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 Yan[3] · Shaorong Yan[1] · Xiaolin Zhou[1] · Albrecht W. Inhoff[4]

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 event-related 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 eye movements were recorded while participants read. Analyses of oculomotor activity revealed shorter viewing durations and fewer refixations on—and fewer regressive saccades to— 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 · Neutral tone · Articulation duration · Syllabic tone · Sentence reading · 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 speech-specific 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 articulation-specific 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;

✉ 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] Department of Psychology, Binghamton University, Binghamton, NY, USA

Springer

Lukatela, Eaton, Sabadini, & Turvey, 2004; Wheat, Cornelissen, Frost, & Hansen, 2010). Furthermore, recordings of eye movements have shown that relatively detailed sub- and 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 when targets were preceded by syllable-congruent primes. Collectively, these studies provide strong and converging evidence for the use of articulation-specific features during the early stages of visual word identification.

The orthography of alphabetic scripts was designed to express the identity and sequence of aurally perceived and spoken phonemes, although the strength of the orthography-to-phonology correlation differs somewhat across languages (Frost, 2012). When alphabetic scripts are read, the extraction of orthographic information could thus be associated with the routine activation of phonemic units, and also with their articulation. Chinese script, by contrast, was not designed to represent individual phonemes (but to represent morphemes), and

the orthographic features of Chinese are less informative than the orthographic features of alphabetic scripts with regard to phonology and articulation. For example, in Chinese one syllable can map onto several different characters that may not have any orthographic overlap, resulting in a high homophone density: Up to 99.76 % of the characters have homophone mates, and on average each character has 11 homophone mates (Tan & Perfetti, 1998). Conversely, characters with similar—or even identical—orthographic properties may also have radically different pronunciations (e.g., 会 can be pronounced as "hui4" or "kuai4," depending on the meaning). It should be noted that a majority of Chinese words are compounds, in that they contain more than one constituent morpheme. Within a particular lexical context, only one pronunciation is licensed for each character, even if the constituent characters are homographs.

Because of these unique properties, the question arises whether the use of phonological knowledge for word recognition is functionally equivalent during the reading of Chinese and of alphabetic scripts. There appears to be some consensus, according to which phonological features of individual Chinese characters are activated early and automatically during character recognition. Homophone primes influence target character recognition, with shorter prime durations than semantic character primes (see Perfetti et al., 2005, for a review), and recordings of eye movements during sentence reading showed facilitation in the processing of target words when parafoveally previewed characters shared syllables or phonetic radicals with the target than when they were unrelated characters (Liu, Inhoff, Ye, & Wu, 2002; Tsai, Kliegl, & Yan, 2012; Tsai, Lee, Tzeng, Huang, & Yen, 2004; Yan, Pan, Bélanger, & Shu, 2015; Yan, Richter, Shu, & Kliegl, 2009; see Tsang & Chen, 2012, for a review). Neuroimaging studies have further provided evidence for phonological activation at the subcharacter level (Hsu, Tsai, Lee, & Tzeng, 2009; Lee, Huang, Kuo, Tsai, & Tzeng, 2010).

There is, however, no converging evidence for the automatic and prelexical use of phonology for the recognition of Chinese multicharacter words (see Tsang & Chen, 2012, for a review). According to one view, phonological activation of individual characters, derived directly from orthography, occurs early and dominates the semantic activation of corresponding morpheme and whole-word meanings (e.g., Perfetti & Tan, 1999; see also Perfetti et al., 2005, for a review; but see Zhou & Marslen-Wilson, 2000, for a critical view of the experimental stimuli). Alternatively, phonological codes may not assume a critical role for the accessing of word meaning (e.g., B. Chen & Peng, 2001; Hoosain, 2002; Wong, Wu, & Chen, 2014; Zhou & Marslen-Wilson, 1999, 2000, 2009). For instance, Wong, Wu, and Chen (2014) found that neither lexical decision nor ERP responses to two-character compound words were influenced by phonologically similar primes when they shared an identical syllable and were

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 high-frequency 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 context-dependent 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 sound-specific 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 Number
N − 1 (Exp. 1)				
Mean	2.58	17.6	2.84	16.6
SD	1.85	5.8	1.24	5.5
N − 1 (Exp. 2)				
Mean	2.69	16.3	2.96	15.0
SD	1.46	5.7	1.26	4.2
Target				
Mean	2.07	16.1	2.03	16.3
SD	0.72	4.0	0.79	3.8
N + 1 (Exp. 2)				
Mean	2.87	16.3	2.76	15.5
SD	1.19	3.9	1.02	4.4

The frequency shows the $\log_{10}$ of the total number of times the word was observed in the corpus. The stroke number shows the sum of the strokes for each constituent character.

Table 2 Mean durations and intensities of the second syllables of the target words, and mean ratings of the different types of norms

	Neutral-Tone Words	Full-Tone Words
Phonetic Measures		
Duration (ms)	233	271
Intensity (dB)	64.6	65.4
Ratings		
Ease of imagination (5-point scale)	3.77	3.66
Age of acquisition (years)	5.81	6.02
Cloze predictability (%)	8	4

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 full-tone 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 full-tone 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 results showed that the neutral- and full-tone target words were equally (hardly) predictable: They were predicted 69 and 37 times out of 855 guesses (8 % and 4 %, respectively) [$F(1, 112) = 1.436$, $p = .233$; see Table 2]. In addition, 17 native speakers of Standard Chinese were recruited to rate the pretarget and target words in terms of familiarity on a 7-point scale. Neither of them showed a significant difference between the neutral- and full-tone sentences, ps > .1, demonstrating that the words in the two different conditions were equally familiar to the native speakers of Standard Chinese.

During the experiment, sentence reading difficulty ratings were obtained after each sentence was read, to determine the success with which the full- and neutral-tone conditions were matched. These ratings indicated that the full sentences with neutral- and full-tone targets were considered equally difficult, as will be shown below.

