Syllabic tone articulation influences the identification and use of words during Chinese sentence reading: Evidence from ERP and eye movement recordings

Yingyi Luo^{1,2} · Ming Yan³ · Shaorong Yan¹ · Xiaolin Zhou¹ · Albrecht W. Inhoff⁴

Published online: 27 August 2015 © Psychonomic Society, Inc. 2015

Abstract In two experiments, we examined the contribution of articulation-specific features to visual word recognition during the reading of Chinese. In spoken Standard Chinese, a syllable with a full tone can be tone-neutralized through sound weakening and pitch contour change, and there are two types of two-character compound words with respect to their articulation variation. One type requires articulation of a full tone for each constituent character, and the other requires a full- and a neutral-tone articulation for the first and second characters, respectively. Words of these two types with identical first characters were selected and embedded in sentences. Native speakers of Standard Chinese were recruited to read the sentences. In Experiment 1, the individual words of a sentence were presented serially at a fixed pace while eventrelated potentials were recorded. This resulted in less-negative N100 and anterior N250 amplitudes and in more-negative N400 amplitudes when targets contained a neutral tone. Complete sentences were visible in Experiment 2, and eve movements were recorded while participants read. Analyses of oculomotor activity revealed shorter viewing durations and fewer refixations on-and fewer regressive saccades to-

⊠ Yingyi Luo yingyi.luo@gmail.com

¹ Center for Brain and Cognitive Sciences and Psychology Department, Peking University, Beijing 100871, China

- ² Research Unit for Brain Science of Language, Inference, and Thought (BLIT) and Faculty of Science and Engineering, Waseda University, Tokyo, Japan
- ³ Department of Psychology, University of Potsdam, Potsdam, Germany

4

target words when their second syllable was articulated with a neutral rather than a full tone. Together, the results indicate that readers represent articulation-specific word properties, that these representations are routinely activated early during the silent reading of Chinese sentences, and that the representations are also used during later stages of word processing.

Keywords Lexical tone \cdot Neutral tone \cdot Articulation duration \cdot Syllabic tone \cdot Sentence reading \cdot Chinese

Knowledge of spoken language generally precedes learning to read, and converging evidence is indicating that readers use sound-defining properties of orthographic patterns during their processing. This occurs in languages with radically different orthographic script types, such as English and Chinese (see Perfetti, Liu, & Tan, 2005, and Rastle & Brysbaert, 2006, for reviews), presumably because sound codes assume a functional role during reading.

In his comprehensive and influential review, Frost (1998) argued that the high speed of sound-code use during visual word recognition indicates that the effective phonological form must be impoverished; that is, it is devoid of speechspecific features and is abstract in nature. This claim has been referred to as the minimality hypothesis, and classical models of phonological information use-for example, the influential dual-route cascade (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001)-share this representational assumption. Recent studies have shown, however, that readers of alphabetic languages use relatively detailed articulationspecific features—such as vowel duration, spoken syllable boundaries, and lexical stress-during visual word identification (Abramson & Goldinger, 1997; Ashby, 2006; Ashby & Clifton, 2005; Ashby & Martin, 2008; Ashby & Rayner, 2004; Ashby, Sanders, & Kingston, 2009; Huestegge, 2010; Lukatela, Eaton, Sabadini, & Turvey, 2004; Wheat, Cornelissen, Frost, & Hansen, 2010). Furthermore, recordings of eye movements have shown that relatively detailed suband supraphonemic information can be gleaned from (parafoveally visible) words before they are directly fixated during silent reading (Ashby & Martin, 2008; Ashby & Rayner, 2004). Contrary to the minimality hypothesis, readers of alphabetic scripts activate detailed sound-specific representations before individual words are fully identified.

Event-related potential (ERP) recordings for individually presented words have also revealed relatively early responding to phonological word properties. In Grainger, Kiyonaga, and Holcomb (2006), briefly shown homophone primes (e.g., brane-BRAIN) yielded less-negative N250 amplitudes on the target word than did the control primes (e.g., brant-BRAIN). This effect, emerging earlier than the canonical semantically related N400 effect (Holcomb & Grainger, 2006), was attributed to the more effective use of target phonology for lexical access with homophone primes. Moreover, ERP recordings have yielded strong evidence for the prelexical use of sub- and supraphonemic features (Ashby, 2010; Ashby & Martin, 2008; Ashby et al., 2009; Wheat et al. 2010). In Ashby et al. (2009), briefly presented-and subsequently masked-pseudoword primes with voiced or unvoiced final consonants (e.g., fap, or faz) were followed by words with either a congruent or an incongruent final consonant articulation (fap-fat [congruent] vs. fap-fad [incongruent]). Prime-target congruency modulated very-early ERP components under these conditions, with the congruent condition yielding less-negative amplitudes within 80 ms of target onset. Effects of supraphonemic features were also obtained with briefly presented (42 ms) masked primes in Ashby and Martin (2008), with less-negative amplitudes within 250-350 ms of target word onset when the prime and a target's first syllable matched (pi### of pilot) than when they mismatched (pi##### of pilgrim). Using a similar paradigm with visually matched items for the congruent and incongruent conditions, Ashby (2010) observed an even earlier effect of syllable congruency, in which the N100 amplitude (the 100- to

120-ms time window) was reduced whenm-201w1nem-2/T1359(an)-i21(o.2460000b[(),)-9997Td[(fi)12(rs)13)-2-Tf.489or owsh## er. br730(an)15(t)]TJ/T111Tf8.48974.227.Je.B/T111Tf8.48994.075.078.07k0(re)-6(y)-4,im21(01)]TJ.227.J/T111Tf8.489759770Td()Tjde(m

presented for 47 ms. Zhou and Marslen-Wilson (2009) postulated a framework in which phonological activation is not necessary for the linking of written forms to meaning. Instead, orthography can directly lead to the activation of both the phonological and semantic representations of individual morphemes, and morphemes and whole words are accessed in parallel. Furthermore, semantic activations can feed back to the corresponding phonological and orthographical representations. This claim is further supported by eye movement findings: Yan et al. (2009) reported a stronger semantic than a phonological parafoveal preview effect, and parafoveal semantic information is obtained very early (Yan, Risse, Zhou, & Kliegl, 2012; Zhou, Kliegl, & Yan, 2013); reliable phonological effects required both long preview durations and the high parafoveal processing efficiency afforded by highfrequency pretarget words (Tsai et al., 2012).

