The role of tonal information during spoken-word recognition in Chinese: Evidence from a printed-word eye-tracking study

Wei Shen[1, 2], Jukka Hyönä[3], Youxi Wang[1, 2], Meiling Hou[1], and Jing Zhao[1, 2]

[1] Institute of Psychological Sciences, Hangzhou Normal University, 2318 Yuhangtang Road, Hangzhou, China, 311121

[2] Zhejiang Key Laboratory for Research in Assessment of Cognitive Impairments, Hangzhou, China

[3] Department of Psychology, University of Turku, Turku, Finland

Author Note

This research was supported by grants from the National Natural Science Foundation of China (31800918) and the Natural Science Foundation of Zhejiang Province (LQ18C090007) awarded to the first author. Jukka Hyönä was funded by an Academy of Finland grant (#315963). Jing Zhao was funded by the National Natural Science Foundation of China (31771229).

Correspondence should be addressed to Wei Shen, 2318 Yuhangtang Road, Institute of Psychological Sciences, Hangzhou Normal University, Hangzhou, China, 311121, email: shen_wei@yahoo.com; or to Jing Zhao, 2318 Yuhangtang Road, Institute of Psychological Sciences, Hangzhou Normal University, Hangzhou, China, 311121. Email: zhaojing561@126.com.

Abstract

Two experiments were conducted to investigate the extent to which the lexical tone can affect spoken-word recognition in Chinese using a printed word paradigm. Participants were presented with a visual display of four words—namely, a target word (e.g., 象限, *xiang4xian4*, "quadrant"), a tone-consistent phonological competitor (e.g., 相册, *xiang4ce4*, "photo album") or a tone-inconsistent phonological competitor (e.g., 香菜, *xiang1cai4*, "coriander"), and two unrelated distractors. Simultaneously, they were asked to listen to a spoken target word presented in isolation (Experiment 1) or embedded in neutral/predictive sentence contexts (Experiment 2), and then click on the target word on the screen. Results showed significant phonological competitor effects (i.e., the fixation proportion on the phonological competitor was higher than that on the distractors) under both tone conditions. Specifically, a larger phonological competitor effect was observed in the tone-consistent condition than in the tone-inconsistent condition when the spoken-word was presented in isolation and the neutral sentence contexts. This finding suggests a partial role of lexical tone in constraining spoken-word recognition. However, when embedded in a predictive sentence context, the phonological competitor effect was only observed in the tone-consistent condition and absent in the tone-inconsistent condition. This result indicates that the predictive sentence context can strengthen the role of lexical tone.

Keywords: Lexical tone; Spoken-word recognition; Sentence context; Chinese; Eye-tracking

Most spoken-word recognition models agree that with the unfolding of spoken stimuli, more than one word candidate sharing partial phonemic information with the target will be shortly activated for competition (Marslen-Wilson & Tyler, 1980; McClelland & Elman, 1986). For example, when listening to a spoken-word “beaker,” lexical candidates sharing a similar initial syllable (e.g., “beetle,” “beef,” “bee”) will be temporarily activated for competition until the word wins. However, these spoken-word recognition models were all proposed on the basis of the findings of non-tonal alphabetical languages, such as English, Spanish, and Dutch. Chinese significantly differs from these languages in that the lexical items have tones. Mandarin Chinese has four different types of lexical tone—namely, high-level, mid-rising, low-dipping, and falling tones (Huang, Liu, Yang, Zhao, & Zhou, 2018; Malins & Joanisse, 2012). One syllable plus different lexical tones can be used to refer to different morphemes. For example, adding a mid-rising tone to the syllable “ma” forms a morpheme “麻” (“*ma2*” /fiber/), whereas adding a low-dipping tone to “ma” forms another morpheme “马” (“*ma3*” /horse/). Not only the segmental information (i.e., the syllable) but also the tonal information (i.e., the lexical tone type) must be recognized to correctly distinguish a spoken Chinese word. To date, the extent to which lexical tone influences spoken-word recognition is not fully understood. The current study aims to fill this gap by investigating the role of tonal information during spoken-word recognition in Chinese using a printed-word paradigm.

Existing evidence of the role of lexical tone in Chinese spoken-word recognition is contradictory. One view holds that the lexical tone plays a weaker role in constraining spoken-word recognition compared with the segmental information (Cutler & Chen, 1997; Taft & Chen, 1992; Y. Tong, Francis, & Gandour, 2008; Ye & Connine, 1999; Yip, 2001). For example, Cutler and Chen (1997) found in an auditory lexical decision task that Cantonese listeners made more errors on nonwords that were different from the real words in tone than on nonwords mismatching real words in segmental information. Moreover, the tone disadvantage was observed in the same–different task, that is, listeners were faster and more accurate in differentiating words that mismatch in segmental information (vowel or constant) than in tone. Critically, for a group of non-native listeners with no knowledge of Cantonese, the response pattern in the same–different task was also very similar, suggesting that the tonal processing only reflect the perceptual process rather than the linguistic one. The authors argued that tone was accessed later than segmental information because of its later availability; thus, the tones were more likely to be processed incorrectly. Ye and Connine (1999) replicated the tone disadvantage effect in a tone–vowel monitoring task in which participants were asked to judge whether a tone–vowel combination was in the target-bearing stimulus. They found that response times to the stimuli (e.g., *ba4*) that mismatched the target (e.g., *a2*) in tone were longer than those to the stimuli (e.g., *bi2*) that contained a mismatching vowel. Sereno and Lee (2015) also observed a weak role of tone in spoken-word recognition. They found a small but significant facilitation effect when the primes shared the same segmental information but different tones with the targets (e.g., *ru3*-r*u4*). However, no priming effect was observed when the primes and targets overlapped only in tonal information (e.g., *sha4*-*ru4*). These results indicate that tonal information does not strongly constrain word activation and the mismatch in tonal information results in a less harmful effect on the spoken-word perception than that in segmental information. In sum, these results suggest that segmental information weighs stronger than tonal information in spoken-word perception. Moreover, the processing disadvantage of lexical tone is interpreted to suggest relatively late availability of tonal information in the acoustic stream, and less information value is carried by lexical tones (Hu Gao, Ma, & Yao, 2012; Ye & Connine, 1999).

On the contrary, studies using high time-resolution techniques have challenged the privileged role of segmental information and proposed lexical tone to play a strong and comparable role to that of segmental information (Lee, 2007; Malins & Joanisse, 2010; Schirmer, Tang, Penney, Gunter, & Chen, 2005; Zhao, Guo, Zhou, & Shu, 2011). For instance, in a direct priming task, Lee (2007) reported that primes sharing the same segmental structure but different tones with the target (e.g., prime: *lou3*, "hug"—target: *lou2*, "hall") did not facilitate target recognition compared with the controls (e.g., control: *pan1*, "climb"—target: *lou2*, "hall") at two inter-stimulus intervals (ISIs), suggesting that the tones can inhibit the activation of word candidates that share different tones with the targets. However, a significant facilitation effect was found when the prime (e.g., *lou3,* "hug") was related to the target (e.g., *jian4zhu0*, "building") via another non-presented word (e.g., *lou2*, "hall"; which was a semantic associate of the target *jian4zhu0*, "building"). However, the priming effect was merely observed at a short ISI (50 ms) and disappeared at a long ISI (250 ms), suggesting that the lexical tone does not constrain word activation at a very early stage. Recently, in an event-related potentials (ERPs) study, Zhao et al. (2011) investigated the processing of monosyllable words in Mandarin Chinese. They presented a target picture initially following a spoken character of the target picture, which was then followed by a second picture. The spoken character was manipulated to completely match (e.g., *bi2–bi2*), tone mismatch (e.g., *bi2–bi3*), or rime mismatch (e.g., *bi2–bo2*) with the target picture. The task was to judge whether the two pictures belonged to the same semantic category. The N400 peak was used as an index of a semantic violation. Zhao et al. (2011) found that tone and rime mismatches elicited larger N400 effects than the complete match condition. More importantly, the amplitudes of the N400 effects in the tone and rime mismatch conditions were almost identical. Thus, the authors conclude that tonal and segmental information play a comparable role in spoken-word recognition in Chinese (see also Schirmer et al., 2005 for a similar finding).

