CrossMark

Visual attention shift to printed words during spoken word recognition in Chinese: The role of phonological information

Wei Shen [1,2,3] · Qingqing Qu [4,5] · Xiuhong Tong [6]

Abstract
The aim of this study was to investigate the extent to which phonological information mediates the visual attention shift to printed Chinese words in spoken word recognition by using an eye-movement technique with a printed-word paradigm. In this paradigm, participants are visually presented with four printed words on a computer screen, which include a target word, a phonological competitor, and two distractors. Participants are then required to select the target word using a computer mouse, and the eye movements are recorded. In Experiment 1, phonological information was manipulated at the full-phonological overlap; in Experiment 2, phonological information at the partial-phonological overlap was manipulated; and in Experiment 3, the phonological competitors were manipulated to share either full overlap or partial-overlap with targets directly. Results of the three experiments showed that the phonological competitor effects were observed at both the full-phonological overlap and partial-phonological overlap conditions. That is, phonological competitors attracted more fixations than distractors, which suggested that phonological information mediates the visual attention shift during spoken word recognition. More importantly, we found that the mediating role of phonological information varies as a function of the phonological similarity between target words and phonological competitors.

Keywords Phonology · Spoken word recognition · Chinese · Printed-word paradigm

Research shows that listeners tend to spontaneously shift their visual attention to referred objects while viewing a scene during spoken word recognition (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). This suggests that spoken language processing affects visual attention. One question of theoretical interestis whether this effect arises in Chinese, a written language that represents phonological information at the syllabic level. The aim of the present study is to examine the extent to which phonological information influences the shift of visual attention to printed Chinese during spoken word recognition in a visual-world paradigm.

The visual-world paradigm is widely used to study spoken word recognition (Tanenhaus et al., 1995). In this paradigm, participants are required to view an array of pictures while listening to spoken target words. The movements of their eyes as they gaze upon pictures are recorded simultaneously. The relationship between visual attention and language processing can be examined by measuring the fixation probability on the corresponding visual objects. Allopenna, Magnuson, and Tanenhaus (1998) showed that participants produced more eye fixations on phonologically related pictures (e.g., "speaker" and "beetle") than unrelated pictures when they heard a target (e.g., "beaker"). This was defined as the phonological competitor effect. The phonological competitor effect supports some spoken word recognition models, such as TRACE and cohort models, in which acoustic signals are assumed to be mapped continuously onto mental lexical representations, and candidates that share overlapping syllables with spoken target words are activated

✉ Xiuhong Tong
   tongxiuhong@gmail.com

1   Institute of Psychological Sciences, Hangzhou Normal University, Hangzhou, China

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

3   Center for Cognition and Brain Disorders, Hangzhou Normal University, 2318 Yuhangtang Road, Hangzhou 311121, China

4   Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, Beijing, China

5   Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

6   Department of Psychology, The Education University of Hong Kong, 10 Lo Ping Road, Tai Po, New Territories, Hong Kong

🪧 Springer

during spoken word recognition (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986).

Moreover, studies conducted using the visual-world paradigm demonstrate that information processing in visual and spoken fields interact and affect our eye-movement behaviors such that the overt attention to visual objects can be driven based on different kinds of information (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Huettig & McQueen, 2007; McMurray, Tanenhaus, & Aslin, 2002). For example, Huettig and McQueen suggested that spoken word recognition is a cascade process, and overt visual attention shift to visual objects can be driven by phonological, shape, and semantic information, and this information is activated in a chronological sequence as spoken words unfold over time. Huettig, Quinlan, McDonald, and Altmann (2006) further argued that lexical representations associated with spoken words and visual objects are activated concurrently and that the overlap of lexical representations (e.g., phonological representations) between visual and auditory modalities causes visual attention shift to visual referents (Huettig et al., 2006).

Related studies use a variant of the visual-world paradigm, in which pictures are replaced with printed words (i.e., the printed-word paradigm). Such studies show that phonologically related printed words also attract more fixations than distractors (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). However, the nature of phonological competitor effects observed in the printed-word paradigm remains controversial. Researchers have continued debating on what information drives these fixations caused by phonological competitors: whether it is driven by phonological information or other information (e.g., orthographic information). Two contrasting hypotheses have been proposed to explain such competitor effects: phonological and orthographic hypotheses (Huettig & McQueen, 2007; Salverda & Tanenhaus, 2010). Phonological hypothesis assumes that it is the representational overlap established in the phonological level that directs the shift of visual attention to printed words (Allopenna et al., 1998; Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Weber, Melinger, & Lara, 2007). For example, Huettig and McQueen (2007) reported that the time course of different types of information processing (i.e., semantic, phonological, and shape information) varies during spoken word recognition. All competitors were found to attract more fixations than the unrelated distractors using the visual-world paradigm (i.e., when participants were presented with pictures). However, if pictures were replaced with printed words, only the phonological competitor effect was seen, wherein the phonological competitor attracted more looks than distractors. The authors argued that only phonological information was relevant in searching for printed words.

However, contrary to the phonological hypothesis, the orthographic hypothesis argues that the representational overlap established in the orthographic level mediates the shift of visual attention to printed words (Salverda & Tanenhaus, 2010). Salverda and Tanenhaus (2010) conducted two experiments to examine the question of whether more fixations caused by

phonological competitors were driven by phonological information or orthographic information with a printed-word version of the visual-world paradigm. In this paradigm, they presented participants with a visual display of four printed words: a target word, a phonological competitor, and two distractors. Participants were required to follow a spoken instruction (e.g., please click on the word "bead"). In Experiment 1, they manipulated the phonological overlap between a target (e.g., "bead") and the competitors to be high (e.g., "bean"/bin/ –"bead"/bid/) or low (e.g., "bear"/bɛr/–"bead"/bid/), and therefore controlled the orthographic overlap across the two conditions (e.g., word pair "bead"–"bean" and word pair "bead"–"bear" shared the same orthographic overlap, namely, they shared the initial three letters "b-e-a"). The authors found no difference in the proportion of fixations between high and low phonological overlap competitors. This finding provided certain evidence against the phonological hypothesis by showing that the phonological information did not mediate visual attention shift during spoken word recognition.

