Journal of Memory and Language 75 (2014) 93–103

Contents lists available at ScienceDirect

Journal of Memory and Language

journal homepage: www.elsevier.com/locate/jml

# Syllable articulation influences foveal and parafoveal processing of words during the silent reading of Chinese sentences

Ming Yan [a,*], Yingyi Luo [b,c], Albrecht W. Inhoff [d]

[a] Department of Psychology, University of Potsdam, Germany
[b] Project Center for Brain Science of Language and Cognition, Graduate School of Education, Hiroshima University, Japan
[c] Center for Brain and Cognitive Sciences and Psychology Department, Peking University, Beijing, China
[d] Department of Psychology, Binghamton University, USA

ARTICLE INFO

*Article history:*
Received 26 June 2013
Received in revised form 9 May 2014
Available online 22 June 2014

*Keywords:*
Articulation duration
Neutral tone
Syllabic tone
Sentence reading
Chinese

ABSTRACT

The current study examined effects of syllable articulation on eye movements during the silent reading of Chinese sentences, which contained two types of two-character target words whose second characters were subject to dialect-specific variation. In one condition the second syllable was articulated with a neutral tone for northern-dialect Chinese speakers and with a full tone for southern-dialect Chinese speakers (neutral-tone target words) and in the other condition the second syllable was articulated with a full tone irrespective of readers' dialect type (full-tone target words). Native speakers of northern and southern Chinese dialects were recruited in Experiment 1 to examine the effect of dialect-specific articulation on silent reading. Recordings of their eye movements revealed shorter viewing durations for neutral- than for full-tone target words only for speakers of northern but not for southern dialects, indicating that dialect-specific articulation of syllabic tone influenced visual word recognition. Experiment 2 replicated the syllabic tone effect for speakers of northern dialects, and the use of gaze-contingent display changes further revealed that these readers processed an upcoming parafoveal word less effectively when a neutral- than when a full-tone target was fixated. Shorter viewing duration for neutral-tone words thus cannot be attributed to their easier lexical processing; instead, tonal effects appear to reflect Chinese readers' simulated articulation of to-be-recognized words during silent reading.



## Introduction

Converging evidence from different visual word recognition paradigms indicates that orthographic, phonological, and lexical representations are automatically activated at different points in time during visual word recognition. Less time is spent looking at words with familiar letter sequences, regular phonology, and frequently encountered lexical forms than at words with unfamiliar orthography, irregular phonology, and rarely encountered forms (see Rayner, 2009, for a review). In alphabetic writing systems, activation of a word's represented phonological form occurs more quickly than activation of its semantic properties when the experimental task requires responding to individual words (see Perfetti & Bell, 1991; Tan & Perfetti, 1999, see also Frost, 1998, and Rastle & Brysbaert, 2006, for reviews), and also when it involves recognition of words during silent reading (Lee, Rayner, & Pollatsek, 1999, 2001).

In his comprehensive and influential review of phonological code use, Frost (1998) argued that the relatively early use of this code during visual word recognition suggests that it is relatively simple and abstract, i.e., that

* Corresponding author. Address: Department of Psychology, University of Potsdam, Karl-Liebknecht-Str. 24/25, 14476 Potsdam-Golm, Germany.
*E-mail address:* mingyan@uni-potsdam.de (M. Yan).

http://dx.doi.org/10.1016/j.jml.2014.05.007
0749-596X/

it consists of a sequence of sound-specific categories (phonemes) that map onto a corresponding sequence of visible letters. In the literature this has been referred to as the minimality hypothesis. The hypothesis implies that readers' use of phonological knowledge for visual word recognition has little in common with listeners' use of acoustic features during auditory word recognition, which does include use of articulation-specific sub- and supra-phonemic variation (e.g., Ranbom & Connine, 2007).

Contrary to the minimality hypothesis, the results of several recent studies on visual word recognition with alphabetic scripts revealed that visual word recognition may involve the use of articulation-specific word properties. In Abramson and Goldinger (1997; see also Lukatela, Eaton, Sabadini, & Turvey, 2004) lexical decisions were shorter for monosyllabic target words that were articulated with a short-duration vowel (e.g., "deaf") than for orthographically matched targets that were articulated with a long-duration vowel ("deal"). Similarly, less time is spent gazing at target words with a short- than with a long-articulation vowel during sentence reading, and this effect disappeared when a secondary articulatory suppression task was added to the reading task (Huestegge, 2010).

Furthermore, Ashby's work (Ashby & Clifton, 2005; Ashby & Martin, 2008; Ashby & Rayner, 2004; Ashby, Sanders, & Kingston, 2009; Ashby, Treiman, Kessler, & Rayner, 2006) provided evidence for the early use of articulation-specific sub- and supra-phonemic features during visual word recognition. Using the eye-movement-contingent display change technique (Rayner, 1975) to manipulate the parafoveally visible preview of a target word *before* the actual target was subsequently fixated, Ashby et al. (2006) showed that readers spent less time gazing at a target when its previously available parafoveal preview contained a vowel with a matching rather than a mismatching articulation duration. Using the same technique, Ashby and Rayner (2004; see also Ashby & Martin, 2008) further showed that readers spent less time viewing target words when the parafoveal preview revealed the intact beginning syllable than when it revealed a shorter or longer beginning letter sequence (preview of "pi###" or pil##, respectively, prior to the fixation of the target "pilot"). Syllabic stress also influenced word viewing durations, as stressed words and words with two syllables received more fixations than unstressed and monosyllabic words (Ashby & Clifton, 2005; Breen & Clifton, 2011). Using the masked priming technique with ERP recordings and prime target sequences with matching or mismatching voicing for final consonants, Ashby et al. (2009) showed that mismatching consonants elicited more negative amplitudes than matching consonants as early as 80 ms after target onset. Together, these findings with English text indicate that articulation-specific features influence visual word identification.

Logographic scripts, such as Chinese, differ fundamentally from alphabetic scripts, as they are not designed to represent sound-specific properties of words. The basic units of Chinese script, characters, are designed to represent syllables and morphemes, and graphemically similar characters can have radically different pronunciations. Nevertheless, there is strong evidence for the use of sound codes during visual word recognition. Effective prime durations are shorter for homophone primes than for matched semantic primes (Tan, Hoosain, & Peng, 1995; Tan, Hoosain, & Siok, 1996; see Perfetti, Liu, & Tan, 2005, for a review), and Chinese readers use parafoveally available phonological information during the subsequent fixation on a previewed target character (Liu, Inhoff, Ye, & Wu, 2002; Tsai, Lee, Tzeng, Hung, & Yen, 2004; Yan, Richter, Shu, & Kliegl, 2009).

Nevertheless, Frost (2012) argued that orthographic transparency has a profound effect on the use of phonological information, and some language-specific features of Chinese script may prevent its readers from early and fast access to phonological information. Chinese characters are formed according to a variety of principles (Feng, Miller, Shu, & Zhang, 2001). The most ancient characters are so-called *pictographs* and *ideograms*, which originated from drawings by ancient cave dwellers or formed through analogy or association. Obviously, these characters were designed to represent meaning rather than pronunciation. The majority of modern Chinese characters are *phonograms* typically consisting of two independent components/radicals, one of which represents the meaning category of the whole character, and the other provides a rough clue regarding pronunciation. However, the articulation of the phonetic radical and of the full character often differ, as only about 30% of phonetic compound characters have the same pronunciation as their phonetic radical (Zhou & Marslen-Wilson, 1999). Frost (2012) pointed out that the use of semantic information could thus precede use of phonological information in phonograms because of their orthographic structure: the semantic radical is generally to the left of the phonetic radical and Chinese sentences are read from left to right.