Our recent work (Yan, Luo, & Inhoff, 2014) suggested that phonological information is used relatively early during the foveal processing of Chinese compound words, and that the phonological code of a compound word was not simply a linear concatenation of the phonological codes of its constituent characters. Moreover, the results indicated that Chinese readers use articulation-specific sub- or supraphonemic word features, as do readers of alphabetic scripts, rather than abstract phonological forms, as has commonly been assumed. The study took advantage of lexically conditioned tonal variation in Chinese speech. Standard (i.e., Mandarin) Chinese is a tonal language in which four full tones are used on syllables to express lexical distinctions in speech. Moreover, it also includes a licensed tonal variant that is to be produced instead of the full-tone form on the same syllable under some conditions. This tonal variant, generally referring to as neutral tone (Chao, 1968), involves (a) shortening of the syllable articulation duration, as compared to its full-tone alternative; (b) a reduction of intensity (Cao, 1986; Lin & Yan, 1980; see H. Wang, 2008, for a review); and (c) the generation of a contextdependent fundamental frequency (i.e., F0) contour (Y. Chen & Xu, 2006). Furthermore, the occurrence of neutral tone is lexically conditional and arbitrary, and it is independent of such phonological constraints as tone sandhi in Chinese or allophonic variability in English. Neutral tones are applied to character-syllables that occupy a noninitial position within a multicharacter compound word (Y. Chen & Xu, 2006), and their use is dialect-specific. For instance, in Standard Chinese, 火 is articulated as the syllable "huo" with full-tone 3 when it is a single-character word (meaning "fire"), the first constituent of the compound 火柴 (meaning "matches"), or the second constituent in the compound 炉火 (meaning "fire in furnace"). But the syllable has to be articulated with a neutral tone when it is the second constituent in another compound, 柴火 (meaning "firewood"); in this case, a full-tone articulation would sound odd to a native speaker of Standard Chinese, but not to a speaker of southern dialects without the use of an equivalent neutral tone. Following Yan et al. (2014), we refer to a compound with a tone-neutralized syllable in Standard Chinese as a *neutral-tone word*. More critically, since most syllables that are articulated with a neutral tone are derivations from their full-tone origins, the corresponding morpheme/character form is thus associated with two different syllable pronunciations, one occurring when the syllable is articulated in isolation or when it is the first character of a compound word, and another occurring when it is articulated as the second syllable of a neutral-tone word.

In Yan et al. (2014), participants read sentences with neutral- and full-tone target words. The results showed that speakers of Standard Chinese spent less time viewing neutral-tone than full-tone words, and that this tonal effect was not observed for speakers of Chinese dialects who used full tones for the articulation of all target word syllables. Articulation-specific variation that was unrelated to a word's morphemic/semantic meaning thus might influence its ease of recognition. This implies that speakers of Standard Chinese did not generate a soundspecific representation of compound words through a linear concatenation of the constituent syllables (Perfetti et al., 2005; Zhou & Marslen-Wilson, 2009). Instead, they generated articulation-specific phonological forms that were lexically conditioned, and this occurred early during visual word recognition.

The effects of lexically conditioned syllabic tone articulation were not clear cut, however. Although speakers of Standard Chinese spent less time viewing neutral- than viewing full-tone words in Yan et al. (2014), which might suggest more effective processing of neutral-tone words, they also obtained less useful information from the next (parafoveal) word when a neutral- than when a full-tone target word was foveally viewed. Since diminished processing of a parafoveal word generally occurs when a fixated word is difficult to process, there is a seeming paradox concerning the effectiveness with which neutral-tone words are processed. To account for these seemingly discrepant findings, Yan et al. (2014) argued that the simulated articulation of a target word could be accomplished earlier for neutral-tone words, and that this accounted for the shorter neutral-tone target-viewing durations. On the other hand, the continued representation of a neutral-tone word could be weaker or less effective than the continued representation of a full-tone word, and this might have resulted in the less effective processing of the next word.

The main goal of the present study was to trace the time course of neutral-tone usage during Chinese compound word processing in order to dissociate early from late neutral-tone effects. In Experiment 1, ERPs were recorded when speakers of Standard Chinese read sentences that contained matched neutral- and full-tone two-character target words. According to Yan et al. (2014), when a target word is processed while being fixated, benefits of neutral-tone usage should be obtained for early ERP components, and costs might be found for later components. To further investigate neutral-tone usage not only before but also after a target word is identified during sentence reading, eye movements were recorded while sentences with neutral- and full-tone target words were read in Experiment 2. Readers of Standard Chinese were expected to spend less time viewing neutral- than full-tone words during first-pass reading, but this should not occur during target rereading if neutral-tone use during the later stage was relatively difficult. In addition, different types of spoken distractors were presented when the target words were viewed. Earlier work (Eiter & Inhoff, 2010; Inhoff, Connine, Eiter, Radach, & Heller, 2004) indicated that these distractors can influence the processing of a recognized word, and their deleterious effects should be greater when the representation of a word is weak or ambiguous.

Experiment 1

If articulation-specific features influenced early stages of word recognition, then the amplitudes of the N100 and N250 components, which have been shown to be sensitive to phonological activations (Ashby & Martin, 2008; Holcomb & Grainger, 2006), should be reduced for neutral- relative to full-tone word processing, assuming that lexical access for the neutral-tone words was more efficient (Yan et al., 2014). Subsequent use of word meaning could also be more effective for neutral-tone words, and this should result in a decreased N400 amplitude for these words. Alternatively, the N400 component could be sensitive to relatively late stages of target processing that were assumed to be more difficult for neutral-tone words in Yan et al. (2014). If this were the case, the amplitude of the N400 could be larger when neutral- than when full-tone words are read.

Method

Participants

A total of 32 students from universities in Beijing (18 female, 14 male) between 19 to 28 years of age (mean = 22) were paid to participate. They were right-handed and were naive regarding the purpose of the experiment. Eight of the participants were excluded from the statistical analysis due to excessive artifacts (see below). Eighty-five additional students were recruited to establish different types of norms for target words. All of the participants for this study were native speakers of Standard Chinese.

Material

Fifty-seven two-character compound words were selected whose second syllable consisted of a consonant-vowel sequence that assumed a neutral-tone articulation in Standard Chinese. For each of these neutral-tone target words, a closely matched full-tone two-syllable word was selected. The first character-syllable of a matched full-tone word was identical to the first character of each neutral-tone word, therefore matching the phonological neighborhood density (i.e., the number of words sharing the same initial syllable); the second characters of the full-tone and neutral-tone word pair were closely matched on lexical and orthographic properties (see Table 1), all Fs < 1. The two types of targets had identical dominant syntactic roles, according to the SUBTLEX-CH database (Cai & Brysbaert, 2010). In addition, the neutral-tone and full-tone words were rated 3.77 (SD = 0.83) and 3.66 (SD= 0.84) on a 5-point scale (by 12 participants) with regard to ease of imagination [F(1, 112) = 1.102, p = .296], and 16 other participants indicated that these words were acquired at the mean ages of 5.81 (SD = 1.22) and 6.02 (SD = 1.07) years [F(1, 112) = 0.939, p = .335], respectively (see Table 2). Besides, the neutral-tone and the full-tone target words were also closely matched with regard to syntactic categories, number of polyphones, morphological structure, and morphemic status (see the Appendix).

Acoustic data were also obtained to verify the tone difference between the two groups of selected words. To establish norms for the articulation of the full- and neutral-tone syllables, ten participants were recorded individually in the speech

 Table 1
 Mean word frequencies for pretarget, target, and posttarget words, together with their mean stroke numbers

	Neutral-Tone Condition		Full-Tone Condition	
	Frequency	Stroke Number	Frequency	Stroke Numbe
N – 1 (Ex	xp. 1)			
Mean	2.58	17.6	2.84	16.6
SD	1.85	5.8	1.24	5.5
N - 1 (Ex	(p. 2)			

Men 2i469(209.529546136258(ge1)]TJ2.96602)2257(20942626331809992487/2600264269

lab at Minzu University of China while reading aloud 114 simple sentences in Standard Chinese. Each sentence contained a target word in the middle position (e.g., the word *shi-huan* in the following example), with a frame, as in:

Ta /	shuo /	Target word [shi-huan] /	zhe-ge /	ci.			
he /	say /	Target word [order around] /	this-CLASSIFIER /	word			
'He said the word [order around].'							

