



The role of foveal and parafoveal word-processing efficiency in reading Chinese

Timo T. Heikkilä¹ · Yongsheng Wang^{2,3,4} · Chuanli Zang⁵ · Xin Li^{2,3,4} · Jukka Hyönä¹

Received: 8 December 2025 / Accepted: 16 August 2026
© The Author(s) 2026

Abstract

This study examined the relationship between the processing efficiency of single-character and two-character words presented in fovea and parafovea and word-level eye-movement parameters in reading Chinese among a heterogenous sample of 70 adult readers (age 18–60 years). Foveal processing efficiency was assessed with performance accuracy in lexical decision, where one-character and two-character words of medium frequency and pseudowords were presented backward-masked for identification using different exposure times (40 ms and 60 ms). Parafoveal-processing efficiency was assessed by presenting similar stimuli for 150 ms to the right and left of the central fixation point (1.35 and 2.27 degrees of visual angle). Measures of foveal and parafoveal-processing efficiency were used to predict eye behavior in reading eight expository texts in varying topics. The Random Forests technique was used to assess the relative importance of individual difference measures in predicting readers' eye-movement patterns. Results were further tested with (generalized) linear mixed-effects ((G)LM) modeling. The results show that the ability to recognize two-character words in fovea best predicts first fixation and gaze durations of words during reading and that both the ability to recognize single-character and two-character words in fovea best predicts word refixation and skipping probabilities during reading as well as the length of between-word saccades. The results are consistent with an analogous study conducted in Finnish, which found that efficiency in foveal word recognition was a better predictor of both foveal and parafoveal processing during reading than efficiency in parafoveal word recognition.

Keywords Individual differences · Reading · Foveal and parafoveal processing · Chinese · Eye tracking

Introduction

Efficient word-recognition ability acts as a foundation for fluent reading. Fluency in reading is based on the interaction between swift word-recognition ability and the ability to integrate word meanings into meaningful sentences and further into a coherent representation of larger segments of text.

In the current study, we probed for the fast activation of lexical representations in readers' minds to assess the extent to which individual word-recognition abilities can serve as predictors of reading fluency among adult Chinese readers. The main goal was to see if variability in reading fluency can be explained with word-recognition abilities measured under strict temporal constraints and from variable visual eccentricities. For this, foveal and parafoveal word recognition skills were measured separately and used as predictors of reading fluency. Reading fluency, in turn, was assessed by readers' eye-movement records when they read expository texts silently for comprehension. We used two lexical decision tasks that measure orthographic and semantic word-recognition abilities to examine their function as possible predictors of foveal and parafoveal word processing during passage reading among adult Chinese readers.

A recent similar investigation conducted in Finnish found that individual differences in foveal word-recognition efficiency was a better predictor of both foveal and parafoveal processing during reading than individual differences in

✉ Timo T. Heikkilä
timo.t.heikkila@utu.fi

¹ Department of Psychology and Speech-Language Pathology, University of Turku, FI-20014 Turku, Finland

² Key Research Base of Humanities and Social Sciences of the Ministry of Education, Academy of Psychology and Behavior, Tianjin Normal University, Tianjin, China

³ Faculty of Psychology, Tianjin Normal University, Tianjin, China

⁴ Tianjin Social Science Laboratory of Students' Mental Development and Learning, Tianjin, China

⁵ Liverpool Hope University, Liverpool, UK

parafoveal word-recognition efficiency (Heikkilä et al., 2024). Finnish is an alphabetic language characterized by relatively long words and shallow orthography. In the current study, we tested the universality of the pattern of results obtained by Heikkilä et al. (2024) in Chinese, a logographic script with vast qualitative differences to Finnish and other alphabetic scripts.

Individual differences in foveal and parafoveal processing during reading

For precise visual information extraction during reading, individuals are able to utilize only the central part of the visual field where visual acuity is at its best. This area can be divided into two functionally different parts known as the fovea and parafovea. Fovea is the central part of the retina with the highest visual acuity, extending about 2 degrees of visual angle around the center of the fixation. Parafovea extends outside the boundaries of the fovea up to 5 degrees on both sides around the central field of view. During reading, full access to word meanings typically require foveal vision, while the parafovea is used for information preprocessing that speeds up subsequent foveal processing of words (for reviews, see Hyönä, 2011; Schotter et al., 2012).

A large number of eye-tracking studies have shown that individual differences in foveal and parafoveal processing capabilities in adults can be explained by the variability in different components of reading skill (for a review, see Heikkilä et al., 2024). These studies have found, for example, that greater print exposure leads to more efficient reading manifested in shorter fixations (e.g., Schmidtke et al., 2018) and that skilled readers have wider perceptual spans than less skilled readers in that they can extract useful visual information from a larger area (e.g., Choi et al., 2015; Veldre & Andrews, 2014; Veldre et al., 2021). Also, better readers skip over words more frequently than poorer readers (e.g., Kuperman et al., 2018; Taylor & Perfetti, 2016).