In agreement with script-specific effects of word recognition, there is considerable experimental evidence according to which Chinese script is optimized for foveal and parafoveal semantic information extraction. For example, direct access from orthography to semantics in Chinese, with phonological mediation bypassed, has been reported with character/word recognition tasks under some circumstances (Chen & Shu, 2001; Zhou & Marslen-Wilson, 2000). Using the error disruption paradigm, Feng et al. (2001) obtained evidence for very early phonological activation for English readers, but not for Chinese readers. Furthermore, a direct comparison of phonological and semantic parafoveal preview effects with Chinese text revealed more robust semantic than phonological preview effects (Yan et al., 2009). Multi-level regression models that included the time line of a parafoveal word preview as predictor showed that preview benefits accrued earlier for useful semantic but not for useful phonological information (Tsai, Kliegl, & Yan, 2012; Yan, Risse, Zhou, & Kliegl, 2012; Zhou, Kliegl, & Yan, 2013).

For Chinese text, the current default assumption regarding a word's represented phonological form appears to be that it is impoverished and relatively abstract, as maintained by the minimality hypothesis. In fact, the somewhat later use of phonological codes during the reading of Chinese could well be associated with the use of abstract phonological knowledge. While word recognition with alphabetic scripts uses articulation-specific sub- and supra-phonemic

detail during early stages of visual word recognition, word recognition with Chinese script could bypass use of these speech-specific properties in favor of more categorical sound-specific representations. Alternatively, the nature of the effective sound code could be independent of script type, and the use of spoken word features for visual word recognition could be obligatory and universal.

So far, there has been no published study on the nature of the effective phonological code during the silent reading of Chinese sentences. To determine the validity of the minimality hypothesis for the recognition of Chinese words under relatively natural reading conditions, articulation-specific features of the to-be-recognized target words were manipulated, and recordings of eye movements were used to index visual word recognition. To rule out a potential confounding of articulation-specific properties with other word properties, words with different articulation properties were carefully matched with regard to visual and linguistic features. Furthermore, speakers of northern and southern dialects read sentences with a subset of target words that had a dialect-specific syllabic tone articulation. According to the minimality hypothesis, neither the articulation-defining properties of words nor readers' dialect should influence target recognition during silent reading. If, however, Chinese readers used articulation-specific features during visual word recognition, then articulatory features should influence visual word identification during silent reading.

The assumption of the universality of spoken language usage would also require evidence for relatively early use of articulation-specific word features among different languages. Recordings of eye movements during reading can be used to obtain indexes that are somewhat sensitive to the time course of information usage. Specifically, processing effects that emerged within the first fixation on a target word are generally assumed to occur relatively early – and earlier than effects that require that a target be re-fixated or re-read (Inhoff, 1984; Pollatsek, Lesch, Morris, & Rayner, 1992). Universality of spoken language usage should thus manifest itself not only in Chinese readers' use of articulation specific word properties, but also in effects of spoken word properties on first fixation duration.

## Experiment 1

Mandarin Chinese (i.e., the standard dialect, which was created on the basis of northern Chinese dialects) is a tonal language. An individual character maps onto a syllable that is articulated with one of four lexical (full) tones when the character or a corresponding single-character word is articulated in isolation. The key experimental manipulation took advantage of principled variation in the tone articulation of the second syllable of two-character words in Mandarin Chinese. Specifically, some syllables are articulated with full tones when read in isolation and when they constitute single-character words in text, but with a simplified and 'neutral' articulation when they are the second constituent character of a two-character word. The articulation of a word's second syllable with a neutral tone does not entail substitution of one full tone with another, a change of syntactic status, or of word meaning in the present study. Instead, it involves the generation of a shorter and lighter (less stressed) variant articulation through the removal - or minimizing- of the tone's prosodic pattern. In other words, the neutral tone syllable is articulated with shorter articulation duration, with less stress, and with lower intensity (Cao, 1986, 1992; Lin & Yan, 1980).

Importantly, the neutralizing of the second syllable is confined to a specific subset of syllables, and it is subject to dialect-specific variation. In Experiment 1, two types of two-character target words with identical first characters, which were fully articulated irrespective of readers' dialects, were used. One type consisted of words whose second characters/syllables had a neutral tone for speakers of northern dialects and, in contrast, a full tone for speakers of southern dialects (henceforth referred to as 'neutral-tone words'). The second type consisted of words whose second syllables were articulated with a full tone across both groups of speakers (referred to as 'full-tone words'). The neutral and full tone characters were carefully matched on a large number of perceptual and linguistic dimensions. If spoken word properties influenced target word identification, then target type (neutral- vs. full-tone words) and dialect type (northern- vs. southern-dialect speaking readers) should influence the silent reading of Chinese, with relatively short viewing durations when neutral-tone targets are read by speakers of northern dialects.

### Method

#### Participants

Thirty-five[1] graduate and undergraduate students from the Beijing Normal University effectively participated in the eye-tracking experiment. According to a categorization of Chinese dialects (Wurm, Tsou, & Bradley, 1987), 17 participants from Heilongjiang, Jilin, Liaoning, Shandong, Hebei, Henan, Shannxi and Sichuan provinces were considered northern-dialect speakers and the remaining 18 participants from Hubei, Hunan, Jiangxi, Jiangsu, Zhejiang, Anhui, Guangdong, Guangxi, Fujian and Hainan provinces were considered southern-dialect speakers.

An additional 20 participants from these provinces (10 native northern-dialect speakers) were recruited for an oral reading study which was used to measure the articulation difference of neutral- and full-tone targets for the two groups of speakers. We carefully chose participants who had remained in their places of birth before moving to Beijing for undergraduate or graduate studies so that their use of a particular dialect was well established. Southern and northern Sinitic dialects are in general not mutually intelligible (Chaoju & van Heuven, 2009), and thus Mandarin is sometimes considered a second language for southern-dialect speakers.[2]

[1] Data from one participant were excluded from analyses due to his low comprehensive rate (i.e., <80%).

[2] Whether different Chinese dialects should be considered as related languages is controversial among linguists. The traditional criteria of mutual intelligibility been called into a question (see Chappell, 2001, for a review). It has been argued that even within a certain dialect group such as Min or Yue, there is a high degree of mutual unintelligibility between subdivisions. In addition, there is a unifying use of the written Chinese language irrespective of dialect.

