Journal of Experimental Psychology: Learning, Memory, and Cognition

Chinese Readers Can Perceive a Word Even When It’s Composed of Noncontiguous Characters

Guojie Ma, Alexander Pollatsek, Yugang Li, and Xingshan Li

CITATION

Ma, G., Pollatsek, A., Li, Y., & Li, X. (2016, August 8). Chinese Readers Can Perceive a Word Even When It’s Composed of Noncontiguous Characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://dx.doi.org/10.1037/xlm0000298

Journal of Experimental Psychology:
Learning, Memory, and Cognition

0278-7393/16/$12.00 http://dx.doi.org/10.1037/xlm0000298

RESEARCH REPORT

Chinese Readers Can Perceive a Word Even When It's Composed of Noncontiguous Characters

Guojie Ma
Shaanxi Normal University and Institute of Psychology,
Chinese Academy of Sciences

Alexander Pollatsek
University of Massachusetts Amherst

Yugang Li and Xingshan Li
Institute of Psychology, Chinese Academy of Sciences

This study explored whether readers could recognize a word composed of noncontiguous characters (a *cross-character word*) in Chinese reading. All 3 experiments employed Chinese 4-character strings ABCD, where both AB and CD were 2-character words. In the cross-character word condition, AC was a word but in the control condition, AC was not a word. A character identification task was employed in Experiment 1 and sentence reading tasks were employed in Experiments 2 and 3. In all 3 experiments, an AC word produced inhibition effects. In Experiment 1, an AC word decreased the accuracy of character B identification, but increased the accuracy of character C identification. In Experiments 2 and 3, an AC word slowed reading on CD, indicating that the cross-character words were activated. These results imply that Chinese character encoding leading to word recognition does not proceed in a strictly serial way from left to right, or is strictly constrained by invisible word boundaries.

Keywords: Chinese reading, word recognition, word segmentation, word competition, character position coding

In most alphabetic languages, interword spaces mark word boundaries that facilitate word perception and recognition (Morris, Rayner, & Pollatsek, 1990; Perea & Acha, 2009; Rayner, Fischer, & Pollatsek, 1998). Readers can use the low-level visual features to segment letter strings into word units. Therefore, in the E-Z reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998), one popular model of reading alphabetic languages, its serial hypothesis on word recognition was also constrained by word boundaries. In contrast, there are no spaces marking word boundaries in Chinese texts. Chinese sentences are composed of contiguous equal-width characters. These characters can constitute words of different lengths. About 6% of Chinese words are one-character words; 72%, 12%, and 10% are two-, three-, and four-character words, respectively (Wei, Li, & Pollatsek, 2013). Despite lacking interword spaces for Chinese readers to segment words in the parafovea (Li, Liu, & Rayner, 2011; Ma, Li, & Pollatsek, 2015), words still have a psychological reality (Bai, Yan, Liversedge, Zang, & Rayner, 2008; Hoosain, 1992; Li, Bicknell, Liu, Wei, & Rayner, 2014; Li, Gu, Liu, & Rayner, 2013). This raises the question of how readers segment the text into words and thus recognize words when they read texts consisting of contiguous Chinese characters.

One possibility is that readers process characters serially according to their order in texts and segment contiguous characters into words in a strictly serial way from left to right. By this strategy, once a character is assigned to a word, it would not be assigned to the following word. However, such a strategy would be deficient in reading texts having overlapping ambiguity where the middle character could constitute a word with the character to its left and another word with the character to its right (Hsu & Huang, 2000; Inhoff & Wu, 2005; Ma, Li, & Rayner, 2014). For instance, there was an overlapping ambiguity "花生长" imbedded in the sentence "花生长在屋后的田里" (Hsu & Huang, 2000). 花 (flower) can be a single-character word and can also constitute a word 花生 (peanut) with the character 生 (birth). Moreover, the middle character 生 can constitute another word 生长 (grow) with the character 长 (growing). If readers always segment words in text in a

---

Guojie Ma, Shaanxi Key Laboratory of Behavior and Cognitive Neuroscience, School of Psychology, Shaanxi Normal University, and Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences; Alexander Pollatsek, Department of Psychology, University of Massachusetts Amherst; Yugang Li and Xingshan Li, Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences.

This research was supported by a grant from the Natural Science Foundation of China (31571125 to Xingshan Li) and a grant from the Fundamental Research Funds for the Central Universities (GK201603121 to Guojie Ma).

Correspondence concerning this article should be addressed to Xingshan Li, Key Laboratory of Behavioral Science, Institute of Psychology, Chinese Academy of Sciences, 16 Lincui Road, Beijing, China 100101. E-mail: lixs@psych.ac.cn

strictly serial way, it would be difficult to resolve all the ambiguities in such kinds of sentences.

Thus, Inhoff and Wu (2005) replaced the serial activation hypothesis with their multiple activation hypothesis. The latter hypothesis assumed that all Chinese characters within the perceptual span, which includes one character to the left of current fixation and 2–3 characters to its right (Chen & Tang, 1998; Inhoff & Liu, 1998), are free to combine into viable word units without directional constraints. Moreover, as the majority of Chinese words are two-character words, there is a useful two-character parsing heuristic: When readers identify four characters ABCD, they would quickly parse all the spatially adjacent sets of two characters (AB, BC, and CD) into potential word units. Inhoff and Wu (2005) provided evidence to support the multiple activation hypothesis during reading Chinese texts with overlapping ambiguity. In their experimental condition, there was a critical region of four Chinese characters ABCD, where AB and CD were two-character words that were consistent with the sentence context and BC was a two-character word that was not consistent with the sentence context. In the two control conditions, the central two characters in the critical four-character regions did not form a word. They found that first-pass reading times (the sum of all fixation durations on a target region before moving to another region) and total reading times (the sum of all fixation durations on a target region including regressions) were longer in the experimental condition than in either of the control conditions. These results indicated that readers did not segment words in a strictly serial way, otherwise the word BC would not be activated in the experimental condition.

