CrossMark

Skipping of Chinese characters does not rely on word-based processing

Nan Lin[1,2] · Bernhard Angele[3] · Huimin Hua[1,2] · Wei Shen[1,2] · Junyi Zhou[1,2] ·
Xingshan Li[1,2]

Abstract Previous eye-movement studies have indicated that people tend to skip extremely high-frequency words in sentence reading, such as "the" in English and "的/de" in Chinese. Two alternative hypotheses have been proposed to explain how this frequent skipping happens in Chinese reading: one assumes that skipping happens when the preview has been fully identified at the word level (word-based skipping); the other assumes that skipping happens whenever the preview character is easy to identify regardless of whether lexical processing has been completed or not (character-based skipping). Using the gaze-contingent display change paradigm, we examined the two hypotheses by substituting the preview of the third character of a four-character Chinese word with the high-frequency Chinese character "的/de", which should disrupt the ongoing word-level processing. The character-based skipping hypothesis predicts that this manipulation will enhance the skipping probability of the target character (i.e., the third character of the target word), because the character " /de" has much higher character frequency than the original character. The word-based skipping hypothesis instead predicts a reduction of the skipping probability of the target character because the presence of the character "的/de" is lexically infelicitous at word level. The results supported the character-based skipping hypothesis, indicating that in Chinese reading

the decision of skipping a character can be made before integrating it into a word.

In sentence reading, people do not fixate on every single word. Word skipping, i.e., moving one's gaze past a word during first-pass reading without fixating it, happens frequently in both alphabetic reading (Carpenter & Just, 1983; Rayner, 1998) and Chinese reading (Inhoff & Liu, 1998). How do readers decide whether to skip a word or not? In alphabetic reading, word-skipping probability is modulated by the length and frequency of the previewed word and the predictability of the upcoming word (Abbott, Angele, Ahn, & Rayner, 2015; Angele, Laishley, Rayner, & Liversedge, 2014; Angele & Rayner, 2013; Rayner, 1998). Even though the effect of a word's predictability on the probability of skipping is well established, Angele and Rayner (2013) found that, at least in English, the influence of contextual information on skipping is rather weak and can be easily overridden by preview information. Using the gaze-contingent display change paradigm (Rayner, 1975), Angele and Rayner (2013) manipulated the preview that was available for a three-letter target word. When the upcoming word appeared to be the definite article "the" (a function word with the highest word-frequency in English vocabulary) readers skip it frequently even if the syntactic sentence context does not allow for an article in that position in the sentence (e.g., "the" as the preview for "ban" in: "The council voted to ban smoking immediately near public buildings"). A follow-up study (Angele et al., 2014) has found a similar effect for skipping previews of short high-frequency content words in infelicitous positions (e.g., "sad" as a preview for "sew" in: "She would always sew her own dresses").

✉ Xingshan Li
lixs@psych.ac.cn

1 CAS Key Laboratory of Behavioral Science, Institute of Psychology, Beijing, China

2 Department of Psychology, University of Chinese Academy of Sciences, Beijing, China

3 Psychology Research Centre, Faculty of Science and Technology, Bournemouth University, Poole, UK

 Springer

Furthermore, it seems that preview information is processed completely independently from contextual information: Abbott et al. (2015) found that manipulating the semantic constraint of the sentence (i.e., the predictability of the word in whose position the preview of "the" appeared; e.g., "die" in "If you are shot in the head you will likely die" vs. "If you are shot in the foot you will likely die") does not reduce skipping of words that look like an infelicitous occurrence of "the." In fact, the effect of predictability on skipping that Abbott et al. (2015) found was completely independent from the "the"-skipping effect. This led to the somewhat counterintuitive result that readers in Abbott et al.'s (2015) experiment were more likely to skip an infelicitous "the"-preview if the context up to the position of the "the"-preview was highly predictive of the actual target word (which was never "the") than if the actual target word was of low predictability. These findings indicate that, in quite a few cases, readers make their word-skipping decision before integrating a previewed word into the sentence, while taking the contextual constraint into account separately (and without checking if the word at the target position actually fit the constraint).