**Table 1**
Word properties.

| | Full tone | Neutral tone |
|---|---|---|
| Example word | 苍鹰 | 苍蝇 |
| Meaning | Goshawk | Fly |
| North pronunciation | cang1-ying1 | cang1-ying0 |
| South pronunciation | cang1-ying1 | cang1-ying2 |
| WF 1 | 18.2 (26.8) | 20.2 (20.3) |
| WF 2 | 11.0 (17.0) | 11.9 (19.9) |
| N. of strokes | 16.0 (3.5) | 16.0 (4.3) |
| AoA | 6.0 (1.0) | 5.7 (1.2) |
| Imageability | 3.6 (0.9) | 3.8 (0.8) |
| HNS | 0.6 (1.0) | 0.5 (0.9) |

Means (and standard deviations, in parentheses) of frequency per million from Beijing Language Institute Publisher (1986; WF1) and from Cai and Brysbaert (2010; WF2), number of strokes, age of acquisition (AoA), imageability and homophony neighborhood size (HNS, i.e., number of homophonic words) of the target words.

Furthermore, three groups of 14, 16 and 30 students were recruited to obtain age-of-acquisition, imageability and predictability estimates for the two types of target words. All participants had normal or corrected-to-normal vision.

*Materials*

Forty-eight pairs of two-character neutral- and full-tone target words were selected. The first characters of both members of the full- and neutral-tone word pair were always identical; the second characters of the word pair were closely matched. As shown in Table 1, there were no differences between the two types of words with respect to frequency, age-of-acquisition (AoA), imageability, homophony, neighborhood size, or visual complexity as indexed by number of strokes [for AoA analysis: $F(1,94) = 2.040$, $p = .156$; all other $F$-values < 1].

The two members of target word pair were designed to differ in their second characters with respect to their susceptibility to dialect-specific variation in articulation. This was tested using twenty participants who read aloud 96 sentences in Mandarin, each containing a target word at the middle position (e.g., the word *fly* in the following example). The sentences were structured as:

Ta / shuo / $_{\text{Target word}}$ [cang-ying] / zhe-ge / ci.
He / say / $_{\text{Target word}}$ [fly] / this-classifier / word.
'He said the word [fly].'

The speakers were recorded individually in the speech lab at Minzu University of China, Beijing. They 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. ProsodyPro, a Praat script (Xu, 2013) was used to perform the initial acoustic analysis (Boersma & Weenink, 2005). Based on the waveform and spectrogram of each sentence, segmentation labels were marked manually to identify the target syllable. Duration and intensity measurements regarding the marked segmentations (i.e., the target syllables), were then automatically extracted from the sentences.

Effects of word type and dialect type on the duration and the intensity of the target word's second syllable were examined. The second syllable was pronounced with a shorter duration in the neutral- than in the full-tone word condition, as shown by a main effect of word type [$F(1,94) = 20.113$, $p < .001$]. In addition, this word type effect was modulated by dialect type [$F(1,94) = 14.219$, $p < .001$], with a larger word type effect for northern-dialect speakers (232 ms vs. 260 ms, $F(1,94) = 32.724$, $p < .001$) than for southern-dialect speakers (261 ms vs. 278 ms, $F(1,94) = 9.221$, $p = .003$). In addition, the intensity of second syllable articulation was weaker in the neutral- than the full-tone word condition [$F(1,94) = 4.900$, $p = .029$]. Word type and dialect type interacted [$F(1,94) = 11.675$, $p = .001$], as intensity differences were significant for northern-dialect speakers [64.9 dB vs. 66.3 dB, $F(1,94) = 9.128$, $p = .003$], but not for southern-dialect speakers [66.0 dB vs. 66.7 dB, $F(1,94) = 1.901$, $p = .171$]. Together, these data suggest that the neutral-tone effect is more evident for northern- than southern-dialect speakers and that readers' native dialect influenced their syllabic tone articulation in Mandarin.[3]

Two sentence frames were constructed for each pair of target words so that there were a total of 96 pairs of experimental sentences. Critically, for each pair of full- and neutral-tone sentences, context preceding the target and at least the first word immediately following the target were identical in order to minimize possible influences of context as well as from parafoveal processing. Sentence contexts were written to be neutral for all target words, and a cloze test indicated that neutral- and full-tone target words were equally constrained, i.e., they were predicted 10 and 9 times, respectively, out of 1440 guesses ($F$-value < 1). All sentences contained 16–24 characters ($M = 20.1$, $SD = 2.0$), and target words were never among the first three or the last three words. Each sentence was only presented once, the two members of a sentence pair being counterbalanced over participants. An example of sentences is shown in Fig. 1.

*Apparatus*

Eye movements were recorded with an EyeLink 1000 system with a 35 mm lens running at 1000 Hz. Single sentences were presented on the vertical position one third from the top of the screen of a 19-inch ViewSonic G90f monitor (resolution, 1024 by 768 pixels; frame rate, 100 Hz). Participants were seated comfortably with a chin rest and a forehead rest at a distance of 56.5 cm from the monitor. The font Song 40 was used with one character equal to 1.4 degrees of visual angle. All recordings and calibrations were done monocularly based on the right eyes and viewing was binocular.

*Procedure*

Gaze position was calibrated with a nine-point grid. Prior to the presentation of each sentence, the eye tracker checked for drift. An extra calibration was performed if the tracker did not detect participants' eyes within a pre-defined window around the initial fixation point.

[3] Our result is slightly different from previous findings which claimed that southern-dialect speakers do not utilize neutral tone at all; the difference might be due to the popularization of Mandarin in recent years.

Full tone condition

他们发现有大量*苍鹰*飞舞在草原上空。

(They saw many *goshawks* hovering above the grassland.)

Neutral Tone condition

他们发现有大量*苍蝇*飞舞在食品加工厂内。

(They saw many *flies* hovering in the food factory.)

**Fig. 1.** A pair of example Chinese sentences using the target words 苍鹰 (goshawk) and 苍蝇 (fly) with identical first character (i.e., 苍). The target words are in italics.

Participants were instructed to read each sentence for comprehension, then fixate a dot in the lower right corner of the monitor, and finally press a button to signal completion of the trial. Of the 96 sentences 30 were selected for being followed by an easy yes–no comprehension questions about the general meaning of the sentences. The same questions were asked to two different versions of the sentence. On average participants correctly answered 93% of all questions (*SD* = 3%).

*Data analysis*

Fixations were determined with an algorithm for binocular saccade detection (Engbert & Kliegl, 2003). A total of 101 (i.e., 3%) trials were removed due to participants' blinks, coughs or body movements during reading or tracker errors. We distinguish between first-fixation durations (FFDs; the first fixation on a word, irrespective of the number of fixations), single-fixation durations (SFDs; cases in which a word was inspected with exactly one fixation), and gaze durations (GDs; the sum of fixations during the first pass reading of the word). In addition, we computed intra-word refixation rates for fixated words. FFDs and SFDs shorter than 60 ms or longer than 600 ms and GDs longer than 800 ms were excluded (i.e., 3% of valid trials). A total of 2886, 2488 and 2886 observations on the target word contributed to the analyses of FFDs, SFDs and GDs.

Effect size estimates are based on a linear mixed model (LMM) for durations and a generalized linear mixed model (GLMM) for skipping and regression rates. Participants and items were entered as crossed random effects using the *lmer* program of the *lme4* package (Bates & Maechler, 2011) in the R environment for statistical computing and graphics (R-Core Development Team, 2012). There were three fixed effects: the main effects of word type and dialect type and their interaction. Estimates twice larger than their standard errors were interpreted as significant at the 5% level; given the number of participants and the large number of observations for each participant, the *t*-statistic in LMMs (i.e., M/*SE*) effectively corresponds to the *z*-statistic. Analyses for untransformed and log-transformed durations yielded the same pattern of significance.