In a later study, Li, Rayner, and Cave (2009) proposed an interactive activation model of word segmentation and recognition in Chinese reading which further casts light on the multiple activation hypothesis. It assumed that all characters in the perceptual span were processed in parallel and that words were processed serially. Li et al.'s model can also explain the data provided by Inhoff and Wu (2005). When the ambiguous four-character string ABCD was presented in readers' perceptual span, all of the words it contained (AB, BC, and CD) would be activated (similar to multiple activation). Compared with an unambiguous string, the ambiguous four-character string contains more word candidates that need more time to settle in the model's word competition level. Therefore, Li et al.'s (2009) model can also explain why the reading time increased when the central two characters in a critical four-character region could constitute a word.

In addition, Li et al.'s (2009) model tried to answer how readers parse contiguous characters into words after multiple activation of an ambiguous four-character string ABCD. It included a left-priority hypothesis: Other words could not win the word competition until the left-hand word has been identified. This serial processing hypothesis was proposed to ensure that Chinese readers could read words in the correct order from left to right. In their series of experiments, they asked Chinese readers to report as many characters as possible after they briefly viewed four characters. Participants could usually report all of the four characters when the four characters constituted a four-character word. However, they could usually report the first two characters and the accuracy of character recognition had a large drop from the second character to the third character if the four characters constituted two words with two characters each (i.e., the first two characters were a word and the last two characters were another word). This word boundary effect suggests that the process of character and word recognition runs sequentially from left to right in Chinese reading and only the left-hand word wins in each round of word competition.

Notice that although the left-priority hypothesis gives a good account of how word segmentation can occur after multiple words' activation in an ambiguous four-character string ABCD, their data were only based on reading four-character strings without overlapping ambiguity. Recently, Ma et al. (2014) found that when overlapping ambiguous strings were used, the left-priority hypothesis did not work. Ma et al. embedded each overlapping ambiguous string ABC (where both AB and BC are two-character words) into one of two sentence frames so that it could be either segmented as AB-C or A-BC depending on the sentence context. They found that Chinese readers were more likely to segment the string as A-BC (rather than AB-C) when the word on the right (BC) had a higher frequency than the left-hand one (AB). Based on these findings, Ma et al. (2014) proposed a *(word) competition hypothesis* to explain this phenomenon: All the words in the perceptual span can be activated and all of these words compete for a single winner, and each of these activated words (the left-hand word AB and the right-hand word BC) has a chance to win the word competition if its activation is high enough. When the word on the right is of higher frequency, it has a better chance to win the word competition.

The competition hypothesis helps our understanding of how Chinese readers segment and recognize words when they read contiguous Chinese characters. Nevertheless, there is a critical question that should be further clarified. The competition hypothesis assumed that all the words in the perceptual span can be activated, but it is not clear whether "all the words" should also include words consisting of noncontiguous characters. In Chinese text, some noncontiguous characters can also constitute words, but rarely form meaningful words in the context. We refer these words as cross-character words. Taking a four-character string ABCD (长期征战 endless war) as an example, both AB (长期 endless) and CD (征战 war) are two-character words composed of contiguous characters. However, the noncontiguous characters A and C also constitute a two-character word AC (长征 long march). This phenomenon can be similarly (but not identically) illustrated by the string *overlook landlady*. This string contains two compound words *overlook* and *landlady*, where the noncontiguous morphemes *over* and *land* constitute the word *overland*. In this study, we explored whether such cross-character words can be activated in Chinese reading.

Studying cross-character words has important implications for models of Chinese word segmentation and recognition. Both the multiple activation hypothesis (Inhoff & Wu, 2005) and competition hypothesis (Ma et al., 2014) were proposed based on studying Chinese texts only including words consisting of contiguous characters. Inhoff and Wu (2005) even clarified that when a four-character string ABCD was identified, readers would use a two-character parsing heuristic to combine each two adjacent characters into a word unit (which meant only the subunits AB, BC, and CD could be possible candidates). In addition, the only model up to now on Chinese word segmentation and recognition (Li et al., 2009) was built on an interactive activation model which assumed that character position coding was accurate. As both the interactive activation and multiple activation models basically share this hypothesis, they are unlikely to predict the activation of a cross-character word. Therefore, studying whether a cross-character word can be activated is important to test and thus potentially improve models of Chinese word processing.

Notice that although no study has been conducted to directly address processing of cross-character words, previous studies on letter/character position coding may have some relevance. In Western languages, many researchers found that a transposed-letter nonword (e.g., jugde) can prime a target word (e.g., JUDGE) more than a substituted-letter nonword (e.g., jupte; see Perea & Lupker, 2003, 2004). The transposed letter effect confirmed flexible coding of letter position (Davis, 2010; Gomez, Ratcliff, & Perea, 2008; Norris, Kinoshita, & van Casteren, 2010; Whitney & Berndt, 1999), even for transposed nonadjacent letters as caniso can prime CASINO (Perea & Lupker, 2004). In Eastern languages, researchers also found similar transposition effects in Japanese and Chinese reading (Gu & Li, 2015; Gu, Li, & Liversedge, 2015; Perea, Nakatani, & van Leeuwen, 2011). Using ABCD to represent four Japanese Kana or Chinese characters, researchers found that a string with the middle two characters BC transposed (i.e., ACBD) can prime the four-character target word ABCD (Gu & Li, 2015; Perea et al., 2011). These results suggest flexible character position coding even occurred in unspaced writing systems. In this case, if readers misperceived character order in processing a four character word ABCD, the cross-character word AC would have a chance to be grouped together.

