a higher Mandarin proficiency in the current study—may not have been exposed to sufficient tonal input to develop robust tonal representations, making it more difficult for them to interpret within-category tonal information in relation to prototypical tonal categories. This may have resulted in decreased target word activation and increased competitor word activation when the pitch contour of the tokens they heard did not closely match that represented for the prototypical tokens. These findings add to the literature on L2 learners' perception of Mandarin tones, suggesting that, like native Mandarin listeners, English listeners do not discard fine-grained tonal details when recognizing Mandarin words, and providing further evidence that L2 learners have difficulty interpreting the phonetic variability of lexical tones (e.g., Chang & Bowles, 2015; Liu & Zhang, 2016; Shen & Froud, 2016, 2018; Wang et al., 1999, 2003). Further research should be conducted with near-native Mandarin L2 learners to determine whether English listeners interpret tonal variability differently when they reach very advanced levels of proficiency in Mandarin.

Like with Mandarin listeners, the current results also raise the question of *why* English listeners, as a group, showed an effect of within-category tonal information for only one of the two target tones tested—in this case, T1 (though see Footnote 12 for how the effects of the Distant and Close conditions were more likely to inhibit word recognition with increased Mandarin proficiency). One possibility is that T1, a level tone, might be easier to learn than T2 because of its stable pitch over time. Both training and acquisition studies have shown that non-native Mandarin speakers produce and perceive T1 with the highest accuracy among the four Mandarin tones (e.g., Hao, 2012; Wang et al., 1999). If T1 is easier to learn, L2 learners may have a greater sensitivity to the variability of it than to that of other tones, making it easier for them to detect a change in early pitch height and corresponding tone shape compared to T2.

Alternatively, a slight change in the pitch onset of the level tone may be more likely to be perceived as a pitch height difference than the same change in the pitch onset of the rising tone. English listeners have been shown to use pitch height differences rather than pitch contour differences to perceive Mandarin tones (e.g., Bent, 2005; Chandrasekaran et al., 2010; Francis, Ciocca, Ma, & Fenn, 2008; Gandour, 1983; Guion & Pederson, 2007; Qin & Jongman, 2016; Qin & Mok, 2013). These findings have been attributed to the functional relevance of pitch height in English, for example in the realization of English stress (e.g., Beckman, 1986; Lieberman, 1960; Shen, 1989; White, 1981; see also Qin, Chien, & Tremblay, 2017). Because all three acoustic realizations of T1 in the present study are sufficiently flat to be categorized as T1 by native Mandarin raters (see Section 2.3), it is possible that the slight change in the onset of these tones was perceived as a pitch-height difference by the L2 learners, thus enabling them to use this information in the recognition of Mandarin words.

To the best of our knowledge, this study is the first to examine whether (and if so, *how*) native Mandarin listeners and L2 learners of Mandarin show different time courses of lexical activation in their use of the within-category tonal information. The present findings suggest that native and non-native Mandarin listeners adopt *qualitatively* different processing routines that likely reflect how phonetically detailed their representations of lexical tones are, with L2 learners' more limited exposure to tonal variability (compared to native Mandarin listeners) possibly contributing to their non-target-like use of tonal information in the recognition of Mandarin words (in the Distant condition). On the basis of these findings, we propose that L2 learners' coarser tonal representations may help explain the difficulties that the L2 learners encounter in the use of tonal information in spoken word recognition.

In closing, it is necessary to acknowledge some limitations of the present study. First, the current findings may be limited in their generalizability given the restricted sample of items that could be used due to constraints placed on the stimuli (i.e., approximant-initial syllables, imageable monosyllabic nouns). For example, it is possible that our results do not generalize to obstruent-initial words, for tone bias reasons (e.g., Weiner & Ito, 2015) and/or due to the later onset of the tone in the word (i.e., which may lead to tone anticipation effects). Further investigation of listeners' use of within-category fine-grained tonal information will be necessary to determine whether the current results generalize to other types of Mandarin words. Second, although GCA is a powerful tool to analyze visual-world eye-tracking data, the autocorrelation of time-series data (i.e., the correlation between data collected from the same individual at different points in time) can lead to overconfidence in the values estimated by the model. We attempted to mitigate this issue by lowering the sampling rate to 50 Hz (for details on Generative Additive Mixed Models as an alternative method for analyzing time-series data, see Nixon, van Rij, Mok, Baayen, & Chen, 2016). We recognize that given this limitation, the results should be interpreted conservatively.