Procedure

Each participant was seated comfortably in a dimly lit and sound-attenuating booth, approximately 100 cm in front of a computer monitor. Participants were asked to avoid eye movements and body movements during sentence presentation. Characters were shown in size 24 font and were displayed in white color on a black background. A trial began with the presentation of a fixation point in the center of the screen for 500 ms, and this was followed by a 200-ms blank interval, followed by the onset of the first word. Each word was shown individually for a fixed duration, 400 ms, at the screen center, and its presentation was followed by a 400-ms interval during which the screen was blank.

Sentences with the two types of target words were presented in a pseudorandomized order, with no more than three consecutive sentences from the same target type condition. A second list was constructed with a reversed sentence order for the 114 experimental sentences, to balance potential sequence or fatigue effects. Participants were randomly assigned to one of the two lists. To focus attention on the extraction of sentence meaning and to check the ease of sentence reading in the neutral- and full-tone conditions, participants were also asked to rate the difficulty of a sentence at the end of each trial. This was accomplished by moving a cursor on a 5-point rating scale that ranged from *very easy* to *very difficult*. These

| *Neutral tone condition*					
Chinese	老板	总是	使唤	秘书	来 帮忙 处理 私事。
Interest Regions		Pretarget	Target	Posttarget	
Pinyin & Tone			shi3 huan		
Literal Translation	Boss	always	order around	secretary	to help conduct private business
Translation	The boss always bossed the secretary to help take care of his private business.				
Full tone condition					
Chinese	会计	经常	使用	算盘	来 核对 账目。
Interest Regions		Pretarget	Target	Posttarget	
Pinyin & Tone			shi3 yong4		
Literal Translation	Accountant	often	use	abacus	to check accounts.
Translation	The accountant often used abacus to check accounts.				

Fig. 1 Sample sentences used in the experiments

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 kΩ. 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 15 contiguous topographic locations (see Fig. 2): left frontal (F1, F3), left fronto-central (FC1, FC3), left central (C1, C3), left centro-parietal (CP1, CP3), left parietal (P1, P3), midline frontal (Fz), midline fronto-central (FCz), midline central (Cz), midline centro-parietal (CPz), midline parietal (Pz), right frontal (F2, F4), right fronto-central (FC2, FC4), right central (C2, C4), right centro-parietal (CP2, CP4), and right parietal

(P2, P4). These 15 groupings were classified into two orthogonal dimensions: one for left-hemisphere, midline, and right-hemisphere locations, and another for the five anterior-to-posterior locations. The neighboring electrodes on lateral sites (left/right hemisphere) were combined in order to avoid a loss of statistical power (Oken & Chiappa, 1986) and to focus on the contrasts between the left and right hemispheres (e.g., Henrich, Alter, Wiese, & Domahs, 2014).
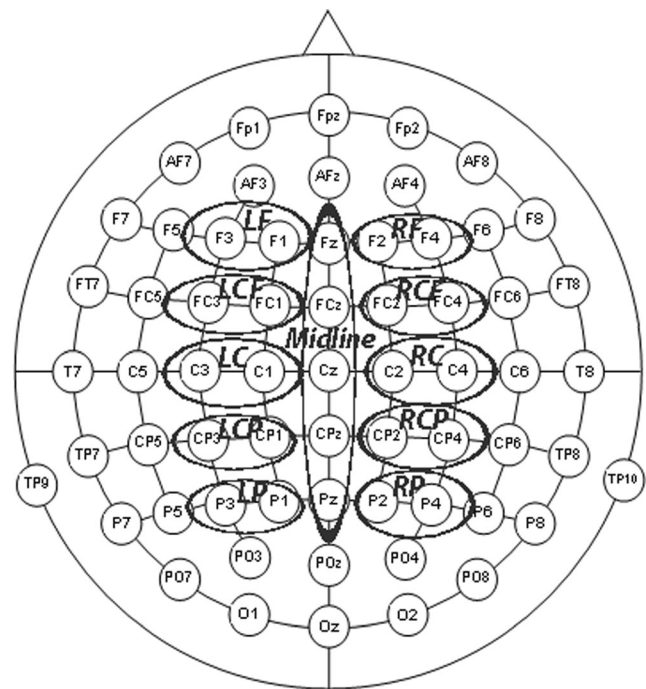
Fig. 2 Electroencephalographic recording sites and regions of interest used in the statistical analyses

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.
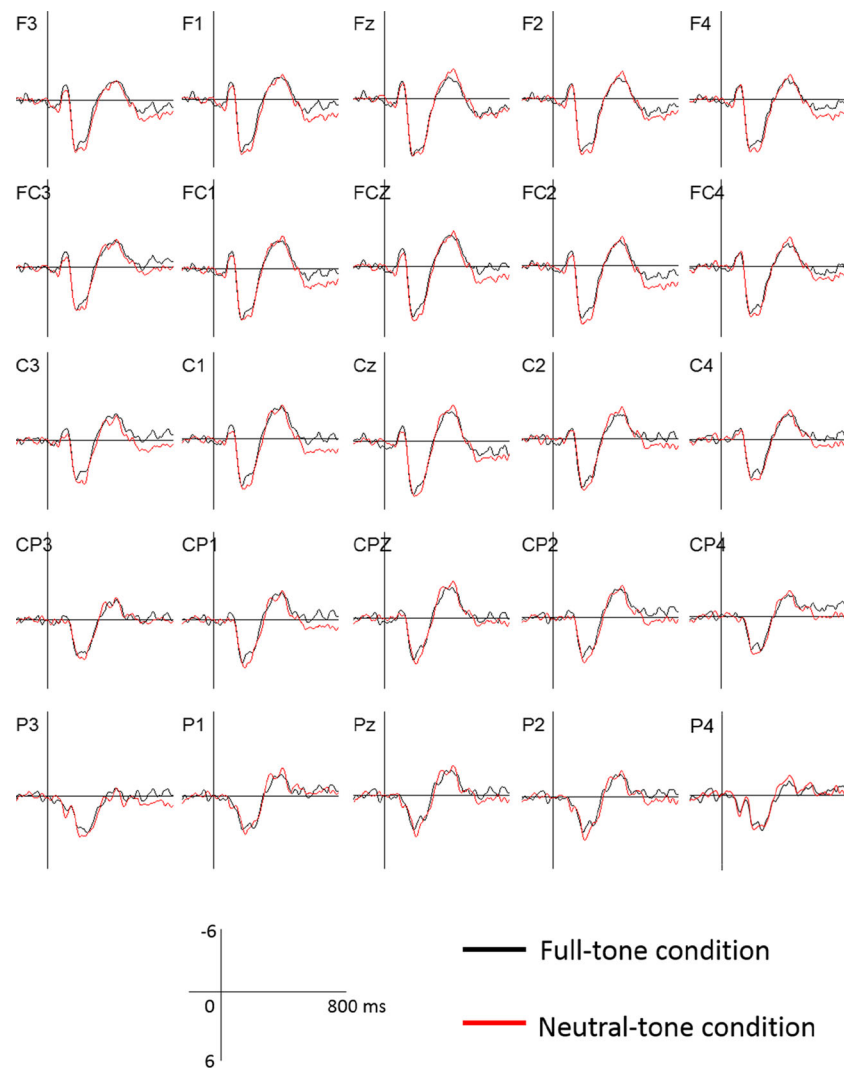
Pretarget words