Skipping also happens frequently in Chinese reading. Previous studies have highlighted several characteristics of skipping of Chinese characters and words. First, Chinese readers do not fixate every character. The probability of fixating one particular character in a sentence is approximately 42.8%, and the average saccade length is 3.15 characters (Li, Bicknell, Liu, Wei, & Rayner, 2014). Second, words comprising two or more characters are also sometimes skipped as a whole, and longer words are skipped less often than short words. For example, Li, Liu, and Rayner (2011) showed that the skipping probability for two-character words was 0.34, whereas the skipping rate for four-character words was 0.07. Third, all characters in a word have similar fixation probabilities during first pass reading independent of their position within the word (Li et al., 2011; Li et al., 2014). Fourth, the reason why Chinese readers do not fixate every single character seems to be that Chinese readers can process more than one character during each fixation. It has been shown that the perceptual span in Chinese reading comprises one character to the left and two to three characters to the right of fixation (Inhoff & Liu, 1997, 1998; Tsai & McConkie, 1995).

Although we know about many aspects of word/character skipping during Chinese reading, the exact mechanism of how to determine when to skip a word or a character is still not clear. Interestingly, a recent study of Chinese reading (Zang et al., in press) has obtained some findings highly similar to those of Angele and Rayner (2013). In their study, Zang et al. investigated whether the preview of a single-character word—" /de"—which is a function word (more precisely, a grammatical particle, indicating, among other things, ownership) and the word with the highest word-frequency in the Chinese vocabulary, can enhance skipping probability. They found that, just as occurs with "the" in English reading, a preview of "的/de" increased the probability of skipping the character, even

when it occurs in a position where it violates syntactic constraints from the sentence context.

What triggers this "automatic" skipping response? One possibility is that readers might process the preview word lexically and then make decision whether it will be necessary to fixate that word to achieve full lexical access. This would be predicted by the E-Z Reader model (Reichle, Pollatsek, & Rayner, 2006). The E-Z reader model assumes that words are processed serially, and the decision on when to move the eyes and where to move the eyes are directly determined by different stages of word processing. The completion of the "familiarity check" stage L1 on the currently fixated word (word n) will initiate a saccade program to move the eyes to next word (word n +1). The currently fixated word will then finish lexical access (stage L2) concurrently with the planning of the upcoming saccade. Attention will shift to word n + 1 if word n can be fully identified during this stage. Therefore, attention is likely to be shifted to word n + 1 when the eyes are still on word n and the lexical processing of word n + 1 is likely to begin (again with the familiarity check stage L1) while the eyes are still on word n. The eye movement control module needs some time to program and prepare the saccade after they receive the saccade program signal (stages M1 and M2). In some situations, the familiarity check (L1) of word n +1 finishes quickly enough that saccade programming still is at a stage where the saccade to word n+1 (now no longer necessary) can still be cancelled (i.e., before the end of stage M1). In this situation, a new saccade to word n+2 (the word to the right of word n+1) is prepared and, as a consequence, word n + 1 is skipped.

Essentially, the oculomotor system "bets" that it will be able to finish identifying the current word before the saccade preparation is complete. If both the currently fixated word and the upcoming word are easy to identify, the model predicts that there will be enough time to identify both words and move attention to the subsequent word. If this happens early enough, the saccade to the upcoming word will be cancelled and a skipping saccade programmed instead. We will refer to this hypothesis as the *word-based skipping hypothesis*.

Another hypothesis has been proposed to explain how skipping happens, especially in Chinese reading. Because there are no inter-word spaces to mark word boundaries, Chinese readers usually do not target any specific position within a word (Li et al., 2011). Instead, some studies have suggested that Chinese readers might use a processing-based strategy to plan their eye movements. During any given fixation, they try to process as much information as possible foveally and parafoveally and then plan their next eye movements to obtain information beyond what they have already processed (Li et al., 2011; Wei, Li, & Pollatsek, 2013). Thus, if the preview character is easy to process (e.g., a high-frequency character, such as "的/de"), the resulting saccade is likely to be long, increasing the chance that the preview character may be skipped. Instead of emphasizing the role of words, the processing-based strategy assumes that the amount

of information that has been processed for upcoming characters is what determines where to move the eyes. We will refer to this possibility as the *character-based skipping hypothesis*.

How can we distinguish between the word-based skipping and character-based skipping hypotheses? If the character-based skipping hypothesis is true, characters that are easy to recognize (e.g., characters with low visual complexity or high frequency) should be skipped more often than those that are difficult to recognize, no matter whether the character to be skipped is a word on its own, such as in all the previously reported experiments, or if it is embedded in another word. However, if the decision to skip a character is based on successful word-level processing, embedding the character in a longer word should lower the probability that it is skipped, because its presence should affect the processing of the embedded word.