5.0 Conclusion

The present study used a visual-world eye-tracking paradigm to investigate whether native Mandarin listeners show sensitivity to within-category tonal information, and whether native Mandarin listeners and English-speaking L2 learners of Mandarin differ from each other in their use of this information the recognition of Mandarin words. The results showed that native Mandarin listeners' target-over-competitor word activation was enhanced in the Distant condition and inhibited in the Close condition (compared to the Standard condition) only when the target word contained T2; by contrast, English listeners' target-over-competitor word activation was inhibited in both the Distant and Close conditions (compared to the Standard condition) as a group when the target word contained T1 and among higher-proficiency L2 learners when the target word contained T2. These findings suggest that native and non-native listeners, who likely differ in the robustness of their representations of lexical tones, may have different abilities to process tonal variability relative to prototypical tonal categories. Further research should investigate whether near-native L2 learners of Mandarin pattern like native Mandarin listeners in their use of within-category tonal information in spoken word recognition.

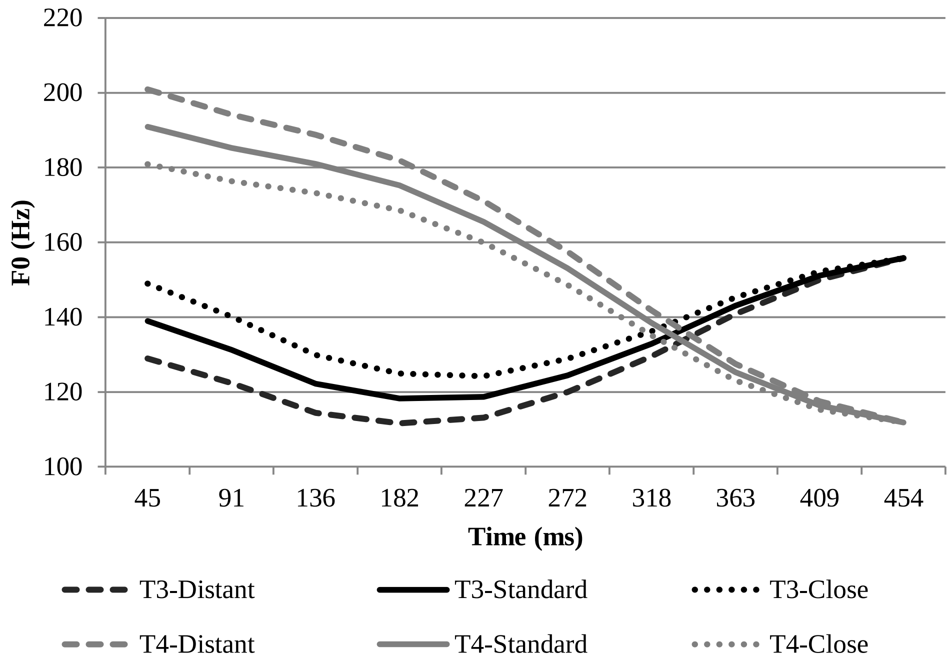
Acknowledgements

The work was supported by the National Science Foundation Doctoral Dissertation Improvement Grant (BCS-1627554) and KU Doctoral Research Fund. Many thanks to Dr. Xiaolin Zhou at Peking University, China, for providing us with lab access; Drs. Allard Jongman, Joan Sereno, Yan Li, James Magnuson, Stephen Politzer-Ahles, and the student members of LING 850 at the University of Kansas for their helpful comments on our project; three anonymous reviewers for their valuable comments on our manuscript; and all of the participants in our study.