N400 No significant effect was found for the analysis of N400 amplitudes, all *t*s < 1.1.

Late component In the 600- to 800-ms time window, a larger positivity was observed for the neutral- than for the full-tone 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.
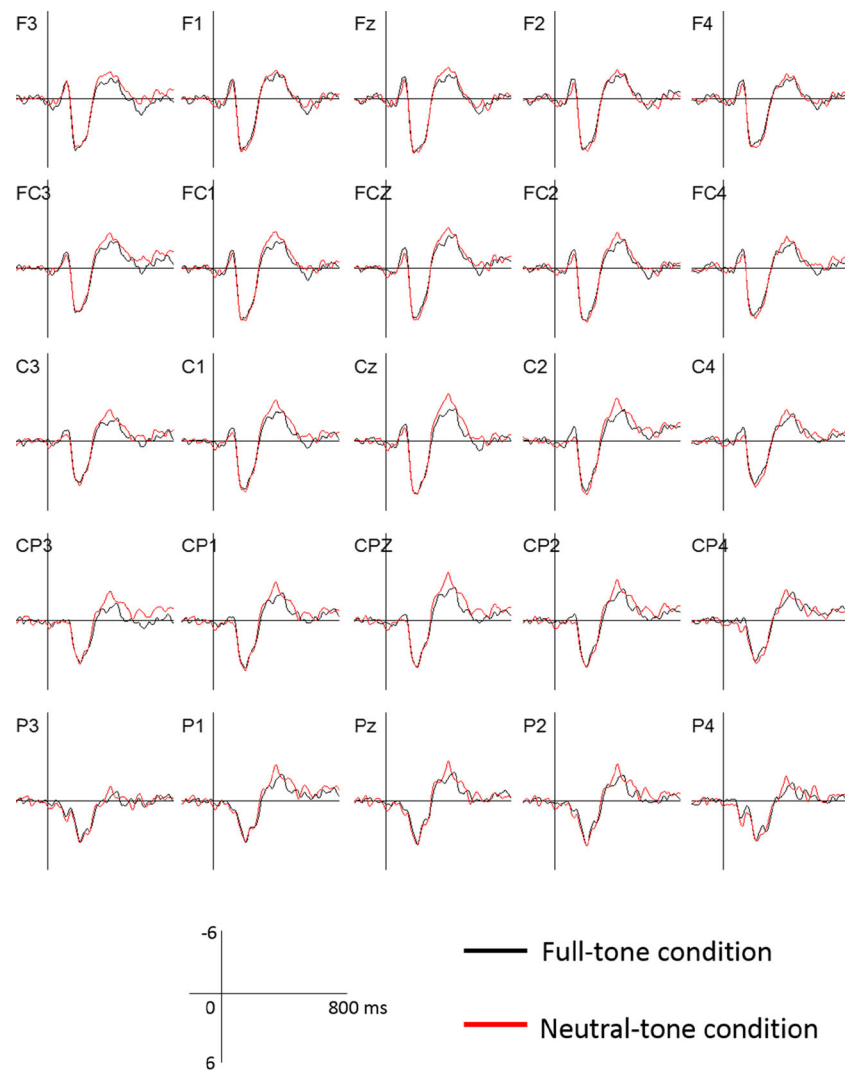
N250 Means for the N250 component are shown as a function of tone type and topographic location in Fig. 6. As can be seen in the depiction of the full waveforms in Fig. 4, the N250 component occurred between a trough at around 200 ms and a spike at around 400 ms, which is consistent with the N250 properties 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, $b = 0.41\,\mu V$, $SE = 0.12$, $t = 3.46$. 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, $b = -0.21\,\mu V$, $SE = 0.10$, $t = 2.12$.

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, $b = -0.32\,\mu$V, $SE = 0.14$, $t = -2.19$. In addition, one topographic effect was reliable, due to increases in negativity from anterior to posterior locations, $b = -0.14\,\mu$V, $SE =$