Malins and Joanisse (2010) conducted an eye-tracking study and found converging evidence. In a visual-world paradigm, participants were presented with a visual display of four pictures—namely, a target, a phonological competitor, and two distractors, while listening to a spoken target word. For each spoken target (e.g., *shu3*, "mouse"), the phonological competitors were manipulated to either share the segmental structure but a different tone (segmental competitor; e.g., *shu1*, "book") or share the onset phonemes and the same tone (cohort competitor; e.g., *shui3*, "water"). They found that both caused delayed fixations to the targets, and the results of the growth curve analysis showed that in the two competitor conditions, the trajectories of fixations to the targets were very similar. Moreover, a comparable competitor effect (i.e., a higher fixation probability on competitors than on unrelated distractors) was observed for the two conditions. These findings suggest that tonal and segmental information are accessed simultaneously and play comparable roles in spoken-word recognition in Mandarin.

Some studies also found that contextual information can modulate the relative importance of tone and vowels in spoken-word perception; yet, the findings are contradictory. Ye and Connine (1999) used a monitoring task to examine the effect of context on tone processing in single-syllable recognition, neutral context, and constraining idiomatic context (i.e., idioms). They found that vowel detection is faster than the tone detection in single-syllable recognition and the neutral context. However, the vowel advantage disappeared in the constraining idiomatic context, with tones being detected faster than vowels. Idioms that provide a highly constraining context are believed to pre-activate the lexical tone, thereby leading to the tone being processed earlier than the vowel. In a similar monitoring task, Liu and Samuel (2007) extended Ye and Connine's (1999) observation of idioms to normal sentences. However, in a lexical decision task, no accuracy difference was found between tone and vowel processing in isolated word recognition or idioms. Moreover, in sentences, vowels were processed more efficiently than tones. Liu and Samuel (2007) concluded that the effect of the context on tone processing is task dependent. Such an effect is observed in tasks tapping into sublexical processing (i.e., the monitoring task), but not in those requiring lexical access (i.e., the lexical decision task).

Recent findings from ERP studies make the effect of the context even more elusive. Hu et al. (2012) instructed their participants to identify whether the last character in four-character idioms is correct. Critically, the last character was manipulated to a mismatch in tone, vowel, or both. The results showed that the vowel mismatch elicited an earlier negative component and a larger N400 effect than the tone mismatch. This finding suggests that the role of tone is weaker than that of vowels even in constraining idiomatic contexts (Huang et al., 2018).

Taken together, the aforementioned studies report contradictory findings concerning the role of lexical tone in spoken-word recognition (i.e., a weak vs. a strong/comparable role). In other words, it is not clear to what extent lexical tone can constrain word recognition. Models of spoken-word recognition (e.g., TRACE, of McClelland & Elman, 1986) assume that the "global goodness of fit" between the acoustic input and the word candidates in the mental lexicon plays an important role in spoken-word recognition. The degree of activation of lexical representations is assumed to be determined by the phonological similarity between the acoustic information. Based on this assumption, the lexical tone in tonal languages may be accessed similarly as segmental information, and tonal similarity affects the degree of word activation. As reviewed above, Ye and Connine (1999) reported that the tone of low similarity to the target tone has longer detection times in a monitoring task compared with that of high similarity. The authors argue that tonal information is accessed in a graded manner in the mental lexicon so that the activation of low-similarity tones is reduced compared with high-similarity tones. One concern with the study is that the explicit tone detection task could have resulted in the use of task-specific response strategies so that the tonal information may be intentionally retrieved from memory (Schirmer et al., 2005). Thus, the finding does not necessarily reveal the mechanism underlying online tone processing. In addition, the study does not shed light on how tonal similarity constrains the word activation in spoken-word perception. In the current study, we aim to further investigate this issue using the eye-tracking technique combined with a printed-word visual-world paradigm.

The present study also aims to investigate whether the role of lexical tone will be modulated by constraining sentence context. As reviewed above, evidence shows that in idioms, tonal information is prioritized over segmental information; however, the findings are inconsistent across different studies (Hu et al., 2012; Huang et al., 2018; Liu & Samuel, 2007; Ye & Connine, 1999). In addition, in natural language comprehension, word recognition occurs more often in sentences than in idioms. Thus, it is of great importance to investigate how tones are accessed in sentence context. As reviewed above, Liu and Samuel (2007) observed tone and vowel processing in the sentence context to be task-dependent. The domination of tone processing in sentences was found in tasks tapping into sub-lexical processing (e.g., monitoring tasks) but not in those requiring lexical access (e.g., lexical decision). Notably, the target word predictability in the sentences used by Liu and Samuel (2007) was relatively low (48%). Thus, relatively low-constraint sentences are probably less likely to preactivate tonal information resulting in a reduced role of tone. Moreover, in that study, participants were asked to perform a nonword detection task (i.e., judging whether a nonword exists) rather than a normal comprehension task. Taken together, whether lexical tone affects spoken-word recognition in the sentence context is unclear. In Experiment 2, we used high-predictable sentences as materials and adopted a paradigm that closely mimics spoken-word recognition taking place in natural language environments to shed light on this issue.

Thus, two visual-world eye-tracking experiments were designed and conducted to examine the role of tonal similarity in spoken-word recognition and whether contextual predictability affects the role of lexical tone in Mandarin Chinese. As a follow-up of Shen, Qu, and Tong (2018), the present study employed a printed-word version of the visual-world paradigm (Huettig, Rommers, & Meyer, 2011; McQueen & Viebahn, 2007). In this paradigm, eye movements of participants on visual words were continuously recorded and time locked to the unfolding of the spoken target words. The paradigm can reveal underlying cognitive mechanisms behind spoken-word recognition in a rather ecological way, as introducing any secondary tasks (e.g., lexical decision) is unnecessary. The printed-word version of the paradigm has a few advantages over the classic picture version. First, unlike the picture version in which all the stimuli are depictions of objects, the printed-word version enlarges the range of material selection (McQueen & Viebahn, 2007). Second, different from the picture version of the paradigm, a naming task is unnecessary for the printed-word version to assess the naming consistency before the eye-tracking experiment. The naming task would pre-activate the name of the stimuli which may increase the activation level of the stimuli in the subsequent eye-tracking experiment (Wang, Wang, & Malins, 2017). In addition, several studies have consistently reported reliable findings regarding phonological processing in spoken-word recognition using the printed-word version (McQueen & Viebahn, 2007; Salverda & Tanenhaus, 2010; Shen, Qu, & Li, 2016).