*Results*

On average, readers skipped the target words on 9% of valid trials and regressed back to the target words on 2% of trials. Neither the main effects nor the interactions were reliable in the two corresponding GLMMs (all *p*-values > .1). As shown in Table 2, effects of word type and dialect type interacted, and the corresponding effect was statistically significant for GDs ($b = -.053$, $SE = .026$, $t = -2.04$), and marginally significant for SFDs ($b = -.038$, $SE = .022$, $t = -1.70$) and refixation probabilities ($b = -.401$, $SE = .023$, $t = -1.77$, $z = .077$). None of the main effects was significant [all abs(*t*-values) < 1.7].

The interaction of dialect type with word type indicates that dialect-specific spoken word properties influenced visual word identification. As predicted, northern-dialect readers spent less time viewing target words when the second syllable was pronounced with a neutral tone rather than a full tone. With paired contrasts, this difference was significant for GDs ($b = .048$, $SE = .018$, $t = 2.58$), and there were similar numerical trends for FFDs ($b = .021$, $SE = .014$, $t = 1.48$) and SFDs ($b = .027$, $SE = .015$, $t = 1.78$). For northern-dialect speakers, there were also marginally more first-pass fixations ($b = .033$, $SE = .019$, $t = 1.75$), and a higher refixation probability ($b = .320$, $SE = .162$, $t = 1.97$, $z = .049$) for full-tone as compared to neutral-tone target words.

In striking contrast to this, the identical set of target words revealed virtually no differences in any of the three fixation duration measurements for southern-dialect readers who articulated both members of the target word pair with full(er)-tone second syllables ($b = -.001$, $SE = .017$, $t = -0.09$; $b = -.012$, $SE = .016$, $t = -0.74$ and $b = -.005$, $SE = .019$, $t = -0.29$, for FFD, SFD and GD analyses, respectively). In addition, there was no difference in refixation probability between word types among southern-dialect readers ($b = -.088$, $SE = .157$, $t = -0.56$, $z = .575$).

*Discussion*

Experiment 1 used dialect-specific variation in the articulation of visual target words to determine the use of spoken language properties for visual word recognition during the silent reading of Chinese. The GDs of northern-dialect speaking readers were shorter for neutral- than for full-tone target words. Critically, Experiment 1 further

**Table 2**
Means (and standard deviations, in parentheses) of skipping probabilities (SP), refixation probabilities (RP), first-fixation duration (FFD), single-fixation duration (SFD), gaze duration (GD) on target words in Experiment 1.

| | Northern | | Southern | |
|---|---|---|---|---|
| | Neutral tone | Full tone | Neutral tone | Full tone |
| SP | 9% (7%) | 9% (7%) | 9% (5%) | 7% (5%) |
| RP | 11.7% (9.0%) | 15.1% (8.9%) | 14.6% (7.8%) | 13.5 (8.8%) |
| FFD | 271 (29) | 276 (29) | 275 (30) | 272 (25) |
| SFD | 271 (30) | 277 (30) | 281 (35) | 276 (30) |
| GD | 301 (33) | 315 (34) | 309 (34) | 308 (41) |

*Note.* Means and standard deviations are computed across participants' means.

revealed equivalent GDs for the two types of target words for speakers of southern dialects, thus implying that neutral- and full-tone targets were effectively matched for visual, graphemic, and linguistic properties. Differences between these two types of words with northern-dialect speaking readers can thus be attributed dialect-specific articulation properties.

Southern dialect speakers are exposed to Mandarin, and they may have learned to associate neutral tones with orthographic properties, as neutral tones can be marked in Pinyin (a shallow, Roman alphabet-based orthography that seeks to represent the sound of Chinese characters) which is used when learning to read. Southern dialect speakers are thus familiar with the use of neutral tone syllables, and this should have diminished the interaction between dialect type of tone type, making the neutral-tone effect equally visible for northern- and southern-dialect speakers. Nevertheless, enduring and principled dialect-specific articulation differences were observed, and they accounted for target viewing during silent reading. As such, Experiment 1 provides strong evidence for universality in the use of articulation-specific features during visual word identification.

The results of Experiment 1 are also in general agreement with related work by Luo, Yan, Yan, Zhou, and Inhoff (submitted for publication) and Ashby and Clifton (2005). In Luo et al., articulation of the second character's syllabic tone influenced target viewing duration in northern-dialect speaking readers. Although the target words in the full- and neutral-tone conditions were carefully matched, the evidence for the use of articulation specific features was based on the comparison of two different word types that were embedded in different sentence contexts, and readers heard a spoken syllable related to the first character of a target word when it was fixated. Target processing could thus have been compromised by contextual bias, the drawing of attention to the irrelevant spoken syllable, or the drawing of attention to speech-specific target properties. Therefore, Experiment 1 extends Luo et al. in that it provides the first direct evidence for the use of articulation specific word features during Chinese word recognition under natural reading conditions. As noted earlier, Ashby and Clifton (2005) observed that words with two stressed syllables (i.e., marked for stress) were fixated more often than their unmarked counterparts. This agrees with our findings, as northern-dialect speakers refixated full-tone words with two fully articulated syllables more often than neutral-tone words that contained only one fully articulated syllable.

Use of articulation-specific knowledge may occur slightly later during the reading of Chinese than during the reading of alphabetic text. Consistent with this, FFDs did not yield a significant interaction between target word type and readers' dialect type. Analogously, Luo, et al. (submitted for publication) obtained word type differences for GDs but not for FFDs, and Huestegge (2010) also obtained a vowel duration (vowel length) effect for GDs but not for either FFDs or skipping rates for German text. Huestegge concluded that "in normal sentence reading vowel length affects lexical processing rather late during lexical processing, presumably around the time of the completion of lexical access" (p. 7). He also argued that the effect of vowel duration on GD indicates that speech-based phonetic information plays a significant role during lexical but not during post-lexical processing during silent reading, as GD is generally assumed to be determined by lexical access (Reichle, Rayner, & Pollatsek, 2003). The findings of Experiment 1 appear consistent with this view. That is, articulation-specific features appear to contribute to visual word recognition when recognition is imminent but not before that, perhaps by providing converging evidence for a particular lexical candidate during a concluding stage of lexical processing.

## Experiment 2

The main results of Experiment 1 showed that the articulation properties of a word influenced its GD during the silent reading of Chinese sentences. The word type effect was not reliable for FFDs, possibly because articulation-specific word properties were used at a relatively late stage of processing, when lexical access was imminent. However, FFDs of northern-dialect speaking readers also tended to be shorter for neutral- than for full-tone targets in Experiment 1, and the difference between GDs, which include target word refixation durations, and FFDs, which did not include refixations, was not clear cut. The claim that articulatory properties influence recognition only when it is imminent must thus be considered tentative, and it was further examined in Experiment 2.