In their Experiment 2, they further investigated the phonological competitor effect by manipulating orthographic overlap between the target words (e.g., "bead") and the phonological competitors to be high (e.g., "bear," which shared three letters "b-e-a" with the target word "bead") or low (e.g., "bare," which shared only one letter "b" with the target word "bead") and therefore controlled the same phonological overlap between targets and phonological competitors simultaneously (e.g., target word "bead"/bid/ had the same phonological overlap with competitors of "bear" /bɛr/ and "bare" /bɛr/). Results of Experiment 2 showed that competitors with high orthographic overlap attracted more fixations than those with low orthographic overlap. The authors interpreted this finding as that orthographic representations rather than phonological information between the target words and the competitors coactivatedbetween spoken words and visual referents (e.g., printed words) mediates the shift of visual attention to visual printed words.

However, the above findings in the study conducted by Salverda and Tanenhaus (2010) could not be considered as robust evidence against phonological hypothesis. It is well-known that graphemes are strongly linked to phonemes, referred to as the grapheme–phoneme correspondence rule, in alphabetical languages. Thus, the sounds in words and the letters used to represent those sounds are relatively difficult to separate. This leads to the fact that the role of phonology and orthography in spoken word recognition becomes difficult to distinguish in alphabetic languages, and this is a natural consequence of using an alphabet (Leck, Weekes, & Chen, 1995; Weekes, Chen, & Lin, 1998; Zhang, Chen, Weekes, & Yang, 2009; Zhang & Weekes, 2009). Therefore, the findings in Salverda and Tanenhaus's (2010) study could not completely disentangle phonological from orthographic information. The contribution of phonological information in their study

was highly likely to be just smaller than that of the orthographic information, which led to the nonsignificant effect of phonological information.

Chinese is a widely known logographic language, which differs noticeably from alphabetic languages. Chinese words map directly into meaning units rather than phoneme units, and the rules of regular or quasiregular lettersounds expressed in all alphabetic languages cannot be found in Chinese (Tan & Perfetti, 1998). The link between orthography and phonology is relatively weak in Chinese. Chinese thus provides a fantastic window to examine the two hypotheses mentioned. Moreover, the role of phonological information in spoken word recognition in Chinese has also been a topic of debate. Some studies have shown that phonological information was activated during word recognition, whereas other studies have found no phonological information competition effect by showing that word recognition could be accomplished directly from orthography to semantics without phonological information (Chen, d'Arcais, & Cheung, 1995; Hoosain, 1991). Further examination of the role of phonological information in Chinese spoken word recognition could greatly contribute in distinguishing phonological and orthographic hypotheses.

The purpose of this research is two-pronged: First, we aimed to revisit the role of phonological information in spoken word recognition using Chinese. We investigate whether phonological information at full-phonological overlap could mediate the shift of visual attention to printed Chinese words. Shen, Qu, and Li (2016) found that both semantic and phonological information could guide visual attention shift to printed words. However, the role of phonological information was not as stable as the semantic information in such study. Shen et al. (2016) found that the phonological competitor effect was only significant in the short-preview condition (i.e., visual display was previewed for 200 ms) in Experiment 2 but was insignificant with another set of experimental materials in the same condition in Experiment 1. Another study conducted by Meng (2014), using a similar experimental design, found no phonological competitor effect when the phonological competitor shared the initial phonological information with target words. Therefore, the role of phonological information in spoken word recognition should be re-examined.

The second aim of this study was to examine the question of whether the fixations directed to printed words are sensitive to the degree of phonological overlap between targets and competitors. Most spoken word recognition models suggest that spoken word recognition is an incremental process (Marslen-Wilson, 1987; McClelland & Elman, 1986). This hypothesis suggests that word candidates that share the matched phonological information with the acoustic information can be activated as the spoken word gradually unfolds. For instance, words such as "beep", "beaker", and "beetle" would be activated upon hearing a segment of "/bi/"

(Allopenna et al., 1998). It would be possible that the phonological information of word candidates, which are partially phonologically similar with spoken target words, can also be activated to direct visual attention shift to printed words. In addition, prior studies have showed that the fixations on visual objects are in proportion to the semantic similarity between spoken words and visual pictures (Huettig & Altmann, 2005; Huettig et al., 2006). For example, Huettig et al. (2006) found that the semantic similarity score between targets and competitors could predict participants' fixation behavior in visual display using a visual-world paradigm. A smaller semantic competitor effect was observed even when targets were less semantically similar to the competitor (e.g., "coat"–"slipper"). If the visual shift to printed words is also sensitive to phonological information, we then speculate that phonological similarity might serve the same role of the semantic similarity as in Huettig et al.'s (2006) study. That is, we expected that the fixation proportions on printed words may vary as a function of the degree of the phonological similarity shared between targets and competitors.

The eye-tracking technique with the printed-word paradigm was adopted to investigate the above two research questions. Eye -tracking technique has at least two advantages over the traditional experimental tasks used in prior spoken word recognition studies (e.g., lexical decision and priming tasks). First, eye-tracking technique has larger ecological validity than the traditional tasks. For example, traditional tasks, such as the lexical decision task, require explicit responses to target stimuli, which may induce different processing strategies. By contrast, eye-movement recording can occur implicitly and in a natural environment. Second, eye-movement technique can be used to reveal the online time course of spoken target word recognition, given that it can accurately provide time-locked and continuous eye-movement activities. As for traditional tasks (e.g., lexical decision task), reaction time and accuracy are used as independent measures, which can only reflect cognitive processing at a certain processing stage.

Moreover, we improved the printed-word paradigm used in prior studies in the following two aspects. First, we presented all spoken target words without sentence context. In prior studies, all spoken target words were embedded in different neutral-context sentences, in which listeners might have to listen to a preceding context before hearing target words. In the preceding context, some words may share the same phonological information with the spoken target words, which may interfere or weaken the phonological competitor effect. We speculate that the phonological interference of context words might decrease the robustness of phonological competitor effects observed in prior studies. Second, we presented target referents in the visual display in this study. In some cases, the visual referents of spoken words would appear in the view field. With the display of target words, the target referent would attract the most fixation relative to other

referents, which provides a better field to examine whether the phonological competitor effect would survive under a more stringent environment.