In a recent investigation, Heikkilä et al. (2024) probed the speed of activation of readers' lexical representations with lexical decision and naming tasks that measured temporal (foveal processing speed) and spatial (parafoveal word-recognition span) efficiency of word recognition and how the observed performance variability in efficiency was linked to reading fluency among adult Finnish readers. The basic assumption behind the study was that fluent reading is dependent on fast foveal and extended parafoveal word-processing skills. For fast word recognition, the word's lexical representation needs to be activated quickly in readers' mind. To measure different components of lexical processing speed, a lexical decision task and two naming tasks (word naming and pseudoword naming) were used. Very short exposure durations of 30–60 ms were used. The lexical decision task was employed to test global

activation during early stages of word processing at the orthographic level, whereas word naming was used to measure the activation of orthographic and phonological representations. Pseudoword naming was used to measure the efficiency of the indirect grapheme-phoneme conversion route. The lexical decision and naming tasks were also employed to assess parafoveal-processing efficiency. The span of parafoveal lexical processing was examined by varying the eccentricity of items (1.33 and 2.24 degrees of visual angle) to the left and right of a fixation cross combined with a 150-ms exposure duration.

The results of Heikkilä et al. (2024) showed that individual variability in word-level eye-movement measures of first-fixation duration, gaze duration and probability of refixation calculated as indices of foveal word processing were all best explained with measures of foveal word-recognition efficiency. This indicates that readers able to recognize words and pseudowords with short exposure duration also spent less time reading individual words in the text.

A more surprising result was that measures of foveal word-processing efficiency were also the best predictors for parafoveal processing during reading as indexed by word skipping and between-word saccade length. This result was explained by the E-Z Reader model (Reichle et al., 1998, 2009), according to which the amount of parafoveal processing is dependent on the processing difficulty of the currently fixated word. Readers capable of fast foveal word recognition experience fewer processing difficulties than those with poorer foveal processing ability. They are also able to shift attention earlier to the next word prior to its fixation, allowing its more extensive parafoveal preprocessing. If a word is then parafoveally recognized, it may be skipped and the following saccade can be programmed to a subsequent word in text. The Heikkilä et al. (2024) results demonstrated that all three measures of foveal-processing efficiency (lexical decision, word naming, and pseudoword naming) were good predictors of parafoveal processing during reading. This in turn suggests that decoding ability and orthographic and phonological skills are all good predictors of parafoveal processing.

The Heikkilä et al. (2024) study was conducted in an alphabetic script of Finnish that has a perfect grapheme-phoneme mapping. The current study was designed to test if a similar pattern of results would emerge in Chinese reading – a logographic script with vast qualitative differences to the Finnish script. This allows us to estimate the universality of the observed pattern of results.

Qualitative differences between the Finnish and Chinese scripts

The written forms of languages exhibit many qualitative differences. The scripts can be categorized as alphabetic, syllabic, and logographic based on the way they map onto

spoken language. In alphabetic scripts this is done via letters mapping onto individual sounds, in syllabic scripts individual symbols correspond to syllables, while in logographic scripts characters convey semantic information without a direct link to the word's spoken form. These script-specific properties require adaptation from the reader in order to read efficiently (Li et al., 2022; Liversedge et al., 2016, 2024).

Finnish is the best example of the alphabetic principle due to its practically one-to-one mapping between graphemes and phonemes. Anyone who has knowledge about how to sound out the Finnish alphabets can in practice read Finnish aloud correctly, even without understanding the word meanings. Chinese, on the other hand, is the best example of a logographic language in exhibiting a very strong connection between individual characters and their meanings but a weaker connection between the characters and their phonological form.

The Finnish and Chinese writing systems are very different. Finnish is written using a set of 29 alphabets from which all words are constructed with the letters written in close proximity to one another. As in other alphabetic languages, word boundaries are demarcated with an empty space between words. On the other hand, Chinese as a logographic language consists of square-shaped characters that are written using a variable number of lines called strokes. All Chinese characters are of equal width and are written back-to-back without any demarcation for word boundaries. The number of strokes in a character can vary between one and 36 among the 7,000 most frequently used characters (Zang et al., 2011). Words are constructed of one or more characters. Thus, there is a lot of variability in the visual complexity of Chinese characters and words.

In Chinese, most words typically contain only one or two characters, and therefore Chinese words do not have much variability in their horizontal extent. In Finnish, however, the visual length of words varies considerably. Average word length in Finnish text is about eight letters. Finnish words can also be long due to the language's rich compounding and inflectional morphology, whereas Chinese does not contain inflections at all and is generally poor in morphology. The differences in horizontal extent and in visual complexity of words in the two languages are relevant factors in reading, since only a limited number of letters or characters can fit in the most accurate part of vision (i.e., fovea) at a time. Since Chinese script is more visually dense than Finnish, more information can fit within the foveal vision in Chinese than in Finnish. According to research, the perceptual span in alphabetic languages like English extends up to 18–20 letters, while in Chinese it is limited to five characters (for a review, see Li et al., 2022).

When Finnish readers encounter an unknown word or a pseudoword, they can read the word aloud by decoding it via grapheme-phoneme correspondence, i.e., by sounding out

the phonemes that correspond to each letter. However, this is not possible in Chinese. Since Chinese characters have no direct link to their phonemes (although some compound characters contain both semantic and phonetic radicals, with the latter providing an approximate pronunciation of the character), Chinese readers need to memorize a large number of characters (approximately 3,000–4,000) in order to achieve fluency in everyday reading.

Alphabetic languages like Finnish also use spaces at the word boundaries that facilitate reading. Word boundaries in alphabetic languages help readers guide their eyes efficiently across consecutive words in the text (Rayner et al., 1998). In contrast, spaces between words are missing in Chinese writing. Thus, word segmentation becomes a significant issue in Chinese reading, as about 80% of words are compound words, that is, comprised of more than a single character (Institute of Language Teaching and Research, 1986). Since there is no word boundary demarcation in the Chinese script, the need for constant and efficient word segmentation from a string of back-to-back written characters becomes a unique requirement not present in alphabetic languages like Finnish. In all, the script dissimilarities discussed above present readers of Finnish and Chinese with differing demands on the visual and cognitive system (Li et al., 2022).