These readers sat before a computer monitor on which the test sentences were displayed using the custom-written recording tool AudiRec. A Shure 58 Microphone was placed about 10–15 cm in front of them. The sampling rate was 48 kHz, and the sampling format was one-channel 16-bit linear.

The duration and intensity of the target words were measured to establish their reliability in capturing the tone neutralization (Y. Wang, 2004). ProsodyPro, a Praat script (Xu, 2013), was used to perform the initial acoustic analysis in Praat (Boersma & Weenink, 2005). On the basis of the waveform and spectrogram of each sentence, segmentation labels were marked manually to identify the boundaries of the target syllable. The duration and intensity measurements for marked segments (i.e., target syllables) were then automatically extracted. The results showed that the second syllable had a shorter articulation duration for neutral-tone than for fulltone words (233 vs. 271 ms), F(1, 112) = 8.003, p < .001, and the intensity of neutral-tone words was also marginally weaker (64.6 vs. 65.4 dB), F(1, 112) = 1.938, p = .055, indicating reliable articulation differences between the two types of target words (see Table 2).

In order to increase the number of sentences, and thus the number of observations, we used between-item sentence frames. Each member of an experimental neutral- and full-tone target word pair was embedded in a different contextually neutral sentence, which yielded 114 experimental sentences (see Fig. 1 for an example). All sentences were relatively short, with 10 to 16 syllables, and the neutral- and full-tone target words occupied matching locations within their

sentences—but in neither the first nor the last position. On average, the target was the third word in the sentence (its position ranged from two to four, SD = 0.7). The neighboring words preceding the targets—that is, the *pretarget words* were statistically not different between the neutral- and fulltone conditions with respect to word frequencies and numbers of strokes (see Table 1), ps > .3. The context preceding the target was constructed so that it would impose few—if any constraints. Thirty participants completed cloze predictability tests for the sentence segments up to the target words, and the ratings yielded identical numeric values of 1.4 (F < 1) for sentences with full- and reduced-tone syllables, indicating that the sentences were relatively easy to read and that articulatory variation did not matter. The next sentence was presented 1, 000 ms after a sentence was rated. The sentences of each list were divided into two equal-sized blocks, and a rest period was offered in between these blocks. Five warm-up sentences were presented at the beginning of each block.

EEG recordings

The electroencephalogram (EEG) scalp sites were selected according to the International 10–20 System, and tin electrodes were mounted in an elastic cap (Brain Products, Munich, Germany). The vertical electro-oculogram (EOG) was recorded supraorbitally of the right eye. The horizontal EOG was recorded from an electrode placed at the outer canthus of the left eye. All EEGs and EOGs were re-referenced offline to the average of the left and right mastoids. Electrode impedance was kept below 5 k Ω . The EEG and EOG were amplified using a 0.016- to 100-Hz bandpass and were digitized with a sampling frequency of 500 Hz. Twenty-five electrodes, which could adequately cover the principal sites of interest (see, e.g., Bridger, Bader, Kriukova, Unger, & Mecklinger, 2012; Scudder et al., 2014) were selected for the analyses. Each of the electrodes was assigned to one of

Data analysis

Trials contaminated by excessive movement artifacts (mean voltages exceeding $\pm 70 \,\mu$ V) were excluded before trials were averaged over the items of a particular condition. On average, 72 % of trials were accepted for the statistical analysis (41 trials for the neutral- and full-tone syllable targets). Loss of data was not evenly distributed across participants, and eight participants were excluded due to large numbers of rejected data points (>50 %).

ERPs for the remaining participants and for each experimental condition were epoched from 200 ms preonset to 800 ms postonset for each target word. The 200-ms preonset interval was chosen for baseline correction. The same pattern of ERP responses was obtained when the average EEG in the 100-ms interval postonset of the target word was used instead for baseline correction. In view of this, we report only the ERP results with the 200-ms preonset interval baseline. The ERP peak amplitude between 80 and 110 ms was measured to index the N100 component, and the averaged amplitudes in the 200- to 300-ms, 350- to 450-ms, and 500- to 700-ms time windows to index the N250, N400, and P600 wave components, respectively. To identify whether any conditional effect observed on target words was merely a carryover of the potential difference induced by the two types of preceding contexts, ERP analysis was also performed for the pretarget word. Two time windows, 350-450 and 600-800 ms, were mainly focused on for the N400 and late components, which might be contaminated by spillover effects from the waveforms observed for the next target word.

ERP recordings were aggregated over experimental items for each participant, and location-specific mean values were computed for each participant in the two syllabic tone articulation conditions. These values were analyzed using a linear mixed-effect model (LMM), implemented using the lme 4.0 library (Bates & Maechler, 2013; version 0.999999-4) in the R system for statistical computing (R Development Core Team, 2014; version 3.0.3). Three fixed factors were entered: Lexical Tone Type (full vs. neutral), Hemisphere location (left, right, midline), and the Anterior-to-Posterior continuum. For the statistical analyses, a difference contrast was used to determine the estimated size of the neutral- versus the full-tone effect. Two orthogonal Helmert contrasts were applied to the Hemisphere factor: a primary contrast that compared the left with the right hemisphere (hemisphere contrast), and a secondary contrast that compared the midline location with the mean of the two lateralized locations. Since there were no categorical differences between the five anterior-to-posterior recording sites, the five regions along the anterior-posterior continuum were coded numerically from 1 to 5, and this predictor was centered to remove potential collinearity. The Subject-specific intercept was used as a random factor, and each time segment was analyzed separately. Additional supplementary analyses were applied to each segment in order to examine potential carryover of an earlier wave component onto the subsequent component. For this, the component in the 600- to 800-ms time window on the pretarget word was added as a covariate (predictor) to the statistical model for the N100 component on the target word, and the N100, N250, and N400 components were also added, respectively, as covariates (predictor) to the statistical models for their following components.

Estimated effect sizes (*b* values), standard errors (*SE*), and significance levels are reported in the text. Due to the large number of trials, the *t* distribution approximated a normal distribution, and all values of t > 1.96 were considered significant. Figures were created using the ggplot2 package (Wickham, 2009).

Results

The full waveforms for the 25 electrodes during the pretarget and target words are shown as a function of syllabic tone articulation in Figs. 3 and 4, respectively.

Pretarget words

N400 No significant effect was found for the analysis of N400 amplitudes, all $t_{\rm S} < 1.1$.

Late component In the 600- to 800-ms time window, a larger positivity was observed for the neutral- than for the fulltone condition, $b = 0.62 \mu V$, SE = 0.27, t = 2.32, even though the pretarget words for the two types of targets did not differ in word frequency or stroke number. We found no other reliable ERP difference in this time window.