To summarize, three experiments in this study were designed to examine the phonological competitor effect in Chinese spoken word recognition. Experiment 1 aimed to examine whether the phonological competitor effect in Shen et al.'s (2016) study could be observed in a more stringent situation. In this experiment, we manipulated the phonological information between target words and competitors such that it would not only be phonologically related[1] but also orthographically unrelated. According to the orthographic hypothesis, if the orthographic information is critical in mediating shifts in visual attention, then only phonologically related competitors would not attract more fixations than distractors; namely, no phonological competitor effect should be observed. Otherwise, a significant phonological competitor effect would be observed, which supports the phonological hypothesis. In Experiment 2, we manipulated the phonological competitor to have partial phonological overlap with target words in order to explore whether partial phonological similarity between spoken words and phonological competitors could also mediate the shift in visual attention to printed words. Competitors with partial phonological overlap with targets were hypothesized to attract more fixations than distractors. In Experiment 3, the degree of phonological overlap between phonological competitors and targets was manipulated to be full-overlap or partial-overlap for each given target word to further investigate whether visual shifts to printed words is sensitive to phonological information. If the phonological competitor effect varied as a function of phonological similarity, then an interaction between the degree of overlap and the competitor type was expected to be observed.

Experiment 1

Method

Participants

A total of 30 undergraduate students (22 females and 8 males) ages 19 to 25 years ($M = 21.20$ years) were recruited from a university in Hangzhou. Each student was paid 10 Yuan (approximately 1.47 U.S. dollars) to participate in this study. All participants were native Chinese speakers. They had normal or corrected-to-normal vision and were right-handed. All were

Fig. 1 Example of a printed-word display. For the spoken target word "番茄"(/fan1qie2/, tomato),the printed-word display consisted of the target word"番茄"(/fan1qie2/, tomato), the phonological competitor word "帆船"(/fan1chuan2/, sailing boat), and two unrelated distractors "插图"(/cha1tu2/, illustration) and "车手" (/che1shou3/, driver) in four different positions of the display

unaware of the purpose of the experiment. This study was approved by the Ethics Committee of the university.

Materials and design

A total of 48 words were selected as target words. Each printed-word display consisted of four printed words: a target word, a phonological competitor that shared the same syllable and tone of the first constituent with the target word, and two unrelated distractors that were neither semantically or phonologically related to the target word (an example, see Fig. 1). The word frequency and number of strokes across the four printed word conditions were matched (all Fs < 1;see Table 1). Phonological competitors were selected to be neither semantically nor orthographically unrelated[2] with the target words. However, the frequency of characters and number of strokes were unmatched across conditions due to the difficulty in material selection. The positions of the four printed words were counterbalanced across trials. Moreover, 48 filler trials, which consisted of four words, were added to the experimental list to avoid participants becoming aware of the relationships between phonological competitors and target words in the critical trials. All spoken targets were recorded in a natural voice by a female native Chinese

---

[1] The phonological relatedness was manipulated for the first character in all three experiments.

[2] In this study, "orthographically unrelated" is defined as having no radical shared between phonological competitor and spoken target words.

Table 1 Lexical properties of experimental materials in Experiment 1

	Target word	Phonological competitor	Distractor	Distractor
Mean word frequency	3.33	3.15	3.38	3.38
Mean stroke number	16.79	16.13	16.46	15.58
Mean frequency of first character	292.34	360.78	810.59	723.80
Mean stroke number of firstcharacter	8.48	8.96	8.50	7.60

Mean word frequency was calculated occurrence per million

speaker. Spoken targets were presented to participants through headphones.

Apparatus

Eye movements were recorded using an EyeLink1000 tracker sampling at a rate of 1000 Hz (SR Research, Mississauga, Ontario, Canada). Experimental materials were presented on a 21-inch CRT monitor with a resolution of 1024 × 768 pixels and a refresh rate of 85 Hz. Participants were asked to place their chins on a chinrest in order to minimize head movement during the experiment. Participants were seated in a chair and positioned 58 cm away from the video monitor. All printed words were shown in 30 Song font in black (RGB: 0, 0, 0) against a white background (RGB: 255, 255, 255). Each character subtended a visual angle of approximately 1.4°. Although viewing position was binocular, eye movement data were collected only from the right eye.

Procedure

After participants entered the lab, they were briefly introduced to the eye tracker. A nine-point calibration was then conducted at the beginning of the experiment, which participants completed by looking at nine white dots presented randomly on the computer screen. The validation error was smaller than 1°. A drift check was performed at the beginning of each trial. Afterwards, a blank screen was presented for 500 ms before the words were displayed. The words were displayed for 200 ms before the targetswere spoken. Participants were instructed to click the spoken target words on the screen with a mouse. Seven practice trials were initially conducted to familiarize participants with the experimental procedure. They then performed 48 experimental trials and 48 filler trials. The trials were presented randomly. The test lasted approximately 15 min.

Results and discussion

We excluded trials (0.6%) with incorrect responses for the analysis. A fixation was defined as a focus on the printed word if it fell within a square of 8× 8cm around the center of a printed word. Figure 2 shows the proportion of fixations to printed words (targets, phonological competitors, and distractors[3]) from 200 ms before the onset of spoken target words and 1,500 ms after the onset of spoken target words (time window was calculated in every 100 ms).

As shown in Fig. 2, fixations on phonological competitors started to diverge from distractors at around 200 ms after the target word was presented. To test whether the difference between phonological competitors and distractors reached statistical significance, we employed logit mixed models to analyze fixation proportion data[4] (Ferreira, Foucart, & Engelhardt, 2013; Jaeger, 2008). One advantage of the logit mixed model over traditional analyses (e.g., *t* test or ANOVA) is that it can evaluate different types of random effects in one model (i.e., random effects of participants and items),which is helpful in decreasing Type I error rates. Moreover, this model is also better in handling situations when the sphericity and homoscedasticity are violated (Cunnings, 2012). Following the procedure of Ferreira et al. (2013), we defined a dependent variable as whether or not a printed word received a fixation in a specific time window; hence, the value "0" means that no fixation was made to the printed word, whereas the value "1' means that a fixation was made to the printed word. Two types of time-window analyses were conducted: (1) the global analysis that begins from 200 ms[5] to 600 ms after the onset of spoken target words, which was time-locked to the display duration of the first character (the average duration is approximately 350 ms) of the spoken target word; (2) the four 100-msbin analysis that starts from 200 ms to 600 ms after the onset of spoken target words.