In the present study, we performed three experiments to explore whether and how a cross-character word can be recognized during Chinese reading. In Experiment 1, participants were asked to report as many characters as possible after they briefly viewed four characters (a similar method was used in Li et al., 2009). For these four characters, the first two characters constituted a word, and the last two characters constituted another word. In the cross-character word condition, the first and third characters also constituted a word, but in the control condition, the first and third characters did not constitute a word. If the cross-character word can be recognized in parallel with the left-hand word and compete with the processing of it, readers could perform better on the third character in the cross-character condition than in the control condition.

In Experiment 2, we embedded the two kinds of four-character strings into the same sentence frame to explore whether the words composed of noncontiguous characters are recognized in sentence reading. In natural reading, it is important to process words in correct order from left to right, because readers have the task of correctly understanding the sentences. If these embedded cross-character words in the critical regions are activated, this activation may conflict with the sentence context and compete with the activation of the left-hand word in the critical region, resulting in delayed understanding of the region. Thus, readers should spend more time locally in the cross-character word condition than in the control condition.

In Experiment 3, we further investigated whether the activation of cross-character word is influenced by its word frequency. We will describe the logic of this experiment later.

Experiment 1

Method

Participants. Eighteen native Chinese speakers from universities near the Institute of Psychology, Chinese Academy of Sciences, were paid to participate in Experiment 1.

Materials and apparatus. One hundred pairs of four-character strings (ABCD) were used in Experiment 1. The first two characters of these strings constituted a two-character word, and the last two characters constituted another word (i.e., both AB and CD were two-character words, but ABCD was never a four-character word or a meaningful sequence of words). Each pair of four-character strings shared the same first word AB (see Table 1). In the cross-character word condition, the first and third characters (AC) constituted a two-character word; in the control condition, the first and third characters (AC) did not constitute a word. Character frequency, stroke numbers, and word frequency between the cross-character word condition and the control condition were matched. All the paired comparisons did not show significant differences on the above measures, $ts < 1.38$, $ps > .169$.

The materials were presented on a 21-in. cathode ray tube monitor (Sony G520) with a resolution of 1,024 × 768 pixels and a refresh rate of 100 Hz. The character strings were shown in 24-point font in black (RGB: 0, 0, 0) on a gray background (RGB: 128, 128, 128). Eye movements were monitored by an Eyelink 1000 eye tracking system with a sampling rate of 1,000 Hz.

Procedure. The eye tracker was calibrated at the beginning of the experiment and was calibrated again when needed. A five-point calibration procedure was used. The maximal error of the validation was 0.5° in visual angle. When participants successfully fixated on a cross in the middle of the screen for 300 ms, a four-character string was presented for 80 ms. The first character was always presented at the position occupied by the fixation cross. A mask then appeared until the next trial started. Partici-

Table 1
Properties of Materials Used in Experiment 1

Property	Cross-character word				Control			
	C1	C2	C3	C4	C1	C2	C3	C4
Character frequency	680 (114)	771 (104)	693 (91)	717 (109)	680 (114)	771 (104)	732 (64)	684 (86)
Stroke numbers	8.7 (.13)	8.7 (.11)	8.5 (.22)	8.3 (.25)	8.7 (.13)	8.7 (.11)	8.5 (.28)	8.5 (.25)
Words	素食 (vegetable diet)		质点 (particle)		素食 (vegetable diet)		助教 (assistant)	
Word frequency	2.1 (.06)		2.2 (.12)		2.1 (.05)		2.2 (.32)	

Note. C1–C4 refer to the first through fourth characters. *SE*s are given in parentheses. In the cross-character word condition, the first character 素 can constitute a word 素质 (quality) with the third character 质. The average frequency of cross-character words is 20 occurrences per million ($SE = 2.9$), which is significantly larger than that of the left-hand word, $t = 6.26$, $p < .001$.

pants were encouraged to report as many characters as possible (speed was not stressed so response times were not analyzed). The experimenter then pressed the "enter" key to start the next trial. One hundred experimental trials were presented in a random order following 10 practice trials. Half were in the cross-character word condition and the other half were in the control condition.

Analysis. Trials in which participants did not report any character or reported only the first character were not included in the analysis; thus, 211 out of 1,980 trials (11.7%) were excluded. We analyzed whether participants reported the third character C (and thus the cross-character word AC) more accurately in the cross-character word condition than that in the control condition.

Results and Discussion

The average accuracy of character recognition for A, B, C, and D (see Figure 1) decreased from left to right (94%, 75%, 51%, and 42%, respectively). This indicates that the efficiency of character recognition decreased from the initial fixation point to the right. Because the accuracy at the word boundary positions (characters B and C) is our major interest, we analyzed the accuracy at these positions using generalized linear mixed models (*lrm* function; Baayen, Davidson, & Bates, 2008; Jaeger, 2008) in the R environment (R Core Team, 2016). The *Z* or *t* values greater than 1.96 were considered significant at the 5% level.