Target words

N100 Peaks for the N100 component are shown as a function of tone articulation and topographic location in Fig. 5. The N100 amplitudes were less negative for neutral- than for full-tone target words, and the effect was reliable, $b = 0.29 \mu$ V, SE = 0.11, t = 2.74. Two topographic effects were also reliable, due to more negative values for the center location than for the two lateral locations, $b = -0.32 \mu$ V, SE = 0.11, t = -2.87, and decreases in negativity along the anterior-posterior axis, $b = 0.46 \mu$ V, SE = 0.04, t = 12.35.

Supplementary analyses with the late-component amplitudes on the pretarget words as a covariate showed that, despite a notable impact from the previous context on the N100 component, $b = 0.12 \mu$ V, SE = 0.02, t = 4.57, N100 effects of tone type remained significant when the influence of pretarget words was removed, $b = 0.51 \mu$ V, SE = 0.25, t = 2.07. This indicates that



Fig. 3 Event-related potential waveforms for the pretarget word

less-negative amplitudes for neutral- than for full-tone words cannot be attributed to carryover from the prior context.

N250 Means for the N250 component are shown as a function of tone type and topographic location in Fig. 6. As can be seen in the depiction of the full waveforms in Fig. 4, the N250 component occurred between a trough at around 200 ms and a spike at around 400 ms, which is consistent with the N250 properties specified in prior work (Hoshino, Midgley, Holcomb, & Grainger, 2010; Morris, Frank, Grainger, & Holcomb, 2007; Timmer & Schiller, 2012). The main effect of lexical tone was highly significant, with less negative values for neutral- than for full-tone targets, $b = 0.34\mu$ V, SE = 0.11, t = 3.00. This was due to a robust interaction of tone articulation with the anterior–posterior axis, $b = -0.25\mu$ V, SE = 0.08, t = 3.12. As can be seen in Fig. 6, the effects of tone type were relatively large for the anterior and mid-anterior locations, and they decreased almost linearly from anterior to posterior locations.

The N250 analysis also revealed two topographic effects. Center recordings were less negative than recordings from the left and right hemicortices, $b = 0.41 \,\mu\text{V}$, SE = 0.12, t = 3.46. The laterality effect (difference between the right- and left-hemisphere recording sites) interacted with anterior–posterior locations. Anterior locations yielded less-negative right- than left-hemisphere values, and this was reversed for posterior locations, $b = -0.21 \,\mu\text{V}$, SE = 0.10, t = 2.12.

To examine potential N100 carryover effects, supplementary analyses with N100 recordings as a covariate revealed a highly reliable influence from the N100, $b = 0.12 \mu$ V, SE =0.04, t = 3.17. Nevertheless, the N250 effects were virtually unchanged: The tone type effect and the interaction of tone type with the anterior–posterior continuum were again highly significant, $b = 0.31 \mu$ V, SE = 0.11, t = 2.75, and $b = -0.24 \mu$ V, SE = 0.08, t = 3.07, respectively.



Fig. 4 Event-related potential waveforms for the target word

N400 Means for the N400 component are shown as a function of tone type and topographic location in Fig. 7. As can be seen, the mean amplitudes were distinctly more negative for neutral- than for full-tone targets, $b = -0.60 \mu$ V, SE = 0.13, t = 4.62. One topographic effect was reliable, with more negative center than lateral amplitudes, $b = -0.67 \mu$ V, SE = 0.14, t = -4.84. No other effect approached significance.

The inclusion of the N250 component as a covariate revealed a highly reliable covariate effect, $b = 0.34 \mu$ V, SE = 0.04, t = 8.12. The extraction of N250 variance further increased the size of the target type effect, $b = -0.72 \mu$ V, SE = 0.13, t = 5.73, and of the center–lateral difference, $b = -0.81 \mu$ V, SE = 0.13, t = 6.06.

P600 P600 amplitudes also yielded a significant effect of lexical tone, with more-negative amplitudes for neutral-tone words, $b = -0.32 \,\mu\text{V}$, SE = 0.14, t = -2.19. In addition, one topographic effect was reliable, due to increases in negativity from anterior to posterior locations, $b = -0.14 \,\mu\text{V}$, $SE = -0.14 \,\mu\text{V}$, SE = -0.1

0.05, t = -2.80. No other effect was reliable. Inclusion of the N400 component in the statistical model revealed substantial carryover, $b = 0.80 \mu$ V, SE = 0.03, t = 28.09. When carryover was factored out, the effect of target type was no longer reliable, $b = 0.14 \mu$ V, SE = 0.10, t = 1.38.

Discussion

EEG recordings revealed less-negative N100 amplitudes for neutral- than for full-tone two-character Chinese compound words, and a corresponding N250 effect for anterior recording locations for native speakers of Standard Chinese. This effect was reversed for the N400 component, which was more negative for neutral-tone targets. Despite the well-matching lexical properties between the two conditions, the reading of pretarget words yielded discrepant ERP responses for the neutral- and the full-tone conditions at the late time window. However, the N100 effect observed on target words could not be solely due to spillover effects from the processing of the





previous context, since it remained significant when the contribution of the ERP waveforms on pretarget words was taken into account. Similarly, the N250 and N400 effects were not due to carryover from the preceding ERP component. The N100 and N250 effects of target type are in empirical disagreement with the minimality hypothesis; they are, however, in line with studies using alphabetic scripts, according to

which visual w the use of articu word recogniti type on N400 of Standard C also during la the word was



regression model

nition during silent reading entails cific features during early stages of





Fig. 7 N400 component. Negative values are plotted upward, and SEs were computed from the residuals of the regression model

The temporal properties of the two early effects of syllabic tone articulation, a broadly distributed N100 effect and an anterior N250 effect, are in general agreement with the timeline of prior work that examined ERPs in response to phonemic and supraphonemic manipulations with English text. As we noted earlier, briefly presented phonetic or syllabic primes that matched or mismatched the beginning phonetic or syllabic segment of English target words yielded robust N100 and N250 effects, with less negativity for matching than for mismatching prime-target pairs (Ashby, 2010; Ashby & Martin, 2008; Ashby et al., 2009).

The topographic distribution of the two early N100 and N250 ERP effects in Experiment 1 can also be reconciled with prior work. The N80-180 effect was not confined to specific recording sites, and broadly distributed early sub- and supraphonemic ERP effects have been reported in the literature (Ashby, 2010; Ashby & Martin, 2008; Ashby et al., 2009). Larger articulation-specific N250 effects for anterior than for posterior recording locations-which appears to differ from the more general N250 effect in Ashby and Martinmatched Grainger et al.'s (2006) topographic effects, in which homophone primes yielded less negative N250 components than control primes for anterior but not for posterior recording locations. Overall, the time course and topographic properties of the two early ERP components in the present study are thus in reasonably good agreement with prior work. Note that the N250 effect here was not attributed to the canonical P200 component (Federmeier & Kutas, 2005) because of the delayed time course as well as the waveform: Whereas the P200 component is usually identified as a "trough" (positivity) centered at around 200 ms, the N250 effect in this study occurred at the ascending limb following the trough.