A base model was initially created for statistical analysis, which included random intercepts for participants and items. The significance of the model was enhanced by adding a fixed factor (i.e., competitor type[6])and by-participant random slopes

---

[3] The fixation proportionofdistractor in Fig. 2 was the average of the fixation proportions of distractors.

[4] Both the logit mixed model analysis used in Ferreira et al.'s (2013) study and the *t*-test analysis used in Huettig and McQueen's (2007) study were performed for data analysis in this study. Results of the two methods were consistent. Given the advantages of the logit mixed model, we only reported the results using the logit mixed model method.

[5] This time window considered the time for a saccade programming, which is about 150–200 ms (Rayner, 1998).

[6] The competitor type included a phonological competitor and a distractor.
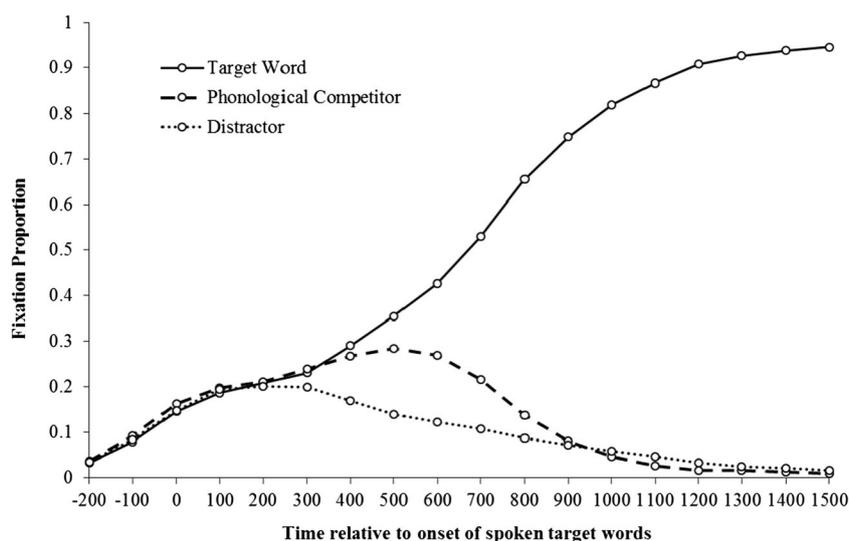
Fig. 2 Fixation proportions to the target words, phonological competitors, and distractors from 200 ms before the onset of the spoken target words in Experiment 1

for competitor type step by step. The ANOVA function in R (Version 3.3.3; R Core Team, 2016) was used to isolate which model improved model fit significantly. In the models, one of the distractors was randomly assigned as a baseline, and phonological competitors were compared with that distractor. In each analysis, random intercepts for participants and items and by-participant random slopes for competitor type were entered as random effects (Barr, Levy, Scheepers, & Tily, 2013). In addition, competitor type was entered for each model as a fixed effect. For the model fit, we used the glmer function from the lme4 package (Version 1.1-7; Bates, Maechler, Bolker, & Walker, 2014) in R. Regression coefficient b, standard error SE, Zvalues, and corrected p values are reported in Table 3. R^2, coefficient of determination, is used as a measurement of effect size (see also Nakagawa & Schielzeth, 2013). Thus, the correlation between the fitted and observed values

was calculated via an r2.corr.mer function developed by Jarrett Byrnes in an R environment.

Global analysis results showed that the model was significantly improved by inclusion of the competitor type, $\chi^2(2) = 37.25$, $p < .001$. However, the addition of by-participant slopes for competitor type failed to significantly improve the model fit, $\chi^2(5) = 0.29$, $p = .99$. The contrast test revealed that the variable phonological competitor attracted significantly more fixations compared with distractor ($b = 0.42$, $SE = 0.08$, $Z = 5.43$, $p < .001$, $R^2 = .07$).

We also performed a time-course analysis of four periods of 100-ms bins from 200 ms to 600 ms after the onset of spoken target words (see Table 2). We used Bonferroni correction to avoid the issue of multiple comparisons. The corrected p values are displayed in Table 2. The inclusion of the competitor type or by-participant random slope of competitor type did not improve

Table 2 Logit mixed model analysis for competitor type in Experiment 1

Time window from spoken target word (ms)	Predictor	Estimate	SE	zvalues	Corrected pvalues
200–300	(Intercept)	−1.32	0.11	−12.07	<.001***
	Distractor	0.06	0.09	0.68	0.49
	Phonological competitor	0.07	0.09	0.82	0.41
300–400	(Intercept)	−1.33	0.09	−13.77	<.001***
	Distractor	0.12	0.09	1.36	0.18
	Phonological competitor	0.28	0.08	3.18	.003**
400–500	(Intercept)	−1.49	0.08	−19.48	<.001***
	Distractor	0.04	0.09	0.48	0.63
	Phonological competitor	0.61	0.09	6.87	<.001***
500–600	(Intercept)	−1.75	0.10	−16.16	<.001***
	Distractor	−0.01	0.11	−0.18	0.85
	Phonological competitor	0.94	0.13	7.39	<.001***

$^{**}p < .01.$ $^{***}p < .001$

model fit for the time window of 200–300 ms ($p >$ .68). However, the model was significantly improved by the competitor type for the other three time windows (χ^2s > 10.19, ps< .006). The phonological competitor diverged from the distractors significantly at 300–400 ms after the spoken word onset ($b = 0.28$, $SE = 0.09$, $Z = 3.18$, $p = .003$, $R^2 = .05$). In addition, the time windows of 400–500 ms ($b = 0.61$, $SE = 0.09$, $Z = 6.87$, $p < .001$, $R^2 = .03$) and 500–600 ms ($b = 0.94$, $SE = 0.13$, $Z = 7.39$, $p < .001$, $R^2 = .07$) showed similar data patterns, wherein the phonological competitor attracted more fixations than distractors, which suggest that the phonological information of spoken target words was activated early in the processing stage during spoken word recognition (at around 300 ms from the onset of the first constituent of the spoken target word in the present study).

In summary, results of Experiment 1 strongly suggest that phonological information does mediate shifts of visual attention to Chinese printed words, which is independent of orthographic information even when performing a relatively stringent task. That is, the phonological hypothesis is supported by findings in Experiment 1, highlighting that phonological information plays an important role in mediating the shift in visual attention to printed words in during spoken word recognition in Chinese.