Models of Chinese reading

As discussed above, Chinese script sets qualitatively different types of demands on a reader than reading alphabetic script like Finnish. The Chinese Reading Model (CRM; Li & Pollatsek, 2020) describes how Chinese readers are able to segment words out of the continuous string of individual characters and efficiently guide their eyes to extract new information. The model assumes two main processing components: a word-processing module and an eye-movement control module. According to the model, all characters visible within the perceptual span are processed in parallel and all possible word representations constituted by these characters are activated. These activated word representations may be made up from overlapping characters, be of different lengths, and therefore start from differing positions within the character string visible in the perceptual span. The model describes word identification as the result of the competition between all the activated word representations. The model assumes word segmentation and word identification to be a unified process. After a word has been identified, its length also becomes apparent allowing (after word recognition) for the eye-movement control module then to guide the eye to make a saccade beyond the word just recognized. The timing of the next eye movement is determined by the processing status of the currently fixated word.

The CEZR model (Liu et al., 2025) is a modification of the E-Z Reader model (Reichle et al., 1998, 2009) adapted to the Chinese script. Contrary to the full parallel competition among characters proposed by CRM (Li & Pollatsek, 2020), the CEZR model suggests that word segmentation is based on familiarity of different possible character groupings in the parafovea. The model proposes that readers target saccades dynamically based on parafoveal word familiarity and processing difficulty. So, while word segmentation in CRM is based on lexical competition between characters, the CEZR explains it via serial attention shifting and familiarity. Character strings that frequently appear together are according to CEZR treated as holistic units, while with strings that are not as common or the characters are transparently separate, the first character in multiple-character strings is processed first.

CRM and CEZR differ in the way they explain parafoveal preview processing. Whereas CRM assumes all characters within the perceptual span to be simultaneously activated, CEZR explains the parafoveal preview process via attention shifting that happens in stages and only one linguistic unit at a time. According to CEZR, when the recognition process of the currently fixated character string has reached the stage where the string is perceived as familiar, attention is directed to the parafoveal processing of the next string. The effectiveness of the preview process is then affected by the familiarity of the upcoming character string.

Current study

In the current study, we wanted to test if a similar pattern would emerge in Chinese reading in the relation between foveal and parafoveal-processing efficiency and reading fluency as observed for Finnish. As argued above, Finnish and Chinese scripts present many qualitative differences ranging from visual properties like words' horizontal extent and word boundary demarcation to mappings between graphemes, meanings, and sounds.

Due to script differences, a direct experimental replication between the two languages is not feasible. This is due to the fact that, for example, naming pseudowords in Chinese is not possible as in Finnish, where this is done by decoding a pseudoword via grapheme-phoneme correspondence rules. Instead, for the current study, we chose two lexical decision tasks that tap into the core word-recognition skills needed in Chinese reading.

As in the Finnish study, we measured individual differences in the quality of readers' lexical representations by presenting words in fovea and parafovea with very short exposure durations (for 40 ms and 60 ms in fovea and 150 ms in parafovea) and with variable visual eccentricities in parafovea (1.35 and 2.27 degrees of visual angle). We used two different word-recognition tasks both in fovea and parafovea

that were selected based on the particular characteristics of the Chinese writing system. In the single-character lexical decision task, subjects had to differentiate between single-character words of medium-frequency range and extremely rare single-character words that work as pseudo-characters since readers will not be able to recognize them. In the two-character lexical decision task, subjects were presented with two medium-frequency characters that either combine or do not combine into a two-character compound word.

The single-character lexical decision task emphasizes processing of orthographic information; readers need to process low-level visual properties (e.g., strokes) in order to determine if they recognize the character. In the two-character lexical decision task, readers need to decide if the combination of two real characters is a real word or a nonword. This task emphasizes semantic processing and readers may be more attentive to whole-word processing. Comparing the recognition of single characters to that of two-character words allows us to examine if the recognition ability of basic meaning units (single characters) or the ability to combine characters into words (similar to word segmentation in text reading) has more predictive power on reading fluency related eye-movement measures. Comparing the predictive power of these tasks also allows us to assess whether the processing of orthographic or semantic information has more predictive power.

The foveal lexical decision tasks employ a very short exposure duration (40–60 ms), backward masking, and with no measure of response times. Participants were told to have ample time to provide the responses. Due to very short exposure durations, these tasks probe for orthographic processing and lexical activation under time pressure to gauge whether readers' perceptual and lexical system is able to reach the threshold for word identification. Our foveal task departs from a standard lexical decision task, where reaction times are used to index processing speed. The absence of reaction time measurements shifts the emphasis from the speed of decision making to word-detection thresholds. The problem related to speed-accuracy tradeoffs is thus avoided; in other words, participants do not have to decide whether they should emphasize speed at the expense of accuracy (or vice versa). In sum, our foveal task indexes the efficiency with which readers can successfully activate lexical representations under stringent stimulus exposure time.

The parafoveal task employs a short constant exposure duration of 150 ms while varying the item eccentricities in the parafovea. The participants are asked to fixate at the center fixation point while recognizing items as they appear randomly in the left and right parafovea. Successful word recognition requires shifting covert attention to the parafoveal stimulus. The task measures the individual efficiency in successfully extracting lexical information in the parafoveal vision. Together the foveal and parafoveal task gauge the

ability for word recognition under temporally and spatially stringent conditions. It should be noted that performance in both tasks can also be affected by individual variation in visual acuity, and in the case of the parafoveal task, also by the ability to shift attention covertly.