Experiment 2 also sought to distinguish between two types of articulation use. As suggested by Lukatela et al. (2004), recognition of a word could involve mentally simulated articulation. According to one view, northern-dialect speaking readers gazed longer at full- than at neutral-tone targets simply because their simulated articulation took more time. The simulation itself was automatic and was equally difficult for the two types of target words. Alternatively, longer GD on full-tone words does not only reflect simulated articulation but also processing difficulty: In general, a word's viewing duration is equated with its ease of processing, the assumption being that investment of more effort would result in longer viewing durations. However, this rationale is circular when the duration of a process is to be distinguished from its processing demands. These two simulation hypotheses could not be teased apart in Experiment 1. Experiment 2 was, therefore, predicated on a different assumption: that processing effort and viewing duration can be dissociated, and it employed an indirect means for an independent assessment of speech simulation effort. Instead of using target viewing duration, it examined how neutral- and full-tone target words influenced the processing of the next (parafoveally visible) word in the text with northern-dialect speaking readers.

Prior empirical and computational work has shown that increased difficulty with the processing of a fixated word diminishes the acquisition of useful information from a spatially adjacent parafoveally visible word (Henderson & Ferreira, 1990; Schroyens, Vitu, Brysbaert, & d'Ydewalle, 1999; Yan, Kliegl, Shu, Pan, & Zhou, 2010). In Henderson and Ferreira's (1990) original study, parafoveal preview of a word was less useful when the fixated word was a low frequency word than when it was a high frequency word

(Experiment 1), and when its syntactic parsing was likely to be incorrect than when it was correct (Experiment 2). Yan et al. (2010) demonstrated a dynamical modulation of parafoveal information extraction beyond the parafoveal word to the right of fixation such that a high frequency but not a low frequency parafoveal word yielded a preview benefit for the subsequent word in the sentence. If articulation of a full-tone target word demanded more effort than that of a neutral-tone target, northern-dialect speaking readers should obtain less useful information from a parafoveally visible word when a full-tone target is fixated than when a neutral-tone target is fixated.

### Method

#### Participants

An independent group of 36 graduate and undergraduate students from the Beijing Normal University with normal or corrected-to-normal vision effectively participated in Experiment 2. They were from the provinces of Heilongjiang, Jilin, Liaoning, Shandong, Hebei, Henan, Shannxi and Gansu, or from the cities of Beijing and Tianjin. That is, all of them were native speakers of northern-dialect.

#### Material, apparatus, procedure and data analysis

The material, apparatus, procedure and data analysis were identical to those of Experiment 1, except that the eye-movement-contingent display-change technique (Rayner, 1975) was adopted to manipulate the visibility of the post-target word during fixations on prior words. That is, an invisible boundary for display change was set between the target word N (i.e., the neutral- and full-tone critical words) and the post-target word N + 1. Prior to boundary crossing, the sentence either contained the intact post-target word (the parafoveal preview condition) or two orthographically matched masking characters (the parafoveal masking condition). The mask was replaced with the correct post-target word [frequency per million: $M = 217.6$, $SD = 346.3$; number of strokes: $M = 15.3$, $SD = 4.2$] when the eyes crossed the boundary (see Fig. 2). Display changes were completed within 13 ms. Sentence reading was again followed by comprehension questions, and these were correctly answered on 28 out of 30 trials, on average ($M = 92\%$, $SD = 5\%$).

As in prior studies (e.g., Yan et al., 2010, 2012), trials with late display changes were excluded (26%) from the analyses of post-target words, because readers were more likely to perceive a display change or flicker on these trials. A total of 91 (i.e., 3%) trials were removed due to tracking errors, blinks, coughs or other body movements. Fixation durations were filtered using the same duration criterion as in Experiment 1, and trials in which readers regressed out of a target word were excluded which removed another 6% of trials. This left 2885, 2467 and 2885 FFD, SFD and GD data, respectively, for the assessment of word type and parafoveal preview effects on target words. Analogous criteria were applied to post-target words. Trials with regressions out of target or post-target words were also excluded (5%), because these trials may reflect incomplete parafoveal or foveal processing of post-target words. Use of the same fixation duration filters resulted in the exclusion of another 3% of trials, and a total of 1794, 1586 and 1794 observations contributed to the analyses of post-target FFDs, SFDs and GDs, respectively.

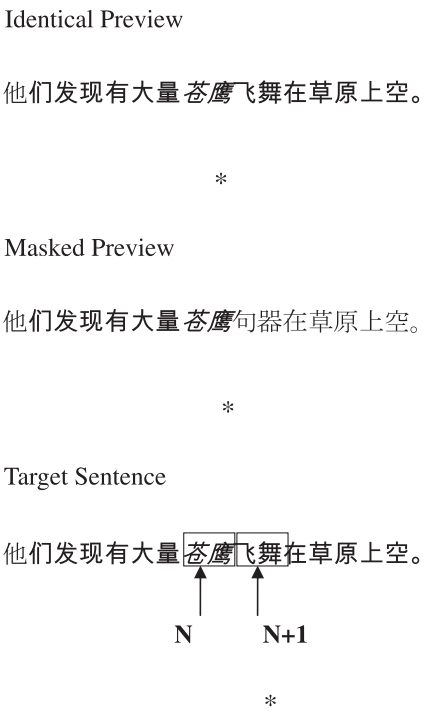


**Fig. 2.** A set of example Chinese sentences using the boundary paradigm. The target sentence is translated as: They saw many *goshawks* hovering above the grassland. The previews (飞舞 or 句器) that are initially displayed in the post-target location are replaced by the correct post-target word (飞舞) as soon as the reader's eyes (as indexed by the asterisks in the figure) cross the invisible boundary located between word N (苍鹰) and N + 1 (飞舞). The pre-boundary target words are in italics.

### Results

#### Target word region

On average, participants skipped 9% of the full and neutral tone target words and regressions to targets occurred on 2% of trials. Neither main effects nor interactions were reliable for these data (all $p$-values > .1).

As shown in Table 3A, and in agreement with Experiment 1, native northern-dialect speaking readers spent less time reading neutral- than full-tone words ($b = .028$, $SE = .010$, $t = 2.91$; $b = .027$, $SE = .011$, $t = 2.57$ and $b = .027$, $SE = .013$, $t = 2.16$, for FFD, SFD and GD data, respectively). Extending Experiment 1, the word type effect was significant for all three viewing duration measures, including FFD. The main effect of parafoveal preview visibility (visible vs. masked) and the interaction of target word type with parafoveal preview were not significant [all abs($t$-values) < 0.5].

#### Post-target word region

Participants skipped 13% of the post-target words, and there was neither a parafoveal preview nor a target type effect on skipping rate (all $p$-values > .1).