Experiment 2

Experiment 2 aimed to determine whether partially activated phonological information during spoken word recognition could also mediate the shift of visual attention to printed words by manipulating the overlap between spoken target words and phonological competitors such that a partial phonological overlap exists. If visual attention was sensitive to the phonological information, even with the partial phonological overlap between spoken target words and phonological competitors, then we would still observe a phonological competitor effect.

Method

Participants

A total of 30 participants (20 females and 10 males) ages 19 to 28 years ($M = 21.37$ years), who were recruited from the same participant pool as Experiment 1 but were different participants, were involved in Experiment 2. Each participant was paid 10 Yuan (approximately 1.47 U.S. dollars). They were all native Chinese speakers with normal or corrected-to-normal vision and were right-handed. This study was approved by the Ethics Committee of the university.

Material

The experimental design was similar to that of Experiment 1, except spoken targets and phonological competitors shared partial phonological information. For example, the target word "冰雹" (/bing1bao2/, hailstone) shared only three phonemes (e.g., /bin/) with the phonological competitor "宾客" (/bin1ke4/, guest). A total of 48 target words were selected. The orthographic similarity between targets and phonological competitors was also carefully controlled. Moreover, word frequency and number of strokes across the four printed words were carefully matched (see Table 3). Table 4 also presents the frequency of character and number of strokes of the first character.

Apparatus and procedure

The apparatus was the same as in Experiment 1.

Results and discussion

Data analyses were conducted in the same manner as in Experiment 1, and 0.2% of all trials with incorrect responses were excluded from the analysis.

As shown in Fig. 3, the fixation curve for the partial-overlap phonological competitors indicated a moderate increase relative to distractors. Global analysis showed that the model fit failed to be improved by inclusion of any factors, $\chi^2(2) = 2.02$, $p = .36$. Similar to Experiment 1, we also conducted a time-course analysis of four periods of 100ms bins from 200 ms to 600 ms (see Table 4). The inclusion of the competitor type or by-participant random slopes for competitor type did not improve model fit for the time windows of 200–300 ms and 300–400 ms (ps> .34). However, the model was significantly improved by the competitor type and by-participant random slope for competitor type for the time windows of 400–500 ms, $\chi^2(5) = 15.17$, $p = .009$, Also, the model was significantly improved by the competitor type for

Table 3 Lexical properties of experimental materials in Experiment 2

	Target word	Phonological competitor	Distractor	Distractor
Mean word frequency	2.31	2.37	2.31	2.37
Mean stroke number	17.60	16.16	17.03	17.43
Mean frequency of first character	219.88	537.44	1,255.29	608.57
Mean stroke number of first character	9.42	8.60	7.60	8.96

Mean word frequency was calculated occurrence per million

Table 4 Logit mixed model analysis for competitor type in Experiment 2

Time window from spoken target word (ms)	Predictor	Estimate	SE	z values	Corrected p values
200–300	(Intercept)	-1.16	0.09	-11.84	<.001***
	Distractor	0.06	0.09	0.74	0.46
	Phonological competitor	−0.06	0.09	−0.66	0.51
300–400	(Intercept)	−1.14	0.07	−15.38	<.001***
	Distractor	−0.02	0.09	−0.26	0.79
	Phonological competitor	0.03	0.09	0.39	.0.69
400–500	(Intercept)	−1.36	0.09	−15.51	<.001***
	Distractor	0.11	0.13	0.83	0.41
	Phonological competitor	0.26	0.09	2.76	.01*
500–600	(Intercept)	−1.47	0.08	−19.09	<.001***
	Distractor	0.08	0.09	0.85	0.39
	Phonological competitor	0.31	0.09	3.41	.001**

$^{*}p < .05.$ $^{**}p < .01.$ $^{***}p < .001$

the time window of 500–600 ms, $\chi^2(2) = 12.73$, $p = .002$. Comparisons between phonological competitors and distractors indicated that partial-overlap phonological competitors attracted more fixations than distractors in time windows of 400–500 ms ($b = 0.26$, $SE = 0.09$, $Z = 2.76$, $p = .01$, $R^2 = .04$) and 500–600 ms ($b = 0.31$, $SE = 0.09$, $Z = 3.41$, $p = .001$, $R^2 = .02$).

Results of Experiment 2 demonstrate that even partial phonological overlap between target words and phonological competitors can also mediate visual attention shifts, which suggests that visual attention is sensitively driven by phonological information.

Experiment 3

Although results of Experiments 1 and 2 support the phonological hypothesis for the phonological competitor effect in Chinese

spoken word recognition and the phonological competitor effect can be even observed at the partial phonological overlap level between target word and phonological competitors, Experiments 1 and 2 were conducted separately. It should be more direct to examine whether the phonological competitor effect varied as a function of phonological similarity in one experiment by using a within-item design. In Experiment 3, we thus directly manipulated the phonological overlap degree between target words and phonological competitors using a within-item design to further investigate the phonological competitor effect. More specifically, for a given spoken target word, two types of phonological competitors were created: one shared full phonological overlap with the target word (hereafter, the full-overlap condition), and the other one shared partial phonological overlap with target word (hereafter, the partial-overlap condition). We expected that the phonological competitor effect would be observed in both
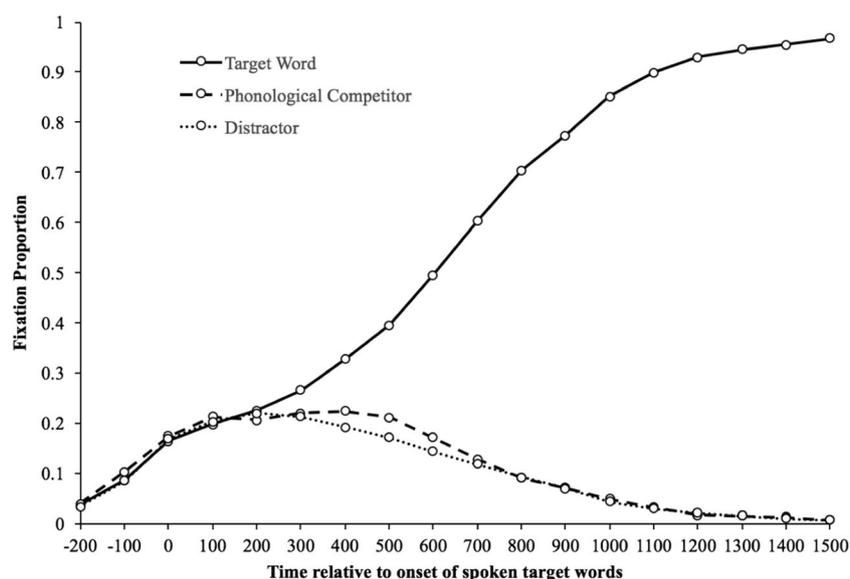
Fig. 3 Fixation proportions to the target words, phonological competitors, and distractors from 200 ms before the onset of spoken target words in Experiment 2

conditions on the basis of the results observed in Experiments 1 and 2.Moreover, if the visual attention shift was sensitive to the phonological information processing, then the full-overlap competitors would attracted more fixations than the partial-overlap competitors in comparison to the distractors. That is, an interaction between overlap degree (full-overlap, partial-overlap) and competitor type (phonological competitor, distractor) was expected to be observed in Experiment 3.