A pool of Chinese word stimuli was selected for the lexical decision tasks to be comparable with frequency ranges used in the Finnish study of Heikkilä et al. (2024). The reading materials used for the extraction of reading fluency measures were the Chinese translation equivalents of the Finnish texts used in Heikkilä et al. (2024). Both language versions of the texts were adopted from a previous trilingual language comparison study (Liversedge et al., 2016, 2024).

To assess reading fluency, we calculated word-level eye-movement measures from the reading data that index both foveal and parafoveal processing during reading. Then we carried out a two-step analysis for each eye-movement measure. In the first step, we used the Random Forests technique (see Kuperman et al., 2018; Matsuki et al., 2016) also used by Heikkilä et al. (2024) to simultaneously assess the relative importance of all individual-difference measures extracted from the different item exposure durations and eccentricities in the foveal and parafoveal-processing efficiency tasks. This allowed us to identify the best predictor variables for each eye-movement measure. In the following step, we then calculated (G)LMM models for the best predictors by adding word length (number of characters), word frequency, and number of strokes as fixed effects in the models. This allowed us to assess how frequency and word length may modulate the effects of the individual difference measures. In addition, for models of word-skipping probability we entered the saccade launch site as an additional fixed effect to model the effects of parafoveal preview distance.

Hypotheses

This study assessed if the role of foveal and parafoveal-processing efficiency in reading is similar during reading Chinese to that in reading Finnish. Moreover, the roles of single-character and two-character word-recognition abilities were assessed in predicting foveal and parafoveal processing during reading.

Considering the significant qualitative differences between the Finnish and Chinese scripts, it may be argued that processing Chinese script could evoke a larger importance for parafoveal processing during reading to achieve fluency. This is due to the need for constant segmentation of characters into words due to the lack of word boundary demarcation. Chinese words, compared to Finnish words, are also very short in length with more visually tightly packed information. Also, the grapheme-phoneme mapping differs greatly between these languages: one-to-one phoneme-grapheme correspondence in Finnish compared to weak

mapping between orthography and phonology in Chinese. Thus, similarity in results for these two qualitatively different scripts related to the role of foveal and parafoveal-processing efficiency in foveal and parafoveal processing in reading would suggest universality in the examined relationship.

Next, we outline specific hypotheses for foveal and parafoveal processing during Chinese reading.

Foveal processing in Chinese reading

An obvious hypothesis regarding individual differences in foveal processing is that foveal-processing efficiency will predict reading fluency estimated by word-based measures of first-fixation duration, gaze duration, and probability of refixation. Efficiency in the foveal lexical decision tasks should translate into fast word processing that exhibits itself in short fixation durations and a lower probability in the need to refixate words. If foveal processing in reading is more strongly determined by the recognition of single characters, the single-character lexical decision task would be a better predictor than the two-character lexical decision task. On the other hand, if combining characters to word meanings is more important, the two-character lexical decision task will outperform the single-character task in predicting fixation durations on words.

Parafoveal processing in Chinese reading

The Heikkilä et al. (2024) study indicated that individual differences in foveal-processing efficiency were the best predictors of parafoveal processing during reading Finnish, as measured by word-skipping probability and saccade length between words. This relation may be due to the fact that readers with efficient foveal-processing skills are able to guide their attention earlier and more efficiently to the forthcoming text before executing the next saccade, leading to heightened parafoveal processing. The same may be true in Chinese reading. However, since the Chinese script contains no visual cues for word boundaries, parafoveal processing might have a stronger role in reading Chinese compared to Finnish. In alphabetic scripts like Finnish, readers can use low-level visual information to guide saccadic programming during reading. Readers of Chinese, on the other hand, need to constantly decide how to combine characters to form word meanings. This may enhance the significance of parafoveal-processing efficiency in reading fluency. If so, the measures of parafoveal-processing efficiency would be good predictors of the probability of word skipping and between-word saccade length used to index parafoveal processing during reading. Moreover, it may be further hypothesized that the ability to parafoveally recognize two-character words would be a

better predictor of parafoveal processing in reading than the single-character task. This is because in Chinese reading it is more important to recognize character combinations than single characters.

Data availability

The data and R-scripts for all the statistical analyses are available in the Open Science Framework repository (<https://osf.io/wnx7d>).

Method

Participants

Seventy native Chinese speakers (18 males, 52 females; average 27.7 years of age, $SD = 12.2$ years, range 18–60 years) from Tianjin Normal University in China took part in the experiment. All participants had either normal or corrected-to-normal vision (i.e., used eye glasses or contact lenses). To increase the heterogeneity of the sample, the participants were recruited among academic staff (i.e., teachers, $n = 10$), non-academic staff (i.e., security guards and janitors, $n = 11$), and a student population ($n = 49$). The age distributions across the groups were as follows: for students $M = 20.6$ years, $SD = 2.8$ years, age range 18–27; for teachers $M = 34.7$ years, $SD = 4.4$ years, age range 28–44; for non-academic staff $M = 52.6$ years, $SD = 4.8$ years, age range 45–60. Four participants were between 55 and 60 years of age. In addition to the above-listed sample, data for eight participants were collected but their data were excluded as the participants either did not complete all the tasks in the experiment (two participants) or exhibited poor quality reading data due to difficulties in calibrating the camera or due to excessive head movements during reading (a total of six participants).