The direction of the early ERP effects in related prior work can thus be used to constrain the interpretation of the present N100 and N250 effects. Specifically, ERP components were less negative with homophonic and matching sub- and supraphonemic primes than with control or mismatching primes (Ashby, 2010; Ashby & Martin, 2008; Ashby et al., 2009; Grainger et al., 2006), which suggests that reduced negativity indexes more effective processing. Analogously, the finding of less negativity for neutral than for full-tone targets in the present study appears to reflect more effective early processing of neutral-tone targets. This early processing could hardly be attributed to the phonological representations at the constituent morpheme/character level, since the two groups of compound words were carefully matched in terms of their constituent morphemes. Rather, it must be due to the rapid use of phonological properties for the full word. As such, the results of Experiment 1 provide direct evidence for Yan et al.'s (2014) key claim, that the initial stages of processing are more effective for neutral- than for full-tone words.

Experiment 1 also revealed more-negative N400 amplitudes for neutral- than for full-tone words, and the direction of these effects appears to be inconsistent with the direction of the N100 and N250 effects. That is, whereas the two early ERP components indicate that less effort was required for the lexical access of neutral-tone targets, the N400 component indicates, by contrast, that neutral-tone target processing required more effort at a later point in time. This seeming reversal of the neutral-tone effects over time is also consistent with Yan et al. (2014), in which the processing of neutral-tone target words diminished the uptake of information from the next words(s), relative to full-tone words.

What accounts for the reversal of our neutral-tone effects? In their comprehensive review of N400 effects, Kutas and Federmeier (2009, 2011) noted that N400 amplitudes increased with items' difficulty and lack of familiarity. Related work suggests that this component may index late stages of word processing, when integrated multimodal lexical representations are constructed from phonological and orthographic forms (Laszlo & Federmeier, 2011), and when semantic processing converges upon a specific word meaning (Wlotko & Federmeier, 2012). Hence, one viable account for more negative N400 amplitudes for neutral-tone words may be conflicting articulations during early and later stages of neutral-tone target processing. Whereas the tonal features of full-tone targets did not differ at the morpheme and wholeword levels, the tonal features of neutral-tone words were level-specific. Integration of the two corresponding representations could thus have been more difficult for neutral-tone targets with incongruent second-syllable articulations at the character and word levels than for full-tone targets with congruent second-syllable articulations at the two levels, resulting in a larger N400 for neutral-tone words.

Our finding for lexical tone neutralization in Standard Chinese is not in accordance with the canonical frameworks of Chinese compound word recognition, in which the form representations of compound words consist purely of those of the individual morphemes (e.g., Taft & Zhu, 1995; Perfetti et al., 2005; Zhou & Marslen-Wilson, 2009). But it generally conforms to models that incorporate interactive activations over the nodes or levels in a hierarchical language network and allow for the mutual influence of the representations at different levels (see Norris, 2013, for a review): facilitation when the representations are congruent, and inhibition when they are not.

Experiment 2

ERP responses to neutral- and full-tone words revealed that native speakers of Standard Chinese could rapidly represent lexically conditioned articulation-specific features of sequentially presented words, and that an initial advantage for neutral-tone words was followed by subsequent processing difficulties. In Experiment 2 we sought to generalize these findings to normal reading conditions, and sought to elucidate the nature of neutral- and full-tone targets' postlexical processing. For this, oculomotor activity was recorded while participants were reading fully visible sentences that contained neutral- and full-tone target words—the assumption being that initial oculomotor responding would reveal neutral-tone benefits, as occurred in Yan et al. (2014). Measures that index later stages of target processing were expected to reveal neutral-tone costs. To discern the nature of the targets' postlexical processing, task-irrelevant auditory distractors were presented when the eyes moved onto the target words (Inhoff et al., 2004)—the assumption being that neutral-tone words would be more susceptible to sound-based distraction because their phonological representations at the morpheme and word levels were incongruent.

In Inhoff et al. (2004), participants heard an irrelevant auditory distractor (AD) word when their eyes moved onto a visual target word during reading. The ADs were either identical, phonologically similar, or dissimilar to the target. Two key findings emerged. During target viewing, deleterious AD effects were smaller when the AD and the target word were identical than when they were nonidentical, indicating that at this point identity-defining information dominated the AD effects. The AD effects differed during posttarget reading. Here, ADs that were phonologically similar to the previously viewed target interfered with posttarget viewing more than the identical and unrelated ADs. This effect was attributed to the interference of phonologically similar ADs with targets' postlexical representations (Eiter & Inhoff, 2010).

Experiment 2 used a variant of Inhoff et al.'s (2004) contingent-speech technique to examine AD effects on the reading of target and posttarget words. Three types of ADs were presented. All were full-tone syllables that were either identical, similar, or dissimilar in articulation to the first (fulltone) syllable of a fixated full- or neutral-tone target. Since we manipulated the relationships between ADs and the targets' first syllables, the same identical, phonologically related, and unrelated full-tone AD could be presented with a corresponding neutral- or full-tone member of a target pair. If the AD effects for spoken syllables were similar to the AD effects for spoken words (e.g., Eiter & Inhoff, 2010; Inhoff et al., 2004), then hearing an identical AD syllable should be less distracting than hearing a phonologically similar or dissimilar syllable when the target was viewed, and having heard a phonologically similar AD should be more distracting than having heard an identical or unrelated AD when the posttarget word was viewed. Moreover, if the postlexical phonological representation of neutral-tone targets was weak because the secondsyllable articulation was incongruent between the morpheme and lexical levels, then neutral-tone targets should be more vulnerable to distraction by a phonologically similar AD during posttarget reading.

Method

Participants A group of 50 undergraduate and graduate students (19 to 28 years old) from universities in Beijing participated. They were all native speakers of Standard Chinese, were naive regarding the purpose of the experiment, and had normal or corrected-to-normal vision. Six were excluded from further analysis due to malfunctions of equipment or problematic responses to the rating task.

Material The same target words were used as in Experiment 1. To obtain steady oculomotor measures, sentences were lengthened so that the critical words were not positioned at the beginning but nearer the middle of the sentence. The target was the eighth word of a sentence, on average. However, to avoid the potential influence of the length of word N-1--that is, the number of constituent characters in Chinese-upon the oculomotor activities responding to word N, the pretarget words were occasionally reselected such that all of them consisted of two characters. The pretarget and posttarget words were closely matched between the neutral- and fulltone conditions in terms of word frequency and number of strokes, all $p_{\rm S} > .1$ (see Table 1). The context following the posttarget words was also elaborated to create more harmonious sentence endings. The length of the sentences now ranged from 13 to 25 characters, with a mean of 18 characters.

Three AD types were used: an identical AD that consisted of the first spoken syllable of a target word, a phonologically similar AD that consisted of a syllable with a similar segmental and tonal structure (e.g., the visual target "shi[3]" was paired with the AD "chi[3]"), and a dissimilar AD that was unrelated to the target (the visual target "shi[3]" paired with the AD "ma[1]"). All syllables were spoken individually by a native male speaker (the third author) of Standard Chinese with clear diction. A directional microphone (RØDE NT1-A) was used to record the words at 16 bits/41.1 kHz. The duration of the spoken syllables ranged from approximately 102 to 300 ms, with a mean of 191 (SD = 29), 194 (SD = 21), and 195 ms (SD = 21) for the identical, similar, and dissimilar types, respectively, Fs < 1.