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
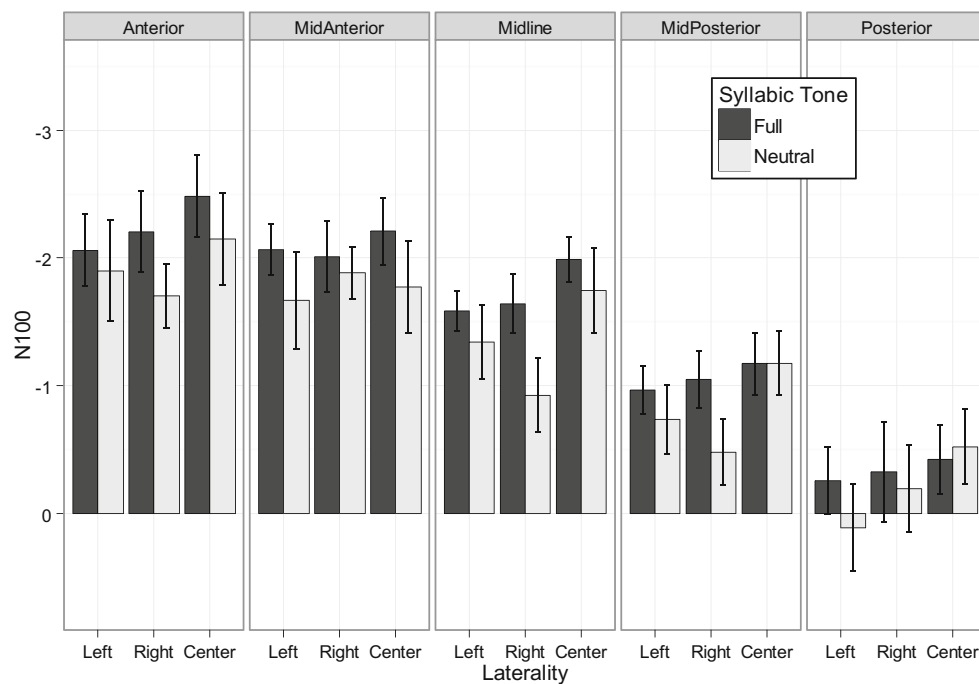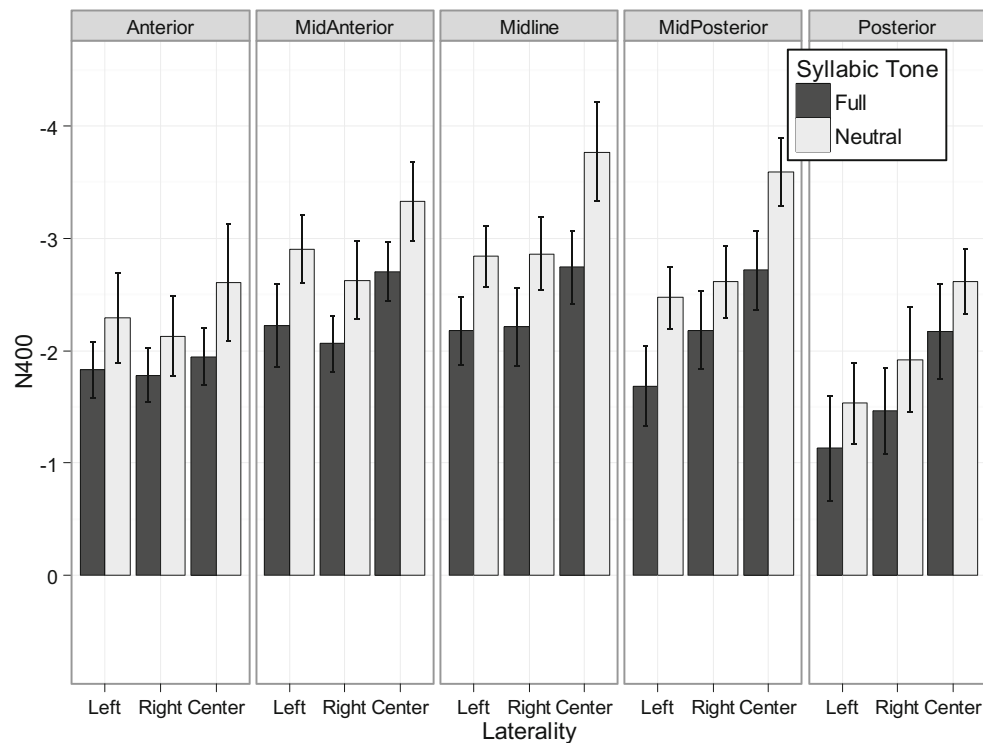
Fig. 5 N100 values. Negative values are plotted upward, and *SE*s were computed from the residuals of the regression model

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 word recognition during silent reading entails the use of articulation-specific features during early stages of word recognition. Moreover, the robust effect of target type on N400 amplitudes indicated that native speakers of Standard Chinese used articulation-specific features also during later stages of word representation when the word was viewed.
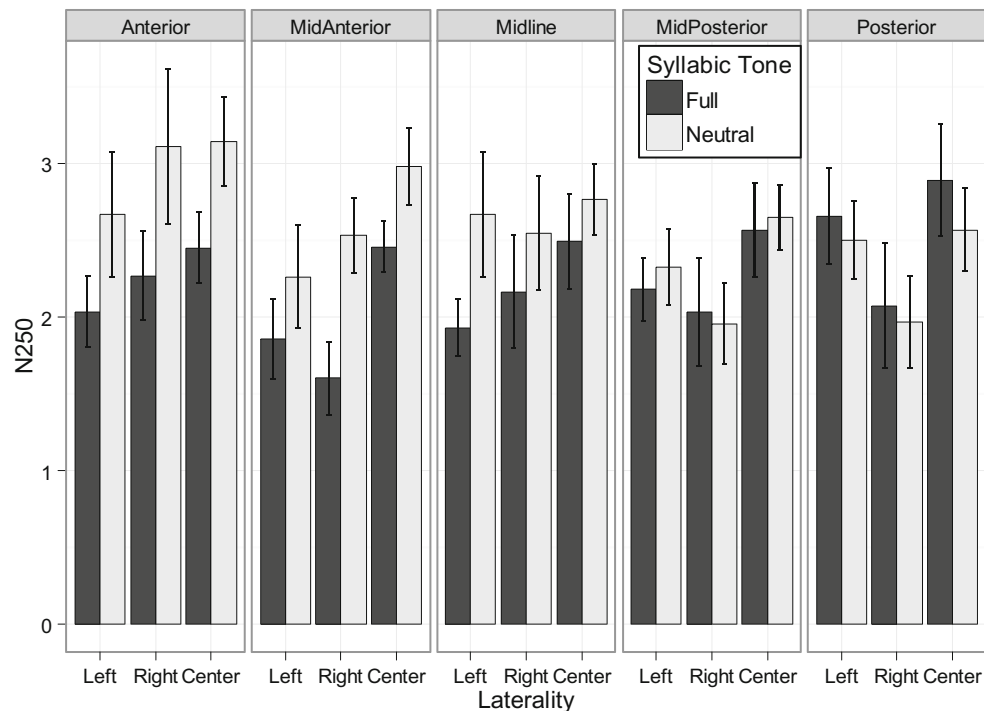
Fig. 6 N250 component. Positive values are plotted upward, and *SE*s were computed from the residuals of the regression model

Fig. 7 N400 component. Negative values are plotted upward, and *SE*s 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 time-line 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 Martin—matched 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 whole-word 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 (full-tone) 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 second-syllable 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 full-tone conditions in terms of word frequency and number of strokes, all $ps > .1$ (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-presentation of the AD.