On average, the accuracy of character recognition in the cross-character condition ($M = .64$, $SE = .02$) was slightly higher than that in the control condition ($M = .61$, $SE = .02$), $b = -0.236$, $SE = 0.069$, Wald $Z = -3.43$, $p < .001$. Next, on average, the accuracy of character recognition on the second character position ($M = .75$, $SE = .02$) was significantly higher than that on the third character position ($M = .49$, $SE = .03$), $b = -0.863$, $SE = 0.069$, Wald $Z = -12.54$, $p < .001$. However, of greatest interest was the significant interaction between the cross-character word condition and the control position, $b = -1.981$, $SE = 0.142$, Wald $Z = -13.91$, $p < .001$. In addition, a simple effects analysis revealed that a word boundary effect only existed in the control condition. In that condition, the accuracy of character recognition was .52 higher for the second character ($M = .88$, $SE = .02$) than for the third character ($M = .36$, $SE = .04$), $b = -1.894$, $SE = 0.105$, Wald $Z = -17.96$, $p < .001$. In contrast, in the cross-character word condition, the accuracy of character recognition for the second character ($M = .63$, $SE = .03$) did not significantly differ from that for the third character ($M = .65$, $SE = .03$), Wald $Z < 1$.

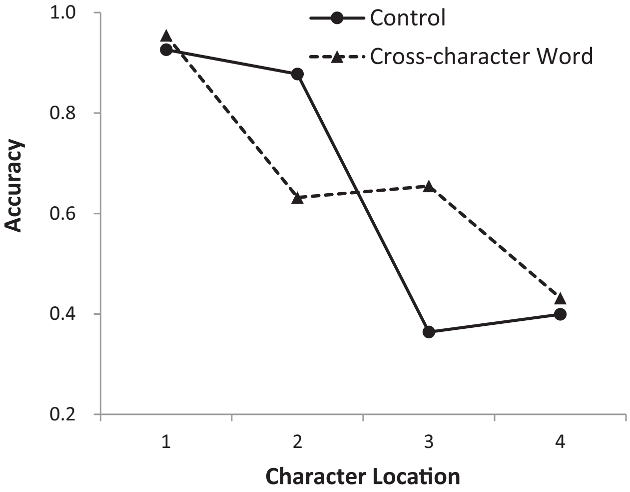


Figure 1. Accuracy of character recognition in Experiment 1.

We found other evidence for the activation of cross-character words. On about 43% of all trials, readers only reported two characters: character A and B or character A and C. We found a significant relation between reporting pattern (reporting character A and B vs. character A and C) and word condition (the cross-character word condition or control condition), $b = -6.914$, $SE = 0.456$, Wald $Z = -15.15$, $p < .001$. Readers reported the combination of characters A and C more often in the cross-character word condition ($M = .10$, $SE = .01$) than that in the control condition ($M < .01$, $SE < .01$), $b = -3.457$, $SE = 0.322$, Wald $Z = -10.71$, $p < .001$. These results clearly showed that the cross-character word AC was activated at least some of the time, even though it broke word boundaries in Chinese reading.

Experiment 2

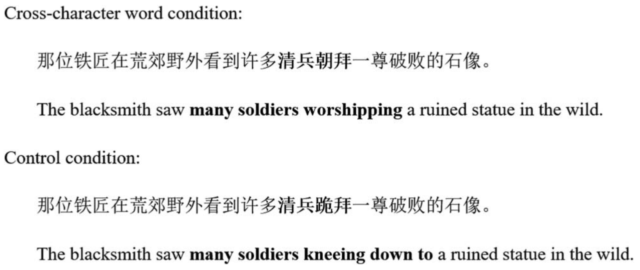
Method

Participants. Twenty-four native Chinese speakers from the same participant pool as that in Experiment 1 were paid to participate in Experiment 2. None of them had participated in Experiment 1. All participants had either normal or corrected-to-normal vision.

Materials and apparatus. Forty-eight pairs of four-character strings ABCD were selected and each pair was embedded into the same sentence frame. The first two characters constituted a word, and the last two characters constituted another word.

Each pair of strings shared the same two-character word AB and the same fourth character D (see Figure 2). In the cross-character word condition (e.g., 清兵朝拜), the first character and the second character constituted a word AB, and the first character and the third character constituted another word AC. However, as in Experiment 1, ABCD was never a word. The frequency of word AC (清朝, Qing Dynasty; $M = 34$, $SE = 1.8$) was higher than that of the word AB (清兵, soldiers in the Qing Dynasty; $M = 2.5$, $SE = 0.29$), $t(47) = 5.24$, $p < .001$. In the control condition (e.g., 清兵跪拜), the characters A and C did not constitute a word. Character frequency, stroke numbers of the third character C and the frequency of word CD were matched between the cross-character word condition (character frequency, $M = 774$, $SE = 55$; stroke numbers, $M = 8.5$, $SE = 0.47$; word frequency, $M = 4.8$, $SE = 0.81$) and the control (nonword) condition (character frequency, $M = 706$, $SE = 96$; stroke numbers, $M = 8.5$, $SE = 0.51$; word frequency, $M = 4.5$, $SE = 0.79$), *t*s < 1.45, *p*s $> .153$.

The naturalness of all sentences was assessed by 12 volunteers on a 7-point scale (1 = *very unnatural*, 7 = *very natural*) to ensure that all sentences were easy to understand. The results showed no significant difference between the cross-character word condition ($M = 6.4$, $SE = 0.2$) and the control condition ($M = 6.3$, $SE = 0.1$), *t*s < 1. The predictability of the word CD in the sentence was close to zero which avoided the influence of predictability on the various eye movement measures. The apparatus was identical to Experiment 1 except that each sentence was displayed in Song



Cross-character word condition:

那位铁匠在荒郊野外看到许多**清兵朝拜**一尊破败的石像。

The blacksmith saw **many soldiers worshipping** a ruined statue in the wild.

Control condition:

那位铁匠在荒郊野外看到许多**清兵跪拜**一尊破败的石像。

The blacksmith saw **many soldiers kneeing down to** a ruined statue in the wild.

Figure 2. Materials used in Experiments 2. The target four-character strings are in bold for the purpose of illustration (the characters were not bold in the experiment).