In the present study, we examined how the preview of the extremely high-frequency character "的/de" embedded in a multi-character Chinese word affects the skipping probability. We chose four-character words as target words, and we used a gaze contingent display change paradigm in which participants were asked to read sentences on a computer screen. Before the eyes crossed an invisible boundary located to the left of the target character (the third character of the target word), the preview was substituted with the character "的/de." After the eyes crossed the boundary, the third character changed to the correct character as in the target word. The above design allowed us to examine whether the presence of a high frequency preview character would make the skipping probability increase when it resulted in lexical violation at word level. We expected that if Chinese readers decide to skip a previewed character based on character level processing and before integrating it into a word, as assumed by the character-based skipping hypothesis, then the previewed "的/de" may make the skipping probability increase; otherwise, if the skipping decision is on the premise of successful word-level processes associated with the preview character, as assumed by the word-based skipping hypothesis, the lexical violations associated with the previewed "的/de" would result in a reduction of skipping probability as readers should fixate, rather than skip, a part of the sentence that is difficult to process.

A key point of the above experimental design was to ensure that the presence of the incorrect preview character "的/de" would result in lexical violation at word level. In this context, it should be noted that the majority of Chinese characters can be used either as a component of a multi-character word or as a single-character word. For example, "方针" ("fangzhen") is a two-character word that means "policy," whereas its component characters can both be used as single-character words: "方" ("fang") means "square," and "针" ("zhen") means "needle." Therefore, if the character "的/de" is embedded into a short target word (e.g., "方的针"), readers may treat the remaining character(s) of the target word and the embedded

character "的/de" as independent single-character words (e.g., "的针" can be interpreted as "a square needle"), which would possibly constitute a semantic or syntactic violation but not necessarily a lexical violation. To minimize this problem, we chose to embed the character "的/de" in a long word where the remaining characters and the embedded character "的/de" hardly can be perceived as single-character words.

We chose four-character words but not longer words as our target words for two reasons. First, according to the Chinese Linguistic Data Consortium (2003) corpus, 2.8% of words are one-character words, 63.9% are two-character words, 17.5% are three-character words, 14.2% are four-character words, and only 1.7% are longer than four characters. Therefore, a typical Chinese word consists of no more than four characters.[1] Second, previous studies have provided adequate evidence that four-character words are represented in Chinese readers' mental lexicon. For example, Li, Rayner, and Cave (2009) found that, for a four-character string, Chinese readers can recognize its third and fourth characters much faster when they constitute a word than when they constitute two unrelated words or a congruent two-word phrase. Gu and Li (2015) found that it is more difficult to detect the transposition between the second and third characters of a four-character string when it is a four-character word than when it is a phrase consisting of two two-character words. Therefore, we used four-character words as target words in the current study.

We chose only four-character words whose first two characters cannot form a word on their own as our target words and substituted the preview of their third characters with the character "的/de." This effectively made an interpretation of "的/de" as a function word impossible. Within this design, the first, second, and last characters of the target word provide rich clues to prime the representation of the target word, minimizing the possibility that readers treat them as independent single-character words. The character "的/de" is embedded in the middle of the target word, preventing it from being perceived as a single-character word preceding or following the target word. The first two characters cannot form a word either by themselves or together with the following character "的/de," ensuring that the presence of the character "的/de" will immediately result in lexical violations at word level even before the last character of the target word is recognized.

Because there is evidence that the probability of skipping a character also is affected by its visual complexity (Liversedge et al., 2014; Ma & Li, 2015) and its grammatical properties (e.g., a function word might be skipped more often than a content word; Carpenter & Just, 1983; cf. Angele et al.,

---

[1] A large portion of four-character Chinese words are four-character idioms. Their role and usage as words are well-accepted. For example, they are included as words in the Contemporary Chinese Dictionary (Chinese Academy of Social Sciences, 2012), which is the most authoritative Chinese language dictionary. Therefore, we did not differentiate four-character idioms from the other four-character words in the present study.