Appendix A: Target, Competitor, and Distracter Words

	Target or Competitor		Distracter 1	Distracter 2
T1-T2 Targets in Experimental Trials	/jā/鸭 ‘duck’	/já/牙 ‘tooth’	/tɕǐŋ/井 ‘well’	/tɕìŋ/ 镜 ‘mirror’
	/jān/烟 ‘cigarette’	/ján/岩 ‘rock’	/tǎu/岛 ‘island’	/tàu/稻 ‘rice’
	/jāŋ/秧 ‘sprout’	/jáŋ/羊 ‘sheep’	/fěi/匪 ‘bandit’	/fèi/肺 ‘lung’
	/jī/衣 ‘clothes’	/jí/姨 ‘aunt’	/ʂǔ/鼠 ‘rat’	/ʂù/树 ‘tree’
	/jīŋ/婴 ‘infant’	/jíŋ/蝇 ‘fly’	/wěi/尾 ‘tail’	/wèi/胃 ‘stomach’
	/ɥān/鸳 ‘Mandarin duck’	/ɥán/圆 ‘circle’	/tʰǔ/土 ‘soil’	/tʰù/兔 ‘rabbit’
T1-T2 targets in Filler Trials	/tʂʰā/叉 ‘fork’	/tʂʰá/茶 ‘tea’	/ɕɥě/雪 ‘snow’	/ɕjàu/ 笑 ‘smile’
	/tʂʰwāŋ/窗 ‘window’	/tʂʰwáŋ/床 ‘bed’	/mǎ/马 ‘horse’	/mài/麦 ‘wheat’
	/ʂī/ 狮 ‘lion’	/ʂí /食 ‘food’	/xwǒ/火 ‘fire’	/xàu/号 ‘trumpet’
	/ɕjōŋ/胸 ‘chest’	/ɕjóŋ/熊 ‘bear’	/tɕjǎn/剪 ‘scissor’	/mjàn/面 ‘noodle’
	/tɕʰjāŋ/枪 ‘gun’	/tɕʰjáŋ/墙 ‘well’	/kǔ/鼓 ‘drum’	/kʰù/裤 ‘pants’
	/tɕʰī/妻 ‘wife’	/tɕʰí/旗 ‘flag’	/wǎn/碗 ‘bowl’	/tswàn/钻 ‘drill’

Appendix B: Tonal Continua of Tone 3 and Tone 4 Used in the Filler Trials



Appendix C: Listeners' Proportions of Target, Competitor, and Distracter Fixations

