Cognitive Psychology 58 (2009) 525–552

Contents lists available at ScienceDirect

Cognitive Psychology

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

On the segmentation of Chinese words during reading

Xingshan Li [a,*], Keith Rayner [b], Kyle R. Cave [c]

[a] Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA
[b] Department of Psychology, University of California, San Diego, USA
[c] Department of Psychology, University of Massachusetts, Amherst, USA

ARTICLE INFO

Article history:
Accepted 26 February 2009
Available online 5 April 2009

Keywords:
Word segmentation
Word recognition
Reading
Chinese

ABSTRACT

Given that there are no spaces between words in Chinese, how words are segmented when reading is something of a mystery. Four Chinese characters, which either constituted one 4-character word or two 2-character words, were shown briefly to subjects. Subjects were quite accurate in reporting the 4-character word, but could usually only report the first 2-character word, demonstrating that word segmentation influences character recognition. The results suggest that even with these simple 4-character strings, there is an element of seriality in reading Chinese words: processing is initially focused at least to some extent on the first word. We also found that the processing of characters that are not consistent with the context is inhibited, suggesting inhibition from word representations to character representations. A simple model of Chinese word segmentation and word recognition is presented to account for the data.

1. Introduction

Words are generally regarded as the basic meaningful unit of language. In English, bottom-up information provided by the spaces between words can be used by readers to segment words. However, in Chinese, there are no spaces between the words. In fact, Chinese readers don't always agree on where the word boundaries are. Yet, it is clear that words are important in reading Chinese, because word frequency and word predictability effects for Chinese readers are comparable to those of readers of English (Rayner, Li, Juhasz, & Yan, 2005; Yan, Tian, Bai, & Rayner, 2006). Chinese sentences consist of characters that vary in complexity, but each character fits within the same sized square region;

* Corresponding author. Fax: +1 615 343 8449.
E-mail address: xingshan@gmail.com (X.S. Li).

0010-0285/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.cogpsych.2009.02.003

these characters are only separated by punctuation marks. How words are segmented in Chinese is an important question that we address in the present article.

In the last decade, a number of studies have examined the issue of Chinese word segmentation. Studies in which spaces were added between Chinese words have generally found no benefit for Chinese reading (Hsu & Huang, 2000a, 2000b; Inhoff, Liu, Wang, & Fu, 1997). However, Bai, Yan, Liversedge, Zang, and Rayner (2008) recently demonstrated that adding spaces between characters interfered with reading, while adding spaces between words yielded reading performance that did not differ from reading with normal non-spaced Chinese text. In many ways, it is remarkable that inserting spaces between words did not hinder reading (given that subjects in the study had a lifetime of experience reading without spaces). Bai et al. also found that inserting spaces between pseudo-randomly chosen character pairs interfered with reading to the same extent as inserting spaces between every pair of characters.

Chen (1999) argued that Chinese word segmentation is automatic and efficient. He asked subjects to search for a Chinese character among distractors, and found that search was faster when the character was embedded in a string of asterisks or a 2-character word than in a 2-character non-word or a string of scrambled characters. Inhoff and Wu (2005) embedded four characters constituting two 2-character words in sentences. In the ambiguous condition, the central two characters also constituted a 2-character word, while in the control condition the central two characters did not constitute a word. They found that gaze duration (the sum of all eye fixations on a word prior to moving to another word) and total viewing time (the sum of all fixations on a word, including regressions) were longer in the ambiguous condition than in the control condition. They argued that this result was inconsistent with what they called the *unidirectional parsing* hypothesis, which assumes that characters are assigned to words in a strictly serial left-to-right process. Rather, they argued that their results were consistent with what they called the *multiple activation* hypothesis, which assumes that all of the possible words that can be combined by the characters falling into the perceptual span are activated.

For Chinese reading, given that there is no bottom-up spacing information to aid in word segmentation, top-down information is likely to be a key factor in segmenting Chinese words. There are two possible hypotheses concerning the use of top-down information by Chinese readers to segment words. We will refer to these as the *feed-forward* hypothesis and the *holistic* hypothesis. The feed-forward hypothesis assumes that the visual information obtained from Chinese characters is initially fed into a character recognition system, and word segmentation follows after character recognition, leading to word recognition. Under this hypothesis, word segmentation does not feed back to the character recognition system. In other words, top-down information does not influence the process of character recognition. On the other hand, according to the holistic hypothesis, word segmentation influences character recognition through feedback. Thus, the various subsystems cooperate to influence word segmentation and word recognition.

In the present study, we further examined word segmentation in Chinese reading. In the crucial conditions of the five experiments, four Chinese characters, which either constituted a 4-character word or two 2-character words, were briefly presented to subjects. Three techniques were used to examine how the visual perception process and character recognition process differed in these two conditions. First, in Experiments 1, 2, 3, and 5, we used a naming paradigm in which subjects had to name the characters after a brief exposure of the characters. If the feed-forward hypothesis is correct, the number of characters that are recognized should not differ between the two conditions. Secondly, in Experiments 1 and 2, we presented a probe at one of the character locations. The response time to this probe should reflect attentional deployment (Cepeda, Cave, Bichot, & Kim, 1998; Hoffman, Nelson, & Houck, 1983; Kim & Cave, 1995; Kramer, Weber, & Watson, 1997; Laberge, 1983; Logan, 1994; Posner, 1980; Tsal & Lavie, 1988). This paradigm was mainly designed to examine whether word segmentation could influence attentional deployment. Object-based attention experiments have demonstrated that reaction times are faster when a probe is presented within the same object as the cue (Egly, Driver, & Rafal, 1994). Previous research has shown that words can be perceived as objects (Prinzmetal, Hoffman, & Vest, 1991; Robertson & Treisman, 2006). For example, Robertson and Treisman (2006) found that a patient with Balint's syndrome, who could only perceive single objects, could identify familiar words (ON and NO) but not the relative location of the two letters (O and N) in the display. If words are considered as objects, then in our experiments there were two objects in the two-word condition, but only one object in the one-word condition. If word segmentation can influence attentional deployment, then theories of object-based attention might predict a large increase in reaction time between the second and third characters (the word boundary) when there were two words. In contrast, such a difference would not be observed if the four characters together constitute a single word. Thirdly, in Experiment 4, a character detection task was used. Subjects were asked to detect a predefined target character among a set of four characters, which could constitute one word or two words. With this method, subjects did not need to remember or report the characters they identified, so the memory load was similar between the one-word condition and the two-word condition. This experiment eliminated the possibility that the results of Experiment 1 and 2 were due to short-term-memory limitations.

Previous studies have indicated that English word processing is holistic (Cattell, 1886; Healy, 1976, 1994; Johnston, 1978; Tao, Healy, & Bourne, 1997). Cattell (1886) showed that when viewing a display for a short amount of time, people could recognize two words if the stimuli were unrelated words, but could only recognize three or four letters if the stimuli were unrelated letters. These results suggest that words were not processed letter by letter, but processed as word units holistically. Other studies provided additional evidence that words are the unit of reading instead of letters (Healy, 1976, 1994; Johnston, 1978; Tao et al., 1997). For example, Healy (1976, 1994) asked subjects to search for a letter in a paragraph of text, and found that searching for a letter is usually difficult if it is in a frequently used word such as "of" or "the". She argued that frequently used words are processed at levels higher than single letters. LaBerge and Samuels (1974) argued that word recognition was an automatic process. These studies all suggest that words are processed in a holistic way, but most of them are based on English words, with non-words as a control. The words were easily parsed in the stimuli because of the spaces between them. There are no such spaces between words in Chinese reading. Chinese readers have to segment the characters into words based on their word knowledge. When word segmentation requires effort and is not explicitly represented in the written text, is word processing still holistic? The current study examines this question.

Chinese differs in another way from English that makes it easier to test the effects of words on character recognition. In Chinese, each character represents one syllable, and reading the word aloud is no different than reading the name of each character aloud. Thus, in Chinese word recognition tasks, subjects can report the entire word when possible, and can report individual characters when they do not perceive the entire word. Unlike English, there are no complications that come from having to choose whether to speak the name of the word or spell it out letter by letter.

In the present study, we used strings of four Chinese characters in all of the experiments. In some conditions, as noted above, the four characters together constituted a single word; in some other conditions, the first two characters constituted a word, while the other characters constituted another word, or did not constitute a word at all. In another condition, none of the characters constituted a word. In these displays, there was no space between the Chinese words, just as there are no spaces between words in regular reading of Chinese. Thus, the words were not easily segmented in the two-word condition. The comparisons between the one-word condition and the two-word condition will show how the holistic nature of word recognition changes when word segmentation is necessary and when it is not.

2. Experiment 1

In Experiment 1, four Chinese characters were presented and the subjects had two tasks. First, they had to indicate if a probe was present. Second, they had to verbally report the characters that were present. In the one-word condition, the four characters constituted a word; in the two-word condition, the first two characters constituted a word, and the last two characters constituted another. According to the feed-forward hypothesis, character recognition accuracy should not differ for these two conditions. According to the holistic hypothesis, the characters belonging to a word should be recognized almost equally accurately, while characters belonging to different words might be recognized differently if one word is more thoroughly processed than the other.

2.1. Method

2.1.1. Subjects

Twenty native Chinese speakers who were graduate students or spouses of students at the University of Massachusetts, Amherst were paid to participate in the experiment. In this and all subsequent experiments, they all had either normal uncorrected vision or their vision was corrected via contact lenses or glasses.

2.1.2. Apparatus

Stimuli were presented on a 19 in. LCD monitor controlled by a Dell PC. Subjects responded by pressing a button on a button box. A chin-rest located 60 cm away from the monitor was used to minimize head movements. In this experiment and subsequent experiments, eye movements were monitored via an EyeLink 2 tracker with eye position sampled at 250 Hz. The primary goal of tracking the eyes was to make certain that subjects were looking where they were supposed to be looking during a given trial. Because subjects did indeed consistently fixate where they were instructed to fixate, we will not discuss eye movements further.

2.1.3. Materials

Four Chinese characters were shown in each of 96 trials.[1] Half of the trials comprised the one-word condition (in which the four characters were a 4-character word, as shown in Fig. 1), the other half comprised the two-word condition (in which the first two characters constituted a word and the last two constituted another word). The average character complexity, measured by strokes per character, did not differ across the two conditions (with 7.4 strokes per character for each condition). The word frequencies for all of the stimulus words were between 5 and 12 per 1,116,417 words,[2] and did not differ significantly between the two conditions ($p > .1$). In the two-word condition, the combination of the two words did not form a meaningful string. By doing this, we tried to exclude higher level contextual effects in this study. Note that in Chinese, the first two characters in a 4-character word usually do not make sense by themselves without being combined with the last two characters. Hence, semantic constraints were not different between the one-word condition and the two-word condition. The four characters were shown at the center of the screen on a single line. Each character fit within a 1° by 1° square. The order of the experimental trials was randomized for each subject. There were 10 practice trials before the formal experimental trials started. In this experiment and all of the following experiments, the materials in the practice trials were different from those used in experimental trials.

2.1.4. Procedure

Characters were presented 300 ms after subjects fixated on a black dot, which occupied the location that would be the center of the first character once it appeared.[3] The characters were presented for 80 ms before they disappeared. Then in 36 of the 48 trials in each condition, a small red square (the attention probe) was presented for 40 ms equally often at one of the four character locations. Subjects had two tasks. The first was to press one of two buttons to indicate whether the red square was present or not. An auditory error signal occurred when subjects pressed the wrong button. The second task was to verbally report the characters that were presented. Subjects were encouraged to report as many characters as possible in this and all subsequent experiments. In Chinese, each character corresponds to a single syllable, and those syllables together make up the word represented by the characters. Naming a word is the same as naming its characters, and characters are pronounced similarly when they stand alone and when they are part of a word. Thus, even when subjects reported words, they had to report each character. Subjects were told that the first task (probe detection) was more

[1] Material used in the experiments are available at http://www.xingshanli.com/Documents/wordSeg_material.doc.