2014), we also added two control conditions to examine whether the relatively low visual complexity and the grammatical properties of "的/de" enhance the skipping probability. Given that the number of function words in Chinese is very limited, there is no other character that matches "的/de" on both the visual and grammatical dimensions. Therefore, we used two separate control conditions. The characters "钓/diao" serves as a control condition that closely matches the subcharacter visual properties of "的/de": They have the same stroke numbers and share the same right radical (i.e., "勺"). The character "乎/hu" serves as another control condition in that its grammatical properties are similar to those of "的/de" (although much less frequent). Both "乎/hu" and "的/de" can be used as single-character function words and can be used as auxiliaries as well as modal particles (Sun, Huang, Sun, Li, & Xing, 1997). The inclusion of the control conditions can help distinguish the two target hypotheses: According to the word-based skipping hypothesis, the "钓/diao" and "乎/hu" previews, such as the "的/de" preview, should disrupt word-level processing and result in a reduction of skipping probability. According to the character-based skipping hypothesis, if the "的/de" preview is skipped more often due to its extremely high frequency but not because of its visual complexity or its grammatical properties, then the skipping probability of the "钓/diao" and "乎/hu" previews should not be significantly different from that of the unsubstituted previews.

Methods

Participants

Twenty-eight students (17 females) from colleges around the Institute of Psychology, Chinese Academy of Sciences participated in the experiment. All participants were native Chinese speakers with normal or corrected-to-normal vision and were naive to the purpose of the experiment.

Apparatus

Participants' eye movements were recorded using an SR Research EyeLink 1000 eye tracker with a sampling rate of 1,000 Hz. Sentences were displayed on a 21-inch CRT monitor with a refresh rate of 150 Hz. Each character subtended a visual angle of approximately 1°.

Materials

Participants read 80 experimental sentences with four experimental conditions. The sentences were no more than 32 characters (range 18-32) and were all single-line sentences (Table 1). Each experimental sentence contained a four-character target word (located at least 5 characters away from

Table 1 An example of experimental stimuli of the four preview conditions

Preview conditions	Character frequency of the preview character	Stroke number of the preview character	Example sentences
Identity	192.5 ± 165.0	8.0 ± 2.4	大学里，几位道貌岸 的 教授对于迟到学生总是白眼相待。
(de)	22797 ± 0	8 ± 0	大学里，几位道貌的 的 教授对于迟到学生总是白眼相待。
(diao)	23 ± 0	8 ± 0	大学里，几位道貌钓 的 教授对于迟到学生总是白眼相待。
(hu)	162 ± 0	6 ± 0	大学里，几位道貌乎 的 教授对于迟到学生总是白眼相待。

The character frequency and stoke number are presented as mean ± standard deviation. Characters in the target region are shown in bold. The English translation of the example sentence is: In the university, a few sanctimonious professors are very scornful to the students who are late. The vertical black line represents an invisible boundary. When the eyes crossed the boundary, the preview character was replaced by the target character

the beginning and the end of the sentence). Frequencies of target words in the Chinese Linguistic Data Consortium (2003) corpus ranged from 0.05 to 6.30 occurrences per million ($M = 0.87$; $SD = 0.88$).

There were four preview conditions for each target word. The first condition was the identity condition, in which the preview was identical to the target word (e.g., 道貌岸然 as the preview of 道貌岸然, which means "sanctimonious"). The second condition was the "的/de" condition, in which the preview of the third character of the target word (i.e., the target character) was substituted with the character "的/de" (e.g., 道貌的然 as the preview of 道貌岸然). The third and fourth conditions were control conditions, in which the preview of the target character was substituted with the characters "钓/diao" (e.g., 道貌钓然 as the preview of 道貌岸然) and "乎/hu" (e.g., 道貌乎然 as the preview of 道貌岸然), respectively. The characters in the control conditions were chosen to investigate whether the characteristics of "的/de" other than its extremely high frequency, specifically its visual complexity (e.g., number of strokes) and its grammatical properties (e.g., a character that is very frequently used as a function word), may enhance skipping. The control conditions are very similar to "的/de" in most aspects, but there are some small differences (e.g., "钓/diao" is a verb and "乎/hu" is visually simpler than "的/de"). Because of this, the findings of the control conditions should be interpreted with caution.

We selected the four-character target words from the Chinese Linguistic Data Consortium (2003) corpus considering the following factors: First, we ensured that the first two

characters of each target word cannot constitute a two-character word so that substituting a different character for the target character would always result in a lexical violation. Second, the mean stroke number of the target characters (M = 7.95; SD = 2.41) was matched to the number of strokes of the character "的/de" (which is 8). Third, the mean character frequency (occurrences per million) of the target characters (M = 192.51; SD = 164.96) was much lower than the character frequency of the character "的/de" (which is 22,797) and no lower than that of the character "钓/diao" or "乎/hu" (which are 23 and 162, respectively).