20-point front. Subjects read the sentences binocularly, but only the right eye was monitored.

Procedure. The eye tracker was calibrated at the beginning of the experiment and was calibrated again when needed. Each participant read six sentences for practice and then 72 sentences for the experiment (including 24 fillers) in a random order. Half of the 48 experimental sentences were in the cross-character word condition and the other half were in the control condition. Participants were asked to read silently and then to answer comprehension questions following one third of the sentences. The questions were used to ensure that participants read the sentences carefully. Each sentence appeared after participants successfully fixated on a character-sized box at the location of the first character of each sentence. After reading a sentence or answering a comprehension question, the participants were asked to press a response button to start the next trial.

Analysis. Accuracy on the comprehension questions was high (94%), indicating that the participants understood the sentences well. Trials were removed when participants blinked more than three times or blinked once while fixating at the target word region, resulting in a loss of 3% of the trials. Fixations with durations longer than 800 ms or shorter than 80 ms (approximately 3% of all fixations) were also excluded from the analysis. We mainly reported five eye movement measures (Rayner, 1998) in the two regions of interest (words AB and CD): (a) first-fixation duration on each of the target regions, irrespective of the number of fixations, (b) gaze duration (the sum of all fixation durations on each of these target regions before moving to another region), (c) total fixation time on each of the target words, including regressions, (d) second-pass reading time (the sum of all fixations on each of the target regions following the initial first-pass time, including zero times), and (e) initial landing position on a two-character region (coded as either 0 or 1 for the first or second character, respectively). The data were analyzed using a linear mixed-effects model by Lme4 package (Version 1.1–12, Bates, Maechler, Bolker, & Walker, 2015) in the R environment (R Core Team, 2016). The t values greater than 1.96 were considered significant at the 5% level.

Results and Discussion

In the word AB region, eye movement measures (see Table 2) on fixation times and landing positions did not show significant differences between the cross-character word and control conditions, $t\text{s} < 1$, except for second-pass reading time, which showed that readers spent more time in reanalyzing the region AB in the cross-character word condition than in the control condition, $b = -36.73$, $SE = 16.55$, $t = -2.22$.

Most importantly, eye movement measures on the later region CD (see also Table 2) indicated a clear difference between the two conditions.[1] Readers spent much more time on CD in the cross-word condition than in the control condition for first-fixation duration, $b = -13.68$, $SE = 6.27$, $t = -2.18$, gaze duration, $b = -29.11$, $SE = 9.29$, $t = -3.13$, and total time, $b = -42.15$, $SE = 12.20$, $t = -3.45$. The only measure in which the difference failed to reach significance was second-pass reading time, $b = -25.92$, $SE = 19.49$, $t = -1.33$. Moreover, the initial landing position data showed a trend that readers initially fixated slightly closer to the beginning of word CD in the cross-character word condition, $b = 0.056$, $SE = 0.031$, $t = 1.82$. Initial landing position on the whole four-character ABCD region did not show significant differences between the cross-character word condition ($M = .65$, $SE = .06$) and the control condition ($M = .59$, $SE = .05$), $b = -0.057$, $SE = 0.042$, $t = -1.33$. Note that most of the eye movement measures on the later region CD were significant, which was in contrast to the pattern in the early region AB. It suggests that the cost of activating the cross-character word on sentence reading is delayed into an integration stage in the critical region.

Experiment 3

Up to now, we demonstrated that the cross-character word can be activated in both a naming task and a sentence reading task, but it is not clear whether the activation only occurs in special circumstances. That is, in Experiment 2, the cross-character word (AC pair) was more frequent on average than the word constituted by the continuous characters (AB pair). Thus, Experiment 2 could not distinguish whether the cross-character word was activated only when the frequency of AC is higher than AB, or that the cross-character word could be activated whenever these two characters constitute a word, regardless of its frequency. In Experiment 3, we manipulated the frequency of the cross-character word to see whether the activation of the cross-character word was modulated by word frequency.

Method

Participants. Twenty-eight native Chinese speakers from the same participant pool as that in Experiment 1 and 2 were paid to participate in Experiment 3. All participants had either normal or corrected-to-normal vision.

Materials and apparatus. Ninety-six pairs of four-character strings ABCD were selected. Each pair of strings shared the same

---

[1] We analyzed reading times on the posttarget two-character region to see whether the cross-character word had a further delayed effect. However, no positive evidence was found in that region. First-fixation duration did not show significant differences between the cross-character word condition ($M = 257$ ms, $SE = 7$ ms) and the control condition ($M = 254$ ms, $SE = 6$ ms), $t < 1$. Total time did not show significant difference in the cross-character word condition ($M = 370$ ms, $SE = 12$ ms) and the control condition ($M = 362$ ms, $SE = 11$ ms) either, $t < 1$. The only exception was gaze duration, which showed a trend that was longer in the cross-character word condition ($M = 298$ ms, $SE = 10$ ms) than the control condition ($M = 281$ ms, $SE = 8$ ms), $b = -16.47$, $SE = 8.43$, $t = -1.95$.