In Experiment 1, we investigated the role of tonal similarity in spoken-word recognition. Participants were presented with a printed-word display including a target word, a phonological competitor, and two distractors, when they simultaneously heard a spoken target word. Critically, the tonal similarity was established by manipulating the lexical tones of the phonological competitors to be either consistent or inconsistent with the targets. Moreover, the segmental information (i.e., phonemes) was matched to be identical between the targets and the competitors to maximize the possibility of observing an effect of lexical tone. If the lexical tone does not play any role in spoken-word recognition, then we expected to observe comparable phonological competitor effects in the tone-consistent and the tone-inconsistent conditions (Shen et al., 2018). Furthermore, if tone plays a strong constraining role in spoken-word recognition (Lee, 2007), then the phonological competitor effect will only be found in the tone-consistent condition and will be absent in the tone-inconsistent condition. Last, if the lexical tone is accessed as a function of tonal similarity in spoken-word recognition, then the competitor effect will be observed in both conditions, but with a large effect in the tone-consistent condition compared with the tone-inconsistent condition. These hypotheses were tested in Experiment 1. Experiment 2 aimed to examine whether the contextual predictability can modulate the role of tone in spoken-word recognition. The design of Experiment 2 was identical to that of Experiment 1 except that the spoken target words were embedded in a neutral or a predictive sentence context. Based on the findings of Ye and Connie (1999), we expected that tonal information would be pre-activated in a predictive sentence context. Furthermore, if the lexical tone can be predicted on-line, then we presumed that the phonological competitor effect should solely be observed in the tone-consistent condition and not in the tone-inconsistent one. This hypothesis was tested in Experiment 2.

Experiment 1

“This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0”.

Method

Participants

A total of 40 participants (25 women and 15 men) aged 19–28 years (mean = 21.18 years) were recruited from a university in Hangzhou, China. They were all Mandarin Chinese speakers[1] and had normal or corrected to normal vision. All participants were given a small fee for their participation. The research protocol reported here was approved by the Ethics Committee of the university.

Materials and design

Of the Chinese disyllable words, 52 were selected as the target stimuli. The target characters were embedded in disyllable words to avoid possible ambiguity as much as due to the use of homophones. Each visual display was composed of four printed words—namely, a target word, a phonological competitor, and two unrelated distractors. For each target word, two phonological competitors were constructed: tone-consistent and tone-inconsistent conditions. The phonological manipulation was only conducted for the first character. In both conditions, the target words shared the same initial syllable with the phonological competitors. In the tone-consistent condition, the target word shared the same tone with the phonological competitor (e.g., target word: 相册 *xiang4ce4*, “photo album”—phonological competitor: 象限 *xiang4xian4*, “quadrant”). In the tone-inconsistent condition, the target word had a different tone with the phonological competitor (e.g., target word: *xiang4ce4*, “photo album”—phonological competitor: 香菜 *xiang1cai4*, “coriander”, see Fig. 1 for the example stimuli in the two conditions). Word frequency and number of strokes were carefully matched across the four printed-word conditions (one-way ANOVAs, *F*s < 1, *p*s > .55; see Table 1).

[1]Some participants speak other Chinese dialects in addition to Mandarin. However, all participants were educated in Mandarin Chinese in early childhood; thus, they were all fluent in listening and speaking Mandarin Chinese.

Sixteen participants were recruited and instructed to judge whether the target words were semantically related to the phonological competitor with 0 (no semantic association) and 1 (with the semantic association). The items scored with 1 were modified and reassessed to ensure that no semantic association exists. A research assistant checked to make sure that the items did not share the same radical to ensure that no orthographic similarity exists between the target word and the phonological competitor.[2] The materials were divided into two lists. Each list included half of the trials in the tone-consistent condition and the other half of the trials in the tone-inconsistent one. Each item was presented only in one condition in each list. Participants were randomly assigned to one of the two lists.

In addition, 52 filler trials were conducted to avoid participants becoming aware of the phonological manipulation of the critical trials. The filler trials were similar to the critical trials, except that no phonological overlap was manipulated between the spoken targets and the visual words.

Apparatus

Eye movements were recorded using an EyeLink1000 Desktop tracker (SR Research Ltd, Ontario, Canada) sampling at 1000 Hz. All the spoken stimuli were presented to the participants through headphones (Sennheiser, PC 230). The visual stimuli were presented on a 21-inch monitor (resolution: 1,024 × 768, refresh rate: 85 Hz). The monitor was located 58 cm away from the participants; 1 cm on the screen subtended a visual angle of approximately 1° at this viewing distance. Although viewing was binocular, only the eye movements of the right eye were recorded and collected.

Procedure

The calibration and validation of the eye tracker were conducted at the beginning of the experiment.

[2]One item was found to share orthographic similarity (i.e., a radical). However, the results did not change after excluding the item. In the Results section, we report the results without this item.

A drift check was performed at the beginning of each trial. Participants were instructed to listen to the spoken instruction, and then, on the computer screen, to click the word corresponding to what they heard through the headphones. Each trial began with a blank screen for 500 ms. After that, a visual display of four printed words was presented on the screen for 200 ms before the presentation of the spoken target word. The visual display disappeared, and the trial ended after the participant responded. The positions of the four printed words were randomized across trials. Each participant performed seven practice trials followed by 52 critical trials and 52 filler trials, which were intermixed and presented in random order. The whole experiment lasted approximately 10–15 minutes.

Results and discussion

Accuracy data

The accuracy of the mouse click was 99.9%, suggesting that participants were paying close attention to the task.

Data coding

Trials with a wrong response (i.e., clicking on a nontarget word) were excluded from the analysis (0.1%). A fixation was defined as being on the printed word if it fell within a square of 8° × 8° centered around the printed word. Fixations that fell outside the words were regarded as fixating on the background of the visual display. Fixations on the words were coded as 0 (not fixed) or 1 (fixed) for every 100-ms bin starting from 200 ms before the onset of the spoken target word.

Eye-movement data

All eye movement data were analyzed using the logit mixed model (Ferreira, Foucart, & Engelhardt, 2013; Jaeger, 2008; Quené & van den Bergh, 2008) in R (Version 3.5.0; R Core Team, 2018). The glmer() function in the lme4 package (Version 1.1-17; Bate, Maechler, Bolker, & Walker,

2015) was used to create the mixed effects model. As the first step, a base model was created with a random intercept for participants and items. Then, the fixed factor "tone consistency" (tone-consistent, tone-inconsistent) and "competitor type" (phonological competitor, distractor) and the random slopes for the fixed factor were gradually added into the model (Barr, Levy, Scheepers, & Tily, 2013; Cunnings, 2012). The anova() function was used for model comparison and to test whether the inclusion of a factor improved the data fit.