Three identical lists, each with the same 114 experimental sentences, were constructed. A target word paired with one type of AD on one list was paired with different AD on another list, with the constraint that one third of full- and neutral-tone targets (n = 19) would be presented with a different AD type on each list. The conditions on each list were pseudorandomized so that no more than three sentences from the same AD and target type condition would appear consecutively.

The setup of the contingent-speech technique was similar to that used in previous studies (e.g., Eiter & Inhoff, 2010; Inhoff et al., 2004). An invisible boundary was set to coincide with the left boundary of the target word's first character within a sentence. Only the first crossing of (or landing on) the invisible boundary initiated the presentation of the AD. Subsequent boundary crossings did not result in the re-

prkHapres(ii)th0(oxa)h0(Koe)tis)2/e(poe) Rpst28(tha)h2(0))1/ef)tas7.67.121/ef)tas7.67.1251999Apntdo508(sn)15(i)18(al)]TJ/T14.723Tf3.8810000

irrelevant—that is, that it did not assist with sentence comprehension and that they would not need to report it.

Measurement and data analysis Pretarget, target, and posttarget viewing were analyzed. From the large number of oculomotor measures that could be extracted (see Inhoff & Radach, 1998; Rayner, 1998), three routinely used viewing duration measures were computed to index the time course of word recognition: first-fixation duration, gaze duration, and total viewing duration. The first-fixation duration consisted of the duration of the first fixation on a word when it was reached from a preceding word in the sentence, irrespective of the number of fixations on the word, and gaze duration comprised a word's first-fixation duration plus the time spent refixating it during first-pass reading until another word was fixated. These

 $two\ measur(a) 383 me77 Inr8(19(e1) 18(y) 6371((t) 14(ed) - 3t(two) 8) - 1d(s) 1x0(f) - 356(th) \\ e8(y) - 3suc(re) 19(rre) \\ s(me77) \\ w56(i8(h2(t)] \\ TJ-.167 \\ TG-.167 \\ TG-$

effect did not reach significance for first-fixation durations (-

was used to determine the time course and nature of the tonal effects. The key findings revealed shorter first-pass viewing durations and lower refixation probabilities for neutral- than for full-tone targets. Oculomotor measures that are sensitive to relatively late stages of lexical processing were also influenced by the tone articulation property, with shorter total viewing durations and fewer incoming regressions for neutral targets. The effects of lexically conditioned syllable articulation on first-pass target viewing replicate Yan et al. (2014), and they are consistent with the results of Experiment 1. They provide further evidence for the early use of articulation-specific features during visual word recognition, and this is inconsistent with the minimality hypothesis.

However, in Experiment 2, the impact of lexical tone type was manifest not only in oculomotor measures that were sensitive to early stages of processing, but also in measures that were sensitive to late stages-that is, regression rate and rereading time-suggesting a rapid and sustained influence of articulation-specific word properties on the representation of target words. The effects of AD syllables on target and posttarget viewing were quite small, and only the comparison of identical with nonindentical ADs yielded a statistically reliable difference that was opposite to the expected findings; that is, posttarget viewing durations were longer when the spoken syllable was identical to a previously fixated target word's first syllable than when it was phonologically similar or dissimilar. Moreover, this reversed identity effect applied equally to neutral- and full-tone targets.

Shorter gaze durations are associated with more effective lexical processing in models of reading (Engbert et al., 2005; Reichle et al., 2009), and the shorter first-pass target viewing durations for neutral- than for full-tone targets indicate that recognition of a two-character target word was more effective when the covert articulation of its second syllable involved the production of a neutral tone. This appears to be analogous to effects of vowel articulation duration with alphabetic text. Here, responses to individually presented words and gaze durations during sentence reading (Abramson & Goldinger, 1997; Lukatela et al., 2004; see also Huestegge, 2010) are shorter when a target word's articulation duration was short ("deaf") than when it was long ("deal"), which was attributed to the use of speech-like codes for lexical access. Specifically, the generation of a speech-like code for lexical access was assumed to take less time when the vowel duration was short-hence, faster lexical access for words with short vowel durations. Analogously, viewing durations may have been shorter for neutral- than for full-tone targets in Experiment 2 because the shorter articulation duration of neutral-tone words resulted in faster generation of a lexical access code.

The first-pass (gaze duration) viewing duration difference between full- and neutral-tone words was 18 ms. When targets' rereading time was included, as indexed by the total viewing duration, the difference was 33 ms, indicating that the advantage of neutral-tone words was not diminished during late stages of target processing. Moreover, the rate of incoming regressions, a relatively direct measure of a target's postrecognition processing (Reichle et al., 2009), indicated that postlexical processing was less demanding for neutral-tone words. Contrary to the results of Experiment 1 and Yan et al. (2014), these findings suggest that the initial benefits for neutral-tone targets were not reversed during later (postlexical) stages of target recognition. Why did postlexical effects in the current study differ from Yan et al. (2014)? In Experiment 2, ADs were presented during target viewing, and this could have changed the dynamics of the targets' postlexical processing. This view is elaborated in the General Discussion.

The pattern of AD effects also did not match expectations. Relative to the identical AD, nonidentical (similar and dissimilar) ADs yielded only a very weak and nonsignificant interference effect during target viewing. Moreover, in seeming contrast to prior findings, posttarget viewing durations revealed longer viewing durations for the identical AD condition than for the two nonidentical ADs, and similar and dissimilar ADs did not differ. Why did the AD manipulation fail to yield some of the expected effects that have been observed in studies with English stimuli (Eiter & Inhoff, 2010; Inhoff et al., 2004)? One potentially critical difference is that the target words and ADs belonged to the same lexical category in earlier work-that is, both were intact words that conveyed matching or mismatching meanings (Eiter & Inhoff, 2010; Inhoff et al., 2004). This was not the case in Experiment 2, where the targets were two-syllable Chinese words, and the ADs were single syllables that were or were not related to the targets' first syllables. Under these conditions, an identical AD syllable might have been perceived as overlapping with-or being merely similar to-the target word, rather than being identical to it. There was even less overlap between the similar and dissimilar ADs and the full target words, and these syllables might have been perceived as dissimilar. If so, the difference between the identical AD and the two nonidentical ADs during posttarget viewing could be viewed as evidence for overlap-or similarity-interference. This view must be considered tentative, however, and requires further investigation.

It should also be noted that the average duration of the spoken AD was typically as long as—or even longer than the target viewing durations in studies with English text, and readers could have heard the AD during posttarget-word viewing on some trials. In Experiment 2, the articulation duration of the monosyllabic distractors was much shorter (mean = 193 ms), and hearing of the AD was generally considerably shorter than the target viewing duration. Thus, the influence of hearing acoustic AD was much more confined to target viewing in the present study. Nevertheless, the target–AD relationship influenced posttarget viewing, indicating that perception of irrelevant—but overlapping or similar—speech interfered with the continued representation

structure may be more difficult and take more time. An analogous account has been offered to explain lexical stress effects during English word recognition. Specifically, longer viewing durations and more refixations for words with more stressed syllables were tentatively attributed to an increase in the difficulty with which suprasegmental phonological units were assembled, rather than to increased speech duration per se (Ashby & Clifton, 2005).