Apparatus An EyeLink 2 K system, with a sampling rate of 2000 Hz and a spatial resolution of better than 0.2 deg, was used to record eye movements during sentence reading. Each sentence was presented in one line at the vertical position one third of the way from the top of a 21-in. CRT screen (1,024 × 768 resolution, frame rate 100 Hz). The font Song-20 was used, with one Chinese character subtending 0.5 deg of visual angle. Participants read each sentence with their head positioned on a chin-and-forehead rest, approximately 80 cm from the screen. All recordings and calibrations were based on the left eye, but viewing was binocular. The experiment was programmed using the EyeLink Experiment Builder software, and Eyelink software was used to separate the continuously sampled eye movement and eye position data into fixations—that is, periods during which the eyes were relatively stationary—and saccades—that is, movements in-between successive fixations. The EyeLink Dataviewer software package was used to extract the oculomotor target and posttarget word-viewing measures. AD stimuli were presented binaurally via a headphone (SONY MDR-V900HD) at a comfortable volume of approximately 60 dB. The Creative X-Fi XtremeGamer soundcard was used to ensure a very low trigger latency with an estimated uncertainty less than 4 ms.

Procedure Prior to the experiment, each participant was asked to read aloud five sentences, of which each included a neutral-tone word, and all participants articulated the word with a neutral tone. After this, participants were calibrated using a nine-point grid. Successful calibration, which yielded a tracking accuracy of better than 0.5 deg of visual angle, was followed by the presentation of a small black cross at the left side of the computer screen at the location of the first character of a to-be-read sentence. The reader was instructed to fixate the marker, to initiate presentation of a sentence with a button press, to read the sentence silently for comprehension, and to terminate its visibility with another button press.

The accuracy of the calibration was visually checked after each trial, and a drift correction and/or recalibration was performed to correct poor tracking accuracies. To encourage reading for meaning, 24 sentences were followed by the presentation of a probe sentence, and the participant was asked to determine whether the content of the probe matched the content of the previously read sentence. As in Experiment 1, readers were also asked to rate the difficulty of each sentence after it was read by pressing a gamepad button at the end of the trial. A 4-point scale was used, with 1 reflecting *most easy* and 4 reflecting *most difficult*. Once more, no differences in the difficulty of sentences were discernible with neutral- and full-tone targets (1.3 vs. 1.3), $F < 1$.

Fifteen sentences were read for practice at the onset of the experiment. Participants were told that they would hear a syllable during sentence reading and that the syllable was task-

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 measures are generally used to index the success with which individual words are recognized. A lexical determination of these viewing duration measures is also assumed in some models of oculomotor control during reading (e.g., Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Warren, & McConnell, 2009). Total viewing durations consisted of a word's gaze duration and the time spent rereading it after other words in the sentence had been viewed. This measure is assumed to be sensitive to lexical processing and to processes that occur after a word has been recognized. In addition, we analyzed the probabilities of making refixations at the target word during first-pass reading and the rate of regressions into the target word. The number of first-pass fixations has been reported to be sensitive to the articulation-specific features during word recognition, with more fixations being launched for two-stressed than for one-stress words (Ashby & Clifton, 2005). Similarly in Chinese, readers made more refixations to full- than to neutral-tone targets (Yan et al., 2014). Incoming regressions are generally assumed to occur when a relatively late stage of visual word use is impeded (Reichle et al., 2009).

Since the AD manipulation was assumed to influence target and posttarget viewing, effects of target type and of AD type were also examined for the fixated posttarget words. The AD was not properly triggered in 1.9 % of all trials, and these trials were consequently eliminated. Approximately 10 % of the target words were skipped during first-pass reading, and analyses of posttarget viewing were made conditional on prior fixation of the target word to avoid a confounding of AD effects with oculomotor effects (i.e., when the target was fixated vs. skipped). Fixated target and posttarget words were also excluded from the analysis when the first saccade into these words was not forward-directed, when it was atypically large (>6 character spaces), and when the duration of the first fixation was shorter than 75 ms or longer than 800 ms. Together, the selection criteria resulted in the exclusion of 9.4 % of the fixated target words and of 19.2 % of the fixated posttarget words. A total of 4,461 target and of 3,924 posttarget words were analyzed.

Separate LMMs were used to analyze pretarget, target, and posttarget viewing. Our analyses of the pretarget words were primarily concerned with potential effects of pretarget context on neutral- and the full-tone target reading. Therefore, individual trial data were entered, and the factor Lexical Tone Type (neutral vs. full) was used as a fixed factor for the analysis of pretarget viewing. Target and posttarget viewing data were analyzed using the fixed factors Lexical Tone Type (neutral vs. full) and AD Type (identical, similar, dissimilar). AD-type effects were examined with two orthogonal Helmert contrasts: a syllable match contrast that compared the identical AD condition against the mean of the two nonidentical (similar and dissimilar) conditions, and an articulation similarity contrast that compared the similar with the dissimilar AD condition. Since the analyzed data were not averaged over items, two crossed random factors were entered in the model, comprising the intercepts for Subjects and Items. As was recommended by Barr, Levy, Scheepers, and Tily (2013), the model also included participant-specific random slopes for each of the two fixed factors and their interactions. The frequency distributions of the three viewing durations were positively skewed, and this was corrected through log transformations. The statistical effect patterns were, however, virtually identical for the transformed and nontransformed data, and effect sizes (b), SEs, and t statistics are reported for the transformed data. A logit-link function was used to analyze regressions into the target words and refixations of the target words. Refixation values were computed by transforming the number of first-pass fixations into binomial values, with 1 representing *one fixation* and 0 representing *more than one fixation* during the first-pass reading. For this analysis, estimated effect sizes are reported in logits with corresponding z values; t and z values >1.96 were considered significant.