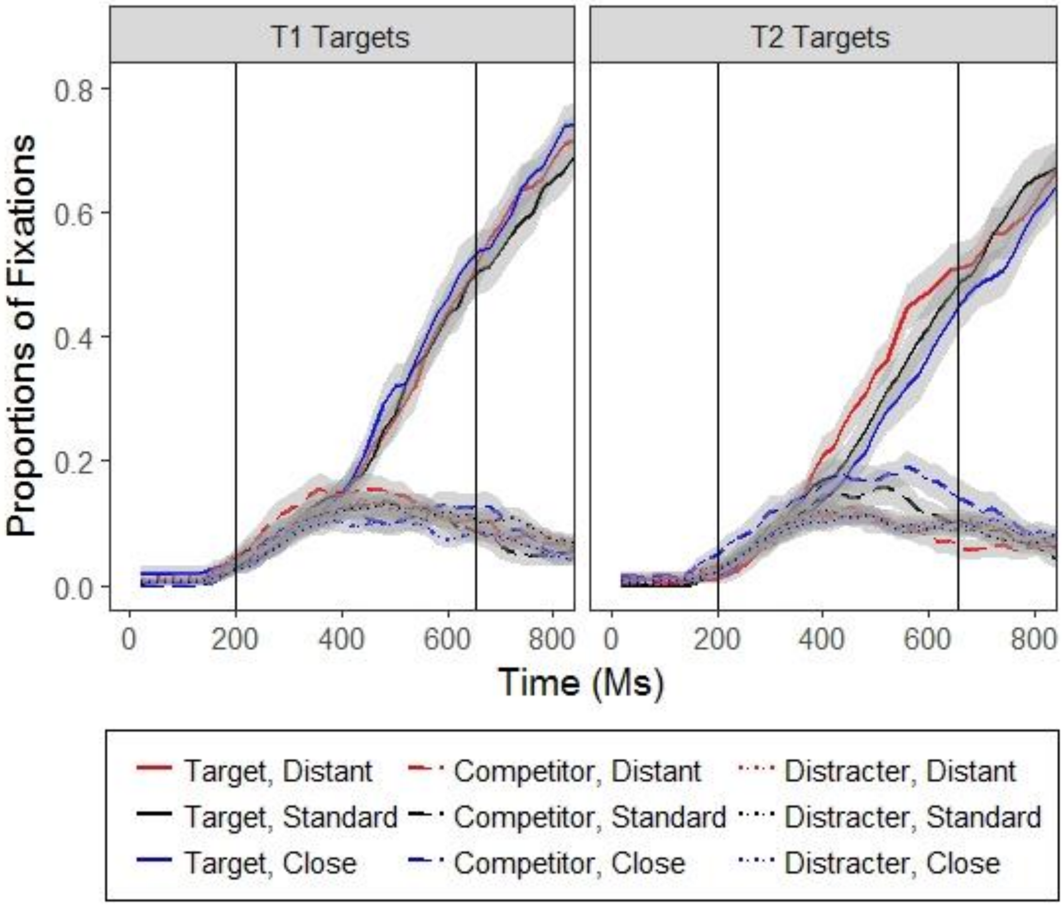


Figure C1. Mandarin listeners' proportions of target, competitor, and distracter fixations in the Distant, Standard, and Close conditions for T1 Targets and T2 Targets; the shaded area represents one standard error above and below the participant mean; the vertical lines represent the onset and offset of the target word (and tone) with a 200-ms delay (time window of analysis)

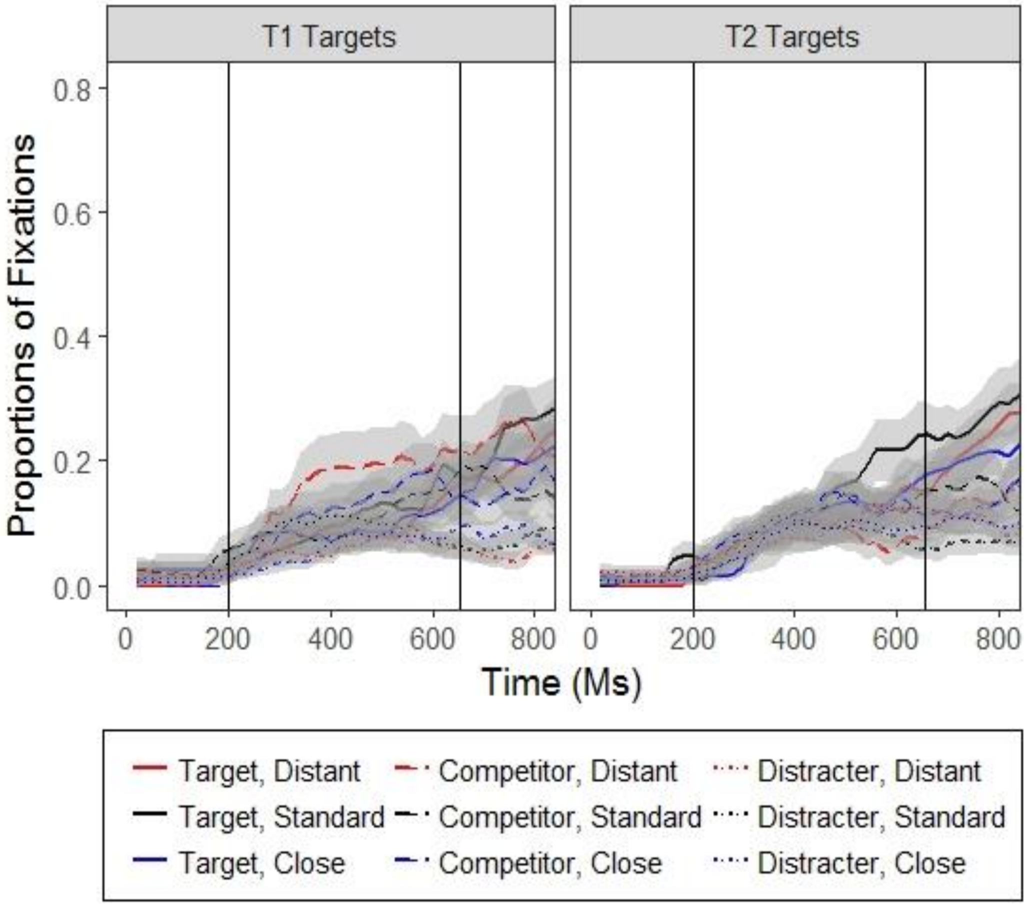


Figure C2. English listeners' proportions of target, competitor, and distracter fixations in the Distant, Standard, Close conditions for T1 Targets and T2 Targets; the shaded area represents one standard error above and below the participant mean; the vertical lines represent the onset and offset of the target word (and tone) with a 200-ms delay (time window of analysis)