Apparatus

Eye movements were recorded with an Eyelink1000 system (SR Research, Canada) at 1,000 Hz using a tower mount. The stimuli were presented on a 24-in. Asus LCD monitor (Model VG248QE; 1,920 × 1,080 resolution; frame rate 144 Hz) controlled by an Intel i7-7700 computer running at 3.6 GHz under a Windows 7 environment. Participants were seated 62.5 cm from the monitor with their head positioned on a chin rest. By default, the right eye of the participants was tracked.

Lexical decision and reading tasks

To assess foveal and parafoveal-processing efficiency, four different tasks were set up. These included two lexical decision tasks to assess foveal-processing efficiency (one task for single-character words and another task for two-character words) and two parafoveal lexical decision tasks to assess parafoveal-processing efficiency (one task for single-character words and another task for two-character words). The items used in the tasks were selected from the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) consisting of 46.8 million character frequencies and 33.5 million word frequencies collected from film and television subtitles.

The mean word and character frequencies, mean number of strokes per character, and the number of items shown to each participant in each task are presented in Table 1. Frequency values are based on word and character frequencies as individual words. The mean number of strokes per character (around eight strokes per character) corresponds quite closely to the corpus data. The mean stroke count for a single-character word in the SUBTLEX-CH corpus (Cai & Brysbaert, 2010) is 10.53 ($SD = 3.78$ strokes). The two-character words had on average around 16 strokes; the average stroke count for two-character words in the same corpus is highly similar (16.49, $SD = 4.74$). The items in all four tasks were presented in Song font (size 21). In all tasks, a single character subtended 1.02 degrees of visual angle.

Chinese has a vast number of characters, for example, the “Zhonghua Zihai” dictionary describes over 85,000 existing Chinese characters, while the commonly used characters constitute only about 3,500 (based on the “List of Commonly Used Characters in Modern Chinese”). Chinese readers do not recognize the vast majority of Chinese characters. Therefore, extremely low-frequency characters are almost never encountered or used in daily life. Such characters were used as pseudo-characters in the lexical decision tasks of the present study.

The items used for single-character foveal and parafoveal tasks included medium frequency single-character words with a medium number of strokes (e.g., 茶 “tea”). A list of single-character pseudowords was created from extremely low-frequency (i.e., practically zero frequency) characters (e.g., 嫫). Due to their rarity, these characters are classified as pseudowords, as readers are highly unlikely to recognize them (e.g., Zang et al., 2020; 2023; Zang, Fu, et al., 2024a, 2024b; Zang, Wang, et al., 2024a, 2024b). Based on the classifications in the “Analysis and Research on the Characteristics of Modern Chinese Characters” (Xing, 2007) the composition of the vast majority of these pseudowords consists of common components. Only a few items contained relatively rare components: central-凹 components in 甲 and 𠂔; deputy central-凹 components in 𠂔 and 𠂔. The rest consisted of commonly used components in Chinese. The

Table 1 Mean word and character frequencies (per million tokens), mean number of strokes in characters, total number of strokes in a word, and number of items shown per subject for words, pseudowords, and nonwords in each of the four different lexical decision tasks. Standard deviations are given in parentheses for frequency and stroke number measures

	Mean word frequency	First character frequency	Second character frequency	Number of first character strokes	Number of second character strokes	Total number of strokes	Number of items shown/subject
Foveal task: one character							
Words	44.34 (7.87)	-	-	8.14 (2.43)	-	-	42
Pseudowords	-	-	-	8.24 (2.30)	-	-	42
Parafoveal task: one character							
Words	42.22 (8.52)	-	-	8.15 (2.29)	-	-	60
Pseudowords	-	-	-	8.12 (2.26)	-	-	60
Foveal task: two characters							
Words	42.87 (8.39)	34.64 ^a	109.36 ^a	8.24 (2.27)	8.02 (1.51)	16.26 (3.15)	42
Nonwords	-	44.34 (7.87)	40.99 (9.28)	8.14 (2.43)	8.10 (2.30)	16.24 (3.76)	42
Parafoveal task: two characters							
Words	43.60 (8.09)	66.63 ^a	33.11 ^a	7.87 (2.16)	8.15 (1.75)	16.02 (3.35)	60
Nonwords	-	42.73 (9.24)	42.99 (7.77)	8.18 (2.25)	8.28 (2.26)	16.47 (2.91)	60

^a These are median frequencies, as there were a few highly frequent characters making mean values unrepresentative

number of strokes of these pseudowords were matched to those of the single-character words. All items in the single-character tasks were free morphemes – characters that can stand alone as words and also form compounds (e.g., 茶 “tea” that can also be present in, e.g., 绿茶 “green tea”). Bound morphemes that cannot occur independently (e.g., 蝴 in 蝴蝶 “butterfly”) were not used.

The word frequencies between the one- and two-character words were matched to be of medium frequency (on average about 43 occurrences per million). The first- and second-character frequencies of the two-character non-words were matched to those of the one- and two-character words. Moreover, the stroke count was matched across the characters used in the stimulus materials (on average, characters contained about eight strokes).

The items used in the two-character foveal and parafoveal tasks included two-character words (e.g., 档案 “files” with individual character meanings of 档 “shelves or official records” and 案 “cases or legal documents”) The frequencies of first and second characters in two-character words show quite large variability mostly due to a few highly frequent characters (e.g., the character “的,” which is a particle that acts much like the apostrophe s (’s) or word “of” in English, has a frequency of 50,155.13 words/million). Thus, in Table 1, median values are reported for the characters of the two-character words. The two-character words were contrasted with two-character nonwords. These nonwords comprised of two medium-frequency characters combined

together to form a nonword (i.e., a character combination that does not combine into a real two-character word (e.g., 波念 where individual characters are 波 “wave” and 念 “to think”).