Table 2
Eye Movement Measures in the Word AB and CD Region in Experiment 2

Measure	Region AB: Cross-character word	Region AB: Control	Region CD: Cross-character word	Region CD: Control
First-fixation duration	272 (9)	273 (8)	290 (7)	277 (6)
Gaze duration	316 (12)	323 (14)	355 (12)	328 (8)
Total time	403 (14)	416 (16)	456 (14)	416 (13)
Second-pass reading time	231 (29)	195 (26)	253 (29)	228 (32)
Initial landing position	.44 (.03)	.40 (.03)	.35 (.03)	.41 (.03)

Note. First-fixation duration, gaze duration, total time, and second-pass reading time were measured in milliseconds. *SE*s are given in parentheses.

two-character word AB and the same fourth character D. On half of the trials, characters AC constituted a high-frequency word. In this condition, the frequency of word AC ($M = 94$ occurrences per million, $SE = 17$) was higher than that of the word AB ($M = 30$ occurrences per million, $SE = 8$), $t(47) = 3.45$, $p = .001$. Character frequency, and stroke numbers of the third character C and the frequency of word CD were matched between the high-frequency cross-character word condition (character frequency, $M = 1{,}105$ occurrences per million, $SE = 181$; stroke numbers, $M = 7.7$, $SE = 0.36$; word frequency, $M = 2.9$ occurrences per million, $SE = .0.67$) and the control condition (character frequency, $M = 1{,}066$ occurrences per million, $SE = 25$; stroke numbers, $M = 8.1$, $SE = 0.38$; word frequency, $M = 2.7$ occurrences per million, $SE = 0.56$), $t\text{s} < 1.12$, $p\text{s} > .268$.

The rest of the AC pairs belonged to the low-frequency cross-character word condition where the frequency of word AC ($M = 0.34$ occurrences per million, $SE = 0.03$) was lower than that of the word AB ($M = 35$ occurrences per million, $SE = 9.32$), $t(47) = -3.78$, $p < .001$. Character frequency, and stroke numbers of the third character C and the frequency of the word CD were matched between the low-frequency cross-character word condition (character frequency, $M = 211$ occurrences per million, $SE = 26$; stroke numbers, $M = 8.7$, $SE = 0.37$; word frequency, $M = 1.3$ occurrences per million, $SE = .0.18$) and the control condition (character frequency, $M = 194$ occurrences per million, $SE = 29$; stroke numbers, $M = 8.6$, $SE = 0.31$; word frequency, $M = 1.4$ occurrences per million, $SE = 0.23$), $t\text{s} < 1$. It is important to note that the mean character frequency of character C was equated in both the high- and low-frequency cross character conditions with the character frequency of character C in the control condition.

Note that we did not manipulate word (cross-character word or nonword) and word frequency (high or low frequency) conditions in the same sentence frame, because it was impossible to get enough materials. Thus, it was a within-subjects but between-items design for the frequency of word AC. As with Experiment 2, the naturalness of all the sentences was rated by 12 undergraduate students who did not participate in the formal study. The data showed no significant difference between the high-frequency cross-character word condition ($M = 6.6$, $SE = 0.1$) and its control condition ($M = 6.4$, $SE = 0.1$), $t\text{s} < 1.42$, as well as between the low-frequency cross-character word condition ($M = 6.4$, $SE = 0.2$) and its control condition, ($M = 6.3$, $SE = 0.1$), $t\text{s} < 1$. The predictability of word CD in the sentence was close to zero which avoided the influence of predictability on eye movement measures. The apparatus was identical to Experiment 2.

Procedure. The same procedure was used as in Experiment 2.

Analysis. The same analysis methods were used as in Experiment 2. The average comprehension accuracy was 96%, which suggested that the readers comprehended all of the sentences very well. Approximately 5% of the trials were excluded using the same selection criterion as in Experiment 2.

Results and Discussion

In Experiment 3, we focused on whether the cross-character word could be activated as in Experiment 2 and whether word frequency would modulate its activation (see Table 3). Thus, we first analyzed the main effect of the cross-character word (vs. the control word) and the interaction between the cross-character word and the frequency of the cross-character word on the critical region CD. We found that the first-fixation duration on this region was significantly longer in the cross-character word condition ($M = 289$ ms, $SE = 7.1$ ms) than that in the control condition ($M = 280$ ms, $SE = 7.6$ ms), $b = -14.78$, $SE = 5.88$, $t = -2.51$, but the interaction was not significant, $b = 12.54$, $SE = 8.31$, $t = 1.51$. Gaze durations in the cross-character word condition were also significantly longer ($M = 336$ ms, $SE = 8.9$ ms) than in the control condition ($M = 322$ ms, $SE = 7.7$ ms), $b = -18.31$, $SE = 8.13$, $t = -2.25$, but there was no significant interaction, $t\text{s} < 1$. Total time was also significantly longer in the cross-character word condition ($M = 493$ ms, $SE = 15.3$ ms) than in the control condition ($M = 478$ ms, $SE = 19.6$ ms), $b = -38.15$, $SE = 13.66$, $t = -2.79$, but the interaction was also significant, $b = -42.15$, $SE = 19.12$, $t = 2.21$.

Further analyses, in which we split the data into the high- and low-frequency word conditions, provided more valuable information. In the high-frequency cross-character word condition, we replicated the main findings in Experiment 2. In the word AB region, similar to Experiment 2, second-pass reading time in the cross-character word condition was significantly longer than that in the control condition, $b = -31.12$, $SE = 10.70$, $t = -2.91$. In addition, readers initially fixated further right to the target region in the cross-character word condition than that in the control condition, $b = -0.061$, $SE = 0.028$, $t = -2.13$. Of greater interest were the data in the word CD region. Readers spent more time in the cross-word condition than in the control condition on first-fixation duration, $b = -14.32$, $SE = 6.18$, $t = -2.32$, gaze duration, $b = -17.78$, $SE = 8.33$, $t = -2.13$, and total time, $b = -36.96$, $SE = 13.83$, $t = -2.67$. Second-pass reading time