Tone neutralization at the word level in the present work should be distinguished from the allophonic variants of spoken words (e.g., "preddy" for "pretty") and from the neutralization or loss of segmental vowel distinctions in casual spoken English (e.g., "libr'y" vs. "library" or "p'lice" vs. "police"). Studies of spoken word recognition have indicated that these allophonic variants of spoken words map directly onto underlying phonological representations (M^cLennan, Luce, & Charles-Luce, 2005), and that both forms are lexically represented. As a result, these variants can be recognized as effectively as their canonical forms, and the frequency of the variant used in spoken language accounts for the ease with which they are recognized in auditory lexical decision tasks (Ranbom & Connine, 2007; Ranbom, Connine, & Yudman, 2009). Moreover, recognition of a simplified spoken-word variant was reported as not being more effective than the recognition of its canonical counterpart (LoCasto & Connine, 2002). In this study, by contrast, tone neutralization was lexically conditioned, and it did not constitute a licensed allophonic articulation of a target word. Moreover, tone neutralization conveyed distinct benefits during early stages of visual word recognition.

We thus attribute the early effects of neutral-tone syllables to the use of articulation-specific features for lexical access. In several models of Chinese word recognition (Perfetti & Tan, 1999; Tan & Perfetti, 1997; Zhou & Marslen-Wilson, 2009), graphemic word forms are assumed to map deterministically-and thus rapidly-onto corresponding phonological forms, which should consist of articulation-specific features according to Experiments 1 and 2. Thus, two word forms contribute to early stages of word recognition, and the phonological-articulatory code influences the ease and speed of lexical access when it conveys useful hints for the identification of a particular word that are not provided by the orthographic code. The benefit of the phonological-articulatory code could be larger for neutral- than for full-tone targets because these phonological hints are available earlier in time to native speakers of Standard Chinese.

The effects of lexical tone properties on early stages of visual word recognition during sentence reading in Experiments 1 and 2 disagree, however, with theoretical conceptions according to which the phonological form of Chinese compound words is assembled from the spoken forms of individual constituent syllables (Perfetti et al., 2005; Zhou & Marslen-Wilson, 2009). Had this been the case, no difference

should have been observed in the early stage of lexical access between neutral- and full-tone target words, or neutral-tone targets should have hampered the initial stages of word recognition because syllable articulations differed at the morpheme and word levels. Consequently, our findings for native speakers of Standard Chinese favor an architecture in which orthographic word forms can inform the articulation of constituent syllables early on, and in which these top-down constraints shape articulation quickly enough to further influence word identification.

In Experiment 2, analyses of regressions to targets and of target rereading times also indicated that neutral-tone targets were processed more effectively during late stages of word recognition. This appears to disagree with Experiment 1 and Yan et al. (2014), in which increased N400 negativities for neutral-tone targets and the diminished extraction of information from the next word when neutral-tone targets were fixated suggested that the late stages of neutral-tone target processing were relatively difficult. How can this discrepancy be reconciled? Did small differences in the materials alter the late stages of target processing across experiments? Although the sentence frames did differ slightly across experiments, ratings of the Experiment 1 and 2 sentence difficulties indicated that sentences with neutral- and full-tone targets were read with equal ease in both experiments. Sentence contexts for the full and neutral targets were also carefully controlled in Yan et al. (2014). Hence, it is unlikely that small differences in the to-beread materials between experiments changed the late stages of target word processing.

Experiment 2 also differed from Experiment 1 and Yan et al.'s (2014) experiments in that ADs were presented during target viewing. The effect of these distractions on early stages of target processing may have been negligible, since early processes may be computationally enclosed (Fodor, 1983). Later stages could have been modulated by hearing of the AD syllables during target viewing. In Yan et al. (2014), the more difficult processing of neutral-tone targets during late stages was assumed to have diminished the uptake of information from the next word. Similarly, the more difficult processing of neutral-tone words during late stages could have diminished their susceptibility to AD effects. That is, the costs that were incurred by more difficult processing of neutral-tone targets could have been offset by the benefits that were derived from diminished AD interference. This would explain why initial processing benefits for neural-tone words were not reversed during the later stages of processing in Experiment 2, and also why the neutral- and full-tone target conditions did not differ during posttarget viewing.

Another potential discrepancy between the experiments appears to be the timelines of the tone neutralization effect. In Experiment 1, ERP peaks differed as early as within 100 ms after onset of the neutral- and full-tone targets; yet, in Experiment 2, first-fixation durations for target words, assumed to index early stages of word recognition, did not reveal a robust corresponding effect. Experiment 1 revealed, however, numerically shorter first-fixation durations, and a similar numeric trend was obtained in Yan et al.'s (2014) first experiment. The neutral-tone advantage was robust, however, for a larger group of Yan et al.'s (2014) participants, and the consistency of the first-fixation duration advantage for neutral-tone targets across experiments and studies indicates that tone neutralization does influence oculomotor measures that index early stages of visual word recognition.

Could early tone neutralization have arisen from a subtle confounding of morphological and/or semantic word properties with tone neutralization? This account seems unattractive in view of Yan et al.'s (2014) findings. As we noted earlier, the eve movements of speakers of Standard Chinese vielded robust tone neutralization effects, but the identical materials did not yield any difference between neutral- and full-tone targets for speakers of Southern dialects who did not use tonal neutralization for any of the target words (Yan et al., 2014). If tonal neutralization was confounded with the words' morphological/semantic properties, then dialect should not have mattered, and speakers of Southern dialects should have shown the same effect pattern as native speakers of Standard Chinese. Future study should further address the commonality and variability of visual word recognition among different dialect speakers, specifically with regard to time course, in order to investigate how different types of information dynamically interact to accomplish the lexical representation.

Together, Experiments 1 and 2 provide converging evidence for the use of detailed spoken-language properties at the whole-word level during visual word recognition, and this appears to involve the use of articulatory gestures or the ease with which suprasegmental phonological units were assembled. The use of spoken-language properties during visual word recognition with alphabetic and Chinese scripts suggests a modification of the universal phonological principle: Rather than using abstract phonological knowledge, readers use articulation-related properties of words during visual word recognition across script types (see also Perfetti, 2003, for a