Appendix D: Results of Growth Curve Analyses

Table D1 ***Growth Curve Analysis on the Difference between Mandarin Listeners' Transformed Proportions of Target and Competitor Fixations***

Effect	**Estimate**	***t***	***p***
(intercept)	0.219	6.497	< .001
Time			
Linear	1.185	8.629	< .001
Quadratic	0.433	7.977	< .001
Condition (Distant)	–0.043	–2.186	.028
Condition (Close)	0.034	1.702	.089
Tone	–0.014	–0.731	.46
Time × Condition (Distant)			
Linear	0.038	0.041	.688
Time × Condition (Close)			
Linear	0.097	1.024	.306
Time × Tone			
Linear	–0.178	–1.877	.06
Tone × Condition (Distant)	0.152	5.452	< .001
Tone × Condition (Close)	–0.166	–5.944	< .001
Time × Tone × Condition (Distant)			
Linear	0.479	3.575	< .001
Time × Tone × Condition (Close)			
Linear	–0.298	–2.221	.026

Note. $\alpha = .05$; $n = 4968$ observations

Table D2 ***Growth Curve Analysis on the Difference between English Listeners' Transformed Proportions of Target and Competitor Fixations***

Effect	**Estimate**	***t***	***p***
(intercept)	0.025	0.777	.44
Time			
Linear	–0.243	–2.030	.05
Condition (Distant)	–0.159	–5.642	< .001
Condition (Close)	–0.059	–2.078	< .038
Tone	0.061	2.161	.03
Time × Tone			
Linear	0.382	4.901	<.001
Tone × Condition (Distant)	0.128	3.204	<.01
Tone × Condition (Close)	–0.024	–0.605	.545

Note. $\alpha = .05$; $n = 3864$ observations

Table D3 ***Growth Curve Analysis on the Difference between Mandarin and English Listeners' Transformed Proportions of Target and Competitor Fixations for Trials with T1 Targets***

Effect	**Estimate**	***t***	***p***
(intercept)	0.229	5.524	< .001
Time			
Linear	1.207	8.365	< .001
Quadratic	0.483	6.381	< .001
Condition (Distant)	–0.043	–2.092	.036
Condition (Close)	0.034	1.630	.01
L1	–0.228	–3.634	< .001
Time × Condition (Distant)			
Linear	0.038	0.384	.70
Time × Condition (Close)			
Linear	0.097	0.980	.33
Time × L1			
Linear	–1.01	–4.616	< .001
Quadratic	–0.347	–3.039	< .001
Condition (Distant) × L1	–0.115	–3.707	.31
Condition (Close) × L1	–0.092	–2.953	< .01
Time × Condition (Distant) × L1			
Linear	–0.407	–2.722	< .01
Time × Condition (Close) × L1			
Linear	–0.408	–2.722	< .01

Note. α = .05; *n* = 4416 observations

Table D4 ***Growth Curve Analysis on the Difference between Mandarin and English Listeners' Transformed Proportions of Target and Competitor Fixations for Trials with T2 Targets***

Effect	**Estimate**	***t***	***p***
(intercept)	0.214	5.940	< .001
Time			
Linear	0.982	6.990	< .001
Quadratic	0.316	5.171	< .001
Condition (Distant)	0.109	5.286	< .001
Condition (Close)	–0.133	–6.413	< .001
L1	–0.153	–2.806	< .01
Time × Condition (Distant)			
Linear	0.517	5.220	< .001
Time × Condition (Close)			
Linear	–0.200	–2.026	.04
Time × L1			
Linear	–0.050	–3.075	< .01
Condition (Distant) × L1	–0.140	–4.498	< .001
Condition (Close) × L1	0.050	1.597	.11
Time × Condition (Distant) × L1			
Linear	–0.515	–3.441	< .001
Time × Condition (Close) × L1			
Linear	0.147	<|1|	.33

Note. $\alpha = .05$; $n = 4416$ observations

References

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

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

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

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

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

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

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

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

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

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

Chao, Y.-R. (1968). *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

Connell K., Hüls S., Martínez-García M.-T., Qin Z., Shin S., Yan H., & Tremblay, A. (2018). English learners' use of segmental and suprasegmental cues to stress in lexical access: An eye-tracking study. *Language Learning*. 68: 635-668. https://doi.org/10.1111/lang.12288

Creel, S. C. (2014). Tipping the scales: auditory cue weighting changes over development. *Journal of Experimental Psychology: Human Perception and Performance, 40*, 1146-1160. https://doi.org/10.1037/a0036057

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

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences, 22*, 680–689.