Procedure

Following a three-point calibration procedure, participants were asked to read each sentence silently for comprehension. Participants read 7 practice sentences and then 80 experimental sentences intermixed with 80 filler sentences. The gaze-contingent boundary paradigm was used to present the stimuli. When the eyes of participants crossed an invisible boundary located left of the target character, the preview character was replaced by the target character on the screen. After the experiment, participants were asked whether or not they noticed anything unusual during this experiment. No participant reported that they noticed the boundary manipulation during the experiment. Among the experimental and filler sentences, 54 sentences were followed by a forced-choice question to ensure that participants read the sentences carefully.

Results

The accuracy of the comprehension questions was high (96.2%), indicating that participants read and understood the sentences well. We eliminated 2.8% of the trials for blinks or track loss, 22.2% of the trials for display changes that completed either late (after fixation onset) or extremely early (50 ms earlier than fixation onset, which is quite rare and may be triggered by very infrequent long saccades), and 3.3% of the trials that were triggered by saccades that terminated to the left of the boundary ("j-hooks"). In sum, 28.3% of the data were eliminated and data from a total of 1,606 trials were included in the following analysis.

We analyzed skipping probability and fixation times on the target character and the pretarget region (the two characters before the boundary). In addition, the launch site of the saccades that landed at or passed the target character was calculated. The analyses were conducted using linear mixed models (LMMs), which were fitted using the *lmer* function from the lme4 package (Version 1.1-7) in the R Environment for Statistical Computing (Version 3.3.1). The preview condition was included as a fixed factor and three treatment contrasts ("的/de" vs. identity; "钓/diao"

vs. identity; "乎/hu" vs. identity) were used to test for the preview effect. Before reporting specific LMM preview contrasts, we conducted an omnibus likelihood-ratio test for the effect of preview to examine the significance of the main effect. We reported the three contrasts between identity preview condition and other preview conditions only when the main effect was significant. For skipping probability, we report regression coefficients (b), standard errors (SE), and z-values from generalized LMMs using a logit-link. For *first fixation duration* (*FFD*; the duration of the first-pass fixation on the target character), *gaze duration* (*GD*; the sum of all fixations on the target character before leaving it, including refixations), and *launch site* (the distance between the last fixation before saccading to the target character and the left side of the target character), we calculated regression coefficients (b), standard errors (SE), and t-values from LMMs on the basis of raw fixation time data. We ran LMMs with participant and item as crossed random effects. Including random slopes for the preview condition over participants and/or items led to convergence failures for most of the dependent variables and did not lead to a change in the pattern of effects observed for those dependent variables where convergence did occur. We therefore report models with random intercepts for participants and items only. We do not report p values, because it is not clear how to compute the degrees of freedom for t-values testing the effect of coefficients in LMMs. However, because we have relatively high numbers of subjects and items, the t-values should be approximately normally distributed and can be interpreted like z-values. Therefore, we use the same two-tailed criterion $|z|$ or $|t| \geq 1.96$ for z and t-values to test for significance test at the 0.05 α-level. The mean skipping probability, fixation times, and launch site of the four preview conditions are shown in Table 2, and the results of the LMM analyses are shown in Table 3.

Pretarget region

The result of the omnibus test on FFD, expressed as a chi-squared statistic, showed that the model including the preview condition as a factor provided a significantly better fit to the data than the model not including it, $\chi^2(3) = 11.51$, $p = 0.009$. However, the contrasts between identity preview condition and other preview conditions showed no significant difference ("的/de" vs. identity: 259 ms vs. 253 ms, $b = 7.65$, $SE = 6.71$, $t = 1.14$; "钓/diao" vs. identity: 243 ms vs. 253 ms, $b = -9.17$, $SE = 6.73$, $t = -1.36$; 乎/hu" vs. identity: 262 ms vs. 253 ms, $b = 12.37$, $SE = 6.64$, $t = 1.86$). The results of the omnibus test on skipping probability and GD were nonsignificant (skipping probability: $\chi^2(3) = 1.62$, $p = 0.66$; GD: $\chi^2(3) = 7.42$, $p = 0.06$).