[2] The source for word frequency was the Chinese Dictionary (National Language Committee, 1997) in which the word frequency count is based on a corpus of 1,116,417 words.

[3] The first character location was specifically chosen to mimic the initial fixation during reading for both the one-word and two-word conditions.

One-word condition: 不知所措

Two-word condition: 急速 切实

Half-word condition: 无所 坏功

Non-word condition: 艾抵积促

Related-word condition: 冷静 分析

Fig. 1. Example stimuli. One-word condition and two-word condition were used in all of the experiments; half-word condition was used in Experiment 3; non-word condition was used in Experiment 3 and 5. Related word condition was used in Experiment 5. The words were underlined.

important than the second one (naming). Only correctly reported words were recorded by the experimenter in this experiment.

2.2. Results

Only trials in which subjects pressed the correct button were included in the analysis; 98 trials out of 1920 trials (5.1%) were thus excluded. Separate Analyses of Variance (ANOVAs) were carried out on word recognition accuracy and the probe reaction time, using subjects (*F1*) as random effects, with condition and probe location as within-subject factors; and with items (*F2*) as random effects, with condition and probe location as between-subject factors.

2.2.1. Word report accuracy

Subjects correctly reported all four of the characters in more trials in the one-word condition (96%; see Table 1) than in the two-word condition (51%), $F1(1,19) = 42.98$, $p < .01$, MSE = .20; $F2(1,64) = 393,92$, $p < .001$, MSE = .01. There was a main effect of probe location only in the subjects analysis, $F1(3,57) = 4.03$, $p < .05$ MSE = .01; $F2(3,64) = 1.63$, $p > .1$, MSE = .01 and an interaction between condition and probe location, $F1(3,57) = 7.02$, $p < .001$, MSE = .01; $F2(3,64) = 2.26$, $p < .1$, MSE = .01. In the one-word condition, accuracy was lowest when the probe was at the fourth character location; in the two-word condition, accuracy of recognizing both words was lower when the probe was at the second or third character location than when it was at the other locations (see Table 1). Note that the effect of probe location was much smaller compared to the effect of condition. The purpose of the probe was to detect attentional deployment, but it had complex effects on character report accuracy. On the one hand, it could attract attention to the probe location; on the other hand, it could mask the character. These effects become more complex for word recognition since they might have

Table 1
Accuracy rate for word report in Experiment 1 in all of the conditions and locations. L1, L2, L3, L4 represents different locations of probe. L*n* represents the trials when there were no probes.

	Accuracy rate					SD error				
	L1	L2	L3	L4	L*n*	L1	L2	L3	L4	L*n*
Experiment 1										
One-word condition	.96	.96	.97	.94	.96	.07	.07	.07	.09	.08
Two-word condition only first word correct	.34	.50	.48	.41	.35	.06	.08	.07	.07	.07
Two-word condition both words correct	.57	.41	.47	.53	.58	.08	.08	.08	.08	.08
Experiment 2										
One-word condition	.94	.96	.93	.87	.74	.02	.01	.02	.03	.06
Two-word condition only first word correct	.52	.61	.68	.56	.52	.06	.05	.04	.05	.05
Two-word condition both words correct	.20	.10	.12	.20	.13	.06	.03	.03	.04	.03

different effects when they are at different locations in a word. Hence, we will not offer an explanation for the surprising probe location effect on report accuracy, and will remove the probe in Experiments 3, 4 and 5 and focus on character report accuracy.

In the two-word condition, when subjects could not report both words correctly, they reported the first word correctly on most trials (86%). It was very rare that only the second word was correctly reported (2%).

2.2.2. Reaction time to the probe

Median reaction time was calculated for each subject at each probe location for each condition. The means across subjects are shown in Fig. 2. There was a main effect of condition [$F1(1,19) = 9.80$, $p < .01$, MSE = 21606.00; $F2(1,64) = 15.37$, $p < .001$, MSE = 2101.41] and a main effect of location [$F1(3,57) = 13.20$, $p < .001$, MSE = 6016.10; $F2(3,64) = 14.77$, $p < .001$, MSE = 2101.41]. Reaction time was longer in the two-word condition (661 ms, s.e. = 39 ms) than the one-word condition (588 ms, s.e. = 24 ms). A separate analysis was done with only locations 2 and 3 because the word boundary in the two-word condition was between them. There was again a main effect of condition [$F1(1,19) = 7.85$, $p < .05$, MSE = 18018.96; $F2(1,32) = 6.69$, $p < 0.05$, MSE = 2523.68], as reaction time was longer in the two-word condition (661 ms, s.e. = 43 ms) than the one-word condition (577 ms, s.e. = 23 ms) and there was a hint of an interaction between condition and location [$F1(1,19) = 3.39$, $p < .1$, MSE = 2290.23; $F2(1,32) < 1$]. Reaction time was longer when the probe was at the third character location (667 ms, s.e. = 44 ms) than at the second character location (654 ms, s.e. = 44 ms) in the two-word condition, while reaction time was shorter when the probe was at the third character location (563 ms, s.e. = 22 ms) than at the second character location (590 ms, s.e. = 26 ms) in the one-word condition.

2.3. Discussion

In this experiment, subjects were shown four Chinese characters in two conditions. In the one-word condition, subjects reported the 4-character word quite well, while in the two-word condition they often could only report the first two-character word. Furthermore, responses to attention probes were faster in the one-word condition than in the two-word condition.

The differences in accuracy demonstrate that the four characters in these two conditions were not processed in the same way. The results are consistent with the predictions of the holistic hypothesis, which claims that word recognition is not a simple feed-forward process, and word segmentation and word recognition processes influence the recognition of the characters.

However, an alternative account of these results could argue that subjects recognized the first two characters in both of the conditions, but guessed the last two characters in the one-word condition based on lexical constraints. Experiment 3 tested this possibility. It should also be noted that Experiment 1 did not include any useful contextual information that would aid in recognizing a word. It is

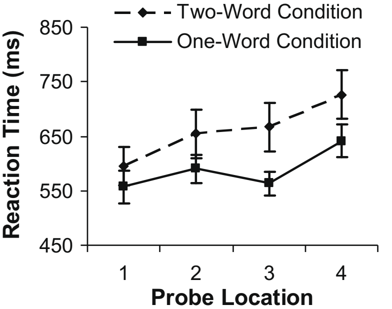

Fig. 2. Median reaction times for probe detection trials with correct responses in Experiment 1.

very likely that contextual information can influence word recognition and word segmentation and this issue will be addressed in Experiment 5 below.

Another possible difference between the one-word condition and the two-word condition was memory load. Subjects only needed to remember one 4-character word in the one-word condition, while they needed to remember two 2-character words in the two-word condition. There are two reasons not to attribute the performance differences to memory load differences. First, if the deficit in character recognition accuracy in the two-word condition was caused by increased working memory load in the two-word condition, we might expect character recognition to be lower for the first two characters in the two-word condition compared with those in the one-word condition. However, we did not find evidence for that. Secondly, short-term-memory capacity is about four items, and even bigger if the items can be organized into meaningful chucks such as words (Cowan, 2000). Hence, memory capacity should not cause problems for remembering and reporting four characters. This problem will be revisited in Experiment 4.

The use of probes in this experiment was triggered partly by Egly et al.'s (1994) evidence for spatial attention and object based attention in a single experiment. They found that reaction time was faster when the probe was presented in the same object as the cue; they also found that reaction time was shorter when the cue was at the same location as the cue than when they were at different locations within the same object. In the current experiment, probe responses were faster in the one-word condition than in the two-word condition, suggesting that whether the four characters constituted one word or two words influenced the detection of the probe. Reaction time was fastest when the probe was at the first character location and increased from left to right, but the pattern of reaction times differed across the two conditions. The increase in reaction time was small in the one-word condition, but was larger in the two-word condition. However, the increase in reaction time between the second and third character locations in the two-word condition was not as large as might be predicted by object-based attention theories. The ability of the probes to reveal spatial attention may have been limited by timing issues: attention was not necessarily at the first word when the probe appeared. Even if attention started by selecting the first word, by the time the probe appeared, it may have moved to the second word on some trials. Hence, it is very hard to draw a conclusion about whether word segmentation can affect attention deployment based on the current results. However, another recent study using a different paradigm did find that word-defined objects can affect attentional deployment. Li and Logan (2008) used the same paradigm as that used in Egly et al. (1994). Instead of using rectangles as objects, Li and Logan defined objects top-down. They presented four Chinese characters that constituted two words, arrayed horizontally or vertically in parallel. The four characters were spaced equally in the four corners of an imaginary square, so there were no bottom-up features that distinguished these two conditions. Using a spatial cuing paradigm, they found that a target character was detected faster by native Chinese readers if it was in the same word as the cued character than when it was in a different word. Because there were no bottom-up factors that distinguished the words, their results suggest that Chinese word segmentation can constrain the deployment of attention.

The effects of word segmentation were more difficult to find in the probe responses in the current experiment, but RT was clearly shorter over all in the one-word condition than in the two-word condition. There are two possible explanations. First, extra time may have been necessary in the two-word condition to switch attention from one word to another, as predicted by object-based attention theories (Duncan, 1984). Second, there may have been some other cognitive cost not associated with attention shifting that was higher in the two-word condition and detracted from probe task performance.

In the current experiment, the probe reaction time and character recognition accuracy reflect different aspects of processing: the probe reaction time reflects an early perceptual stage of character recognition that depends on spatial attention, while character recognition accuracy reflects the outcome of the entire perceptual stage and the actual recognition of the character. Because the accuracy measure is not sensitive to timing in the way that the probe measure is, the accuracy data appear to be more informative in the current study than the probe data. From this point on, we will mainly focus on character report accuracy.

3. Experiment 2

In the two-word condition of Experiment 1, about half of the subjects could only report one word, but the other half could report both words. Did the subjects who reported both words correctly do so because they are fast at perception and character recognition? If they were simply faster than those who could report only one word, then we should be able to find a presentation time that is short enough so that subjects could recognize only the first word correctly in the two-word condition, but could still recognize the word in the one-word condition. The processing time of each subject in Experiment 2 was thus adjusted by varying the exposure time of the words and by introducing a masking pattern that followed the presentation of the Chinese characters.

3.1. Method

3.1.1. Subjects

Eighteen subjects from the same subject pool as Experiment 1 participated in this experiment. None of them had participated in Experiment 1.

3.1.2. Apparatus and materials

These were the same as Experiment 1.

3.1.3. Procedure

At the beginning of the experiment, there were five warm-up trials, the purpose of which was to make subjects familiar with the procedure of the experiment. Then there were 30 adjustment trials to determine the exposure time for each subject using a staircase procedure. All of the adjustment trials had four characters making up two words. In this experiment and all of the following experiments, the materials in the warm-up trials and adjustment trials were different from those used in experiment trials. The exposure time was increased one frame (about 11.75 ms) if a subject could not report any word for two trials in the last five trials; the exposure was decreased by one frame if a subject correctly reported both words for two trials in the last five trials. The purpose of this adjustment was to set the task difficulty such that subjects were able to report one word but not both of the words in the two-word condition in most cases. After the exposure time was determined, there were 10 practice trials, which included both conditions.