Foveal lexical decision tasks

Two different foveal lexical decision tasks were administered: a single-character and a two-character task. In the single-character task, words and pseudowords, and in the two-character task, words and nonwords were presented with backward-masking for identification. In both tasks, a fixation cross was shown in the center of the screen for 1 s before each item. This was then followed by an item displayed on the screen for a duration of 60 ms, 40 ms, or 20 ms. The items were placed in the center of the screen over the place where the fixation cross had previously been shown. The items subtended 1.02 and 2.04 degrees of visual angle in the single-character task and two-character task, respectively. The stimulus exposure durations of 60 ms, 40 ms, and 20 ms were chosen to be the same as in the earlier study conducted in Finnish (Heikkilä et al., 2024). After the stimulus, a visual mask (※) was presented for 30 ms in the position where the stimulus was shown (i.e., ※ for the single-character and ※※ for the two-character version of the task). After the mask, a blank screen was presented as long as the participant gave an answer by pressing a button on the keyboard. After that, the next trial was initiated.

In the single-character version, the task of the participant was to press a *yes* button if they recognized the word (character) or to press a *no* button if they did not recognize the word (character), or if there was not enough time to see the character. For simplicity and to maintain comparability with the two-character task, we refer to the character or pseudo-character in the single-character task as a word or pseudoword. In the two-character version, the task was to press a *yes* button if the two-characters together formed a word and to press a *no* button if the two-characters did not combine to form a word (or if there was not enough time to see the characters). In both tasks, the participants were advised to answer as accurately as possible and were told that reaction times were not measured. In the parafoveal tasks, responding was analogous to the above description.

Three blocks of items with 60-ms, 40-ms, and 20-ms exposure durations (in this order) were presented for both tasks with each block containing 42 items (21 words and 21 pseudowords). The order of items was randomized and the same randomized list was used for all participants. In both versions of the task, six practice trials were shown before the first block with a 60-ms exposure duration. Exposure durations of 60 ms and 40 ms were selected for the analysis to have comparable measures to those in the Heikkilä et al. (2024) study. Also, performance accuracy in the foveal single-character task with 20-ms exposure duration was in practice at chance level (54% accuracy).

Parafoveal lexical decision tasks

Two different parafoveal lexical decision tasks were administered: a single-character and a two-character task. In the single-character task – words and pseudowords and in the two-character task – words and nonwords were presented for 150 ms on the screen to the right and left of a fixation cross. Before each item, a fixation cross was presented in the middle of the screen for 1 s. Next an item was shown randomly on the screen on either the right or the left side of the cross followed by a blank screen. Participants were told to keep their gaze fixated on the cross and try to perceive the item from the parafovea. If the participants tried to fixate on the item, due to saccadic latency it was very unlikely that this would have been possible due to the short exposure duration of 150 ms.

As this study closely mimics the experimental procedure in the Heikkilä et al. (2024) study and in it no active central fixation enforcement was used in the parafoveal tasks (e.g., using an invisible boundary between the fixation cross and the parafoveal item to detect unwanted saccades towards the target), it is worthwhile to compare the previously obtained results to a protocol using this enforcement. The lexical decision results of Heikkilä et al. (2024) compare favorably with the data of Veldre et al. (2023a, b), who reported very

similar performance between two English lexical decision experiments using 100-ms exposure duration, where maintaining central fixation was enforced in one and not enforced in the other. With enforced central fixation, the accuracies on left 2°, 1°, and right 1° and 2° eccentricities with three- to eight-letter words and pseudowords (constructed by replacing one consonant with another) were 65%, 72%, 82%, and 78%, respectively. With unenforced fixation combined with 300-ms exposure durations, the accuracies were much higher: 84%, 84%, 86%, and 84% for left 2°, 1°, and right 1° and 2° eccentricities, respectively. With 150-ms exposure duration, Heikkilä et al. (2024) reported lexical decision accuracies of 61%, 71%, 85%, and 75% for left 1.33°, 2.27° and right 1.33° and 2.27°, respectively. Although the exposure durations between the two studies differ, the reported accuracies in Heikkilä et al. much closer resemble the 100-ms enforced fixation condition than the unenforced 300-ms condition of Veldre et al. This suggests that even without active central fixation enforcement during the parafoveal lexical decision, the parafoveal task of the present study with a 150-ms exposure duration should reliably capture genuine parafoveal processing.

In both tasks, the items were shown randomly on the right and left sides of the center fixation cross with variable eccentricities. Each task comprised three blocks of items with eccentricities of 1.35, 1.81, and 2.27 degrees of visual angle from the fixation point to the nearest edge of the item. The eccentricities to the farthest edge of the item were 2.37, 2.83, and 3.29 in the single-character task and 3.39, 3.85, and 4.31 in the two-character task, respectively. As the parafovea extends 5 degrees to both sides from the center of fixation, the items were within the limits of the parafoveal vision. The item eccentricities were selected as close as possible to the values used in the Heikkilä et al. (2024) study. The single-character and two-character items in the current study subtended 1.02 and 2.04 degrees of visual angle, respectively. The parafoveally presented five-letter words in the Heikkilä et al. (2024) study subtended 1.82 degrees of visual angle.