As shown in Table 3B, the interaction of parafoveal visibility and target word type was significant for FFDs and GDs

**Table 3**
Means (and standard deviations, in parentheses) of skipping probabilities (SP), first-fixation duration (FFD), single-fixation duration (SFD), gaze duration (GD) on target (A) and post-target words (B).

| | Neutral tone | | Full tone | |
|---|---|---|---|---|
| | Identical preview | Masked preview | Identical preview | Masked preview |
| *(A)* | | | | |
| SP | 9% (8%) | 9% (7%) | 9% (7%) | 8% (7%) |
| FFD | 263 (26) | 262 (32) | 273 (38) | 272 (41) |
| SFD | 263 (27) | 264 (33) | 273 (39) | 273 (43) |
| GD | 297 (28) | 298 (40) | 306 (41) | 308 (49) |
| *(B)* | | | | |
| SP | 13% (9%) | 13% (11%) | 15% (10%) | 13% (11%) |
| FFD | 262 (32) | 258 (38) | 258 (34) | 269 (36) |
| SFD | 261 (32) | 259 (41) | 257 (34) | 269 (39) |
| GD | 288 (45) | 282 (44) | 284 (39) | 297 (51) |

*Note.* Means and standard deviations are computed across participants' means.

($b$ = .055, $SE$ = .025, $t$ = 2.18 and $b$ = .064, $SE$ = .031, $t$ = 2.05), and it was marginally significant for SFDs ($b$ = .049, $SE$ = .027, $t$ = 1.84). None of the main effects was significant [all abs($t$-values) < 1.7]. Contrary to the assumption that simulated articulation of full-tone target words would increase recognition difficulty and diminish preview benefits, these interactions revealed that readers acquired more parafoveal information and had larger preview benefits when the second syllable of the target word was articulated with a full tone ($b$ = .034, $SE$ = .018, $t$ = 1.96; $b$ = .038, $SE$ = .019, $t$ = 2.04 and $b$ = .043, $SE$ = .022, $t$ = 2.01, for FFD, SFD and GD analyses, respectively) rather than a neutral tone [all abs($t$-values) < 1.5].[4]

### *Discussion*

Experiment 2 pursued two main goals: to determine whether articulation-specific features of words are used during early stages of visual word recognition, and whether the simulated articulation of a to-be-recognized word occurs effortlessly. To these ends, we recruited only northern-dialect speakers for the reading of sentences with full- and neutral-tone target words. Replicating Experiment 1, GDs revealed robust differences between the two types of words. Furthermore, Experiment 2 revealed a significant word type effect for FFDs, and it showed that longer articulation durations for full-tone words did not diminish the effectiveness with which useful information was obtained from the next word in the text. To the contrary, the viewing of full-tone words yielded larger rather than smaller parafoveal preview benefits.

Comparison of FFD effects across experiments revealed shorter durations for neutral- than full-tone targets in both experiments. Similarly, vowel duration influenced FFDs in Huestegge (2010), and FFDs were similar to the syllabic tone articulation effects in our Experiment 2. The consistency of numeric FFD effects thus indicates that articulation-specific properties influence early stages of visual word recognition, and that this occurs independent of script type. The statistical significance of FFD effects could depend on experimental power. Overall, Experiment 2 included more than twice the number of northern-dialect speaking readers than Experiment 1, and this diminished standard error size.

Preview benefits in Experiment 2 were surprisingly small when compared to prior work. Two factors may have contributed to the relatively small size of the effect: The word frequency of the target was relatively low and the visual complexity of the target was relatively high. For example, on average, according to the Modern Chinese Word Frequency Dictionary, the pre-boundary words in Yan et al. (2009) occurred 183 times per million and they contained 14 stokes; in the current study, the pre-boundary target words occurred only 19 times per million and they contained 16 strokes. Both factors increased processing difficulty and thus presumably decreased preview effects (Henderson & Ferreira, 1990; for Chinese evidence, see Luo, Yan, & Zhou, 2013 and Yan et al., 2010). Therefore, it is reasonable to assume that the readers in the present study would obtain relatively little parafoveal information and that differences between preview conditions would be small. In spite of this, preview benefits for post-target words differed between the neutral- and the full-tone target words, with smaller benefits in the neutral-tone condition. Potential sources for this effect are considered in the General Discussion.

## General discussion

The current study examined the use of articulation-specific word properties during the silent reading of Chinese sentences. Potentially confounding effects of linguistic word properties, ease of comprehension, or task demands were excluded by using the identical set of linguistic materials and tasks in two different word articulation conditions. This was accomplished through the use of target words with dialect-specific articulation properties. The second characters of these targets were assumed to bear a neutral-tone articulation for speakers of northern dialects and a full(er) tone for speakers of southern dialects. Recordings of eye movements in Experiment 1 revealed that northern-dialect speakers spent less time gazing at neutral-tone targets than speakers of southern dialects, and that the two groups of speakers did not differ when they read full-tone words. In addition, Experiment 2 showed that the effect of word type emerged during the very first fixation on the target words, suggesting that articulation influenced early stages of visual word recognition, and that it influenced the success with which useful information could be obtained from the upcoming (parafoveally visible) word.

Similar to Experiment 1, Filik and Barber (2011) examined how dialect-specific differences between British speakers influenced the reading of identical alphabetic text. Participants read poems whose final (target) word either did or did not rhyme with prior text, depending on a reader's dialect. Dialect type influenced oculomotor responses to the target under these conditions, with more outgoing

[4] There was a significant three-way interaction in SFD ($b$ = .108, $SE$ = .054, $t$ = 2.01) when using launch site as a covariate, indicating that the preview benefit was strongest with near launch site and with full-tone target words.

regressive saccades, longer regression path durations, and longer second pass reading times when it did not rhyme in a particular dialect. Rhyming did not influence target word's FFD or GD, however, and corrective saccades out of a target word and the following re-inspection of prior sentence context seemed to occur after the target word was identified. This suggests that the rhyming of a word within a particular context influenced the success with which it could be embedded in available context after it was recognized (see also Luo & Zhou, 2010). As such, their results can be reconciled with the view that word recognition does not use articulation-specific word properties, as maintained by the minimality hypothesis. Experiments 1 and 2 thus extend Filik and Barber (2011), as the dialect-specific articulation of a word influenced its recognition, contrary to the hypothesis. Furthermore, regressions out of target words were relatively rare in the current study, indicating that articulation-specific effects were not a byproduct of post-lexical processing difficulties.

Effects of word-specific articulation features on visual word recognition are well established for alphabetic script. Vowel articulation duration has been shown to influence the classification and naming and individually presented visual target words (Abramson & Goldinger, 1997; Lukatela et al., 2004), and also the viewing of target words during silent sentence reading (Ashby & Clifton, 2005; Huestegge, 2010) with shorter processing durations when the articulation duration of a target word was short than when it was long. ERP recordings during the recognition of words, written in alphabetic scripts, further showed that a word's articulation influenced its early stages of processing (Ashby & Martin, 2008; Ashby et al., 2009; Wheat, Cornelissen, Frost, & Hansen, 2010).

The use of articulation specific properties irrespective of script type suggests that theoretical conceptions that sought to account for articulation-specific effects with alphabetic script may also apply to Chinese script. Effects of spoken vowel duration for (alphabetic) words were used by Lukatela et al. (2004) to argue that processing of language by eye is not visual but gestural, and that theories of spoken and visual language comprehension are inseparable (see also Liberman, 1996). Key to their account was that visual word identification involves the generation of articulatory gestures that provide access to lexical knowledge. Readers thus responded more effectively to alphabetic words with short- than with long-duration vowels because the simulated articulation of a short vowel took less time and resulted in an earlier accessing of lexical knowledge. Similarly, viewing durations of northern-dialect speaking Chinese readers were shorter for neutral- than for full-tone targets because the simulated articulation of a neutral-tone target had lower intensity and took less time.