After the words were presented in each trial, a mask (with 0 ms ISI) covered the spaces occupied by the four characters and one character before and one after. Each of the six parts of the mask was a diagonal grid the size of a single character. All of the other aspects of the experiment were exactly the same as that in Experiment 1.

3.2. Results

Only trials in which subjects pressed the correct button were included in the analysis; thus 57 trials out of 1296 trials (4.4%) were excluded.

3.2.1. Exposure duration

The minimal exposure duration was 47.1 ms, the maximum was 129.4 ms, and the average exposure duration was 77.8 ms (with a standard deviation of 31.8 ms).

3.2.2. Word report accuracy

The rate at which subjects correctly reported the word in the one-word condition (see Table 1) was much higher (92%) than that at which subjects correctly reported both words in the two-word condition (16%), $F1(1,17) = 729.26$, $p < .001$, MSE = .03; $F2(1,64) = 1176.02$, $p < .001$, MSE = .01. There was also an interaction between condition and location, $F1(3,51) = 5.81$, $p < .01$, MSE = .01; $F2(3,64) = 4.20$, $p < .01$, MSE = .01. The pattern of the interaction was similar to that in Experiment 1. In the one-word condition, accuracy was lower when the probe was at the fourth character location; in the two-word condition, accuracy of recognizing both words was lower when the probe was at the second and the third character locations than the other locations (see Table 1). In the two-word condition, when subjects could not report both words correctly, they reported the first word correctly on most trials (71%).

3.2.3. Reaction time to the probes

The average median reaction times are shown in Fig. 3. Reaction time was submitted to an ANOVA with condition and probe location as within-subject factors. There was a significant main effect of condition [$F1(1,17) = 6.94$, $p < .05$, MSE = 20834.88; $F2(1,64) = 25.65$, $p < .001$, MSE = 1305.75]. The reaction time was longer in the two-word condition (mean 579 ms, s.e. = 40 ms) than in the one-word condition (516 ms, s.e. = 23 ms). In a separate analysis including only locations 2 and 3, there were no main effects or interaction (*ps* > .1).

3.3. Discussion

When a mask was introduced and the exposure duration was adjusted for each subject, all of the subjects reported the word in the one-word condition quite well, while they could generally only report the first word in the two-word condition. In Experiment 1, there were some subjects who could report both words in the two-word condition. However, Experiment 2 suggests that they did not employ a different strategy. Instead, they could recognize both words because their processing speed was fast.

The reaction time pattern difference to the probe between the second character and the third character disappeared in this experiment. Reaction time when the probes were located at different locations differed less in comparison with that in Experiment 1. This might reflect different attentional deployment patterns between these two experiments. The exact reason for this difference is not clear to us. One possible reason is that attentional deployment might be interfered with by the mask, which was shown shortly after the stimuli were presented. In Experiment 1, subjects could use information stored in iconic memory to recognize characters and detect probes, while in Experiment 2, iconic memory was interrupted by the mask. Thus, subjects might have used a strategy to deploy attention over a wider range to collect more information for the probe detection task before the mask was presented. Given that the probe results were not particularly informative, we did not include the probe task in Experiments 3–5.

4. Experiment 3

Experiment 3 was designed for the following three purposes. First, we wanted to explore the extent to which subjects guessed the whole word based on the first two characters in the one-word condition in Experiment 1. In Experiment 1, the words in the one-word condition were reported much better than in the two-word condition. It may be argued that subjects only recognized the first two charac-

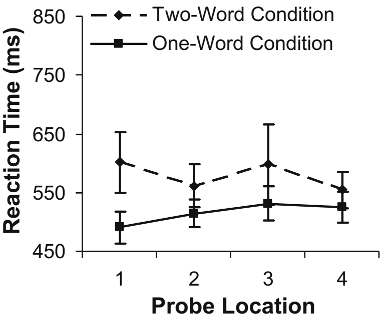

Fig. 3. Probe detection reaction times from Experiment 2.

ters and that based on this they guessed the whole word without any accurate perception of the last two characters. Second, in Experiment 3, the time that subjects took to name the characters or words was directly measured. In Experiments 1 and 2, the speed at which subjects recognized characters or words was reflected only indirectly in the response time to a probe. In Experiment 3, the start time of naming the characters was used as a direct measure of the recognition time of these characters. Third, the accuracy of character recognition was measured in the different conditions. In Experiments 1 and 2, only word recognition accuracy was recorded. We noticed that in some trials some characters were reported without the full word being reported. The accuracy of character recognition was recorded in Experiment 3 to explore the relation between the recognized character and the character location of the four characters.

In this experiment, in addition to the two conditions we used in Experiment 1, the following two additional conditions were introduced. In the half-word condition, the first two characters were part of a 4-character word, but the last two characters were randomly selected characters that did not make sense when combined with the first two characters. In the non-word condition, the four characters were randomly selected characters that did not make sense when combined together.

4.1. Method

4.1.1. Subjects

10 subjects from the same subject pool as Experiment 1 participated in this experiment. None of them had participated in either of the earlier experiments.

4.1.2. Apparatus and materials

The same apparatus was used as in the prior experiments, with the addition of a microphone to measure voice onset time. Word onsets were identified with a routine from the Experiment Builder software that also controlled all other aspects of the experiment. The materials for the one-word condition and the two-word condition were the same as Experiment 1. The materials from the 36 trials with a probe in Experiment 1 were used in this experiment, combined with new materials for the two new conditions. As in Experiment 1, the character frequency and character complexity of the stimuli were balanced among conditions. The word frequencies of the words from which the first two characters in the half-word condition were abstracted were comparable with the word frequencies of the words in the one-word condition. There were 36 trials in each condition.

4.1.3. Procedure

The procedure was similar to Experiment 2 except for the following differences. First, there was no mask after the characters disappeared. Second, there were no probes, so subjects did not need to press a button. Third, a microphone was used to record the start times of subjects' naming of the characters or words. Fourth, the experimenter recorded whether each character was reported correctly or not. The experimenter initiated the next trial by pressing a key.

4.2. Results

The exposure duration was 11.8 ms for all of the subjects except one, which was 47.1 ms. Note that there was no mask in this experiment, and while the exposure duration was short, subjects could do the task quite well.

4.2.1. Accuracy

The accuracy rate (see Fig. 4) of character recognition was submitted to an ANOVA with condition and location as within-subject factors in the subject analysis (*F1*), with location as a within item factor and condition as a between item factor in the item analysis (*F2*). Accuracy decreased in the following order: one-word condition, two-word condition, half-word condition, non-word condition [$F1(3,27) = 98.41$, $p < .001$, MSE = .01; $F2(3,140) = 108.88$, $p < .001$, MSE = 0.05]. Post-hoc contrasts revealed significant differences between the one-word condition and the two-word condition ($F1(1,27) = 24.21$, $p < 0.001$, MSE = 0.01; $F2(1,140) = 187.51$, $p < 0.001$, MSE = 0.01), and between the

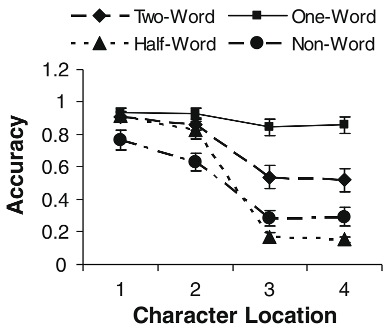

Fig. 4. Accuracy of character recognition in Experiment 3.

two-word condition and the half-word condition ($F1(1,27) = 26.07$, $p < 0.001$, MSE = 0.01; $F2(1,140) = 295.54$, $p < 0.001$, MSE = 0.01). The accuracy rate also decreased as character location varied from left to right [$F1(3,27) = 72.10$, $p < .001$, MSE = .03; $F2(3,420) = 357.76$, $p < .001$, MSE = .02]. Post-hoc contrasts revealed that the difference between location 2 and location 3 was significant ($F1(1,27) = 44.47$, $p < 0.001$, MSE = 0.028, $F2(1,420) = 838.58$, $p < 0.001$, MSE = 0.02). The difference between location 1 and location 2 was only significant in the item analysis ($F2(1,420) = 36.52$, $p < 0.001$, MSE = 0.02). Finally, there was an interaction between condition and character location [$F1(9,81) = 29.5$, $p < .001$, MSE = .01; $F2(9,420) = 34.31$, $p < .001$, MSE = .02]. Accuracy decreased most sharply from left to right in the half-word condition; medium in the two-word condition and the non-word condition; and less so in the one-word condition. The accuracy rate for the first two characters was lower for the non-word conditions than all of the other three conditions. As in the earlier experiments, we were interested in the accuracy at the second and the third locations when comparing the one-word condition and the two-word condition. An analysis including only locations 2 and 3, and only the one-word condition and the two-word condition, yielded a main effect of condition [$F1(1,9) = 18.15$, $p < .01$, MSE = .02; $F2(1,70) = 57.48$, $p < .001$, MSE = .02] and a main effect of location [$F1(1,9) = 32.23$, $p < .001$,MSE = .01; $F2(1,70) = 10918$, $p < .001$,MSE = .01]. There was also an interaction between location and condition [$F1(1,9) = 34.96$, $p < .001$, MSE = .004; $F2(1,70) = 36.86$, $p < .001$, MSE = .01], confirming that there was a larger drop in accuracy between the two locations when a word boundary was present there.

Overall the characters were more likely to be correctly recognized in the one-word condition (82.8%) than in the two-word condition (40.0%), the half-word condition (6.4%) or the non-word condition (9.7%), $F1(3,27) = 100.4$, $p < .001$, MSE = .01; $F2(3,140) = 232.12$, $p < .001$, MSE = .02.

The pattern of the half-word condition was quite interesting, so it was compared with the non-word condition separately. The accuracy rates of the half-word condition and the non-word condition were submitted to an ANOVA. Accuracy in the half-word condition was higher in the first two character locations (87%), but lower in the last two character locations (16%) in comparison to the non-word condition (70% for the first two characters and 29% for the last two characters) [$F1(3,27) = 15.4$, $p < .001$, MSE = .01; $F2(3,210) = 12.19$, $p < .001$, MSE = .03]. In the half-word condition, subjects sometimes guessed that the four characters were a word, but this only occurred in a few trials (mean 8% of the trials, s.e. of 4%, max 14%). When subjects did not guess a word, they usually only reported the first two characters. To test whether the half-word condition actually produced more errors in the two character locations that did not match the word, a separate analysis included only locations 3 and 4. Accuracy in the half-word condition (16%) was significantly lower than that in the non-word condition (29%) [$F1(1,9) = 14.63$, $p < .01$, MSE = .01; $F2(1,70) = 9.41$, $p < .01$, MSE = 0.06].

4.2.2. Naming time

An ANOVA on the time to start reporting the characters yielded a main effect of condition, $F1(3,27) = 29.66$, $p < .001$, MSE = 11755.94; $F2(3,140) = 18.58$, $p < .001$, MSE = 55150.02. The time to start reporting the characters was shorter in the one-word condition (818 ms, s.e. = 43 ms) than the two-word condition (1024 ms, s.e. = 56 ms), $F1(1,27) = 8.35$, $p < .001$, MSE = 11755.94; $F2(1,140) = 32.59$, $p < 0.001$, MSE = 55150.02. The two-word condition was faster than the half-word condition (1123 ms, s.e. = 65 ms), $F1(1,27) = 16.58$, $p < .001$, MSE = 11755.94; $F2(1,140) = 9.43$, $p < 0.01$, MSE = 55150.02. The half-word condition was faster than the non-word condition (1263 ms, s.e. = 75 ms), $F1(1,27) = 35.88$, $p < .001$, MSE = 11755.94; $F2(1,140) = 32.02$, $p < 0.001$, MSE = 55150.02.