Method

Participants

A total of 30 participants (22 females and 8 males) ages 18 to 25 years (M =20.67 years) were recruited from the same participant pool as Experiments 1 and 2, and they were paid to join the experiment. Each participant was paid 15 Yuan (approximately 2.27 U.S. dollars). All were native Chinese speakers with normal or corrected-to-normal vision and were right-handed. This study was approved by the Ethics Committee of the university.

Material

The experiment was a 2 (overlap degree: full-overlap and partial-overlap) × 2 (competitor type: phonological competitor and distractor) within-participant/item design. The overlap degree between target words and phonological competitors were manipulated to be either high (full-overlap condition) or low (partial-overlap condition). In the full-overlap condition, the phonological competitors shared high phonological overlap (four to five phonemes) with target words. In the partial-overlap condition, the phonological competitors shared low phonological overlap (three phonemes) with target words. For example, the full-overlap competitor 姓氏(/xing4shi/, surname) shared fourphonemes with the given target杏仁(/ xing4ren2/,almond), whereas the partial-overlap competitor 信纸(/xin4zhi3/, letter paper) shared threephonemes with the given target 杏仁 (/xing4ren2/,almond).

A total of 104 trials consisting of a target word, a phonological competitor, and two unrelated distractors were constructed as experimental trials. For half of the trials, the phonological competitors were under the full-overlap condition, and the other half were under the partial-overlap condition. No orthographic similarity or semantic similarity was shared between target words and phonological competitors. Two distractor words were selected to avoid any association (semantic/phonology/orthography) between targets, and distractors were controlled. The printed words in each set were matched carefully in terms of word frequency and number of strokes ($Fs<1$; see Table 5).

Apparatus

The apparatus was the same as in Experiment 1.

Procedure

The procedure was similar to that in Experiment 1, with the following exception. The 104 experimental trials were divided into two counterbalanced lists with 52 trials in each list. Each list consisted of 26 trials with partial-overlap phonological competitors and 26 trials with full-overlap phonological competitors. Participants were randomly assigned to one list. In addition, anther 52 trials were included as filler trials to prevent participants from becoming aware of the experimental design. All trials were presented randomly. The whole experiment lasted approximately 15 minutes.

Results and discussion

Data analyses were conducted in the same manner as in Experiments 1 and 2. Trials with incorrect responses (0.38%) were excluded from the analysis.

Figure 4 depicts fixation proportions of each printed word in both the full-overlap condition and the partial-overlap condition. A base model including random intercepts for participants and items was created first. The model was then enhanced by adding the fixed factors competitor type, overlap degree, the interaction of the two fixed factors, and finally by-participants random slopes for the fixed factors sequentially.

The global analysis showed that the model was significantly improved with inclusion of the competitor type,$\chi^2(2) = 28.54$, $p< .001$. The contrast test showed that the phonological competitors attracted more fixations than the distractors ($b = 0.28$, $SE = 0.07$, $Z = 3.79$, $p< .001$, $R^2 = .06$). However, the addition of other factors failed to improve the model fit significantly ($ps> .61$).

Table 5 Lexical properties of experimental materials in Experiment 3

|	Target word	Full-overlap competitor	Partial-overlap competitor	Distractor	Distractor
Mean word frequency	1.93	1.94	1.95	1.93	1.95
Mean stroke number	16.26	16.13	15.21	15.40	16.31
Mean frequency of firstcharacter	324.08	368.65	570.28	1,132.11	560.14
Mean stroke number of firstcharacter	9.46	8.75	8.23	7.40	8.46

Mean word frequency was calculated occurrence per million
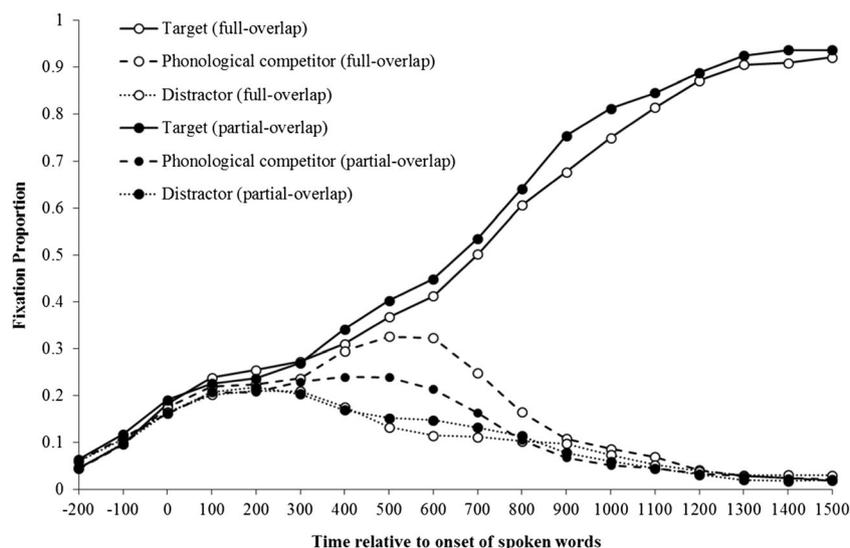
Fig. 4 Fixation proportions to the target words, phonological competitors, and distractors from 200 ms before the onset of spoken target words in full-overlap and partial-overlap conditions, respectively, in Experiment 3

The time-course analyses showed that the model was significantly improved with inclusion of the competitor type across all the time windows (χ^2s > 15.92, ps < .001). The interaction between competitor type and overlap degree was significant in the time window of 500–600 ms, $\chi^2(2) = 14.46$, $p < .001$. The simple effect analysis showed that the phonological competitor effect was observed in both the full-overlap condition ($b = 1.13$, $SE = 0.13$, $Z = 8.96$, $p < .001$, $R^2 = .08$) and the partial-overlap condition ($b = 0.61$, $SE = 0.13$, $Z = 4.88$, $p < .001$, $R^2 = .03$). Further analysis revealed that the phonological competitors in the full-overlap condition attracted more fixations than did those in the low-overlap condition ($b = 0.43$, $SE = 0.11$, $Z = 3.78$, $p < .001$, $R^2 = .12$), while there was no fixation proportion difference on distractors between the two conditions ($Z = 1.58$, $p = .12$; see Fig. 5 for the interaction effect).