Results

Overall, participants correctly responded to 94 % of the probe sentences, indicating that they were reading for meaning. The mean first-fixation durations, gaze durations, and total viewing durations for target and posttarget words are shown as a function of lexical tone and AD type in Table 3.

Pretarget word None of the oculomotor measures yielded a significant neutral- versus full-tone difference: $b = 0.020$ ms, $SE = 0.017$, $t = 1.41$, for first-fixation durations (264 vs. 258 ms); $b = 0.029$ ms, $SE = 0.026$, $t = 1.12$, for gaze durations (319 vs. 308 ms); and $b = 0.003$ ms, $SE = 0.033$, $t = 0.11$, for total viewing durations (369 vs. 368 ms). This indicates that the processing of pretarget words was equivalent in the full- and neutral-tone target conditions.

Target word First-pass and total viewing durations were shorter for neutral- than for full-tone targets. Although the

Table 3 Experiment 2: mean first-fixation durations, gaze durations, and total viewing durations (in milliseconds) for the target and posttarget words as a function of the target's tone articulation and the irrelevant auditory-distractor (AD) type
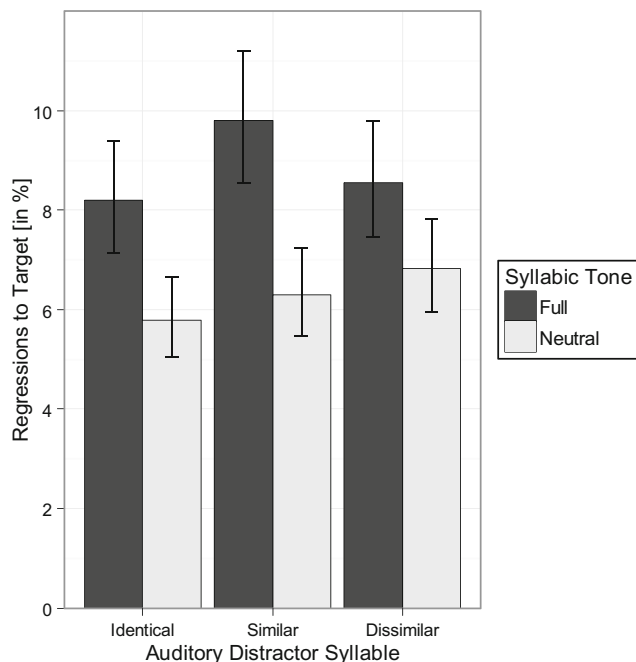
	Neutral-Tone Condition			Full-Tone Condition		
	Identical AD	Similar AD	Dissimilar AD	Identical AD	Similar AD	Dissimilar AD
Target Word						
First-fixation duration	268 (6.2)	276 (7.1)	274 (5.3)	277 (6.3)	278 (5.6)	277 (6.5)
Gaze duration	313 (8.8)	321 (10.2)	325 (8.3)	338 (11.3)	339 (9.7)	336 (9.9)
Total viewing duration	361 (13.3)	373 (14.2)	373 (10.5)	394 (13.7)	406 (13.9)	400 (12.7)
Posttarget Word						
First-fixation duration	272 (5.3)	266 (6.2)	262 (5.0)	275 (6.2)	268 (6.8)	269 (5.8)
Gaze duration	320 (8.5)	312 (7.6)	301 (8.2)	329 (9.1)	314 (8.0)	312 (9.5)
Total viewing duration	371 (10.8)	369 (10.3)	357 (12.3)	389 (12.3)	378 (12.0)	374 (13.4)

Standard deviations are shown in parentheses.

effect did not reach significance for first-fixation durations (–4 ms), $b = –0.018$ ms, $SE = 0.013$, $t = 1.41$, it was reliable for gaze durations (–18 ms) and total viewing durations (–31 ms), $b = –0.046$ ms, $SE = 0.020$, $t = 2.24$, and $b = –0.072$ ms, $SE = 0.025$, $t = 2.88$, respectively. Additional analyses of regressions to the target, shown in Fig. 8, also revealed a significant effect of lexical tone, with fewer regressions to neutral- than to full-tone targets (6.3 % and 8.8 %, respectively), $b = –.352$ [logits], $SE = .171$, $z = 2.06$. Neutral-tone targets also received fewer first-pass fixations than full-tone targets (1.20 and 1.24, respectively). Although small, the effect was marginally significant, $b = .188$ [logits], $SE = .103$, $z = 1.83$, $p < .1$. It replicates a

Fig. 8 Mean regression rates to the target after a subsequent word in the sentence had been viewed. The syllabic tone articulation SE from the mixed model was used to compute the 95 % confidence intervals

corresponding finding in Yan et al. (2014) and is consistent with the effect of lexical stress in Ashby and Clifton (2005), in which words with one stress received fewer fixations than those with two stresses.

The analysis of AD effects revealed numerically shorter durations in the identical than in the two nonidentical AD syllable conditions, but the size of the syllable match effect was quite small for first-fixation durations and gaze durations (3 and 5 ms, respectively), and did not approach significance (both t values < 1.5). The estimated effect size was larger and marginally reliable for total viewing durations (11 ms), $b = –.009$ ms, $SE = .005$, $t = 1.85$, $p < .1$. No other AD effect approached significance, all t values < 1.5.

Posttarget word region The target's lexical tone did not influence any of the three posttarget viewing duration measures, all t values < 1.4. AD type influenced posttarget viewing, with longer—not shorter—viewing durations in the identical condition than in the two nonidentical AD conditions. The corresponding syllable match contrast was significant for first-fixation durations (8 ms), gaze durations (14 ms), and total viewing durations (10 ms): $b = .008$ ms, $SE = .004$, $t = 2.21$; $b = .013$, $SE = .005$, $t = 2.78$; and $b = .011$, $SE = .005$, $t = 2.21$, respectively. The similarity contrast—that is, the difference between the similar and dissimilar AD conditions—was negligible (2, 6, and 13 ms, respectively) and not reliable, t values < 1.5, for all three viewing duration measures. None of the remaining effects, including the interactions of tone type with AD type, approached significance, all t values < 1.5. That is, the tone articulation the target's second syllable did not modulate the effects of an AD syllable.