With alphabetic text, vowel duration was linguistically informative, and a simulated articulation could have distinguished otherwise similar lexical alternatives. In contrast to this, simulation of a Chinese target word's syllabic tone with a full- or neutral-tone articulation did not distinguish between lexical alternatives, as both target articulations denote the same word. Experiments 1 and 2 thus provide novel evidence for the use of articulation-specific word features during visual word identification, as they influenced word recognition even when they did not distinguish between lexical alternatives.

Rather than being exclusively gesture-based, the accessing of Chinese word meaning could also involve the use of graphemic knowledge. 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; in these models, use of phonological forms conveys benefits because they provide additional and converging evidence for the accessing of word meaning. Given that shorter GDs normally reflect more effective lexical access (Engbert, Nuthmann, Richter, & Kliegl, 2005; Rayner, 2009; Reichle et al., 2003), the results of Experiments 1 and 2 would be consistent with a slightly modified version of this code-convergence conception according to which activation of a word's graphemic form instantiated a simulated articulation which required less time when the gesture involved production of a neutral rather than a full syllabic tone. The converging gestural information could be established earlier for neutral-tone words resulting in a faster recognition than for full-tone words.

The faster accessing of neutral-tone words (for speaker of northern dialects) need not imply, however, that their processing required in general less effort. Parafoveal preview benefits were reduced when neutral tone targets were fixated, and prior work suggests that this occurred because some stage(s) of word recognition was more difficult for neutral than for full tone words. Recordings of northern-dialect speaking readers' ERPs during the recognition of neutral- and full-tone words are consistent with this account, as early benefits for neutral-tone words were followed by subsequent processing costs in Luo et al. (submitted for publication). Approximately 100 ms as well as 250 ms after target word onset, wave peaks were less negative for neutral- than for full-tone targets, in this study, indicating that neutral-tone words were processed more effectively. N400 peaks, by contrast, were more negative for neutral-tone targets, suggesting processing difficulties. Early benefits for neutral-tone targets were attributed to a faster simulated articulation that provided benefits during the convergence of lexical forms. Subsequent costs were attributed to the lower familiarity of neutral tone forms or to potential conflict between lexical(full) and neutral tone articulation features that may occur during late stages of word processing, and this diminished preview benefits.

Larger preview benefits for the full- than the neutral-tone condition in Experiment 2 are also consistent with an alternative account, according to which the duration of a parafoveal preview determines the extent of its processing. In this scheme, longer viewing duration for full-tone targets provided more time for the acquisition of information from parafoveal words, and this yielded a larger benefit. The effect of longer preview duration was twofold, however: on the one hand, it provided more opportunity for obtaining useful information from identical previews of post-target words. On the other hand, it also provided more opportunity for the encoding of diverging information from unrelated previews. Yan et al. (2012) and Kliegl, Hohenstein, Yan and McDonald (2013) proposed that

although unrelated masking previews are normally considered a neutral baseline for the specification of 'parafoveal preview benefits' in the literature, these previews do have deleterious consequences for the processing of target words. That is, preview benefits that are computed from a comparison of an unrelated with an identical preview condition consist, in fact, of a mixture of genuine facilitation from identical word previews and of preview cost, due to interference from unrelated previews. This process-duration account has been used to explain larger preview benefits in oral than in silent reading in Chinese (Pan, Laubrock, & Yan, submitted for publication): fixations are longer in oral reading, providing more opportunity for parafoveal processing.

Similarly, in Experiment 2, the longer viewing duration for full- than for neutral-tone target words provided more opportunity for a deeper processing of identical previews and of unrelated masks. Together, useful information obtained from previewed post-target words and costs incurred from previewed masks yielded larger preview benefits for full- than for neutral-tone targets. However, as can be seen in Table 3B, this increased benefit is largely due to differential effects of post-target masking for in the full- and neutral tone target conditions. When post-target previews were masked, subsequent fixation durations on post-target words were longer in the full- than the neutral-tone target condition, with differences of 11 ms for FFD, 10 ms for SFD, and 15 ms for GD. When post-target previews were visible, although target word condition mattered less, the effect occurred in the expected direction: fixation durations on post-target words were slightly shorter in the full- than for neutral-tone target condition, with differences of 4 ms for FFD, 3 ms for SFD and 4 ms for GD.

According to the process-duration account, longer viewing durations for full-tone targets thus primarily reflected the 'deeper' use of unrelated post-target masks when post-target words were subsequently fixated. This suggests that both process duration and ease of processing could have contributed to larger post-target preview benefits in the full tone target condition. Longer full-tone target viewing durations provided more opportunity for the extraction and subsequent use post-target previews, and this influenced processing primarily when post-target previews were "difficult", i.e., when they consisted of two masking characters rather than the post-target word.

Together, the results of Experiments 1 and 2 provide compelling evidence against the minimality hypothesis. Readers use articulation-specific properties during visual word identification, even for scripts such as Chinese which were not designed for the capture of sound- and articulation-related word properties. The use of articulation-specific features during visual word recognition thus appears to be universal among readers.

## Acknowledgments

This research was funded by Deutsche Forschungsgemeinschaft Grant KL 955/18 and was partially supported by Grant-in-Aid for Scientific Research on Innovative Areas (23118001 & 23118003; Adolescent Mind & Self-Regulation) from the Ministry of Education, Culture, Sports, Science and Technology in Japan. We thank Wei Zhou and Bei Wang for their efforts during data collection and Jane Ashby for her comments on an earlier version of the paper. Data and R scripts used in this paper are available at the Potsdam Mind Research Repository (http://read.psych.uni-potsdam.de/PMR2/).

## 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.

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.

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.

Ashby, J., Treiman, R., Kessler, B., & Rayner, K. (2006). Vowel processing during silent reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 32*, 416–424.

Bates, D. M., & Maechler, M. (2011). *lme4: Linear mixed-effects models using S4 classes* (R package, Version 0.999375-42) [computer software]. <http://cran.r-project.org/web/packages/lme4>.

Beijing Language Institute Publisher (1986). *Modern Chinese Word Frequency Dictionary*. Beijing: Beijing Language Institute Publisher [in Chinese].

Breen, M., & Clifton, C. Jr., (2011). Stress matters: Effects of anticipated lexical stress on silent reading. *Journal of Memory and Language, 64*, 153–170. http://dx.doi.org/10.1016/j.jml.2010.11.001.

Boersma, P., & Weenink, D. (2005). <http://www.fon.hum.uva.nl/praat/>.

Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One, 5*(6), e10729. http://dx.doi.org/10.1371/journal.pone.0010729.

Cao, J. (1986). Putonghua qingsheng yinjie texing fenxi. *Applied Acoustics, 5*(4), 1–6 [in Chinese].

Cao, J. (1992). On neutral-tone syllables in Mandarin Chinese. *Canadian Acoustics, 20*(3), 49–50.

Chaoju, T., & van Heuven, V. J. (2009). Mutual intelligibility of Chinese dialects experimentally tested. *Lingua, 119*, 709–732.