Figure 2 shows that fixations on the phonological competitors were more than those on distractors for approximately 400 ms after the onset of the spoken target words for both conditions. The point of divergence between phonological competitors and targets was close to the offset of the first character of the spoken targets (the average duration of the first character was 346 ms), suggesting immediate access of tonal information once available. The logit mixed model revealed a significant interaction between "tone consistency" and "competitor type" from 500 to 800 ms (χ^2s > 46.52, *p*s < .001). For the time window of 500–600 ms, the contrast analysis showed that fixations on phonological competitors were more than those on distractors in the tone-consistent (phonological competitor: M = 0.35; distractor: M = 0.13; b = 1.35, SE = 0.11, Z = 12.25, p < .001) and the tone-inconsistent (phonological competitor: M = 0.24; distractor: M = 0.17; b = 0.44, SE = 0.11, Z = 4.17, p < .001) conditions. More importantly, more fixations were made to the phonological competitor in the tone-consistent condition than that in the tone-inconsistent condition (b = 0.57, SE = 0.09, Z = 5.78, p < .001). We calculated the coefficient determination, R^2, which is used as a measure of effect size, to further confirm the interaction effect (Nakagawa & Schielzeth, 2013). The effect size of the phonological competitor effect in the tone-consistent condition (R^2 = 0.11) was larger than that in the tone-inconsistent one (R^2 = 0.04). Unexpectedly, fixations on the distractors in the tone-consistent condition were found to be fewer than that in the tone-inconsistent condition ($b = -0.32$, $SE = 0.08$, $Z = -3.86$, $p < .001$). The two distractors were identical in both conditions. Hence, the difference was not due to the experimental materials, but likely because the fixations to visual referents in the visual-world paradigm are not independent. Thus, additional fixations to one visual referent (e.g., phonological competitor) result in less fixations to other referents (e.g., distractors) in the same time window.

The results of Experiment 1 showed a significant phonological competitor effect not only in the tone-consistent condition but also in the tone-inconsistent one. These findings stand against the view of the strong role of lexical tone which would predict no effect in the tone-inconsistent condition (Lee, 2007; Schirmer et al., 2005; Zhao et al., 2011). Moreover, the tone-consistent condition has a larger phonological competitor effect than the tone-inconsistent one. This finding further demonstrates that the lexical candidates sharing the same or different tones with the target were all activated temporarily during the spoken-word perception, but with different weights in terms of tonal similarity. Taken together, the findings of Experiment 1 revealed that tonal information cannot restrict the number of word candidates such that the activation of words with different tones from the targets cannot be inhibited. This result suggests that the lexical tone plays a partial role in affecting the degree of word activation based on tonal similarity. Words with great tonal similarity (i.e., tone-consistent competitors) with targets were activated to a higher degree than those with low tonal similarity (i.e., tone-inconsistent competitors). This point is further discussed in detail in the General Discussion section.

Experiment 2

Experiment 2 aimed to examine the following: (1) whether the partial role of lexical tone found in isolated word recognition in Experiment 1 can be generalized to sentence comprehension and (2) whether a predictive sentence context can influence the role of lexical tone in spoken language

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

comprehension. More specifically, we investigated the extent to which a sentence context can predict the lexical tone of the target word. In Experiment 2, spoken target words were embedded in neutral or predictive sentence contexts rather than presented in isolation. The lexical tone in the neutral sentence context would mimic its role observed in isolated word recognition. Thus, similar phonological competitor effects as observed in Experiment 1 were expected in the neutral context condition. Furthermore, if the tonal information can be predicted by the constraining context, then the phonological competitor effect will only be observed in the tone-consistent but not in tone-inconsistent condition.

Method

Participants

A total of 40 participants (27 women and 13 men) aged 18–25 years (mean age = 19.93 years) were recruited from the same participant pool as in Experiment 1. None of them participated in Experiment 1. All participants were native Mandarin speakers and had normal or corrected-to-normal vision. All were given a small fee for their participation.

Material and design

The experimental design was a 2 (context type: neutral sentence context, predictive sentence context) × 2 (tone consistency: tone-consistent, tone-inconsistent) × 2 (competitor type: phonological competitor, distractor) within-participants design.

Forty-eight target words were selected as the critical stimuli (20 of them were from Experiment 1). Each target word was paired with tone-consistent and tone-inconsistent phonological competitors. As in Experiment 1, each visual display consisted of a target word, a phonological competitor, and two distractors. Word frequency and number of strokes were carefully matched across the four conditions (one-way ANOVAs; word frequency: *F*s < 1, *p*s > .95; number of strokes: *F*s < 1.79, *p*s > .15; see Table 2).

The target words were embedded in a neutral and predictive sentence frame. The predictability of target words was evaluated via a cloze test (Ito, 2019). A total of 40 (none of them took part in the eye-tracking experiment) were provided with the sentence frame up to the target word (not including the target word). They were instructed to write down the first word that came to their mind. The predictability score in the neutral sentence context and predictive sentence context was 0% and 85% (*SD* = 14%), respectively (see Table 3 for an example). In addition, the total of filler sentences with no word being predictive from the sentence context was 48.

All critical materials were divided into four lists according to a Latin-square design. Participants were randomly assigned to one list with no item repetition.

Apparatus

The apparatus was identical to that in Experiment 1.

Procedure

The calibration and validation were identical to that in Experiment 1. Following the procedure of Ito, Pickering, and Corley (2018), the printed words were previewed for 1000 ms before the spoken target word. Using a computer mouse, participants were instructed to click on the printed word that corresponded to what they heard in the sentence. The printed words remained on the screen until the spoken sentence was completed. Each participant performed 10 practice trials followed by 48 critical trials and 48 filler trials that were intermixed and presented in random order. The entire experiment lasted approximately 20–25 minutes.

Results and discussion

Accuracy data

The accuracy of the mouse click was 99.8%, indicating that the participants paid close attention to the task.

Data coding

The data coding procedure was identical to that in Experiment 1.

Eye-movement data

The eye-movement data were analyzed using the same statistical model (i.e., logit mixed model) as in Experiment 1. The results were reported separately for neutral and predictive context conditions.

Neutral context condition

In Fig 3, the proportions of fixations are plotted for the target word, the phonological competitor, and the distractors for the tone-consistent and tone-inconsistent conditions. Figure 3 shows that the proportion of fixations on the phonological competitor started to diverge from the distractors from approximately 300 ms after the onset of the spoken target word. The divergence point aligned approximately with the offset of the first character of the target words (the average duration of the first character was roughly 267 ms), indicating that the tonal information was used to constrain the spoken recognition once available. The results of the logit mixed model showed a significant interaction between “tone consistency” and “competitor type” from 300 to 600 ms after the onset of the spoken target word (χ^2s > 10.07, *p*s < .006; see Fig. 5, for the interaction effect). For the time window of 500–600 ms, the contrast analysis showed that fixations on the phonological competitors were more than those on the distractors in the tone-consistent condition (phonological competitor: *M* = 0.27; distractor: *M* = 0.09; *b* = 1.32, *SE* = 0.18, *Z* = 7.33, *p* < .001) and in the tone-inconsistent condition (phonological competitor: *M* = 0.14; distractor: *M* = 0.11; *b* = 0.34, *SE* = 0.19, *Z* = 1.78, *p* = .07). More importantly, more fixations were made to the phonological competitor in the tone-consistent than those in the tone-inconsistent condition (*b* = 0.86, *SE* = 0.16, *Z* = 5.36, *p* < .001). Furthermore, the proportion of fixations on the distractors was lower in the tone-consistent than in the tone-inconsistent condition (*Z* = −0.90, *p* = .37).

Predictive context condition

Figure 4 depicts the fixation proportions on the target word, the phonological competitor, and the distractors in the tone-consistent and tone-inconsistent conditions. As shown in Fig. 4, the fixation curve of the target word starts to separate from the phonological competitor and the distractors even before the onset of the spoken target word. This result suggests that participants predicted the spoken target words in advance in the predictive context condition, demonstrating that our manipulation of context predictability was successful. More importantly, the phonological competitor started to diverge from the distractors at approximately 300 ms before the onset of the spoken target word. Thus, we analyzed whether there were any effects at each 100-ms time window from 300 ms before the onset of the target word to the onset of the spoken target word (i.e., −300–0 ms).