4.3. Discussion

This experiment confirmed the results of the first two experiments: characters were more easily recognized in the one-word condition than in the two-word condition. A direct measure of the time to start verbally reporting the characters indicated that naming time was shorter in the one-word condition than the other conditions.

In the half-word condition, when the first two characters were part of a 4-character word, but the other two characters were not, subjects rarely guessed these four characters as a word. This demonstrates that subjects generally did not guess the word in the one-word condition based only on the information of the first two characters independently of their perception of the third and the fourth characters. This finding therefore seems to exclude the possibility that the recognition difference between the one-word condition and two-word condition was due to purely guessing based on the first two characters. Perception of the third and the fourth characters must play some role in the recognition of these characters. Though a pure guessing account is excluded, there is also clear evidence indicating that the perception of the first two characters can influence the recognition of the last two characters. In the half-word condition, the accuracy rate of the last two characters was lower than that in the non-word condition. Taken together, the present results support a holistic hypothesis of Chinese character recognition and word recognition.

The accuracy rate for the first two characters in the non-word condition was lower than that in all of the other conditions (in which the first two characters were a word or part of a word). This result suggests that character recognition can be facilitated if the characters are part of a word; a conclusion consistent with that from experiments on the word superiority effect (Reicher, 1969; Wheeler, 1970).

In the one-word condition and the two-word condition, on most trials, all of the recognized characters were part of a word (Table 2). However, in some trials some characters were reported without the full word being reported. These characters were distributed across all four locations, but characters located more to the left were more likely to be recognized. If characters were recognized in a purely serial fashion, we would have expected that they were recognized one by one from left to right, but this was not what we observed. When two characters were recognized on the same trial, they were not always from neighboring locations. These results show that characters may be recognized to some extent in parallel, with efficiency decreasing gradually from left to right.

Table 2
Percentage of different patterns of character recognition in Experiment 3. Recognition represents which characters were recognized: 1 is correct, 0 is incorrect.

Recognition	Conditions				Recognition	Conditions			
	Two word	Single word	Half word	Non-word		Two word	Single word	Half word	Non-word
0 0 0 0	3.3	4.7	4.4	10.6	1 0 0 0	2.8	0.6	8.6	14.2
0 0 0 1	0.6	0.0	0	1.1	1 0 0 1	1.9	0.6	0.8	3.1
0 0 1 0	0.8	0.3	0	1.1	1 0 1 0	0.6	0	0.8	1.9
0 0 1 1	1.9	0.3	0	3.1	1 0 1 1	1.4	0	0.8	2.2
0 1 0 0	0.6	0	0.8	2.8	1 1 0 0	31.7	7.5	59.8	31.1
0 1 0 1	0	0	0.3	2.5	1 1 0 1	5.0	1.7	6.7	6.1
0 1 1 0	0.6	0.3	0.8	0.6	1 1 1 0	7.5	0.3	7.8	8.3
0 1 1 1	1.1	0.6	0.3	1.7	1 1 1 1	40.0	82.8	6.4	9.7

In the current experiment, the material in the one-word condition and the two-word condition were different in their character transition probabilities. The character transition probability[4] of character 3 given character 1 was higher for the one-word condition (0.0618, s.e. = 0.0327[5]) than the two-word condition (0.0007, s.e. = 0.0002), and the character transition probability of character 3 given character 2 was higher for the one-word condition (0.0265, s.e. = 0.0129) than the two-word condition (0.0006, s.e. = 0.0002). An alternative explanation of the results is that they have little to do with whether or not there was a word boundary, but rather are due to differences in character transition probability. According to this account, the recognition of the third character should be easier if the character transition probability is higher. Though the current experiment was not designed to test this hypothesis, an additional analysis of the results of the two-word condition was carried out to determine the effect of transitional probability. In the two-word condition, the third character had nothing to do with the lexical representation of the first word. However, the character transition probability varied from 0 to 0.006 both for character 3 given character 1 and for character 3 given character 2. Furthermore, the variance of character recognition accuracy was fairly large across items. If character transition probability played an important role in the results we observed, we would expect a higher recognition rate for the stimuli with higher character transition probabilities, hence a significant correlation between these two factors. For each item, we calculated the percentage of correct recognition of character 3 (across subjects), then we calculated the correlation of character accuracy with character transition probability.[6] The correlation was 0.01 for character 3 given character 1 and 0.08 for character 3 given character 2; neither was significantly different from zero ($p > 0.1$, see Fig. 5 for a scatterplot). In a similar analysis conducted for the one-word condition, the correlation was 0.08 for character 3 given character 1 and 0.06 for character 3 given character 2 ($p > 0.1$). Given that all of the correlations are small, and the word-boundary effects are robust and big, it is very unlikely that the word-boundary effect is caused by the transition probabilities. Furthermore, the transition probability could not explain the effects on recognition accuracy for the first character. According to the transition probability hypothesis, the recognition accuracy of character 1 should not differ across all of the conditions. However, the character 1 recognition accuracy in the non-word condition was significantly smaller than the other conditions ($t(9) = 3.53$, $p < 0.01$). Hence, it is unlikely that character transition probability can explain the character recognition difference between the one-word condition and the two-word condition.

5. Experiment 4

In the first three experiments, the task, or one of the tasks, was to report the words/characters. Perhaps it could be argued that there was no real difference in character recognition between the one-word condition and the two-word conditions, but that the differences observed in the first three experiments were only because subjects were more likely to forget the characters in the second word in the two-word condition before they could report them. To exclude this possibility, in Experiment 4 the task was changed to a search task in which subjects had to detect whether or not a character was present in the array. If the difference between the one-word condition and the two-word condition still shows the same pattern as in the first three experiments, the memory account can be excluded.

5.1. Method

5.1.1. Subjects

Twelve subjects from the same subject pool as Experiment 1 participated in this experiment. None of them had participated in any prior experiment.

---

4 Based on a corpus of 838,803,906 characters from http://ccl.pku.edu.cn:8080/ccl_corpus/jsearch.

5 The SD was large because there were five high values (ranging from 0.1 to 0.9). All of the other values were smaller than 0.02.

6 One item (which influenced the correlation by more than 0.2 and was identified as an outlier by the Systat software) was excluded from analysis when calculating the correlation between the character 3 accuracy and the transition probability of character 3 given character 2 in the two-word condition. Two items were excluded from analysis in the one-word condition since the transition probability of character 3 given character 1 was extremely large (0.93 and 0.76) compared to the other items (less than 0.15).

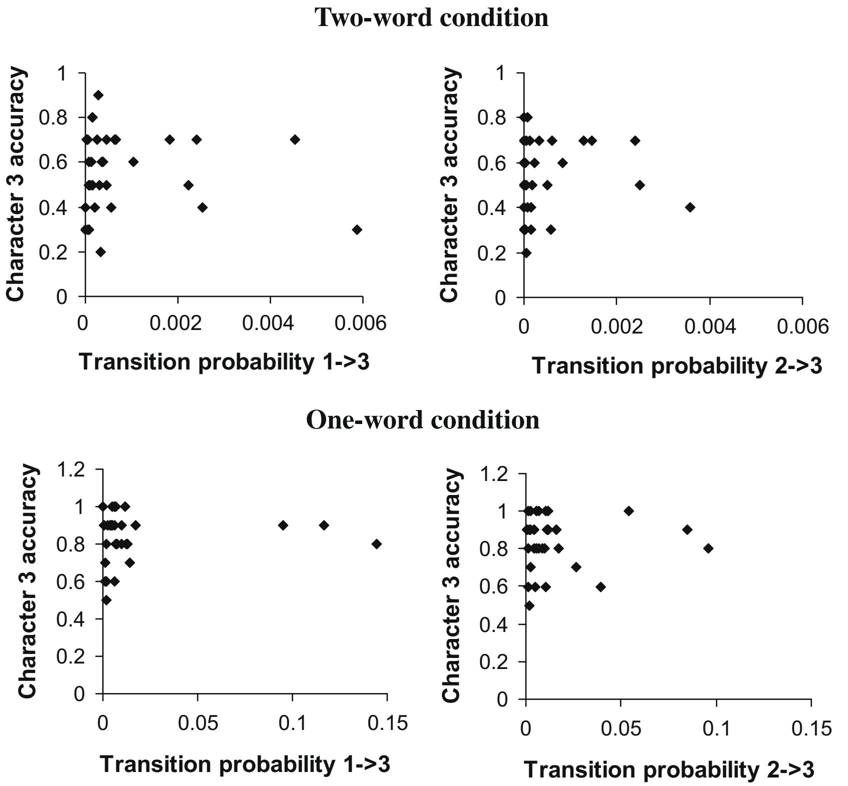

Fig. 5. Scatterplots of the relation between character accuracy and transition probability of Experiment 3. The top panels are for the two-word condition, and bottom panels are for the one-word condition.

5.1.2. Apparatus and materials

The apparatus was similar to that of Experiment 1. There were two conditions in this experiment, the one-word condition and the two-word condition, and each consisted of 72 trials. The target was present in half of trials. For those target-present trials, the target was equally likely to be at any of the four locations. As before, character frequency, character complexity, and word frequency were balanced between the two conditions. In the adjustment trials, four randomly selected characters were used in each trial.

5.1.3. Procedure

Each trial began with the presentation of the target character for 1.5 s. This was followed by a fixation point, which was presented for 1 s at the location that would be the center of the first character once it appeared. Then the four characters were presented for the duration determined during the adjustment trials for that subject. Subjects were asked to press one of the two buttons to indicate whether the target was present or not. The next trial started 1 s after subjects made the response.

There were 15 warm-up trials, 95 adjustment trials, 10 practice trials, and 144 experiment trials. During the adjustment trials, exposure time was adjusted using the staircase procedure so that the correct rate was about 85%. The exposure duration was fixed to this duration during all of the experimental trials for each subject.

5.2. Results

The data of one subject was excluded from analysis because of extremely long response times. The average reaction time for the other subjects was 746 ms (with a s.e. of 128 ms).[7] The exposure duration for most subjects was 11.8 ms, with the exception of two subjects, whose exposure duration was 58.8 ms.

5.2.1. Accuracy

Accuracy rates (see Fig. 6) for only the target-present trials were submitted to an ANOVA, with conditions and location as within-subject factors. Subjects were more accurate in the one-word condition than in the two-word condition [$F1(1,10) = 9.80$, $p < .05$, MSE = 0.01; $F2(1,64) = 8.42$, $p < .01$, MSE = .006], and accuracy decreased from left to right [$F1(3,30) = 5.31$, $p < .01$, MSE = .01; $F2(3,64) = 4.73$, $\eta_p^2 = .18$, $p < .01$, MSE = .006]. Post-hoc contrasts revealed that the difference between character location 2 and character location 3 was significant ($F1(1,30) = 7.86$, $p < 0.01$, MSE = 0.01; $F2(1,64) = 32.86$, $p < 0.001$, MSE = 0.01). Accuracy decreased more rapidly from left to right in the two-word condition than in the one-word condition [$F1(3,30) = 8.41$, $p < .01$, MSE = .004; $F2(3,64) = 4.15$, $p < .01$, MSE = .006]. Specifically, accuracy dropped from location 2 to location 3 in the two-word condition, but not in the one-word condition. A separate analysis including only locations 2 and 3 yielded a marginally significant main effect of condition [$F1(1,10) = 3.29$, $p < .1$, MSE = .08; $F2(1,32) = 6.48$, $p < .05$, MSE = .006] and a main effect of location [$F1(1,10) = 9.83$, $p < .05$, MSE = .01; $F2(1,32) = 8.46$, $p < .01$, MSE = .006]. More importantly, there was an interaction between location and condition [$F1(1,10) = 7.82$, $p < .05$, MSE = .01; $F2(1,32) = 6.48$, $p < .05$, MSE = .006], confirming the large drop in accuracy associated with the word boundary.