Table 2 Condition means for the pretarget region and the target character

	Type of preview			
	Identity	(de)	□(diao)	(hu)
Pretarget region				
Skip probability	21% (3.2%)	24% (3.2%)	24% (4.1%)	21% (3.2%)
First fixation duration	253 (8.13)	259 (8.74)	243 (8.65)	262 (8.01)
Gaze duration	282 (11.96)	285 (10.26)	263 (10.46)	280 (10.48)
Target character				
Skip probability	58%(4.2%)	65%(2.8%)	58%(3.7%)	61%(3.4%)
First fixation duration	223(7.00)	232(8.46)	245(8.11)	243(9.64)
Gaze duration	224(7.12)	232(8.46)	253(10.80)	247(9.54)
Launch site	2.53(0.16)	2.71(0.16)	2.50(0.17)	2.45(0.13)

Standard errors are shown in parentheses

Target character

The results of the omnibus test on skipping probability, FFD, and GD were all significant (skipping probability: $\chi^2(3)$ = 10.90, p = 0.012; FFD: $\chi^2(3)$ = 16.49, p < 0.001; GD: $\chi^2(3)$ = 9.80, p = 0.020), whereas the result on launch site was nonsignificant, $\chi^2(3)$ = 5.26, p = 0.154. Readers skipped the previews of "的/de" more often than the identical previews (0.65 vs. 0.58; b = 0.37, SE = 0.15, z = 2.44). The other preview conditions showed no significant difference from the identity condition ("钓/diao" vs. identity: 0.58 vs. 0.58; b = 0.04, SE = 0.15, z = 0.24; "乎/hu" vs. identity: 0.61 vs. 0.58; b = 0.23, SE = 0.15, z = 1.53). To account for any possible effects of target word frequency, we conducted a supplementary analysis with both target-word frequency and the preview manipulation as fixed effects. The result pattern was highly consistent with that of the main analysis ("的/de" vs. identity: 0.65 vs. 0.58; b = 0.41, SE = 0.15, z = 2.68; "钓/diao" vs.

identity: 0.58 vs. 0.58; b = −0.01, SE = 0.15, z = −0.07; "乎/hu" vs. identity: 0.61 vs. 0.58; b = 0.26, SE = 0.15, z = 1.70).

In all display change experiments where the preview condition strongly influences skipping probability, it is important to keep in mind that the fixation time measures on the target word or character reflect only the (minority of) cases where the target word or character was fixated and not skipped. When the target character was fixated, readers' FFD and GD on the target character were shorter in the identity condition than in the "钓/diao" and "乎/hu" conditions (FFD: "钓/diao" vs. identity: 245 ms vs. 223 ms; b = 31.94, SE = 8.34, t = 3.83; "乎/hu" vs. identity: 243 ms vs. 223 ms; b = 25.12, SE = 8.63, t = 2.91; GD: "钓/diao" vs. identity: 253 ms vs. 224 ms; b = 31.99, SE = 8.53, t = 3.75; "乎/hu" vs. identity: 247 ms vs. 224 ms; b = 29.53, SE = 8.82, t = 3.35). There was no significant difference between the " /de" and identity preview conditions in FFD and GD (FFD: 232 ms vs. 223 ms, b = 11.15, SE = 8.80, t = 1.27; GD: 232 ms vs. 224 ms, b = 9.72, SE = 9.00, t = 1.08).

Table 3 Linear mixed-effects model analyses for the pretarget region and the target character

	SP			FFD			GD			Launch site		
	b	SE	z	b	SE	t	b	SE	t	b	SE	t
Pretarget region												
de vs. identity	0.11	0.18	0.64	7.65	6.71	1.14	5.17	9.43	0.55	-	-	-
diao vs. identity	0.15	0.18	0.82	-9.17	6.73	-1.36	-18.63	9.45	**-1.97**	-	-	-
hu vs. identity	-0.05	0.18	-0.30	12.37	6.64	1.86	1.85	9.36	0.20	-	-	-
Target character												
de vs. identity	0.37	0.15	**2.44**	11.15	8.80	1.27	9.72	9.00	1.08	0.16	0.10	1.56
diao vs. identity	0.04	0.15	0.24	31.94	8.34	**3.83**	31.99	8.53	**3.75**	-0.04	0.09	-0.43
hu vs. identity	0.23	0.15	1.53	25.12	8.63	**2.91**	29.53	8.82	**3.35**	-0.06	0.10	-0.63

Each column represents a model fit to one of the dependent variables

SP skipping probability; *FFD* first fixation duration; *GD* gaze duration; *b* regression coefficient; *SE* standard error; *t* or *z* test statistic (*b/SE*)

Cells marked in bold represent $|t| \geq 1.96$

Discussion

We investigated the influence of substituting the preview of the third character of a four-character word with the character " /de" on skipping probability. The preview of the character " /de" significantly enhanced the skipping probability compared with the unsubstituted preview, despite its appearance in the preview resulting in lexical violations at word level. This suggests that readers may decide to skip a previewed character mainly based on how easily it can be processed at character level. Word-level processing does not seem to play a large role in skipping decisions, if any.