The results of the logit mixed model showed a significant interaction between “tone consistency” and “competitor type” for the time window of 200–300 ms before the onset of spoken target word (χ^2 = 5.53, *p* = .06; see Fig. 5, for the interaction effect). A further contrast analysis showed that more fixations were made to the phonological competitor (*M* = 0.23) than to the distractors (*M* = 0.19) in the tone-consistent condition (*b* = 0.29, *SE* = 0.15, *Z* = 1.89, *p* = .058). By contrast, the difference disappeared in the tone-inconsistent condition (phonological competitor: *M* =

0.18; distractor: $M = 0.21$; $Z = -1.23$, $p = .22$). In addition, the proportion of fixations on the phonological competitor was higher in the tone-consistent than in the tone-inconsistent one ($b = 0.30$, $SE = 0.16$, $Z = 1.94$, $p = .052$); no difference was found on the distractors ($Z = -1.36$, $p = .17$). Last, no interactions were found in other time windows (χ^2s < 3.96, *p*s > .14).

In summary, Experiment 2 has two main findings. First, the results of the neutral context condition replicated the partial role of lexical tone observed in the isolated word recognition in Experiment 1 by demonstrating that tonal similarity affects the degree of word activation. That is, the tone-consistent condition has a larger phonological competitor effect than the tone-inconsistent one. Second, the marginally significant result of the predicable context condition provides some evidence that a constraining sentence can influence the role of lexical tone. To be precise, a predictive context can strengthen the role of lexical tone, different from the case with a neutral context or without contextual information. In predictive contexts, the phonological information of spoken target words can be predicted from the prior sentence, with the predicted information including segmental and tonal information. Furthermore, the phonological representation of the spoken targets is vague and underspecified in the neutral context condition, considering that a neutral context cannot provide efficient clues on the target identity. This point is discussed further in the General Discussion section.

General discussion

The present study conducted two experiments to investigate the role of lexical tone in the recognition of Chinese spoken-words in the isolation (Experiment 1) and sentence contexts (Experiment 2). The printed-word version of the visual-world paradigm was employed to study these questions. Both experiments observed a significant phonological competitor effect (i.e., phonological

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

competitors attracted more fixations than distractor words). Furthermore, the findings of the current study demonstrated that tonal similarity can influence the access of lexical tone. Tones with great similarities were activated to a high degree, whereas tones with low similarity were activated to a small degree. This notion can explain the observation of a large phonological competitor effect in the tone-consistent condition and a small one in the tone-inconsistent condition. Experiment 2 provided further evidence for the role of tonal similarity in recognizing spoken-words in a neutral sentence context. In addition, we found some evidence that the role of the lexical tone could be strengthened by a predictive context compared with the neutral one.

The findings of the current study have two novel contributions to the existing literature. First, the study demonstrates that the lexical tone plays a partial role in constraining spoken-word recognition and is processed in terms of tonal similarity when spoken-words are recognized in isolation and neutral sentence contexts. The results of prior studies regarding the role of lexical tone were contradictory, with some studies proposing a weak role of lexical tone, whereas others obtaining a strong role of lexical tone (Cutler & Chen, 1997; Lee, 2007; Malins & Joanisse, 2010; Taft & Chen, 1992; Ye & Connine, 1999; Zhao et al., 2011). The phonological competitor effect observed in the tone-consistent and tone-inconsistent conditions suggests that phonological representations with all lexical tones can be activated during spoken-word recognition. The observation of a large competitor effect in the tone-consistent condition and a small competitor effect in the tone-inconsistent condition further indicates that the lexical tone determines the degree of word activation as a function of tonal similarity. The tone-consistent competitor shared high tonal similarity with the target so that its lexical representation was activated to a high degree, thus leading to additional attention allocated to its visual referent. Moreover, when the competitor did not share tonal similarity (i.e., an inconsistent lexical tone) with the target, the lexical representation was still activated to a certain extent. However, the degree of activation was low, so that few fixations were made to its visual referent. This pattern of results is also manifested in the spoken sentence comprehension in Experiment 2, when nonpredictable sentence contexts were used. The preceding neutral context did not provide any useful information about the identity of target words; thus, the lexical tone played a similar role in spoken-word recognition as in isolated word recognition examined in Experiment 1.

Second, the present study demonstrates that the role of lexical tone is sensitive to context predictability in that the role of lexical tone is strengthened in a predictive context. Results of Experiment 2 showed that the phonological competitor effect was only observed in the tone-consistent condition and not in the tone-inconsistent condition. We assume that when the preceding context is highly constraining, the contextual information can predict the target word identity; thus, segmental information and lexical tone are preactivated. In this situation, more attention was paid to the phonological competitor in the tone-consistent condition. Moreover, the phonological competitor effect was not observed with an inconsistent tone, for which the phonological representation may be activated to a lesser degree or even be inhibited.

Existing models of spoken-word recognition (TRACE, of McClelland & Elman, 1986; cohort, of Marslen-Wilson & Tyler, 1980; shortlist, of Norris & McQueen, 2008; and neighborhood activation model, of Luce & Pisoni, 1998) are constructed to explain word recognition in non-tonal languages. Thus, they do not incorporate a specific role for lexical tone in their modeling architecture. Hence, they cannot explain how suprasegmental information (i.e., tonal information) influences spoken-word recognition in nontonal languages (e.g., Chinese). Recently, the TTRACE model, was modified based on the TRACE model, as proposed by X. Tong, McBride, and Burnham (2014) to account for

“This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0”.

tone processing in spoken-word recognition in Cantonese. The lexical tone is incorporated in the model as a “toneme” unit, which integrates with phonemes. TTRACE also postulates that the activation of words in the mental lexicon varies as a function of the degree of overlap in terms of segmental and suprasegmental dimensions in spoken-word recognition. The results of the current study provide supporting evidence for this hypothesis. Specifically, the tone-consistent competitor sharing high tonal similarity with the target word resulted in a larger competitor effect compared with that of not sharing tonal similarity with the target. Our results provide further support for this model by demonstrating that the role of lexical tone was observed in not only isolated word recognition but also in the neutral sentence context.

In addition, we observed that a constraining context can strengthen the role of lexical tone. TTRACE proposes three processing levels—namely, the word, phoneme-toneme, and feature levels. However, the impact of the context is not included in the model. We assume that the effect of the context on the activation of lexical tone may be accomplished via top–down feedback from the word level to the phoneme-toneme level. The predictive context may increase the degree of target word activation at the word level, which then leads to excitatory feedback to the phoneme-toneme level. The strongly activated phoneme and toneme units then feed excitatory activation back to the word level so that competitors with the same tonal information are activated to a high degree. In the visual-world paradigm, this event becomes apparent as additional visual attention is allocated to tone-consistent competitors, resulting in a significant competitor effect. Simultaneously, the inhibitory mechanism should pay attention to the phoneme-toneme units inconsistent with the target being inhibited. Hence, the inhibitory bottom-up feedback from the phonemes-toneme level to the word level results in the inhibition of competitors with the different tonal information. This event may

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

explain the absence of a competitor effect in the tone-inconsistent condition in the predictive context condition in Experiment 2.