5.2.2. Reaction time

The median reaction time (see Fig. 6) for each subject was submitted to an ANOVA, with conditions and locations as within-subject factors. Responses were shorter in the one-word condition than in the two-word condition [$F1(1,10) = 6.9$, $p < .05$, MSE = 11350.31; $F2(1,64) = 7.58$, $p < .01$, MSE = 11554.63;]. There was a main effect of location [$F1(3,30) = 5.1$, $p < .01$, MSE = 11420.63; $F2(3,64) = 5.26$, $p < .01$, MSE = 11554.63]. Post-hoc contrasts revealed that reaction time increased from location 2 to location 3 ($F1(1,30) = 5.16$, $p < 0.05$, MSE = 11420; $F2(1,64) = 22.98$, $p < .001$, MSE = 11554.63).

5.3. Discussion

Unlike the previous experiments, subjects did not need to explicitly remember the character (or report it) in this experiment. If the one-word condition recognition benefit found in the previous experiments arose only because of working memory limitations that caused subjects to forget the second word when they reported the first word, we would have expected that the detection accuracy would be similar for the two-word condition and the one-word condition. Instead, the character detection task again produced a word-boundary effect in this experiment. Accuracy did not change much as a function of the four character locations in the one-word condition; however, accuracy dropped significantly from location 2 to location 3 in the two-word condition. Response time was faster in the one-word condition than in the two-word condition. Although the character detection task did not require subjects to perceive all four characters, the word boundary affected performance. These results demonstrate that the word-boundary effect observed in the previous experiments did not appear simply because subjects forgot the second word when reporting the first one.

[7] The excluded subject had an average reaction time of 1130 ms.

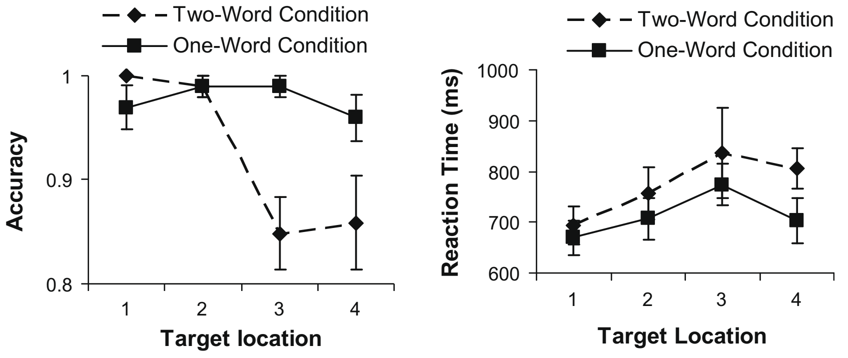

Fig. 6. Accuracy (left) and reaction times (right) for detecting the target character in Experiment 4.

6. Experiment 5

In the first four experiments, we attempted to control for contextual effects. Specifically, the two words in the two-word condition were selected such that they did not make sense when combined together. However, it is known that contextual information influences the processing of Chinese words. Rayner et al. (2005) found that Chinese readers made shorter fixations on words that were highly predictable from the prior context than on words that had low predictability. Experiment 5 was designed to explore contextual influences on word segmentation and recognition. This experiment repeated most of the conditions from Experiment 3 except the half-word condition, which was replaced by a related-word condition. In the related-word condition, there were two 2-character words, which were closely related (e.g., 美满婚姻, which means happy marriage). We were interested in whether context could overcome the large drop in accuracy between the two locations when a word boundary was present there.

6.1. Method

6.1.1. Subjects

Ten subjects from the same subject pool as Experiment 1 participated in this experiment. None of them had participated in earlier experiments.

6.1.2. Apparatus and materials

The apparatus was similar to that of Experiment 3. The materials were exactly the same as that used in Experiment 3 except that the half-word condition was replaced by the related-word condition. In the related-word condition, the two words were closely related. The two words occurred together an average of 45 times (s.e. = 51) in the character corpus (see footnote 3).

6.1.3. Procedure

The procedure was exactly the same as that used in Experiment 3.

6.2. Results

The mean exposure duration was 21.2 ms (SD 10.8 ms, ranging from 12 ms to 35 ms), which was a bit longer than that in Experiment 3 (presumably due to individual differences).

6.2.1. Accuracy

The accuracy rates (see Fig. 7) for character recognition were submitted to an ANOVA with condition and location as within-subject factors. Accuracy decreased in the following order: one-word con-

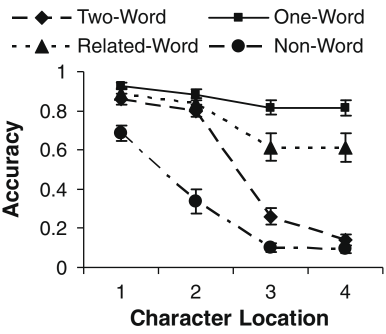

Fig. 7. Accuracy of character recognition in Experiment 5.

dition, related-word condition, two-word condition, non-word condition [$F1(3,27) = 145.7$, $p < .001$, MSE = .02; $F2(1,140) = 3983.61$, $p < .001$, MSE = .04]. Post-hoc analyses revealed that the difference between the closest pairs were significant ($F1$s$(1,27) = 8.71$, 29.24, 26.24, all $ps < 0.01$; $F2$s $(1,140) = 59.54$, 139.88, 244.04, all $ps < 0.001$). The accuracy rate also decreased as character location varied from left to right [$F1(3,27) = 137.2$, $p < .001$, MSE = .01; $F2(3,420) = 946.52$, $p < .001$, MSE = .02]. Post-hoc contrasts revealed a significance between location 1 and location 2 ($F1(1,27) = 13.33$, $p < 0.01$, MSE = 0.01; $F2(1,420) = 141.81$, $p < 0.001$, MSE = 0.02) and between location 2 and location 3 ($F1(1,27) = 58.92$, $p < 0.001$, MSE = 0.01; $F2(1,420) = 638.13$, $p < 0.001$, MSE = 0.02). Finally, there was an interaction between condition and character location [$F1(9,420) = 58.81$, $p < .001$, MSE = .02]. Accuracy decreased most sharply from left to right in the two-word condition, less so in the related-word condition, and even less so in the one-word condition.

We were interested in accuracy at the second and the third locations when comparing the one-word condition and the related-word condition. An analysis including only locations 2 and 3, and only the one-word condition and the related-word condition, yielded a main effect of condition [$F1(1,9) = 12.28$, $p < .01$, MSE = .01; $F2(1,70) = 12.28$, $p < .01$, MSE = .06] and a main effect of location, significant only in the items analysis [$F1 < 1$; $F2(1,70) = 80.24$, $p < .001$, MSE = .01]. There was also an interaction between location and condition [$F1(1,9) = 11.57$, $p < .01$, MSE = .01; $F2(1,70) = 24.96$, $p < .001$, MSE = .01], confirming that there was a larger drop in accuracy between the two locations when a word boundary was present. The same comparison between the two-word condition and the related-word condition also yielded a main effect of condition [$F1(1,9) = 27.49$, $p < .005$, MSE = .01; $F2(1,70) = 24.41$, $p < .001$, MSE = .06], a main effect of location [$F1(1,9) = 72.06$, $p < .001$, MSE = .02; $F2(1,70) = 245.66$, $p < .001$, MSE = .02] and an interaction between location and condition [$F1(1,9) = 49.09$, $p < .001$, MSE = .01; $F1(1,70) = 41.67$, $p < .001$, MSE = .001].

6.2.2. Naming time

An ANOVA on the time to start reporting the characters yielded a main effect of condition [$F1(3,27) = 34.76$, $p < .001$, MSE = 9598.28; $F2(3,140) = 41.90$, $p < .001$, MSE = 23898.29]. The time to start reporting the characters was shortest in the one-word condition (722 ms, s.e. = 30 ms), followed by the related-word condition (749 ms, s.e. = 26 ms), then the two-word condition (915 ms, s.e. = 53 ms), and was the longest in the non-word condition (1118 ms, s.e. = 59 ms). The difference between the one-word condition and the related-word condition was significant in the item analysis ($F1(1,27) < 1$; $F2(1,140) = 5.83$, $ps < 0.05$, MSE = 23898.29). The difference between the related-word condition and the two-word condition, and between the two-word condition and the non-word condition were both significant ($F1$s$(1,27) = 29.12$,42.80, $ps < .001$; $F2$s$(1,140) = 62.47$, 98.70, $ps < .001$).

In Experiment 3, we analyzed the correlation between the recognition accuracy of character 3 in the two-word condition and the transition probability of character 3 given character 1 and character 2. We did the same thing for the related-word condition in Experiment 5. In the related-word condition of Experiment 5, the transition probability for character 3 given the combination of characters 1 and 2 was high enough (varied from 0.0007 to 0.1136, mean = 0.0187, s.e. = 0.0040) that we were able to include it in the correlational analysis as well. The character transition probability was 0.0046 (s.e. = 0.0012) for character 3 given character 1, and was 0.0032 (s.e. = 0.0008) for character 3 given character 2. For each item, we calculated the percentage of correct recognitions of character 3 (across subjects), then we calculated the correlation of character accuracy with character transition probabilities.[4] None of the three correlations was significantly different from zero ($p > 0.1$, see Fig. 8 for a scatterplot).

6.3. Discussion

In Experiment 5, we found that when the two words in the related-word condition were meaningful when combined together, there was still a larger drop in accuracy across the word boundary than in the one-word condition. This suggests that the context does not completely overcome the influence of word boundaries on character recognition. In the related-word condition, the characters in the second word were recognized worse than those in the one-word condition but better than those in the two-word condition, which indicates that context influences character recognition.