Table 3
Eye Movement Measures in the Word AB and CD Region in Experiment 3

	High-frequency word AC				Low-frequency word AC			
	Region AB		Region CD		Region AB		Region CD	
Measure	Cross-character word	Control	Cross-character word	Control	Cross-character word	Control	Cross-character word	Control
First-fixation duration	284 (11)	284 (9)	298 (7)	283 (8)	268 (7)	274 (8)	280 (8)	278 (8)
Gaze duration	327 (15)	336 (11)	345 (10)	325 (9)	310 (9)	324 (10)	329 (10)	318 (8)
Total time	499 (27)	483 (17)	522 (17)	488 (20)	428 (17)	447 (19)	463 (19)	468 (19)
Second-pass reading time	147 (13)	118 (13)	142 (13)	140 (16)	94 (11)	107 (14)	129 (15)	121 (15)
Initial landing position	.49 (.02)	.42 (.03)	.45 (.03)	.45 (.02)	.44 (.02)	.47 (.03)	.41 (.02)	.42 (.02)

Note. First-fixation duration, gaze duration, total time, and second-pass reading time were measured in milliseconds. *SE*s are given in parentheses.

and initial landing position did not show significant differences between the two conditions, ts < 1.

In the low-frequency cross-character word condition, there was no evidence indicating the activation of the cross-character word AC. In the AB region, there were no significant differences between the cross-word and control conditions, ts < 1, except that gaze duration in the cross-character word condition was slightly shorter than in the control condition, $b = 13.62$, $SE = 8.07$, $t = 1.68$. In the CD region, the results were similar. There were no significant differences between the cross-word and control conditions, ts < 1.3. Therefore, we can infer that word frequency of the cross-character word AC modulates the process of word competition and the cross-character words with high frequency are more likely to be activated. Again, these effects were not due to the differences of frequency between the C characters in the cross-character word condition and the control condition, because these were equated.

General Discussion

In the current study, we conducted three experiments to investigate whether a Chinese word constituted by noncontinuous characters could be activated during reading. In Experiment 1, participants attempted to name the characters after they briefly viewed four Chinese characters ABCD. Readers were more likely to report the characters A and C when AC constituted a word (in the cross-character word condition) than when they did not (control condition). In addition, the accuracy of character recognition of the third character C did not differ from that of the second character B and readers frequently reported only the character combination of AC in the cross-character word condition. These data are one piece of evidence for the activation of cross-character words.

In Experiment 2, a sentence reading task also provided evidence for the activation of cross-character words, as the fixation durations were longer in the CD region in the cross-character word condition than in the control condition. It appeared to be an inhibition effect in the cross-character word condition as more time was needed in processing a critical region. These results were not difficult to understand because natural sentence reading requires readers to try to process words accurately in their correct order. In the previous study by Inhoff and Wu (2005), the activation of overlapping words BC in ABCD also lengthened reading times on the critical region ABCD, where AB and CD were two-character words consistent with sentence context but the characters B and C also constituted a word inconsistent with sentence context. Therefore, the activation of a cross-character word in natural reading that competes with the words that are consistent with the sentence context results in longer reading time than in the control condition.

In Experiment 3, we replicated the main finding of Experiment 2 and further found that the frequency of word AC modulated the activation of the cross-character word. Moreover, we demonstrated that the inhibition effects observed in Experiment 3 (e.g., the greater fixation times in the CD region in the cross-character condition than in the control condition) were modulated by word frequency and not the frequency of the C character. That is, when the frequency of the cross-character word AC was high, it was more likely to compete with the left-hand word AB. In sum, these three experiments jointly revealed that readers could recognize a word consisting of noncontiguous characters in Chinese reading, while word frequency played a critical role in this process. These data indicate that word recognition is not solely dependent on the contiguity of component characters which has to be taken into account by any model of word segmentation and recognition in Chinese reading.

Up to now, one formal model of Chinese word segmentation proposed by Li et al. (2009) would have difficulty predicting our data. Li et al.'s model was built on the interactive activation theory of McClelland and Rumelhart (1981) which assumed accurate character position coding. This meant that word recognition was dependent on the contiguity of component characters. In contrast, the present study showed that word recognition is not always constrained by physical word boundaries. Although no previous study has investigated the activation of a cross-character word, its activation was not difficult to understand in light of the competition hypothesis (Ma et al., 2014), which assumes that all the possible words in the perceptual span can be activated and thus enter into word competition, so that the second character of the cross-character word could be reported with a higher frequency than the second character of the "normal" word in Experiment 1 and reading times in the critical region could be lengthened by a cross-character word in both Experiments 2 and 3.

In their study, Inhoff and Wu (2005) said that "characters that form common two-character words could stand out because they typically occur together," and thus the central two-character high-frequency word BC in ABCD should be more likely to be activated than its low-frequency counterpart. However, the relatively low power of their post hoc comparison did not reveal the role of word frequency. In a subsequent study by Ma et al. (2014), when the word frequencies of both AB and BC in an overlapping ambiguous string ABC were manipulated, they found the higher frequency word BC won the competition over the lower frequency word AB a significant number of times. Because higher frequency words are more available in the mental lexicon (Forster & Chambers, 1973), it is reasonable to posit that the cross-character word AC breaks through the constraints of visual acuity and word boundaries to win the word competition a reasonable number of times.

In addition, this study has some implications for theories of character position coding in Chinese reading. First, as we illustrated in the introduction, a transposed character effect suggests flexible character position coding (Gu & Li, 2015; Gu et al., 2015). The present study provided additional evidence for uncertainty in character position coding, in light of the activation of the cross-character word. Notice that we did not transpose any characters in this study; thus, it potentially provided a new way to study character position coding in Chinese reading. Second, there are no spaces between Chinese words and the narrow space between contiguous characters does not provide any clue for word boundaries. Thus, the constraints of word boundaries as that in models of letter position coding in spaced writing systems need to be reconsidered when modeling character position coding in Chinese reading. Third, this study revealed that the activation of cross-character words is modulated by word frequency and thus cannot be explained solely by uncertain character position coding.