TTRACE (X. Tong et al., 2014) assumes that tone incorporated as tonemes is integrated with phonemes. For example, a specific tone (e.g., low-rising) may be represented in the mental lexicon integrated with the different segmental information (e.g., "shu3," "ma3," "xi3"). This assumption can explain the concurrent activation and inhibition of competitors observed in Experiment 2 of the present study. However, our study only examined conditions where targets and competitors shared phonemes and tones. Hence, we cannot exclude the possibility that, under certain conditions, the tone may also be represented as a separate node, as assumed in Ye and Connine's (1999) model. This question needs to be further investigated.

Notably, the interaction effect found in the neutral context condition in Experiment 2 occurred 200 ms later than that in the isolated word recognition in Experiment 1. The later effect may be due to interference of the preceding context in the neutral context condition. We made a great effort to ensure that the preceding context contained no homophonic characters with the critical target character. However, some contextual words may overlap with the target character in onset, rime, or tone. This event may have increased to some extent the activation threshold of target words, thereby holding up the execution of fixations to the competitors to delay the occurrence of the interaction effect in the neutral context. This notion may also explain the marginally significant competitor effect obtained in the tone-inconsistent condition in the neutral context.

The current study has the following limitations. The tonal similarity was defined categorically as either identical or different to maximize the likelihood of observing the effects of lexical tone. However, notably, tones resemble one another to a different degree. For instance, the mid-rising tone shares more acoustic similarity with the low-dipping tone than with the other tones (Gandour, 1983; Kiriloff, 1969). Therefore, additional studies need to be conducted to further examine the effect of tonal similarity on spoken-word perception. In addition, in the present study, we only examined the condition where the target words fully overlapped in segmental information with the competitor words. Thus, based on the present results, it is not known whether the role of the lexical tone in spoken-word recognition would be sensitive to the degree of overlap in segmental information. Specifically, whether lexical tone would constrain spoken-word recognition when competitors share partial segmental information (i.e., onsets or rimes) or no segmental information at all (i.e., tone only) with targets is unclear. Some studies investigated the independent role of tone when no segmental information is shared; however, the findings are mixed. Lee (2007) found that prime words sharing only the same tonal information with target words did not facilitate the processing of target words in an auditory lexical decision task. On the contrary, Malins and Joanisse (2010) found in an eye-tracking study that more fixations were made to the tonal competitor (e.g., *tu3*, "dirt") that overlapped only in tonal information with the target word (e.g., *mi3*, "rice") than to the distractors. Additional studies are necessary to further investigate the issue. Moreover, the tonal manipulations in the current study were conducted on the first syllables of disyllabic words. Thus, it remains unclear whether the obtained findings can be generalized to situations where tonal information of the second syllable or both syllables are manipulated. Last, notably, the interaction effect and the contrast analysis of predictive context just missed statistical significance. Thus, some caution is warranted in interpreting these results. However, they encourage further research on the modulatory role of the processing lexical tone context in Mandarin and other tonal languages.

In sum, the findings of the current study provide clear evidence for a partial role of lexical tone in constraining spoken-word recognition both when words are recognized in isolation and in neutral sentence contexts. Furthermore, the role of lexical tone is modulated by contextual predictability so that a predictive context can strengthen the role of lexical tone.

The experiments reported in this article were not preregistered. The data and the materials are available upon request; requests may be sent via email to the corresponding author at shen_wei@yahoo.com.

References

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255-278.

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

Cunnings, I. (2012). An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369-382.

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

Ferreira, F., Foucart, A., & Engelhardt, P. E. (2013). Language processing in the visual-world: Effects of preview, visual complexity, and prediction. *Journal of Memory and Language*, *69*(3), 165-182.

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

Hu, J., Gao, S., Ma, W. & Yao, D. (2012). Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology. 49*(9), 1179-1190.

Huang, X., Liu, X., Yang, J. C., Zhao, Q., & Zhou, J. (2018). Tonal and vowel information processing in Chinese spoken-word recognition: An event-related potential study. *Neuroreport, 29*(5), 356-362.

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual-world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151-171.

Ito, A., Pickering, M. J., & Corley, M. (2018). Investigating the time-course of phonological prediction in native and non-native speakers of English: A visual-world eye-tracking study.

Journal of Memory and Language, *98*, 1-11.

Ito, A. (2019). Prediction of orthographic information during listening comprehension: A printed-word visual-world study. *Quarterly Journal of Experimental Psychology, 72*(11), 2584-2596.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*(4), 434-446.

Kiriloff, C. (1969). On the auditory perception of tones in Mandarin. *Phonetica, 20*, 63-67.

Lee, C. (2007). Does horse activate mother? Processing lexical tone in form priming. *Language and Speech*, *50*(1), 101–123.

Liu, S., & Samuel, A. (2007). The role of Mandarin lexical tones in lexical access under different contextual conditions. *Language and Cognitive Processes, 22*(4), 566 – 594.

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

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

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.

Marslen-Wilson, W., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition*, *8*(1), 1-7.

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

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken-words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661-671.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution*, *4*(2), 133-142.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. Psychological Review, 115, 357–395.

Quené, H., & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413-425.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (Version 3.5.0) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of Experimental Psychology Learning Memory & Cognition, 36*(5), 1108-1117.

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

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

Shen, W., Qu, Q., & Li, X. (2016). Semantic information mediates visual attention during spoken-word recognition in Chinese: Evidence from the printed-word version of visual-world paradigm. *Attention, Perception, & Psychophysics,78*(5), 1267-1284.

Shen, W., Qu, Q., & Tong, X. (2018). Visual attention shift to printed words during spoken-word recognition in Chinese: The role of phonological information. *Memory & Cognition*, *46*(4), 642-654.

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

Tong, Y., Francis, A. L., & Gandour, J. T. (2008). Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes*, *23*(5), 689-708.

Tong, X., McBride, C., & Burnham, D. (2014). Cues for lexical tone perception in children: Acoustic correlates and phonetic context effects. *Journal of Speech Language & Hearing Research, 57*(5), 1589-1605.

Wang, X., Wang, J., & Malins, J. G. (2017). Do you hear ‘feather’ when listening to ‘rain’? Lexical tone activation during unconscious translation: Evidence from Mandarin-English bilinguals. *Cognition, 169*, 15-24.

Ye, Y., & Connine, C. M. (1999). Processing spoken Chinese: The role of tone information. *Language & Cognitive Processes,14*(5-6), 609-630.

Yip, M. C. (2001). Phonological priming in Cantonese spoken-word processing. *Psychologia, 44*(3), 223-229.

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.

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

Table 1

Lexical properties of the experimental materials in Experiment 1

	Target word	Phonological competitor (tone-consistent condition)	Phonological competitor (tone-inconsistent condition)	Distractor 1	Distractor 2
Mean word frequency	1.68 (1.19)	1.67 (1.31)	1.63 (1.44)	1.64 (1.07)	1.60 (1.16)
Mean number of strokes	16.21 (4.14)	16.13 (4.05)	15.15 (3.43)	15.85 (4.31)	16.02 (3.93)

Note. Mean word frequency was calculated as occurrences per million. The Standard deviations are in brackets.

Table 2

Lexical properties of the experimental materials in Experiment 2

	Target word	Phonological competitor (tone-consistent condition)	Phonological competitor (tone-inconsistent condition)	Distractor 1	Distractor 2
Mean word frequency	2.70 (2.27)	2.48 (2.66)	2.78 (2.22)	2.62 (2.32)	2.66 (2.16)
Mean number of strokes	16.92 (4.77)	16.50 (4.38)	15.13 (3.71)	16.38 (4.41)	16.67 (3.40)

Note. Mean word frequency was calculated occurrence per million. The Standard deviations are in brackets.