7. General discussion

In the experiments reported here, given limited exposure time, most subjects could report the 4-character word in the one-word condition, while they often could not report the second two-character word in the two-word condition. When a mask followed immediately after the words, we determined an exposure duration for each subject at which they could recognize the 4-character word correctly in the one-word condition, but could only recognize the first two-character word in the two-word condition. Occasionally, subjects did report some characters that were not part of any presented word. The probability of characters being recognized generally decreased from left to right. Sometimes the recognized characters were not located at adjacent locations. However, one

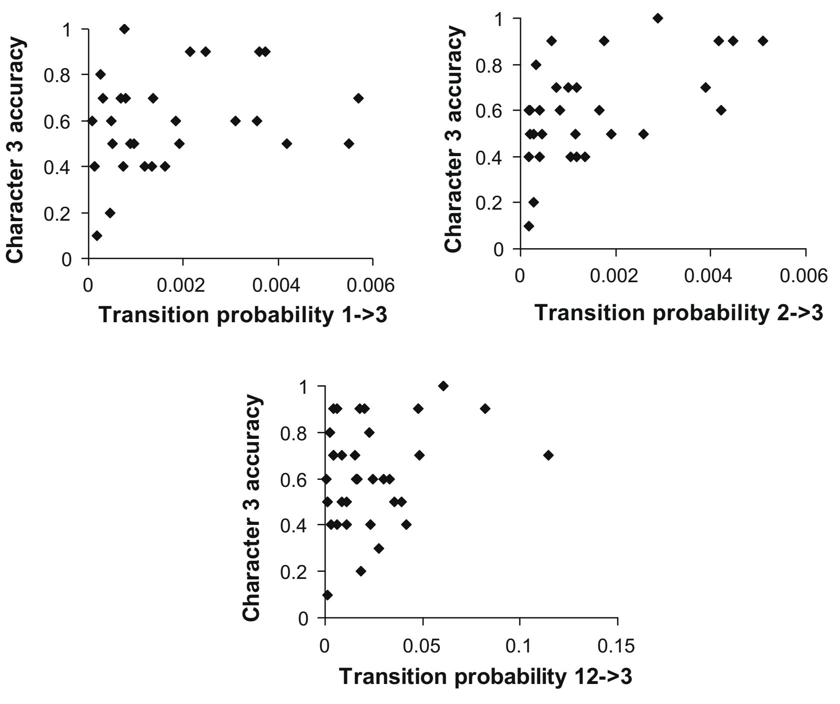

Fig. 8. Scatterplots of the relation between character accuracy and transition probability of Experiment 5.

rather important finding was that in the two-word condition, there was generally a rather sharp drop-off between the second and third characters. The results of Experiment 4 showed that the word-boundary effect still existed when the task was to search for a specific character in the 4-character string. This indicates that the word-boundary effect was not due to a memory load difference. Experiment 5 showed that in the two-word condition, subjects were more likely to report the second word if it fit together with the first word semantically. However, there was still a marked drop in performance between characters 2 and 3 when the two words went together in comparison to the one-word condition.

Overall, the results are consistent with the view that the segmentation and recognition of Chinese words involves an interaction of top-down and bottom-up processes. Although the subjects could recognize the 4-character word quite well in the one-word condition, they usually could only recognize the first word in the two-word condition. This finding demonstrates that the four characters in these two conditions were not processed in the same way. Character recognition was not a simple bottom-up process; instead, the results of the word segmentation process influenced character recognition.

It may be argued that subjects recognized only the first two characters in both conditions and that they could guess the rest of the characters in the one-word condition but not in the two-word condition. We doubt this explanation for the following reasons. First, if subjects had to guess the third and fourth letters in the word in the one-word condition, it should take extra time, which is not apparent in the reaction times. Subjects were very confident and quick when reporting the word in the one-word condition; they were much more hesitating and slow in the two-word condition. Both the reaction time to probes (Experiments 1 and 2) and word reporting time (Experiment 3) were faster in the one-word condition than in the two-word condition. Though the reaction time to the probe in Experiments 1 and 2 was not a direct measure of word recognition time, it presumably does reflect something about the speed of word processing. Because subjects were asked to report the characters in each trial, they might struggle to recognize the characters before pressing the button to avoid forgetting. The word onset time in Experiment 3, which was the time for response as measured by the voice key, was a more direct measure of word recognition time, and it also showed that recognition was faster in the one-word condition than in the two-word condition. Second, in the half-word condition in Experiment 3, where the last two characters were not the characters predicted by the first two characters, subjects rarely guessed the last two characters.

While we argued against the view that the third and the fourth characters could be guessed purely on the knowledge of the first two characters, this study does provide evidence that the recognition of the third and the fourth characters is influenced by the recognition of the first two characters. The critical evidence comes from the half-word condition in Experiment 3, where the first two characters were part of a 4-character word, but the last two characters were randomly selected characters. We found that the recognition of the last two characters was even worse than in the non-word condition. These results support the holistic hypothesis. The activated lexical representation could influence character recognition. The recognition of the third and the fourth characters was facilitated if they were consistent with the activated lexical representation, but inhibited if they were not.

An alternative explanation of the deficit of character 3 and 4 recognition accuracy in the two-word condition is the cognitive load hypothesis. Only one word needs to be processed in the one-word condition, while two words need to be processed in the two-word condition. Hence, cognitive load is larger in the two-word condition than in the one-word condition. We suspect that cognitive load does not account for the accuracy data observed in our experiments. If cognitive load influences character recognition, we would expect that it would influence the recognition of all of the characters, and not just characters 3 and 4. Note that the cognitive load hypothesis rests on the assumption that the characters are processed in parallel so that there is a difference in the cognitive load across letter locations.

The results obtained in the present studies are consistent with the view that words are processed serially when reading, at least in the lexical access stage, as suggested by the E–Z Reader model. Indeed, Rayner, Li, and Pollatsek (2007) recently implemented a version of the E–Z Reader model for Chinese. In the model, it was assumed that words (rather than characters) are the primary processing unit when reading Chinese. The words were processed serially from left to right. The current results are consistent with these ideas. The first word in the four characters could be recognized very accurately; whether it is a two-character word or a 4-character word. This suggests that a word is processed as a unit. In the two-word condition, subjects often could not report the second word even though they could report the first word very accurately, especially if the display time was short. This suggests that the word on left (the fixated word) is accessed earlier than the words at right during Chinese reading.

Although these results are consistent with serial word recognition, they do not require that the process be completely serial. The results indicate that in early processing of a two-word sequence, the fixated word is processed much more fully and efficiently than the word to the right. However, it is not clear whether the processing of the second word can only begin after the processing of the first word is finished. Instead, word processing might be partly parallel and partly serial. Early on, processing may be focused primarily on the fixated word, while the processing of the second word also gets off to a slow start. As time passes and the processing of the fixated word nears completion, the processing would then shift more to the second word. Thus, the current results argue against a purely parallel account in which both words are processed equally from the beginning, but they can be explained by either a purely serial account, or by a parallel account with a focus of processing that shifts serially over time from one word to the next.

The question of serial vs. parallel processing can be asked at the level of characters as well as at the level of words. The results of the current study suggest that Chinese characters are processed not purely serially or purely in parallel. If the characters were processed serially, we would expect that characters would be recognized one by one from left to right. If they are processed purely in parallel, we would expect that the characters at all of the four locations are processed equally. Instead, the results suggested that the characters are processed in parallel, but the processing efficiency decreases from left to right. In Experiment 3, character 1 was recognized more accurately in the two-word condition, in which it constitutes a word with character 2, than in the non-word condition, in which it does not constitute a word with character 2. This shows that the processing of character 2 could influence character 1, and is not consistent with purely serial character processing. The finding that character recognition accuracy decreases from left to right in the non-word condition of Experiment 3 suggests that the characters at different locations are not processed equally efficiently, as they would be in a purely parallel system.

Previous studies on English word processing showed that English words are processed in a holistic way (Cattell, 1886; Healy, 1976, 1994; Johnston, 1978; Tao et al., 1997). The present study showed that Chinese words are also processed in a holistic way, even when the word is presented with other characters without spaces separating them, so that word segmentation is necessary. The accuracy of recognizing the first two characters in the two-word condition was much higher than that in the non-word condition.

The response time to a probe located at one of the character locations was generally faster in the one-word condition than in the two-word condition. If attention in the two-word condition was allocated only to the first word, we would expect much longer reaction times for probes at character 3 than for those at character 2, but there is little evidence for such an attention effect. Prior research on the relation between word recognition and visual attention in English reading is quite complicated. Some have argued that attention is necessary for word recognition (McCann, Folk, & Johnston, 1992; Risko, Stolz, & Besner, 2005; Stolz & McCann, 2000), while others have argued the contrary (Brown, Gore, & Carr, 2002; Carr & Pollatsek, 1985). Though Li and Logan (2008) argued that Chinese word segmentation could affect attention deployment, the stimuli in their study were not in a natural reading context. Whether Chinese word segmentation in reading can affect visual attention deployment is an open question.

8. A model of Chinese word segmentation

As summarized above, we argued that Chinese word segmentation and Chinese word recognition is an interactive process involving top-down and bottom-up factors. Based on the findings from the present experiments, we propose a mathematical model of Chinese word segmentation and recognition. This model borrows some assumptions of the Interactive Activation (IA) model of McClelland and Rumelhart (1981), which assumed that English word recognition is an interactive process involving multiple levels (a visual feature level, a letter level, and a word level). These factors are retained in the current model. The purpose of the current modeling endeavor is to provide a framework for Chinese word segmentation that could account for the findings from the current experiments and other major Chinese reading findings. It provides us with a tool to understand the Chinese word segmentation problem.

One main purpose of the IA model was to account for the word superiority effect in English (Reicher, 1969; Wheeler, 1970). Though later Parallel Distributed Processing (PDP) models (Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989) extended the IA model, they cannot be used directly to explain Chinese word segmentation and recognition for the following reasons. First, there are no spaces between words in Chinese reading. The IA model and the later PDP models only recognized single 4-letter English words. The model of Chinese word recognition has to rely on some additional assumptions to deal with word segmentation. Second, the number of Chinese characters (approximately 5000) is much larger than the number of English letters (26). Third, Chinese characters are much more complex than English letters. Thus, the model described below is not intended to replicate the IA model in Chinese. We have to make new assumptions to account for the differences between Chinese and English.

In the current model, there are multiple levels of processing when recognizing a Chinese word (see Fig. 9). The first level is a visual perception module that abstracts visual features from the stimulus. Because of eccentricity, the efficiency of the perception of the characters decreases gradually from left to right. The second level is a character recognition module, which recognizes characters using perceptual information from the first level and feedback information from the word recognition level. In this level, there are multiple character recognizers, which work in parallel. The third level is the word segmentation and recognition level. This level receives information from both the character recognizers and the lexicon.

Word recognition was implemented as a process of evidence combination. Each character provides evidence for word recognition. If two characters provide consistent evidence for the same word, the activation of the corresponding word is higher. On the other hand, if two characters do not provide consistent evidence for the same word, the activation of the corresponding words is smaller. The word recognition level provides feedback information to the character recognition level. If a character is a part of a word with high activation, it will receive more evidence from the word recognition level.

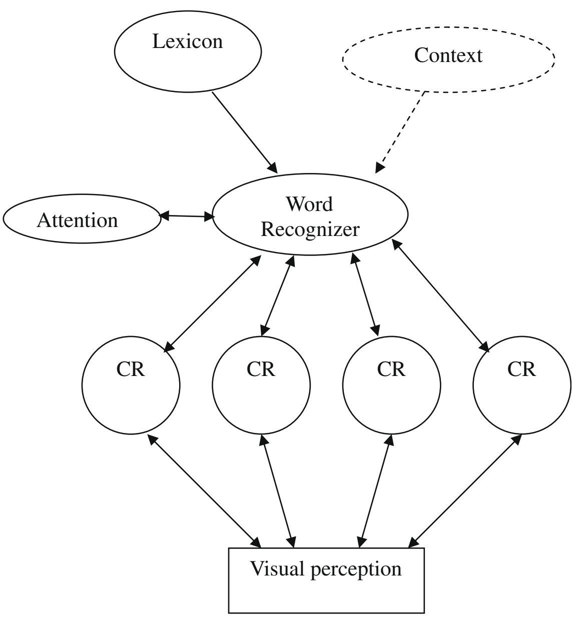

Fig. 9. The framework of a word segmentation and recognition model. CR refers to character recognizer.

Hence it should be recognized faster. If a character recognizer does not receive any feedback information from word recognition level, it can still recognize the character based on bottom-up information, but the recognition of that character will be slower. The character recognition level provides information to the word recognition level, and the word recognition level provides feedback information to the character recognition level. Thus, the model is a dynamic interactive system. In the simulations, the process described above runs for 100 iterations.