Discussion

The two types of lexical tones were examined under relatively normal reading conditions in Experiment 2, and an AD manipulation in conjunction with different oculomotor measures

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 re-reading 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 of a target. Contrary to our predictions, none of the AD effects were, however, a function of target tone type. The independence of the AD and target type effects is also considered in the General Discussion.

General discussion

In the present study, we examined the influence of articulation-specific word properties on early and late stages of visual word processing during Chinese sentence reading. The results of two different experimental approaches converged, both showing that variation in the articulation of syllabic tone influences the early stages of visual word processing for native speakers of Standard Chinese. In Experiment 1, syllabic tone articulation influenced the N100 and anterior N250 ERP components, which were less negative for neutral- than for full-tone targets, and in Experiment 2, neutral-tone words received shorter gaze durations and fewer refixations than full-tone words. The lexical tone property also influenced the subsequent processing of a word. In Experiment 1, the N400 component was more negative for neutral- than for full-tone targets, and neutral-tone targets received shorter total viewing durations and fewer incoming regressions in Experiment 2.

Effects of spoken-word properties on visual word identification are relatively well established for alphabetic text. Briefly presented visual primes with matching sub- and supraphonemic properties yielded less-negative early ERP components than did primes with mismatching properties (Ashby & Martin, 2008; Ashby et al., 2009; Wheat et al., 2010), and naming latencies, lexical decision times, and viewing durations for target words were shorter when the word contained a vowel with a short rather than a long articulation duration (Abramson & Goldinger, 1997; Huestegge, 2010; Lukatela et al., 2004). Moreover, work with alphabetic text has shown that sub- and supraphonemic properties of words are extracted from words before they are fixated during silent reading (Ashby & Martin, 2008; Ashby & Rayner, 2004). The novel contribution of the present experiments is their converging demonstration of early articulation-specific word recognition effects with Chinese script. This is of theoretical significance, because it provides additional evidence against the minimality hypothesis through the use of a morpho-syllabic script that was *not* designed to represent the properties of spoken language. The representations of tone neutralization during the recognition of Chinese words indicates, therefore, that the use of articulation-specific word features for lexical processing is independent of script type, and it may be a fundamental venue for the accessing of represented lexical knowledge for all readers with spoken language skills. Indeed, the articulation of orthographically identical words can be subject to regional variation, and this has been used to account for differences in word reading. For instance, the reading of identical poems was found to be influenced by readers' dialects (Filik & Barber, 2011). Similarly, the words with or without neutralized tone in Standard Chinese were responded to differently at both the behavioral and neural levels by native speakers of Standard Chinese, but not by speakers who did not have the contrast of neutral- and full-tone words in their dialects (Yan et al., 2014). These findings indicate not only that the word identification process uses articulatory features during silent reading but also that this usage reflects a cross-linguistic universal processing mechanism.

The less-negative N100 and anterior N250 components for neutral- than for full-tone targets, and their shorter first-pass viewing duration and lower refixation rate, indicate that the initial stage of processing of neutral-tone words was easier and took less time. Since similar effects of articulation-specific word features on visual word recognition have been obtained with alphabetic text, theoretical accounts of the effect may also generalize across script types. Lukatela et al. (2004, p. 162) outlined a representational account for English words, according to which vowel duration is lexically represented with other word features. The vowel representation of words with a long articulation duration has a more complex representation of features, and recognition of long-articulation vowels could take more time because it demands the accessing of a more complex representation. Lukatela et al. also offered a related theoretical account, according to which represented orthographic forms are automatically mapped onto articulatory gestures. These gestures are similar to overt speech, in that they comprise the setting of dynamic parameters for the generation of acoustic features. They differ from overt speech in that their parameterization involves covert simulation of speech rather than actual engagement of the articulatory effector system. According to the gestural account, recognition of neutral-tone targets was more effective because the simulated articulation of a neutral-tone target during reading was simpler and required less time than the simulated articulation of a full-tone target.

Yet another related account is that tone neutralization influences phonological synthesis, because it alters the metrical structure of a Chinese word. According to J. Wang (1997), a full-tone syllable constitutes a metrical foot by itself, and a full-tone syllable together with the following neutral-tone syllable also forms a single metrical foot. A full-tone target, however, contains two metrical feet, and assembly of its metrical

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 (McLennan, 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-be-read 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 eye movements of speakers of Standard Chinese yielded 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 similar view).

Appendix

It is possible that the lexical ambiguity of the targets could confound the tone-type effect. To avoid this, the neutral- and full-tone target words used in this study allowed for neither ambiguous pronunciations nor ambiguous readings at the word level. At the morphemic level, the second character of five of the target words could have two pronunciations in the neutral-tone condition, and five such words were also present in the full-tone condition. For the other target words, the

second characters were all monophones. Regarding the syntactic categories, the neutral-tone target and its full-tone counterpart shared all possible syntactic categories in 45 out of the 57 word pairs, and shared the dominant category in all word pairs. In addition, they were all assigned to the dominant category in the sentence stimuli.

We also carefully matched the neutral- and the full-tone target words in terms of morphology, using the following indices: (a) morphological structure of the disyllabic compound, which features the syntactic/argument relationship between the two constituent morphemes; (b) morphemic status (Taft, Liu, & Zhu, 1999), which defines the word regarding whether or not the constituent morpheme could be used as monomorpheme, and whether the compound word meaning was transparent as the aggregation of the morphemic meaning or was opaque. The numbers of target words in the different categories were comparable between the neutral- and the full-tone conditions, as are shown in the following table.