Our results support the character-based skipping hypothesis and did not support the word-based skipping hypothesis. The word-based skipping hypothesis assumes that the skipping decision is based on successful word-level processing associated with the preview character. In the experimental preview conditions of the current study, substituting the target character with the unrelated characters "的/de," "钓/diao," or "乎/hu" should disrupt the processing of the target word, which, according to the word-based skipping hypothesis, should result in a reduction of skipping probability. However, this was not the case. Readers showed higher skipping probability in the "的/de" condition than in the identity condition and showed no significant difference in skipping probability between the identity, " /diao," and "乎/hu" conditions.

The results are consistent with the predictions of the "character based" skipping hypothesis. The character-based skipping hypothesis assumes that Chinese readers decide to skip a previewed character based on character level processing and before integrating it into a word. Thus, they predicted that a character will be more likely to be skipped the easier it can be processed (just as the character "的/de" in the current study). This is exactly what we observed in the current study.

The results of the current study is consistent with the processing-based hypothesis of eye movement control (Li et al., 2011; Wei et al., 2013), which assumes that Chinese readers plan their saccades to process as much information as possible at a given fixation and move their eyes beyond the recognized characters. Thus, if the characters in the parafovea are easy to process, readers should be able to process more characters and, as a consequence, make longer saccades. In the current study, the character "的/de" is easy to process, thus it is more likely to be processed with parafoveal vision and thus is more likely to be skipped.

Our results also provided evidence on the relationship between parafoveal processing and expectations of what the next character should be based on lexical representations (the character-level expectation) in terms of skipping decisions: We found that automatic skipping of "的/de" previews occurred even though "的/de" was incompatible with the character-level expectation based on the preceding characters of a same word. In our experiment, the preceding characters were highly predictive of the target character—in fact, for each target word, the third character was the only legal continuation after the first two characters. Therefore, the preview character "的/de" violated not only the word knowledge of readers but also went against any expectations that they might have formed for the preview character based on the preceding characters. This result is to some extent similar to Abbott et al.'s (2015) results, who found that the-skipping was not reduced in cases where the target word was highly predictable. Therefore, it seems that in both English and Chinese reading, parafoveal processing overrides violations of expectation (at character level or at word level) when it comes to skipping decisions. Note that this does not mean that there is no effect of expectations and predictability at all, but when the expectations based on previous processing and the results from current parafoveal processing disagree, it seems that the skipping decision is determined by the parafoveal preview rather than previously generated expectations.

Finally, we used two additional control conditions to test what type of features of "的/de" other than its extremely high frequency may explain the increase in skipping probability. The preview of the character "钓/diao," whose visual complexity is highly similar to that of the character "的/de," did not enhance the skipping probability in comparison with the unsubstituted preview, indicating the skipping-probability enhancement for "的/de" is not due to its low visual complexity. There was a nonsignificant increase in skipping probability when the preview was the function-word character "乎/hu" compared with the identical preview condition. If this effect was significant, it might be taken to indicate that the grammatical properties of "的/de" may contribute to the tendency of readers to skip it. However, it should be noted that the relative high skipping probability of "乎/hu" also may result from its visual simplicity: it is a single-component character with fewer strokes than "的/de" (6 vs. 8) and the actual target characters (6 vs. 7.95; $t(79) = 10.96$, $p < 0.001$).

In addition to the skipping probability, we also analyzed the fixation times of the pretarget region and target characters. For the pretarget region, the conditions with incorrect previews did not show longer fixations than the identity preview condition did. Therefore, the observed skipping probability differences between the "的/de" preview condition and the identity preview condition on the target characters cannot be explained according to the fixation time on the pretarget region. The FFDs and GDs on the target character was significantly longer in the "钓/diao" and "乎/hu" conditions than in the identity condition. However, there was no significant difference in fixation times between the identity and "的/de" conditions. If the difference is due to a classic preview benefit effect, it should be equal across the three comparisons between the identity condition and the conditions with display change. Because of this, the fixation-time difference between conditions might mainly reflect disruptions from incorrect

preview information rather than preprocessing of the identical character. Because the character "的/de" has extremely high frequency, its processing might be more automatic and attract less attention than the processing of "钓/diao" and "乎/hu" and thus resulted in less disruption. Risse and Kliegl (2014) observed a similar effect where the frequency of a preview word influenced fixation times after crossing the boundary.