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

Table 3

Examples of experimental sentences used in Experiment 2

Sentence type	Example sentence
Neutral context sentence	事实上在大多数情况下**剪刀**是不允许带上飞机的
	(In fact, in most cases **scissors** are not allowed onboard)
Predictive context sentence	那位老裁缝手里拿着一把**剪刀**正在认真地剪裁布料
	(The old tailor is carefully cutting the cloth with a pair of **scissors**)

Note. Words in boldface were the target words; the English translations are given in brackets. For the spoken target word "剪刀" (*jian3dao1*,scissors), the printed-word display consisted of the target word "剪刀" (*jian3dao1*,scissors), a phonological competitor word in the tone-consistent condition "简体" (*jian3ti3*, simplified style) or in the tone-inconsistent condition "奸细" (*jian1xi4*, spy) and two unrelated distractors "门路" (*men2lu4*, route), and "海岛" (*hai3dao3*, islands).

"This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0".

Fig. 1

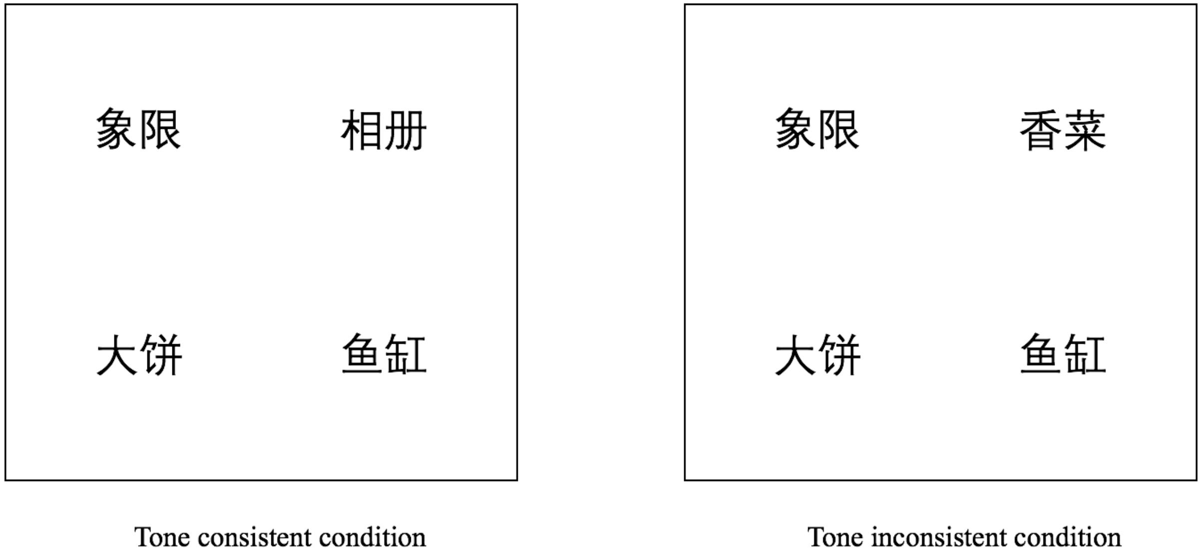

Fig. 1. An example of a printed-word display in Experiment 1. For the spoken target word "象限" (*xiang4xian4*, "quadrant"), the printed-word display consisted of the target word "象限" (*xiang4xian4*, "quadrant"), a phonological competitor word in the tone-consistent condition (left panel) "相册" (*xiang4ce4*, "photo album") or in the tone-inconsistent condition (right panel) "香菜" (*xiang1cai4*, "coriander") and two unrelated distractors "大饼" (*da4bing3*, "pancake"), "鱼缸" (*yu2gang1*, "fish tank") presented in the four corners of the display

“This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0”.

Fig. 2

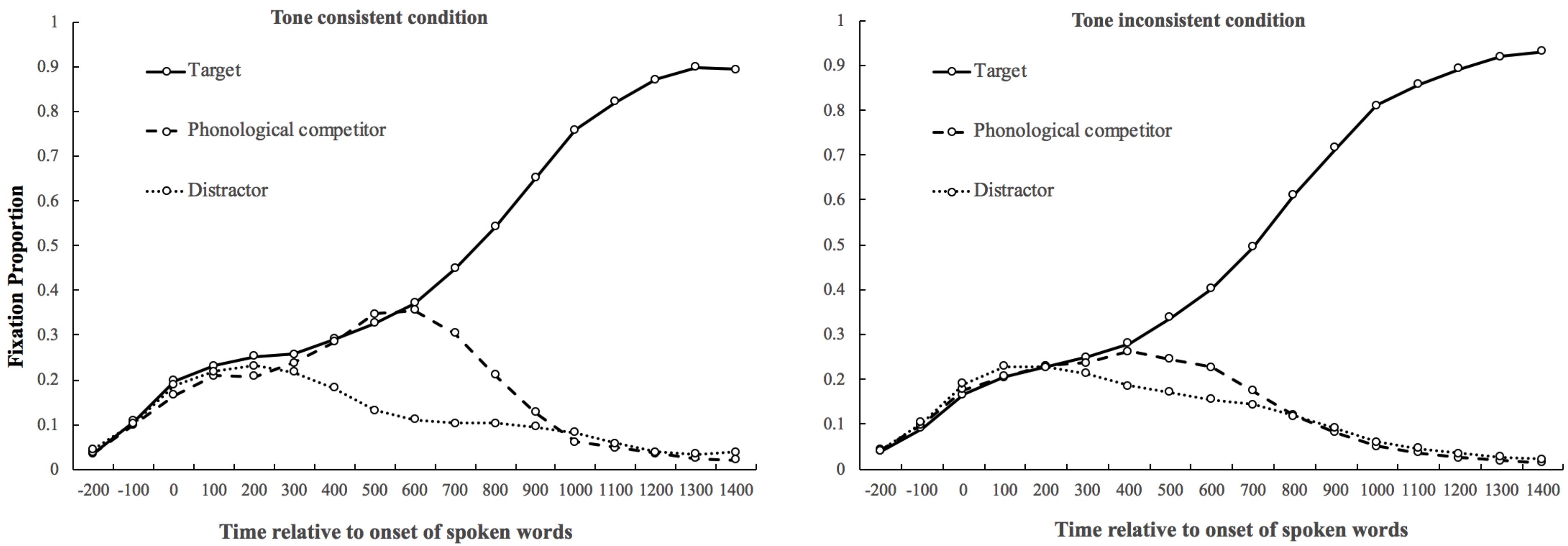

Fig. 2. Proportion of fixations on the target word, the phonological competitor, and the distractors from 200 ms before the onset of the spoken target word in Experiment 1. Tone-consistent condition (left panel), tone-inconsistent condition (right panel)

“This is a post-peer-review, pre-copyedit version of an article published in Memory & Cognition. The final authenticated version is available online at: http://dx.doi.org/10.3758/s13421-020-01070-0”.