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

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

Goldinger, S. D. (1991). On the nature of talker variability effects on serial recall of spoken word lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 152–162.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1166–1183.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Goldinger, S. D., Kleider, H. M., & Shelley, E. (1999). The marriage of perception and memory: Creating two-way illusions with words and voices. *Memory & Cognition*, *27,* 328–338.

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

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

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

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

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

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego: Academic Press.

Johnson, K. (2007). Decisions and mechanisms in exemplar based phonology. In M. J. Sole, P. Beddor, & M. Ohla (Eds.), *Experimental approaches to phonology: In honor of John Ohala* (pp. 25–40). Oxford: Oxford University Press.

Jongman, A., Qin, Z., Sereno, J., & Zhang, J. (2017). Just noticeable differences (JNDs) for pitch direction, height, and slope for Mandarin and English listeners. *The Journal of the Acoustical Society of America, 142*, EL163. http://doi.org/10.1121/1.4995526

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

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

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

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference grammar*. Berkeley, CA: University of California Press.

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

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

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

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

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

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

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

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

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

Nixon, J. S., van Rij, J., Mok, P., Baayen, R. H., & Chen, Y. (2016). The temporal dynamics of perceptual uncertainty: eye movement evidence from Cantonese segment and tone perception. *Journal of Memory and Language*, *90*, 103-125.

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

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

Qin, Z., Chien, Y., & Tremblay, A. (2017). Processing of word-level stress by Mandarin-speaking second language learners of English. *Applied Psycholinguistics, 38 (3)*, 541–570. http://doi.org/10.1017/S0142716416000321

Qin. Z., & Jongman, A. (2016). Does second language experience modulate perception of tones in a third language? *Language and Speech, 59*, 318 338. http://doi.org/10.1177/0023830915590191

Qin. Z., & Mok, P. (2013). Discrimination of Cantonese tones by speakers of tone and non–tonal languages. *Kansas Working Papers in Linguistics, 34*, 1–25. https://doi.org/10.17161/KWPL.1808.12864

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

Shen, J., Deutsch, D., & Rayner, K. (2013). On–line perception of Mandarin Tones 2 and 3: Evidence from eye movements. *Journal of the Acoustical Society of America*, *133*, 3016–3029.

Shen, G., & Froud, K. (2016). Categorical perception of lexical tones by English learners of Mandarin Chinese. *The Journal of the Acoustical Society of America*, *140*, 4396–4403.

Shen, G., & Froud, K. (2018). Electrophysiological correlates of categorical perception of lexical tones by English learners of Mandarin Chinese: An ERP study. *Bilingualism: Language and Cognition*, 1–13. https://doi.org/10.1017/S136672891800038X

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

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

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

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

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

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

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

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

Wayland, R., & Li, B. (2008). Effects of two training procedures in cross-language perception of tone. *Journal of Phonetics, 36*, 250–267.

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

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

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

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

Yuan, B. (2009). Non–permanent representational deficit and apparent target–likeness in second language: Evidence from wh–words used as universal quantifiers in English and Japanese speakers' L2 Chinese. In Snape, N., Leung, Y.–K. L., & Sharwood Smith, M. (Eds.). *Representational Deficits in SLA: In honour of Roger Hawkins* (pp. 69–103). Amsterdam: John Benjamins Publishing.

Zhang, C. (2018). Online adjustment of phonetic expectation of lexical tones to accommodate speaker variation: A combined behavioural and ERP study. *Language, Cognition and Neuroscience*, *33*, 175–195.

Zhang, C., & Chen, S. (2016). Toward an integrative model of talker normalization. *Journal of Experimental Psychology: Human Perception and Performance, 42*, 1252–1268.

Zhang, C., Pugh, K. R., Mencl, W. E., Molfese, P. J., Frost, S. J., Magnuson, J. S., Peng, G., & Wang, W. S-Y. (2016). Functionally integrated neural processing of linguistic and talker information: An event-related fMRI and ERP study. *NeuroImage, 124*, 536–549.