In summary, our results are compatible with the hypothesis that skipping of high-frequency characters, such as "的/de" is triggered by character-level preview processing rather than by word-level processing. The preview of the character "的/de" led to an increased skipping probability despite the word-level lexical violations it causes and despite its very low predictability. Our findings indicate that Chinese readers frequently make the decision of skipping a high-frequency character before integrating it into a word.

References

Abbott, M. J., Angele, B., Ahn, Y. D., & Rayner, K. (2015). Skipping syntactically illegal the previews: The role of predictability. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 41*(6), 1703-1714.

Angele, B., Laishley, A. E., Rayner, K., & Liversedge, S. P. (2014). The effect of high- and low-frequency previews and sentential fit on word skipping during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(4), 1181-1203.

Angele, B., & Rayner, K. (2013). Processing the in the parafovea: Are articles skipped automatically? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(2), 649-662.

Carpenter, P. A., & Just, M. A. (1983). What your eyes do while your mind is reading. In K. Rayner (Ed.), Eye movements in reading: Perceptual and language processes (pp. 275–307). Hillsdale: Academic Press.

Chinese Academy of Social Sciences (2012). 现代汉语词典(第六版) [The Contemporary Chinese Dictionary], 6th. Beijing: The Commercial Press.

Chinese Linguistic Data Consortium. (2003). 现代汉语通用词表 [Chinese lexicon] (CLDC-LAC-2003-001). Beijing, China: Tsinghua University, State Key Laboratory of Intelligent Technology and Systems, and Chinese Academy of Sciences, Institute of Automation.

Gu, J., & Li, X. (2015). The effects of character transposition within and across words in Chinese reading. *Attention, Perception, & Psychophysics, 77*(1), 272–281.

Inhoff, A. W., & Liu, W. (1997). The perceptual span during the reading of Chinese text. In H.-C. Chen (ed) The cognitive processing of Chinese and related Asian languages. Hong Kong: Chinese University Press.

Inhoff, A. W., & Liu, W. (1998). The perceptual span and oculomotor activity during the reading of Chinese sentences. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 20-34.

Li, X., Bicknell, K., Liu, P., Wei, W., & Rayner, K. (2014). Reading is fundamentally similar across disparate writing systems: a systematic characterization of how words and characters influence eye movements in Chinese reading. *Journal of Experimental Psychology: General, 143*(2), 895-913.

Li, X., Liu, P., & Rayner, K. (2011). Eye movement guidance in Chinese reading: Is there a preferred viewing location? *Vision Research, 51*(10), 1146–1156.

Li, X., Rayner, K., & Cave, K. R. (2009). On the segmentation of Chinese words during reading. *Cognitive Psychology, 58*(4), 525–552.

Liversedge, S. P., Zang, C., Zhang, M., Bai, X., Yan, G., & Drieghe, D. (2014). The effect of visual complexity and word frequency on eye movements during Chinese reading. *Visual Cognition, 22*(3-4), 441–457.

Ma, G., & Li, X. (2015). How character complexity modulates eye movement control in Chinese reading. *Reading and Writing, 28*(6), 747-761.

Rayner, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology, 7*(1), 65-81.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372-422.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E–Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*(1), 4-22.

Risse, S., & Kliegl, R. (2014). Dissociating preview validity and preview difficulty in parafoveal processing of word n + 1 during reading. *Journal of Experimental Psychology: Human Perception and Performance*, 40(2), 653–668.

Sun, H., Huang, J., Sun, D., Li, D., & Xing, H. (1997). Introduction to language corpus system of modern Chinese study. In M. Hu (Ed.), Paper collection for the fifth world Chinese teaching symposium (pp. 459–466). Beijing: Peking University Press.

Tsai, C. H., & McConkie, G. W. (1995). The perceptual span in reading Chinese text: A moving window study. Paper presented at the Seventh International Conference on the Cognitive Processing of Chinese and Other Asian Languages, Hong Kong.

Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect outgoing saccade length in Chinese reading. *Vision Research, 80*, 1-6.

Zang, C., Zhang, M., Bai, X., Yan, G., Angele, B., & Liversedge, S.P. (in press). The skipping of the structural particle de (的) in Chinese reading. *The Quarterly Journal of Experimental Psychology*.