Fig. 3

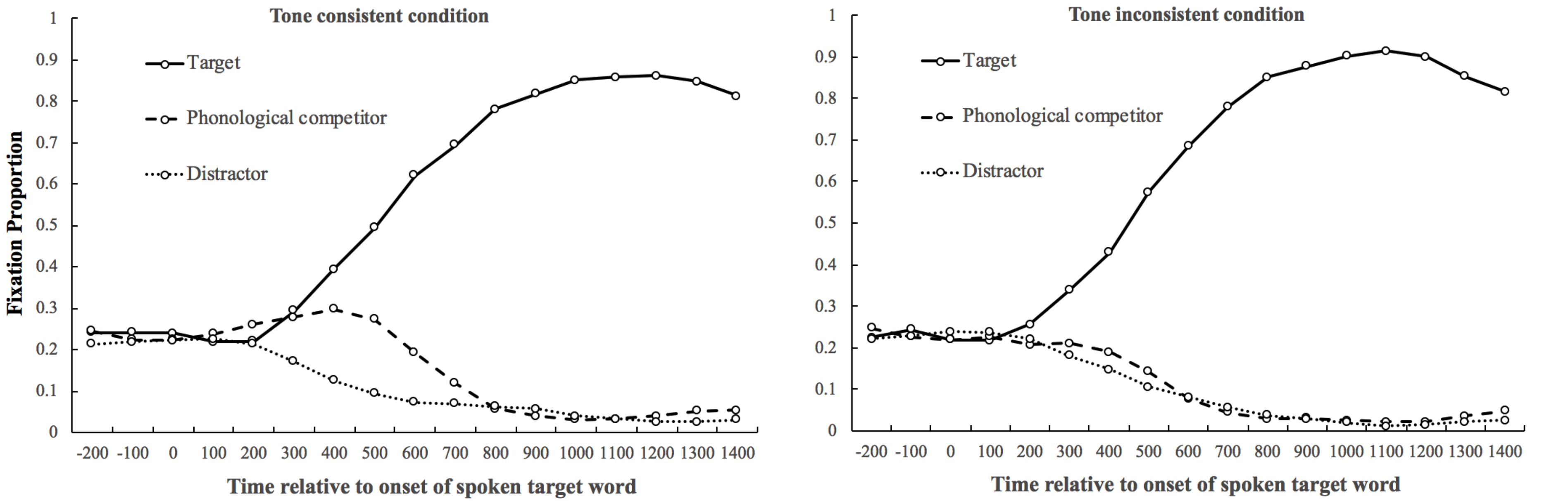

Fig. 3. Proportion of fixations on the target word, the phonological competitor, and the distractors from 200 ms before the onset of the spoken target word in the neutral context condition in Experiment 2. Tone-consistent condition (left panel), tone-inconsistent condition (right panel)

Fig. 4

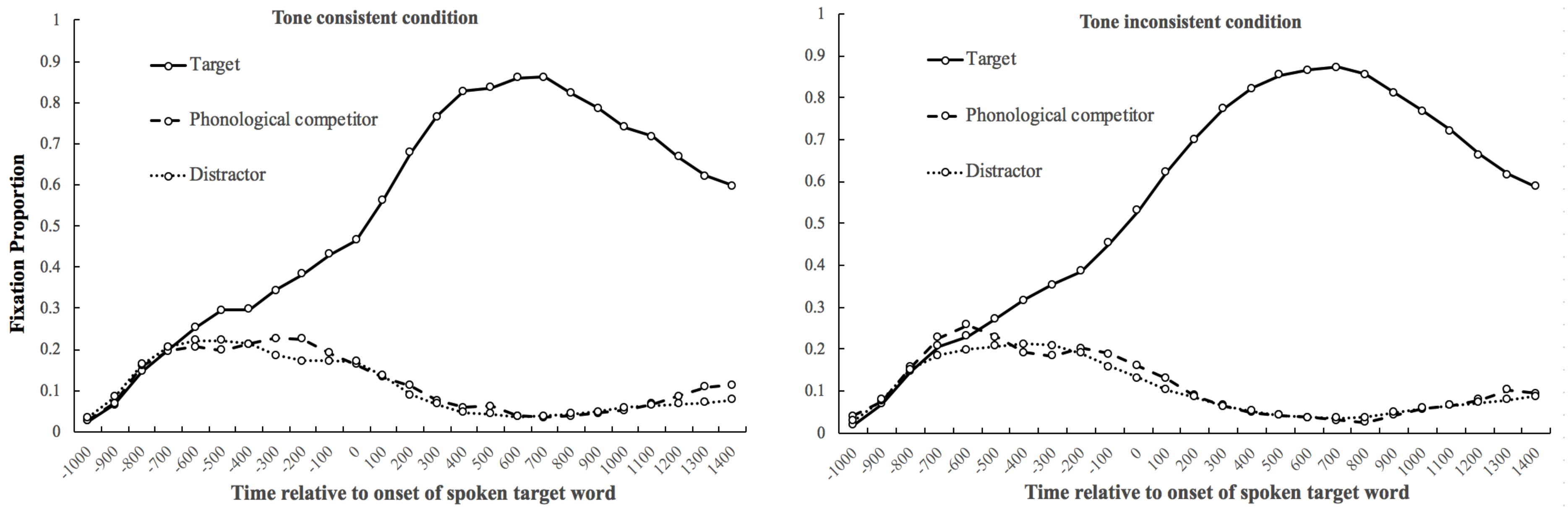

Fig. 4. Proportion of fixations on the target word, the phonological competitor, and the distractors from 1000 ms before the onset of the spoken target word in the predictive context condition in Experiment 2. Tone-consistent condition (left panel), tone-inconsistent condition (right panel)

Fig. 5

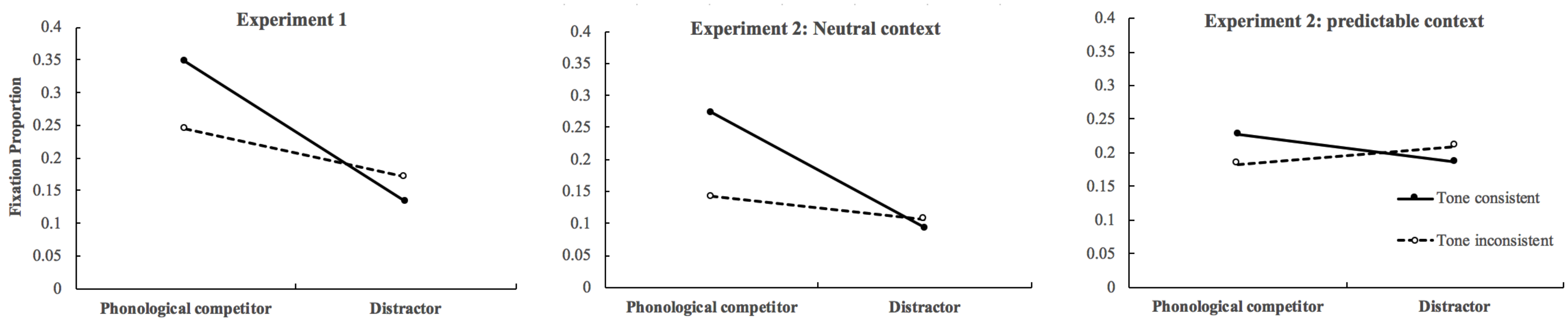

Fig. 5. The interaction effect found between tone consistency and competitor type in Experiment 1(left panel), in the neutral context condition of Experiment 2 (middle panel), and in the predictive context condition of Experiment 2 (right panel)