	Neutral-Tone	Full-Tone
Morphological Structure		
Modifier core	25	28
Coordinate	22	19
Subject predicate	4	2
Verb object	3	3
Verb complement	1	3
Prefix root	1	1
Ambiguous	1	1
Morphemic Status		
Monomorpheme (2nd morpheme)	29	29
Nonmonomorpheme	28	28
Transparent	41	41
Opaque	16	16

References

Abramson, M., & Goldinger, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics, 59,* 1059–1068. doi:10.3758/BF03205520

Ashby, J. (2006). Prosody in skilled silent reading: Evidence from eye movements. *Journal of Research in Reading, 29,* 318–333.

Ashby, J. (2010). Phonology is fundamental in skilled reading: Evidence from ERPs. *Psychonomic Bulletin & Review, 17,* 95–100. doi:10.3758/PBR.17.1.95

Ashby, J., & Clifton, C., Jr. (2005). The prosodic property of lexical stress affects eye movements in silent reading: Evidence from eye movements. *Cognition, 96,* B89–B100.

Ashby, J., & Martin, A. E. (2008). Prosodic phonological representations early in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 224–236. doi:10.1037/0096-1523.34.1.224

Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements. *Language and Cognitive Processes, 19,* 391–426.

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 Experimental 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 vowel length on gaze durations in silent and oral reading. *Journal of Eye Movement Research, 3*(5), 5:1–18. Retrieved from www.jemr.org/online/3/5/5

Inhoff, A. W., Connine, C., Eiter, B., Radach, R., & Heller, D. (2004). Phonological representation of words in working memory during sentence reading. *Psychonomic Bulletin & Review, 11,* 320–325. doi:10.3758/BF03196577

Inhoff, A. W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29–54). Oxford, UK: Elsevier Inc.

Kutas, M., & Federmeier, K. D. (2009). N400. *Scholarpedia, 4,* 7790.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62,* 621–647. doi:10.1146/annurev.psych.093008.131123

Laszlo, S., & Federmeier, K. D. (2011). The N400 as a snapshot of interactive processing: Evidence from regression analyses of orthographic neighbor and lexical associate effects. *Psychophysiology, 48*(2), 176–186.

Lee, C.-Y., Huang, H.-W., Kuo, W.-J., Tsai, J.-L., & Tzeng, J.-L. O. (2010). Cognitive and neural basis of the consistency and lexicality effects in reading Chinese. *Journal of Neurolinguistics, 23,* 10–27. doi:10.1016/j.jneuroling.2009.07.003

Lin, M., & Yan, J. (1980). Beijinghua qingsheng de shengxue xingzhi. *Dialect, 3,* 166–178.

Liu, W., Inhoff, A. W., Ye, Y., & Wu, C. (2002). Use of parafoveally visible characters during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 28,* 1213–1227. doi:10.1037/0096-1523.28.6.1213

LoCasto, P. C., & Connine, C. M. (2002). Rule-governed missing information in spoken word recognition: Schwa vowel deletion. *Perception & Psychophysics, 64,* 208–219.

Lukatela, G., Eaton, T., Sabadini, L., & Turvey, M. T. (2004). Vowel duration affects visual word identification: Evidence that the mediating phonology is phonetically informed. *Journal of Experimental Psychology: Human Perception and Performance, 30,* 151–162. doi:10.1037/0096-1523.30.1.151

McLennan, C. T., Luce, P. A., & Charles-Luce, J. (2005). Representation of lexical form: Evidence from studies of sublexical ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 31,* 1308–1314. doi:10.1037/0096-1523.31.6.1308

Morris, J., Frank, T., Grainger, J., & Holcomb, P. J. (2007). Semantic transparency and masked morphological priming: An ERP investigation. *Psychophysiology, 44,* 506–521.

Norris, D. (2013). Models of visual word recognition. *Trends in Cognitive Sciences, 17,* 517–524.

Oken, B. S., & Chiappa, K. H. (1986). Statistical issues concerning computerized analysis of brainwave topography. *Annals of Neurology, 19,* 493–494.

Perfetti, C. A. (2003). The universal grammar of reading. *Scientific Studies of Reading, 7,* 3–24.

Perfetti, C. A., Liu, Y., & Tan, L. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112,* 43–59. doi:10.1037/0033-295X.112.1.43

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 Experimental 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.

Wlotko, E. W., & Federmeier, K. D. (2012). Age-related changes in the impact of contextual strength on multiple aspects of sentence comprehension. *Psychophysiology, 49,* 770–785.

Wong, A. W.-K., Wu, Y., & Chen, H.-C. (2014). Limited role of phonology in reading Chinese two-character compounds: Evidence from an ERP study. *Neuroscience, 256,* 342–351.

Yan, M., Luo, Y., & Inhoff, A. W. (2014). Syllable articulation duration influences foveal and parafoveal processing of words during the silent reading of Chinese sentences. *Journal of Memory and Language, 75,* 93–103.

Yan, M., Pan, J., Bélanger, N. N., & Shu, H. (2015b). Chinese deaf readers have early access to parafoveal semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41,* 254–261. doi:10.1037/xlm0000035

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Readers of Chinese extract semantic information from parafoveal words. *Psychonomic Bulletin & Review, 16,* 561–566. doi:10.3758/PBR.16.3.561

Yan, M., Risse, S., Zhou, X., & Kliegl, R. (2012). Preview fixation duration modulates identical and semantic preview benefit in Chinese reading. *Reading and Writing, 25,* 1093–1111.

Zhou, W., Kliegl, R., & Yan, M. (2013). A validation of parafoveal semantic information extraction in reading Chinese. *Journal of Research in Reading, 36*(Suppl. 1), S51–S63.

Zhou, X., & Marslen-Wilson, W. (1999). Phonology, orthography, and semantic activation in reading Chinese. *Journal of Memory and Language, 39,* 579–606.

Zhou, X., & Marslen-Wilson, W. (2000). The relative time course of semantic and phonological activation in reading Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 1245–1265. doi:10.1037/0278-7393.26.5.1245

Zhou, X., & Marslen-Wilson, W. (2009). Pseudohomophone effects in processing Chinese compound words. *Language and Cognitive Processes, 24,* 1009–1038.