The current model differs from the IA model in the following dimensions. First, the words may have different lengths. Second, there is an attention control module, which selects the left-most word to be the current focus of recognition. Third, the speed of perception of Chinese characters is assumed to decrease as the eccentricity of the characters increases. Though this model inherits some assumptions of the IA model, the current model is different from the IA model because of these different assumptions. Experiment 5 showed that context could influence recognition of the third and the fourth characters. While the current simulation did not include contextual influences, there is a place in the model for such effects to be added (see Fig. 9). Most likely, contextual information influences word processing at a semantic level. More data are needed before we can make assumptions about how contextual information influences word segmentation and word recognition. To keep the model simple, it does not include levels above the lexical level, and hence the contextual effects were not implemented in the current simulation. The structure of the model is shown in Fig. 9. The detailed structure of the model is described below.

8.1. Visual perception module

The visual perception module provides input for character recognition. There is one perception channel for each character location. At each channel, there is a character recognizer for each possible Chinese character. At a given time t, the evidence accumulated at character recognizer j at location i supporting the character shown at that location is

$$pb_{ij}(t) = \int (\beta + noise)dt \times eccentricity_i \text{ if character } j \text{ is shown at location } i \tag{1}$$

For simplicity, we assumed that

$$pb_{ij}(t) = 0 \text{ if character } j \text{ was not shown at location } i.$$

β is a parameter that represents the speed of perceptual evidence accumulation. Noise is Gaussian noise with an amplitude of δ. pb is always greater than or equal to 0. For simplicity, in the simulation we assumed that the perception module does not provide evidence for the characters that were not shown at that location, and many of the later calculations for characters that are not present can be left out of the simulation. Note that the absence of evidence supporting a specific character does not mean that the character is ruled out for that location. In the experiments, the subjects were instructed to fixate on a fixation point located at the first character location. The rapid perceptual decrease with the increase of eccentricity is represented by a parameter for eccentricity in the model.

$$eccentricity_i = e^{-\gamma \times i} \tag{2}$$

γ is a parameter to be fit during the simulation. It describes how quickly perception efficiency decreases as eccentricity of the character increases. i is the character location. $i = 1$ for the left character, and $i = 4$ for the right character.

8.2. Character recognizer

In the model there is a character recognizer for each possible character at each character location, although in the simulation we only included recognizers that were directly relevant for the stimuli we were testing. Character recognition is a process of combining evidence from the perception module and the word recognition module. The more evidence that a character accumulates, the more likely it is that character will be recognized. The amount of evidence is measured by a belief function. (Belief functions will be considered more fully below.) For the node at the ith location, the belief function is

$$pc_i = 1 - (1 - pb_i)(1 - pt_i) \tag{3}$$

pb is the perceptual evidence described in Eq. (1), and *pt* is feedback information from the word recognition level that will be described in Eq. (6).

A character is considered to be recognized when

$$pc_i > \omega 1 \tag{4}$$

where $\omega 1$ is a threshold that is set to 0.95. *pc* determines which character should be reported. The character is reported by the subjects when *pc* reaches the threshold $\omega 1$. As in the IA model, the present model assumes that the character report is the only output from the model that determines responses.

8.3. Word recognizer

Word recognition was also modeled as evidence combination. This level receives information from both the character recognizers and the lexicon, which has all of the information about known words. Evidence from all of the four characters could contribute to word recognition. Evidence from different characters was combined using the Dempster–Shafer Theory (Shafer, 1976). The Dempster–Shafer Theory is a mathematical theory of evidence that is a generalization of the Bayesian theory of subjective probability. The Dempster–Shafer Theory can be used to calculate the probability of an event based on the combination of different pieces of evidence. In the Dempster–Shafer Theory, each set of possibilities is assigned a mass. The sum of the masses of all of the subsets of a set is called the belief function, and the resulting value is the amount of belief that directly supports a given hypothesis.

Evidence from different sources were combined using Eq. (5). Suppose that there is one source of evidence represented by the set $\{A_1, A_2, \ldots, A_m\}$, with masses $m(A_1)$, $m(A_2), \ldots, m(A_m)$, and another source of evidence represented by the set $\{B_1, B_2, \ldots, B_n\}$, with masses $m(B_1)$, $m(B_2), \ldots, m(B_n)$. The combination of the two sources of evidence generates a set $C = \{C_1 = A_1 \cap B_1, C_2 = A_1 \cap B_2, \ldots, C_{m*n} = A_m \cap B_n\}$. The joint mass of the two sources of evidence is

$$m(C_k) = \frac{\sum_{A_i \cap B_j = C_k} m(A_i)m(B_j)}{1 - \sum_{A_i \cap B_j = \phi} m(A_i)m(B_j)} \tag{5}$$

For the specific question at hand, each character recognizer provides evidence to the word recognizer. For a given character, the set of interest is

$$\{\mathrm{W}_i, \Omega\}$$

W_i includes all of the words that include the character at a specific location with some limitations introduced by the attention module (see below). The attention module facilitates characters to the left more than characters to the right, so that the words tend to be recognized from left to right. The mass of this set is pc_i as defined in Eq. (3). Ω is all of the possible words. The mass of Ω is $1 - pc_i$, which represents the amount of evidence that is not committed to any word. Eq. (5) was used to combine evidence from a pair of sources. By using Eq. (5) repeatedly, the evidence from all of the four characters could be combined.

For example, evidence from the following two characters 美 and 好, which could constitute the word 美好 (which means goodliness) could be combined. The set of interest corresponding to 美 is

$$A = \{A1 = [\text{美好}, \text{美丽}, \text{美国} \ldots], A2 = \Omega\} \quad \text{and } m(A1) = pc_{\text{美}},\ m(A2) = 1 - pc_{\text{美}}$$

The set of interest corresponding to 好 is

$$B = \{B1 = [\text{美好}, \text{友好}, \text{良好} \ldots], B2 = \Omega\} \quad \text{and } m(B1) = pc_{\text{好}},\ m(B2) = 1 - pc_{\text{好}}$$

The combination of these two sources of evidence generates the set of interest,

$$C = \{C1 = A1 \cap B1 = [\text{美好}], C2 = A1 \cap A2 = A1 \cap \Omega = A1, C3 = A2 \cap B1 = \Omega \cap B1 = B1, C4 = A2 \cap B2 = \Omega \cap \Omega = \Omega\}$$

In the simulation, the corresponding calculations are:

$$m(C1) = m(A1) * m(B1)$$
$$m(C2) = m(A1) * m(B2)$$
$$m(C3) = m(A2) * m(B1)$$
$$m(C4) = m(A2) * m(B2)$$

In this example, as $pc_{美}$ and $pc_{好}$ increase, the belief corresponding to the word 美好 ($m(C_1)$) increases, and the belief of all of the other sets decreases. When the belief of a set that only includes one word reaches a threshold (fixed at 0.95 in simulation), the word is considered to be recognized. The result from the belief function described in Eq. (5) is used to produce a mass for the relevant word, which is then fed back to the character recognition level described in Eq. (6).

8.4. Attention module

The purpose of the attention module is to make sure that words are recognized in order from left to right. The attention module has a record of a reference location, corresponding to the first character of the next word. Because of the effects of the attention module, only the words whose first character matches the character at the reference location can be activated. When a word is recognized, the reference location transfers to the next character to the right of the recognized word. If the reference character is recognized, but the belief function of all of the words is low (below 0.2), it was assumed that the character is not part of any word. In this situation, the reference character transfers to the next character.

8.5. Feedback from word recognizer to character recognizer

The results at the word recognizer level provide evidence to the character recognizer level. The evidence character i receives from the word recognition level is

$$pt_i = \sum_{C_j \in C} m(C_j) * \frac{n_j}{N_j} \tag{6}$$

where n_j represents the number of words that include char_i; and N_j is the total number of all of the words in the set C_j.

8.6. Simulation[8]

Because Experiment 3 includes all of the conditions in Experiments 1, 2, and 4, the results of Experiment 3 were simulated. The mean predicted accuracies from 100 simulated subjects were compared with the observed accuracies of Experiment 3. The parameters that generated the least difference between the predicted data and experimental data were selected. The values of the selected parameters are shown in Table 3. For each simulated trial, the simulation ran for 100 iterations before the character that was recognized at each location was determined from the character recognition level described in Eq. (4).

Results are shown in Table 4. From the results, we can see that the model predicts the pattern of the results of Experiment 3 quite well. Characters were recognized very well at all of the four locations in the one-word condition. Characters were recognized very well at the first two character locations and relatively poorly at the third and fourth character locations in the two-word condition.

In the half-word condition, characters were recognized quite well at the first two character locations, and much worse at the third and fourth character locations. These last two locations were worse in the half-word condition than the non-word condition. This pattern suggests that the word representation is able to inhibit recognition of characters inconsistent with that word. This inhibition is pre-

[8] Source code can be found at http://www.xingshanli.com/Documents/wordSeg_code.rar.

Table 3
Parameters used in the model.

Parameter	Meaning	Value
β	Speed of information accumulation at the visual perception level	0.013
γ	Eccentricity factor	0.20
δ	Std dev of Gaussian noise	0.014
ω	Threshold of character recognition	0.95 (fixed)

Table 4
Observed accuracy of Experiment 3 and accuracy predicted by the model.

	Observed				Predicted			
	Ch1	Ch2	Ch3	Ch4	Ch1	Ch2	Ch3	Ch4
Two-word	.91	.86	.54	.52	.98	.84	.56	.43
One-word	.93	.93	.84	.86	.1.00	1.00	.97	1.00
Half-word	.92	.83	.17	.15	.99	.79	.26	.09
Non-word	.77	.63	.29	.29	.88	.63	.32	.10

dicted by our model, but not by the McClelland and Rumelhart model, which has no inhibitory connections from the word level to the letter level. In the non-word condition, recognition was worse than in the other three conditions at almost all of the locations (but better than the half-word condition at the third and fourth character locations). There is also a tendency toward a decrease in performance from left to right.

8.7. Model summary

The model does a good job of explaining the results we obtained in the present experiments. The following points explain how the model achieves its performance. First, in both the word recognizer module and the character recognition module, the word or character is recognized more efficiently when it receives consistent evidence from different sources. For example, the belief function of a word is higher if all of the characters provide evidence supporting it, as in the one-word condition. Second, feedback from the word level to the character level facilitates character recognition. The character recognizers in the one-word condition could receive feedback information, so they recognized characters faster, so that more characters were recognized in the third and fourth character locations. Third, the attention module makes sure that when two words are present, they are recognized serially.

The extent to which words are processed serially or in parallel has been debated for some time. In this model, characters are processed in parallel at the character recognition level, while words are recognized serially. Only one word can be recognized at a time. The attention module makes sure that words are lexically processed serially. This approach is consistent with that stance of some models of English reading, such as the E–Z Reader model (Reichle, Pollatsek, Fisher, & Rayner, 1998), and with the Chinese version of E–Z Reader (Rayner et al., 2007; see also Wu, Slattery, Pollatsek, & Rayner, 2008, for a model of Chinese word segmentation in the context of the E–Z Reader model).

8.8. Word segmentation in the model

The model reflects our understanding of the Chinese word segmentation problem. In this model, word recognition and word segmentation are not distinguishable. Only when the word is recognized is it segmented. The question of whether Chinese word segmentation or Chinese word recognition happens first is something like a chicken and egg problem. On the one hand, word recognition requires words to be segmented; on the other hand, word segmentation sometimes requires semantic information from words. This model demonstrates a possible solution to this problem.