- Ashby, J., Sanders, L. D., & Kingston, J. (2009). Skilled readers begin processing subphonemic feature by 80 ms during visual word recognition. *Biological Psychology*, 80, 84–94.
- Barr, D. J., Levy, R., Scheepers, C. S., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., & Maechler, M. (2013). lme4: Linear mixed-effects models using S4 classes (R package version 0.999999-4). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://cran. r-project.org/package=lme4
- Boersma, P., & Weenink, D. (2005). Praat: Doing phonetics by computer [Computer program]. Retrieved from www.praat.org
- Bridger, E. K., Bader, R., Kriukova, O., Unger, K., & Mecklinger, A. (2012). The FN400 is functionally distinct from the N400. *NeuroImage*, 63, 1334–1342.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS ONE*, 5(e10729), 1–8. doi:10.1371/journal.pone.0010729
- Cao, J. (1986). Putonghua qingsheng yinjie texing fenxi. Applied Acoustics, 5, 1–6.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.
- Chen, B., & Peng, D. (2001). The time course of graphic, phonological and semantic information processing in Chinese character recognition. Acta Psychologica Sinica, 33, 1–6.
- Chen, Y., & Xu, C. (2006). Production of weak elements in speech: Evidence from f0 patterns of neutral tone in standard Chinese. *Phonetica*, 63, 47–75.
- Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. C. (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256. doi:10.1037/ 0033-295X.108.1.204
- Eiter, B. M., & Inhoff, A. W. (2010). Visual word recognition during reading is followed by subvocal articulation. *Journal of E perimental Psychology: Learning, Memory, and Cognition, 36*, 457–470.
- Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review*, 112, 777–813. doi:10.1037/0033-295X.112. 4.777
- Federmeier, K. D., & Kutas, M. (2005). Aging in context: Age-related changes in context use during language comprehension. *Psychophysiology*, 42, 133–141.
- Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PLoS ONE*, 6, e25782. doi:10.1371/ journal.pone.0025782
- Fodor, J. A. (1983). *The modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.
- Frost, R. (2012). A universal approach to modeling visual word recognition and reading: Not only possible, but also inevitable. *Behavioral* and Brain Sciences, 35, 310–329.
- Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin*, 123, 71–99.
- Grainger, J., Kiyonaga, K., & Holcomb, P. J. (2006). The time-course of orthographic and phonological code activation. *Psychological Science*, 17, 1021–1026.
- Henrich, K., Alter, K., Wiese, R., & Domahs, U. (2014). The relevance of rhythmical alternation in language processing: An ERP study on English compounds. *Brain and Language*, 136, 19–30.
- Holcomb, P. J., & Grainger, J. (2006). On the time course of visual word recognition: An event-related potential investigation using masked repetition priming. *Journal of Cognitive Neuroscience*, 18, 1631– 1643. doi:10.1162/jocn.2006.18.10.1631
- Hoosain, R. (2002). Speed of getting at the phonology and meaning of Chinese words. In H. S. R. Kao, C.-K. Leong, & D.-G. Gao (Eds.),

Cognitive neuroscience studies of the Chinese language (pp. 129–142). Aberdeen, Hong Kong: Hong Kong University Press.

- Hoshino, N., Midgley, K. J., Holcomb, P. J., & Grainger, J. (2010). An ERP investigation of masked cross-script translation priming. *Brain Research*, 1344, 159–172. doi:10.1016/j.brainres.2010.05.005
- Hsu, C.-H., Tsai, J.-L., Lee, C.-Y., & Tzeng, O. J. (2009). Orthographic combinability and phonological consistency effects in reading Chinese phonograms: An event-related potential study. *Brain and Language*, *108*, 56–66. doi:10.1016/j.bandl.2008.09.002

Huestegge, L. (2010). Effects of rowel length on gaze durations in silent and oral reading. *Jour al of Eye 1 comment Preearch*, 2(5), 5:1–18. Retrieved from www. **an org** of ne/3 5/

Inhoff, A. W., Connine, C., Certori, and h. 2004 h Uev.D. (2004). Phonological representation of words in working me hory during sentence reading.

- Perfetti, C. A., & Tan, L. (1999). The constituency model of Chinese word identification. In J. Wang, A. Inhoff, & H.-S. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 115–134). Mahwah, NJ: Erlbaum.
- R Development Core Team. (2014). R: A language and environment for statistical computing (Version 3.0.3). Vienna, Austria: R Foundation for statistical computing. Retrieved from www.R-project.org
- Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory* and Language, 57, 273–298. doi:10.1016/j.jml.2007.04.001
- Ranbom, L. J., Connine, C. M., & Yudman, E. M. (2009). Is phonological context always used to recognize variant forms in spoken word recognition? The role of variant frequency and context distribution. *Journal of E perimental Psychology: Human Perception and Performance*, 35, 1205–1220. doi:10.1037/a0015022
- Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognition*, 53, 97–145. doi:10.1016/j.cogpsych.2006.01.002
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124, 372–422. doi: 10.1037/0033-2909.124.3.372
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1–21. doi:10.3758/PBR.16.1.1
- Scudder, M. R., Federmeier, K. D., Raine, L. B., Direito, A., Boyd, J. K., & Hillman, C. H. (2014). The association between aerobic fitness and language processing in children: Implications for academic achievement. *Brain and Cognition*, 87, 140–152.
- Taft, M., & Zhu, X. (1995). The representation of bound morphemes in the lexicon: A Chinese study. In L. B. Feldman (Ed.), *Morphological aspects of language processing* (pp. 293–316). Hillsdale, NJ: Erlbaum.
- Taft, M., Liu, Y., & Zhu, X. (1999). Morphemic processing in Chinese. In J. Wang, A. Inhoff, & H.-S. Chen (Eds.), *Reading Chinese script: A* cognitive analysis (pp. 91–114). Mahwah, NJ: Erlbaum.
- Tan, L. H., & Perfetti, C. A. (1997). Visual Chinese character recognition: Does phonological information mediate access to meaning? *Journal* of Memory and Language, 37, 41–57. doi:10.1006/jmla.1997.2508
- Tan, L. H., & Perfetti, C. A. (1998). Phonological codes as early sources of constraint in reading Chinese: A review of current discoveries and theoretical accounts. *Reading and Writing*, 10, 165–200.
- Timmer, K., & Schiller, N. O. (2012). The role of orthography and phonology in English: An ERP study on first and second language reading aloud. *Brain Research*, 1483, 39–53.
- Tsai, J. L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in traditional Chinese reading. *Acta Psychologica*, 141, 17–23. doi:10.1016/j.actpsy.2012.06.004
- Tsai, J., Lee, C., Tzeng, O., Huang, D., & Yen, N. (2004). Use of phonological codes for Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain and Language*, 91, 235–244.

- Tsang, Y., & Chen, H. (2012). Eyemovement control in reading: Logographic Chinese versus alphabetic scripts. *PsyCh Journal*, 1, 128–142.
- Wang, J. (1997). The representation of the neutral tone in Chinese Putonghua. In J. Kao & N. Smith (Eds.), *Studies in Chinese phonology* (pp. 157–183). Berlin: Mouton de Gruyter.
- Wang, Y. (2004). The effects of pitch and duration on the perception of the neutral tone in standard Chinese. *Acta Acustica*, 29, 453–461.
- Wang, H. (2008). Non-linear phonology of Chinese: The phonological structure and single-character phonology in Chinese (Rev. edition). Beijing, China: Peking University Press.
- Wheat, K. L., Cornelissen, P. L., Frost, S. J., & Hansen, P. C. (2010). During visual word recognition, phonology is accessed within 100 ms and may be mediated by a speech production code: Evidence from magnetoencephalography. *Journal of Neuroscience*, 30, 5229–5233.
- Wickham, H. (2009). Ggplot2: Elegant graphics for data analysis. New York, NY: Springer.
- Wlottk329929001(tk)2te20107(,)-408(H)-8715(lo0239TeF0(g)--306(f16(on)ei(:)-312(H)-20107(g)--306(f16(g)--306(g)--306(f16(g)--30