Chappell, H. (2001). Synchrony and diachrony of Sinitic languages: A brief history of Chinese dialects. In H. Chappell (Ed.), *Sinitic grammar: Synchronic and diachronic perspectives* (pp. 3–28). Oxford: Oxford University Press.

Chen, H.-C., & Shu, H. (2001). Lexical activation during the recognition of Chinese characters: Evidence against early phonological activation. *Psychonomic Bulletin & Review, 8*(3), 511–518.

Engbert, R., & Kliegl, R. (2003). Microsaccades uncover the orientation of covert attention. *Vision Research, 43*(9), 1035–1045. http://dx.doi.org/10.1016/S0042-6989(03)00084-1.

Engbert, R., Nuthmann, A., Richter, E., & Kliegl, R. (2005). SWIFT: A dynamical model of saccade generation during reading. *Psychological Review, 112*, 777–813.

Feng, G., Miller, K., Shu, H., & Zhang, H. (2001). Rowed to recovery: The use of phonological and orthographic information in reading Chinese and English. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 27*, 1079–1100.

Filik, R., & Barber, E. (2011). Inner speech during silent reading reflects the reader's regional accent. *PLoS One, 6*(10), e25782. http://dx.doi.org/10.1371/journal.pone.0025782.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123*, 71–99.

Frost, R. (2012). Towards a universal model of reading. *Behavioral and Brain Sciences, 35*, 263–279. http://dx.doi.org/10.1017/S0140525X11001841.

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 417–429.

Huestegge, L. (2010). Effects of vowel length on gaze durations in silent and oral reading. *Journal of Eye Movement Research, 3*(5), 1–18.

Inhoff, A. W. (1984). Two stages of word processing during eye fixations in the reading of prose. *Journal of Verbal Learning and Verbal Behavior, 23*, 612–624.

Kliegl, R., Hohenstein, S., Yan, M., & McDonald, S. A. (2013). How preview space/time translates into preview cost/benefit for fixation durations during reading. *The Quarterly Journal of Experimental Psychology, 66*, 581–600. http://dx.doi.org/10.1080/17470218.2012.658073.

Lee, H. W., Rayner, K., & Pollatsek, A. (1999). The time course of phonological, semantic, and orthographic coding in reading: Evidence from the fast-priming technique. *Psychonomic Bulletin & Review, 6*, 624–634.

Lee, H. W., Rayner, K., & Pollatsek, A. (2001). The relative contribution of consonants and vowels to word identification during reading. *Journal of Memory and Language, 44*, 189–205.

Liberman, A. M. (1996). *Speech: A special code*. Cambridge, MA: The MIT Press.

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

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.

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 & Performance, 30*, 151–162.

Luo, Y., Yan, M., Yan, S., Zhou, X., & Inhoff, A. W. (2014). Syllabic tone articulation influences word identification during Chinese sentence reading: Evidence from ERP and eye movement recordings (submitted for publication).

Luo, Y., Yan, M., & Zhou, X. (2013). Prosodic boundaries delay the processing of upcoming lexical information during silent sentence reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 39*(3), 915–930. http://dx.doi.org/10.1037/a0029182.

Luo, Y., & Zhou, X. (2010). ERP evidence for the online processing of rhythmic pattern during Chinese sentence reading. *NeuroImage, 49*(3), 2836–2850.

Pan, J., Laubrock, J., & Yan, M. (2014). Parafoveal processing in silent and oral Chinese reading (submitted for publication).

Perfetti, C. A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and masked priming. *Journal of Memory and Language, 30*, 473–485.

Perfetti, C. A., Liu, Y., & Tan, L. H. (2005). The lexical constituency model: Some implications of research on Chinese for general theories of reading. *Psychological Review, 112*(1), 43–59.

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.

Pollatsek, A., Lesch, M., Morris, R. K., & Rayner, K. (1992). Phonological codes are used in integrating information across saccades in word identification and reading. *Journal of Experimental Psychology: Human Perception and Performance, 18*(1), 148–162.

Ranbom, L., & Connine, C. M. (2007). Lexical representation of phonological variation. *Journal of Memory and Language, 57*, 273–298.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognition, 53*, 97–145.

Rayner, K. (1975). The perceptual span and peripheral cues during reading. *Cognitive Psychology, 7*, 65–81.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology, 62*, 1457–1506.

R-Core Development Team (2012). *R: A language and environment for statistical computing*. Wien: R Foundation for Statistical Computing. Vienna, Austria. ISBN 3-900051-07-0 <http://www.R-Project.org>.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Schroyens, W., Vitu, F., Brysbaert, M., & d'Ydewalle, G. (1999). Eye movement control during reading: Foveal load and parafoveal processing. *Quarterly Journal of Experimental Psychology, 52A*, 1021–1046.

Tan, L. H., Hoosain, R., & Peng, D. L. (1995). Role of early presemantic phonological code in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 43–54.

Tan, L. H., Hoosain, R., & Siok, W. W. T. (1996). Activation of phonological codes before access to character meaning in written Chinese. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 865–882.

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.

Tan, L. H., & Perfetti, C. A. (1999). Phonological and associative inhibition in the early stages of English word identification: Evidence from backward masking. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 382–393.

Tsai, J. L., Kliegl, R., & Yan, M. (2012). Parafoveal semantic information extraction in Traditional Chinese reading. *Acta Psychologica, 141*, 17–23. http://dx.doi.org/10.1016/j.actpsy.2012.06.004.

Tsai, J. L., Lee, C. Y., Tzeng, O. J. L., Hung, D. L., & Yen, N. S. (2004). Use of phonological codes from Chinese characters: Evidence from processing of parafoveal preview when reading sentences. *Brain and Language, 91*, 235–244.

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*(15), 5229–5233.

Wurm, S. A., Tsou, B., & Bradley, D. (Eds.). (1987). *Language atlas of China*. Hong Kong: Longman.

Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. In *Proceedings of Tools and Resources for the Analysis of Speech Prosody (TRASP 2013)* (pp. 7–10). France: Aix-en-Provence.

Yan, M., Kliegl, R., Shu, H., Pan, J., & Zhou, X. (2010). Parafoveal load of word n + 1 modulates preprocessing of word n + 2. *Journal of Experimental Psychology: Human Perception and Performance, 36*(6), 1669–1676. http://dx.doi.org/10.1037/a0019329.

Yan, M., Richter, E. M., Shu, H., & Kliegl, R. (2009). Chinese readers extract semantic information from parafoveal words during reading. *Psychonomic Bulletin & Review, 16*, 561–566.

Yan, M., Risse, S., Zhou, X., & Kliegl, R. (2012). Preboundary duration modulates semantic preview benefit for word n + 1 and n + 2 in Chinese reading. *Reading and Writing, 25*(5), 1093–1111. http://dx.doi.org/10.1007/s11145-010-9274-7.

Zhou, W., Kliegl, R., & Yan, M. (2013). A validation of parafoveal semantic information extraction in reading Chinese. *Journal of Research in Reading*. http://dx.doi.org/10.1111/jrir.1556.

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, & Cognition, 26*, 1245–1265.

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