In both tasks, six practice items were shown before the first block with an eccentricity of 1.35 degrees. After the practice, each of the three blocks contained 60 items with 30 items shown in the right and 30 items shown in the left side. The items in both sides comprised 15 words and 15 pseudowords. The order of the items was the same for all participants and was based on a single randomized list. The eccentricities of 1.35 and 2.27 were selected for the analysis to have comparable measures to those in the Heikkilä et al. (2024) study. The data for the 1.81-degree condition (as well as the 20-ms condition in the foveal task) was gathered to achieve procedural comparability with the Finnish study, where this condition was also tested but not analyzed. The 1.81-degree condition was also dropped from the analyses

for this study to achieve close comparability with the analysis pipeline between the two studies.

Reading task

In the reading task, participants read silently eight popular science texts on different topics for comprehension while their eye movements were registered. The texts were adopted from the study of Liversedge et al. (2016) and were shown in Song font (size 22). A single character subtended 1.07 degrees of visual angle. The texts were short expository texts of various popular science topics including vintage car racing, football, crushing oil from nuts and seeds, sheep farming, aging effects caused by sugar, tipping culture, health benefits of walking, and wind energy. The texts were divided in two to six pages (between 32 and 110 words in each page). The texts were between 143 and 409 words in length.

To encourage the participants to read for comprehension, two reading comprehension questions were presented after each text. The questions were statements, to which the participants answered with either a *yes* or a *no* response depending on whether the statement was true or false with respect to the text content. On average, the participants answered 13.5 out of 16 questions correct (84.6%).

Procedure

Before the experiment, each participant's sex and age were registered.

The study began by presenting the single-character foveal lexical decision task followed by the two-character foveal lexical decision task. At the start of both tasks, written instructions were presented. Answering was done by pressing a marked key on the keyboard. Participants were asked to keep their fingers on the keyboard during the task.

Next, the reading task was administered. This began by presenting the participants with written instructions. They were encouraged to read the texts silently at their own pace in a manner that would allow them to understand the texts. They were informed that after each text they would need to answer two reading comprehension questions by pressing a button on the keyboard (questions were formulated as *yes* or *no* statements). Before reading, a 9-point calibration procedure was performed to set up the eye tracker. The calibration error threshold used was a maximum of 1 degree of visual angle for a single calibration point and a maximum average of .5 degrees for the 9 points. Next each text was presented one page at a time. Participants pressed the spacebar to move to the next page. Before each page, a drift correction point was shown in the upper left corner of the screen. After the participant had fixated the drift correction point the experimenter released the next page. This way the participant's gaze could be directed to the beginning of the text before each page was

presented. Participants were allowed to reread the texts as much as they liked but they were unable to return to previous pages.

After the reading task, the single-character parafoveal lexical decision task followed by the two-character parafoveal lexical decision task were presented. Each task began with written instructions that emphasized that the participants should look at the fixation cross and try to perceive the appearing items in the parafovea without making a fixation to the appearing stimuli.

Variables

Measures of foveal and parafoveal word-processing efficiency

To assess performance in the lexical decision tasks, a mean discriminability index (d') was calculated for word detection in each task using the *psycho* package in R (Makowski, 2018). The hit rate was defined as the rate of correct detections of the word items and the false alarm rate was defined as the rate of false positives in pseudoword (the single-character task) or non-word (the two-character task) detection (i.e., number of pseudowords or non-words categorized falsely as words). The d' indices for single-character and two-character foveal lexical decision tasks were calculated separately for each block (60 ms and 40 ms) of each participant. The d' indices for single-character and two-character parafoveal lexical decision tasks were calculated separately for items shown in the right and left visual field in each block of both tasks (eccentricities of 1.35 and 2.27 degrees).

Word-level eye-movement measures and predictors

To assess reading behavior, word-level eye-movement measures of first-fixation duration, gaze duration, probability of refixation, word skipping probability, and length of saccade between words were calculated from the data. First fixation duration refers to the duration of the first fixation the word received during the first-pass reading (i.e., the first time when a saccade was launched to the word from a word preceding it in the text). Gaze duration was calculated as the summed duration of all the fixations the word received between the time it is first entered from a previous word and before it is exited to the right or left. Probability of refixation was calculated as a dichotomous variable based on whether a word received either only one or more than one fixation during the first-pass reading. Word skipping probability was calculated as a dichotomous variable based on whether or not a word was fixated or skipped during the first-pass reading. Between-word saccade length was calculated as the distance (degrees of visual angle) between the first fixation of the word and the last fixation of the previous word. This

was calculated only for the first-pass reading when the word was first entered from the previous word.

The titles presented before each text (eight titles shown per participant) were excluded from the eye-movement data along with a total of 58 text screens (2.4% of all screens) where data quality was either too low due to head movements during reading or showed excessive data loss. The data for first-fixation duration, gaze duration, probability of refixation, and saccade length between words (143,937 instances) were first filtered to contain only the data on first-pass reading (62,011 instances). From these data, the data for the first word in each line were filtered out (4.5%), as were the data for words with logarithmic frequencies below 0 and above 10 per million (10.4%) and words split in two lines (1.94%). Further, for the gaze duration and probability of refixation measures, gaze durations shorter than 60 ms or longer than 800 ms (2.4%) were filtered out. Also, for the first-fixation duration measure, fixations shorter than 60 ms and longer than 800 ms (2.2%) were filtered out. For between-word saccade length, saccades longer than 13 degrees in visual angle (11.7%) and regressive saccades (12.0%) were removed. This resulted in a dataset of 50,941 instances for gaze duration and probability of refixation, 51,084 instances for first-fixation duration, and 47,667 instances for between-word saccade length (82.1%, 82.4, and 76.9% of the first-pass reading data, respectively).