The significant interaction effect observed between overlap degree and competitor type provides direct and firm statistical
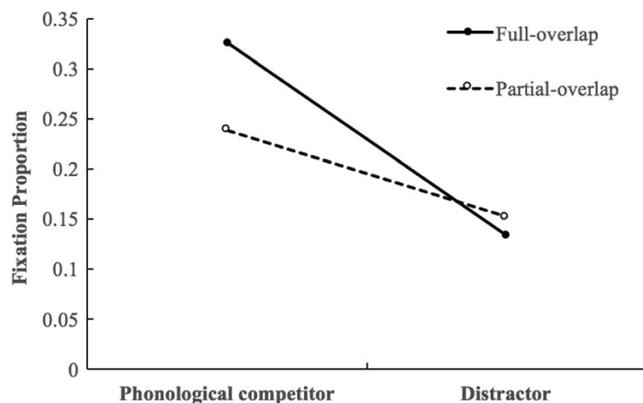
Fig. 5 Interaction effect observed between overlap degree and competitor type in the time window of 500–600 ms in Experiment 3

evidence for the hypothesis that the fixation proportion to printed words varies as a function of the degree of phonological overlap in Chinese spoken word recognition. The more phonological overlap shared between targets and competitors, the larger the phonological competitor effect to be observed, further suggesting that visual attention to printed words is sensitive to phonological information.

General discussion

In the present study, we designed three experiments to investigate to what extent phonological information could mediate visual attention shifts to printed Chinese words during spoken word recognition by using a target-present version of the printed-word paradigm. Phonological competitors attracted more fixations than distractors in all three experiments; that is, a phonological competitor effect was observed, which is independent of orthographic information. Moreover, a phonological competitor effect was found given a full phonological overlap (Experiment 1) and a partial phonological overlap (Experiment 2). Experiment 3 provided further evidence that the phonological competitor effects vary with the degree of phonological similarity, namely, phonological competitors shared more similarity to attract more fixation proportions.

The current findings have two novel contributions. First, we replicated a pure phonological competitor effect as shown by Shen et al. (2016), but observed it in a more stringent environment. As previously noted, the phonological effect observed in alphabetical languages was confounded by orthographic information (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Salverda and Tanenhaus (2010) argued that orthographic information, rather than phonological

information, drive the overt attention shift, which considerably neglected the role of phonological information. In this study, we separated phonology from orthography by manipulating phonological competitors to be phonologically related and orthographically unrelated with target words. A significant phonological competitor effect suggests that phonological information is not only accessed during spoken word recognition but also plays an important role in mediating visual attention shift. Our findings provide direct evidence that the match established at the phonological level drives visual attention shift to printed referents as assumed by the phonological hypothesis. In addition, we found that phonological competitor effect is a robust rather than a weak effect, as claimed by Shen et al. (2016). As previously noted, the decreased robustness observed by Shen et al. (2016) may be attributed to interference from context words. In this study, target words were presented without sentence context. This manipulation could reduce possible interferences from context words that share the same phonological information with target words. Moreover, we presented target words in a visual display, which provided a fierce competitive environment considering that targets could attract increased attention. Under such conditions, we still observed phonological competitor effect, suggesting that phonological competitor effect is stable and robust in Chinese spoken word recognition.

Second, we demonstrated that visual attention shift to printed words is sensitive to phonological information so that the phonological information is used to map the spoken targets onto the printed Chinese words gradually. In Experiment 2, we extended phonological competitor effect from a large unit (i.e., the full phonological overlap in Experiment 1) to a small unit (i.e., sharing three phonemes between phonological competitors and target words). We found that partially overlapping phonological competitors attracted more fixations than distractors from 400 ms to 600 ms after the onset of spoken target words. The effect obtained in Experiment 2 was much smaller and was activated for a very short time compared with the results observed in Experiment 1, which showed some indication that the phonological competitor effect was varied with the degree of phonological similarity to a certain extent. In Experiment 3, we directed manipulated the overlap degree of phonological information between spoken target words and competitors to be full-overlap or partial-overlap in a within-item experimental design. Phonological competitor effects were observed in both full-overlap and partial-overlap conditions. In particular, we found that higher fixation proportion was observed on full-overlap competitors than the partial-overlap competitors in the time window of 500–600 ms. Collectively, the findings of current study indicate that phonological competitor effect functions as the degree of phonological similarity.

Additionally, results of the three experiments provide further evidence for the hypothesis that spoken word recognition is an incremental processing rather than a holistic processing, as claimed in the cohort and TRACE models (see Malins et al., 2014). According to the cohort (Marslen-Wilson, 1987) and TRACE models (McClelland & Elman, 1986), the spoken information is mapped continuously onto the mental lexical representations with acoustic information. In three experiments of this study, we observed that phonological representations at different levels (such as mapping from partial to full phonological overlap) were all activated as the spoken target word unfolded over time. Moreover, what we showed beyond these models is that these activated phonological representations can also direct visual attention shift to printed words in the visual display.

Notably, we used two-character words as stimuli in the present study, given the large number of homophones in Chinese. As shown in Tables 1 and 4, the frequencies of the first characters across experimental conditions were uncontrolled statistically in experiments due to the difficulty in material selection, which could confound our results. However, a study conducted by Yan, Tian , Bai and Rayner (2006) found that character frequency was modulated by word frequency. Thus, character frequency effect was only observed for words with low frequencies and absent for words with high frequencies. In the present study, the means of word frequency of the target words, phonological competitors, and distractors in three Experiments were 2.52, 2.47, and 2.55 per million, respectively. Based on a Chinese Linguistic Data Consortium (2003), the percentage of words that are more frequent than 2.52, 2.47, and 2.55 are 22.04%, 22.35%, and 21.84%, respectively. Therefore, the relatively high-frequency words used in this study might minimize the influence of the first character frequency on the results.