In sum, the activation of cross-character words revealed by this study has the following theoretical implications. First, it indicates that Chinese character encoding leading to word recognition does not proceed in a strictly serial way from the left to the right, or strictly constrained by invisible word boundaries, because the local word context and other top-down information can modulate character recognition. Second, the character-to-word grouping process in Chinese reading is not only constrained by character contiguity, as a cross-character word which crosses word boundaries can be activated and compete with a word composed of noncontiguous characters. These mechanisms should be taken into account by any model of word segmentation and recognition in Chinese reading.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. http://dx.doi.org/10.1016/j.jml.2007.12.005

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). Reading spaced and unspaced Chinese text: Evidence from eye movements. *Journal of Experimental Psychology: Human Perception and Performance, 34,* 1277–1287. http://dx.doi.org/10.1037/0096-1523.34.5.1277

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67,* 1–48. http://dx.doi.org/10.18637/jss.v067.i01

Chen, H. C., & Tang, C. K. (1998). The effective visual field in reading Chinese. *Reading and Writing: An Interdisciplinary Journal, 10,* 245–254. http://dx.doi.org/10.1023/A:1008043900392

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review, 117,* 713–758. http://dx.doi.org/10.1037/a0019738

Forster, K. I., & Chambers, S. M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627–635. http://dx.doi.org/10.1016/S0022-5371(73)80042-8

Gomez, P., Ratcliff, R., & Perea, M. (2008). The overlap model: A model of letter position coding. *Psychological Review, 115,* 577–600. http://dx.doi.org/10.1037/a0012667

Gu, J., & Li, X. (2015). The effects of character transposition within and across words in Chinese reading. *Attention, Perception & Psychophysics, 77,* 272–281. http://dx.doi.org/10.3758/s13414-014-0749-5

Gu, J., Li, X., & Liversedge, S. P. (2015). Character order processing in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 41,* 127–137. http://dx.doi.org/10.1037/a0038639

Hoosain, R. (1992). Psychological reality of the word in Chinese. In H. C. Chen & O. J. L. Tzeng (Eds.), *Language processing in Chinese* (pp. 111–130). Amsterdam, the Netherlands: North-Holland. http://dx.doi.org/10.1016/S0166-4115(08)61889-0

Hsu, S. H., & Huang, K. C. (2000). Interword spacing in Chinese text layout. *Perceptual and Motor Skills, 91,* 355–365. http://dx.doi.org/10.2466/pms.2000.91.2.355

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 20–34. http://dx.doi.org/10.1037/0096-1523.24.1.20

Inhoff, A. W., & Wu, C. (2005). Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition, 33,* 1345–1356. http://dx.doi.org/10.3758/BF03193367

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. http://dx.doi.org/10.1016/j.jml.2007.11.007

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: A systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General, 143,* 895–913. http://dx.doi.org/10.1037/a0033580

Li, X., Gu, J., Liu, P., & Rayner, K. (2013). The advantage of word-based processing in Chinese reading: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39,* 879–889. http://dx.doi.org/10.1037/a0030337

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research, 51,* 1146–1156. http://dx.doi.org/10.1016/j.visres.2011.03.004

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58,* 525–552. http://dx.doi.org/10.1016/j.cogpsych.2009.02.003

Ma, G., Li, X., & Pollatsek, A. (2015). There is no relationship between the preferred viewing location and word segmentation in Chinese reading. *Visual Cognition, 23,* 399–414. http://dx.doi.org/10.1080/13506285.2014.1002554

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 1046–1059. http://dx.doi.org/10.1037/a0035389

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375–407. http://dx.doi.org/10.1037/0033-295X.88.5.375

Morris, R. K., Rayner, K., & Pollatsek, A. (1990). Eye movement guidance in reading: The role of parafoveal letter and space information. *Journal of Experimental Psychology: Human Perception and Performance, 16,* 268–281. http://dx.doi.org/10.1037/0096-1523.16.2.268

Norris, D., Kinoshita, S., & van Casteren, M. (2010). A stimulus sampling theory of letter identity and order. *Journal of Memory and Language, 62,* 254–271. http://dx.doi.org/10.1016/j.jml.2009.11.002

Perea, M., & Acha, J. (2009). Space information is important for reading. *Vision Research, 49,* 1994–2000. http://dx.doi.org/10.1016/j.visres.2009.05.009

Perea, M., & Lupker, S. J. (2003). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition, 31,* 829–841. http://dx.doi.org/10.3758/BF03196438

Perea, M., & Lupker, S. J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51,* 231–246. http://dx.doi.org/10.1016/j.jml.2004.05.005

Perea, M., Nakatani, C., & van Leeuwen, C. (2011). Transposition effects in reading Japanese Kana: Are they orthographic in nature? *Memory & Cognition, 39,* 700–707. http://dx.doi.org/10.3758/s13421-010-0052-1

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124,* 372–422. http://dx.doi.org/10.1037/0033-2909.124.3.372

Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research, 38,* 1129–1144. http://dx.doi.org/10.1016/S0042-6989(97)00274-5

R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105,* 125–157. http://dx.doi.org/10.1037/0033-295X.105.1.125

Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect outgoing saccade length in Chinese reading. *Vision Research, 80,* 1–6. http://dx.doi.org/10.1016/j.visres.2012.11.015

Whitney, C., & Berndt, R. S. (1999). A new model of letter string encoding: Simulating right neglect dyslexia. *Progress in Brain Research, 121,* 143–163. http://dx.doi.org/10.1016/S0079-6123(08)63072-1

Received February 26, 2015
Revision received May 2, 2016
Accepted May 3, 2016 ■