In Chinese, word boundaries are sometimes ambiguous. The first one or two characters of a multiple-character word sometimes also constitute another word. For example, 老板娘 (which means landlady) is a word. The first two characters constitute another word (老板, which means boss). This kind of ambiguity sometimes causes confusion during reading. When facing this ambiguity, the model presented above predicts that the word with more characters is always more likely to be selected in the word recognition level. When the model sees the three characters 老板娘, the word recognition level tends to parse them as a 3-character word (老板娘) instead of a two-character word (老板), since the word 老板娘 receives extra evidence from the character "娘" in addition to the evidence from "老板". That is to say, the model always selects the longest possible word in text. This prediction is consistent with an experimental study reported by Wu, Slattery, Pollatsek, and Rayner (submitted for publication). Subjects read sentences containing either a high or low frequency 3-character target word (ABC) in which AB was also a word. First pass fixation time measures in region AB were influenced by the frequency of ABC. The results of that study suggested that the target word ABC was activated even though AB was also a word. In some reading situations, in which alternative segmentation is needed, higher level processing must be involved, which exceeds the scope of the current model.

The interactive activation model presented in the current article is not the only model that can account for the data reported in the current results. Other models, such as the Fuzzy Logical Model of Perception (FLMP; Massaro, 1979, 1989; Massaro & Cohen, 1991) could also explain the data we reported in this article. According to FLMP, the activity of word recognition does not influence the sensory activity of the character recognition level. Instead, in FLMP character recognition depends on the information from both the character activity and the word activity. The more information a decision system receives, the more likely that character is recognized. In the current experiments, the recognizers of the third and fourth characters in the one-word condition receive information from both the word level and the character level; they should be recognized more readily than in the two-word condition, in which they could only receive information from the character recognition level. The interactive account and the independent account share the view that character recognition could be influenced by the word recognition level. The difference between these two kinds of models is in the way through which word recognition influences the perception of characters. Interactive models, such as FLMP, assume that word activity can influence the sensory information in character recognition, while independent models, such as FLMP, assume that the word recognition can only influence the decision process. Either model predicts that word knowledge can influence the accuracy of character report.

The experimental data from the current study cannot distinguish between these models. The IA Model and the FLMP model agree that processing at the word recognition level can influence character recognition. The IA model claims that the interactive procedure is continuous and can influence the character recognition level before the decision stage, while the FLMP model claims that the interaction occurs in the decision stage. Based on the current experiments, we concluded that word recognition and segmentation was not a simple feed-forward process; instead, processing at the word segmentation level influenced character recognition. Our conclusions from the experimental data are consistent with the common ground between the IA model and the FLMP model. We choose to use the interactive assumption in the model presented above because (1) there are plenty of feedback connections from the high-level processing areas to lower level areas (V1, V2; Crick & Asanuma, 1986; McClelland, Mirman, & Holt, 2006), and (2) Li and Logan (2008) found that knowledge could define an object that could influence attentional deployment, which suggested that knowledge could influence perception. As with these English word recognition studies, more work is needed to distinguish between these two categories of models.

There are some other factors that potentially influence Chinese word segmentation that were not implemented in the model. First, we did not consider phonology. Perfetti, Liu, and Tan (2005) described a model of Chinese single-character word naming that demonstrated the important role of phonology in word naming. It is possible that phonology also plays a role in word segmentation and recognition in multiple-character words. Secondly, as we demonstrated in Experiment 5, contextual information likely influences Chinese word segmentation and word recognition. However, we have not included the influence of contextual information (such as semantic processing) on word segmentation in the current version of the model. In Chinese, word segmentation can in some cases not be settled without semantic information. The following sentence provides one example: 鲜花生长在后院里. (It can be understood as "Flower is grown in the yard", or "the peanut is grown in the yard" depending on different semantic context.) The four characters 鲜花生长 could be segmented into (鲜花)(生长) or (鲜花生)(长) depending on semantic information. In these rare conditions, Chinese word segmentation has to depend on semantic processing. Finally, we did not consider alternative ways to deal with time restrictions in performing the recognition task. The exposure durations in all of the experiments reported in this paper were limited so that subjects could not report all of the characters. In the current version of the model, we chose parameters under the assumption that the model would not have enough time to recognize all of the characters. However, subjects might use special strategies when facing time pressure (Kello & Plaut, 2003; Waters & Seidenberg, 1985). For example, subjects might choose to report the most active character or word even when its activity has not reached the threshold under time pressure. More experimental studies and modeling studies are necessary to address this question. Compared to English reading, Chinese reading is not well understood. We believe that the model presented in the current article enhances that understanding even with these limitations.

Acknowledgments

This research was supported by Grant HD26765 from the National Institute of Health and by a grant from Microsoft. We thank Tom Carr, Chuck Clifton, Albrecht Inhoff, Simon Liversedge, Gordon Logan, Robert Goldstone, and Jennifer Stolz for their advice and helpful suggestions on an earlier draft of the paper, and we especially thank Albrecht Inhoff for suggesting Experiment 5.

References

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*, 1277–1288.

Brown, T. L., Gore, C. L., & Carr, T. H. (2002). Visual attention and word recognition in stroop color naming: Is word recognition "automatic"? *Journal of Experimental Psychology: General, 131*, 220–240.

Carr, T. H., & Pollatsek, A. (1985). Recognizing printed words: A look at current models. In D. Besner, T. G. Waller, & G. E. MacKinnon (Eds.), *Reading research: Advances in theory and practice* (pp. 1–82). Orlando, FL: Academic Press.

Cattell, J. M. (1886). The time it takes to see and name objects. *Mind, 11*, 63–65.

Cepeda, N. J., Cave, K. R., Bichot, N. P., & Kim, M. S. (1998). Spatial selection via feature-driven inhibition of distractor locations. *Perception & Psychophysics, 60*, 727–746.

Chen, J. Y. (1999). Word recognition during the reading of Chinese sentences: Evidence from studying the word superiority effect. In J. Wang, A. W. Inhoff, & H. C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 239–256). Mahwah, NJ: Erlbaum.

Cowan, N. (2000). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Crick, F. H. C., & Asanuma, C. (1986). Certain aspects of the anatomy and physiology of the cerebral cortex. In J. L. McClelland & D. E. Rumelhart (Eds.). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2, pp. 333–371). Boston, MA: MIT press.

Duncan, J. (1984). Selective attention and the organization of visual information. *Journal of Experimental Psychology: General, 113*, 501–517.

Egly, R., Driver, J., & Rafal, R. D. (1994). Shifting visual-attention between objects and locations: Evidence from normal and parietal lesion subjects. *Journal of Experimental Psychology: General, 123*, 161–177.

Healy, A. F. (1976). Detection errors on word the – Evidence for reading units larger than letters. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 235–242.

Healy, A. F. (1994). Letter detection – A window to unitization and other cognitive processes in reading text. *Psychonomic Bulletin & Review, 1*, 333–344.

Hoffman, J. E., Nelson, B., & Houck, M. R. (1983). The role of attentional resources in automatic detection. *Cognitive Psychology, 15*, 379–410.

Hsu, S. H., & Huang, K. C. (2000a). Effects of word spacing on reading Chinese text from a video display terminal. *Perceptual and Motor Skills, 90*, 81–92.

Hsu, S. H., & Huang, K. C. (2000b). Interword spacing in Chinese text layout. *Perceptual and Motor Skills, 91*, 355–365.

Inhoff, A. W., Liu, W., Wang, J., & Fu, D. J. (1997). Use of spatial information during the reading of Chinese text. In D. L. Peng, H. Shu, & H. C. Chen (Eds.), *Cognitive research on Chinese language* (pp. 296–329). Beijing: Shan Dong Educational Publishing.

Inhoff, A. W., & Wu, C. L. (2005). Eye movements and the identification of spatially ambiguous words during Chinese sentence reading. *Memory & Cognition, 33*, 1345–1356.

Johnston, J. C. (1978). A test of the sophisticated guessing theory of word perception. *Cognitive Psychology, 10*, 123–154.

Kello, C. T., & Plaut, D. C. (2003). Strategic control over rate of processing in word reading: A computational investigation. *Journal of Memory and Language, 48*, 207–232.

Kim, M. S., & Cave, K. R. (1995). Spatial attention in visual search for features and feature conjunctions. *Psychological Science, 6*, 376–380.

Kramer, A. F., Weber, T. A., & Watson, S. E. (1997). Object-based attentional selection – Grouped arrays or spatially invariant representations?: Comment on Vecera and Farah (1994). *Journal of Experimental Psychology: General, 126*, 3–13.

LaBerge, D. (1983). Spatial extent of attention to letters and words. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 371–379.

LaBerge, D., & Samuels, S. J. (1974). Toward a theory of automatic information processing in reading. *Cognitive Psychology, 6*, 293–323.

Li, X. S., & Logan, G. D. (2008). Object-based attention in Chinese readers of Chinese words: Beyond Gestalt principles. *Psychonomic Bulletin & Review, 15*, 945–949.

Logan, G. D. (1994). Spatial attention and the apprehension of spatial relations. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1015–1036.

Massaro, D. W. (1979). Building and testing models of reading processes. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 111–146). New York: Longman.

Massaro, D. W. (1989). Test between the TRACE model and the fuzzy logical model of speech perception. *Cognitive Psychology, 21*, 398–421.

Massaro, D. W., & Cohen, M. M. (1991). Integration versus interactive activation: The joint influence of stimulus and context in perception. *Cognitive Psychology, 23*, 558–614.

McCann, R. S., Folk, C. L., & Johnston, J. C. (1992). The role of spatial attention in visual word-processing. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 1015–1029.

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Science, 10*, 363–369.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception. 1. An account of basic findings. *Psychological Review, 88*, 375–407.

National Language Committee (1997). Chinese dictionary: Character and word frequency statistics report, Taiwan (in Chinese).

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*, 43–59.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*, 56–115.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Prinzmetal, W., Hoffman, H., & Vest, K. (1991). Automatic processes in word perception – An analysis from illusory conjunctions. *Journal of Experimental Psychology: Human Perception and Performance, 17*, 902–923.

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review, 12*, 1089–1093.

Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E–Z reader model of eye movement control to Chinese readers. *Cognitive Science, 31*, 1021–1034.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology, 81*, 275–280.

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.

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review, 12*, 119–124.

Robertson, L. C., & Treisman, A. (2006). Attending to space within and between objects: Implications from a patient with Balint's syndrome. *Cognitive Neuropsychology, 23*, 448–462.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96*, 523–568.

Shafer, G. (1976). *A mathematical theory of evidence*. Princeton, New Jersey: Princeton University Press.

Stolz, J. A., & McCann, R. S. (2000). Visual word recognition: Reattending to the role of spatial attention. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 1320–1331.

Tao, L., Healy, A. F., & Bourne, L. E. (1997). Unitization in second-language learning: Evidence from letter detection. *American Journal of Psychology, 110*, 385–395.

Tsal, Y., & Lavie, N. (1988). Attending to color and shape – The special role of location in selective visual processing. *Perception & Psychophysics, 44*, 15–21.

Waters, G. S., & Seidenberg, M. S. (1985). Spelling-sound effects in reading time-course and decision criteria. *Memory & Cognition, 13*, 557–572.

Wheeler, D. (1970). Processes in word recognition. *Cognitive Psychology, 1*, 59–85.

Wu, J., Slattery, T. J., Pollatsek, A., & Rayner, K. (submitted for publication). Eye movement control and Chinese word segmentation.

Wu, J., Slattery, T. J., Pollatsek, A., & Rayner, K. (2008). Word segmentation in Chinese reading. In K. Rayner, D. Shen, X. Bai, & G. Yan (Eds.), *Cognitive and cultural influences on eye movements* (pp. 303–314). Tianjin: Tianjin Peoples Press/Psychology Press.

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