The present study was the first attempt to explore the phonological competition effect at a partial phonological overlap level between phonological competitors and spoken target words with the printed-word paradigm in Chinese (as in Experiments 2 and 3). Prior studies showed that phonological competitor effects can be observed at different phonological levels (e.g., rhyme level; Allopenna et al., 1998). For instance, Allopenna and colleagues found that when recognizing spoken target words (e.g., "beaker"), both cohort competitors (e.g., "beetle") and rhyme competitors (e.g., "speaker") attracted more fixations than the distractors using a typical visual-world paradigm. In this study, however, we did not specify on which level (onset, rime, or coda) the phonological competitor effect might emerge. We manipulated the phonological overlap between target and phonological competitor to share three phonemes. We did this because no phonological competitor effect was observed when there was one phoneme (e.g., "/d/") shared between targets (e.g., "灯泡", /deng1pao4/, lamps) and phonological competitors (e.g., "盾牌", /dun4pai2/, shield) in our pilot study. We speculated that the null difference might be attributed to the smaller degree of phonological

overlap between the phonological competitors and spoken targets. Our results support our hypothesis—that is, a significant phonological competitor effect was obtained when the overlap between targets and phonological competitor was more than one phoneme.

Moreover, Mandarin Chinese has four different lexical tones; an identical syllable with different tones can represent different characters (Zhou & Marslen-Wilson, 1994). Previous studies have shown that lexical tones are accessed during spoken word recognition in Chinese (Nixon, Chen, & Schiller, 2015). However, we only selected materials with the same tonal information shared between spoken target words and phonological competitors to maximize the possibility for observing the phonological competitor effect. Nevertheless, future studies should fine-tune the design to investigate how different-level information (e.g., character frequency, the overlap numbers between phonological competitors and spoken target words, lexical tone) influence the phonological competitor effect in spoken Chinese word recognition.

In summary, findings of the present study provide robust evidence for the hypothesis that phonological information mediates shifts of visual attention to printed words in Chinese spoken word recognition, which is independent of orthographic information processing. Furthermore, we found that the mediating role of phonological information in spoken word recognition functions through the degree of phonological similarity between spoken target words and phonological competitors.

References

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

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.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *lme4: Linear mixed-effects models using Eigen and S4* (Version 1.1-7) [Computer software]. Retrieved from http://cran-r.project.org/package=lme4

Chen, H. C., d'Arcais, G. B. F., & Cheung, S. L. (1995). Orthographic and phonological activation in recognizing Chinese characters. *Psychological Research*, *58*(2), 144–153.

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology*, *6*, 84–107.

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

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 & Cognitive Processes*, *16*, 507–534.

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.

Hoosain, R. (1991). *Psycholinguistic implications for linguistic relativity: A case study of Chinese*. London, UK: Psychology Press.

Huettig, F., & Altmann, G. T. (2005). Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm. *Cognition*, *96*(1), B23–B32.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482.

Huettig, F., Quinlan, P. T., McDonald, S. A., & Altmann, G. T. (2006). Models of high-dimensional semantic space predict language-mediated eye movements in the visual world. *Acta Psychologica*, *121*(1), 65–80.

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

Leck, K. J., Weekes, B. S., & Chen, M. J. (1995). Visual and phonological pathways to the lexicon: Evidence from Chinese readers. *Memory & Cognition*, *23*(4), 468–476.

Malins, J. G., Gao, D., Tao, R., Booth, J. R., Shu, H., Joanisse, M. F., … & Desroches, A. S. (2014). Developmental differences in the influence of phonological similarity on spoken word processing in Mandarin Chinese. *Brain and Language*, *138*, 38–50.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*(1), 71–102.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, *10*(1), 29–63.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition*, *86*(2), B33–B42.

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.

Meng, X. (2014). *The influence of phonological information on spoken word recognition in Mandarin Chinese: Evidence from eye movements* (Master's thesis; in Chinese). Retrieved from www.cnki.com

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.

Nixon, J. S., Chen, Y., & Schiller, N. O. (2015). Multi-level processing of phonetic variants in speech production and visual word processing: Evidence from Mandarin lexical tones. *Language, Cognition and Neuroscience*, *30*(5), 491–505.

R Core Team. (2016). *R: A language and environment for statistical computing* (Version 3.3.3) [Computer software]. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org/

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Salverda, A. P., & Tanenhaus, M. K. (2010). Tracking the time course of orthographic information in spoken-word recognition. *Journal of*

Experimental Psychology: Learning, Memory, and Cognition, *36*(5), 1108–1117.

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 the visual-world paradigm. *Attention, Perception, & Psychophysics*, *78*(5), 1267–1284.

Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in Chinese word identification: A review of current discoveries and theoretical accounts. In C.-K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 11–46). Dordrecht, The Netherlands: Springer.

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

Weber, A., Melinger, A., & Lara Tapia, L. (2007). The mapping of phonetic information to lexical representation in Spanish: Evidence from eye movements. In J. Trouvain & W.J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS 2007)* (pp. 1941–1944). Dudweiler, Germany: Pirrot.

Weekes, B. S., Chen, M. J., & Lin, Y. B. (1998). Differential effects of phonological priming on Chinese character recognition. In C.-K. Leong & K. Tamaoka (Eds.), *Cognitive processing of the Chinese and the Japanese languages* (pp. 47–68). Dordrecht, The Netherlands: Springer.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology*, *97*(2), 259–268.

Zhang, Q., Chen, H. C., Stuart Weekes, B., & Yang, Y. (2009). Independent effects of orthographic and phonological facilitation on spoken word production in Mandarin. *Language and Speech*, *52*(1), 113–126.

Zhang, Q., & Weekes, B. S. (2009). Orthographic facilitation effects on spoken word production: Evidence from Chinese. *Language & Cognitive Processes*, *24*(7/8), 1082–1096.

Zhou, X., & Marslen-Wilson, W. (1994). Words, morphemes and syllables in the Chinese mental lexicon. *Language & Cognitive Processes*, *9*(3), 393–422.