The data on word-skipping probability (39,898 instances) were filtered as follows. Since the texts were presented in multiple lines, the first word in each line was discarded from the analysis (.15% of the whole data). This was done because readers are unable to have a parafoveal preview of these words; thus, skipping line-initial words is not indicative of parafoveal word recognition. Instead, typically the saccade launched from the last word of a line towards the first word of next line may fall short, causing the fixation to land on the second word of the line. This is usually followed by a regressive fixation to the first word and results in the line-initial word to be considered as skipped. In contrast, line-final words can be previewed parafoveally, so possibly skipping them would reflect meaningful lexical processing and thus they were not filtered out. Also, we did not exclude any fixations on words surrounding punctuation. Further, all the skipping data for words with frequencies below 0 and above 10 (log/million tokens) were filtered out (14.51%), as were the data where a word was split over two lines (.85%). After filtering, a total of 84.54% (33,729 instances) was left to be used in the analysis of word-skipping probability.

Word length in number of characters (a factor with two levels: one or two characters per word) and logarithmic frequency per million tokens were used as word-level predictor variables in (G)LMM analyses. As the texts mostly consisted of one- and two-character words, all words longer than two characters were excluded from the analysis. In addition,

number of strokes as included in all models as a main effect and saccade launch site was entered as a predictor in the analyses of word-skipping probability. The launch site of the incoming saccade was defined as the distance (in characters) from the word beginning.

Statistical considerations and analysis

The analysis procedure followed closely the method used in Heikkilä et al. (2024) to achieve comparability with the two studies. Exposure durations of 60 ms and 40 ms in foveal-processing efficiency and eccentricities of 1.35 and 2.27 degrees in parafoveal-processing efficiency were selected to be included in the analyses. These values compare closely with values used in the foveal and parafoveal lexical decision tasks used in Finnish by Heikkilä et al. (2024), where foveal exposure durations were exactly the same and eccentricities in the parafoveal task were 1.33 and 2.24 degrees, respectively. This left us with two parafoveal measures for each visual field (left and right) for both single-character and two-character tasks (i.e., eight parafoveal measures in total) and two measures for each exposure duration for both single-character and two-character tasks (i.e., four foveal measures in total). These 12 foveal and parafoveal efficiency measures were used as predictors in the Random Forests analysis of each eye-movement measure.

We used the Random Forests analysis technique and analysis protocol used previously by Kuperman et al. (2018), Matsuki et al. (2016) and Heikkilä et al. (2024) to simultaneously assess the relative importance of all 12 foveal and parafoveal efficiency measures in explaining variation in the word level eye-movement data. The analysis was carried out by using the party-package (version 1.3–13; Hothorn et al., 2006; Strobl et al., 2007, 2008) in R. Random Forests construct a set of decision trees, each trained on a bootstrap sample of the data and a randomly selected subset of predictors. The predictions from these trees are then aggregated, producing stable estimates of the contribution of each predictor. This approach is particularly well suited for datasets with many collinear or strongly correlated predictors – as in the present study, where correlations among predictors ranged from $r = .31$ to $.81$ (predictor correlations are listed in the Appendix, Table 5) – because the random sampling of predictors at each split reduces the impact of multicollinearity and prevents any single predictor from dominating the model. In contrast, attempting to model all or a selection of the predictors simultaneously using for example linear mixed-effects modelling would result in over-parameterized models with inflated variance estimates and unstable coefficient estimates. The Random Forest technique also allows for selecting a subset of the best predictor variables for further testing, as modelling all 12 predictors individually would be impractical. The technique allows for evaluating

the relative importance of multiple, intercorrelated predictors, as is the case in the current study.

Following the procedure, two Random Forests models were run for each eye-movement parameter by using the 12 processing efficiency measures as predictors in each model. The two models run for each eye-movement parameter had a varying number of simultaneous predictors (value of the *mtry* parameter) in each defined as full integer values between the square root of the number of predictors and number of predictors divided by three. This resulted in three and four predictors for each of the two models calculated for every eye-movement measure. The number of decision trees in each model was set as 1,000. Mean variable importance scores were calculated from values obtained from the two models run for each eye-movement parameter. These mean variable importance scores were then ordered by rank and plotted for each eye-movement measure. The best predictors for each eye-movement measure were selected by visually inspecting the variable importance plot of each eye-movement parameter and determining a cutoff point in each plot, where a gap was detected in the rank ordered list of importance scores (for further details, see Kuperman et al., 2018). After that, (G)LMM analyses were carried out for these smaller sets of predictor variables deemed as best predictors for each eye-movement parameter. These analyses were carried out to confirm the capacity of the selected predictors to predict eye movements in reading.

Each model for first-fixation duration, gaze duration, probability of refixation, and between-word saccade length had a fixed-effects structure with a three-way interaction of efficiency $\times$ word frequency $\times$ word length in characters and the main effect for number of strokes. The random structure (before possible pruning) included a random level for items (words) with an intercept and a random level for subjects with random slopes for word frequency, word length, and number of strokes. Each model for skipping probability had a fixed-effects structure including two three-way interactions of efficiency $\times$ word length $\times$ launch site and efficiency $\times$ word frequency $\times$ launch site and the main effect for number of strokes. The random structure included a random level for items (words) with an intercept and a random level for subjects. First subject-level random slopes for word length and launch site were included to form a baseline model, after which the slopes for word frequency and number of strokes were added one at a time. The random effects of all models were pruned as outlined in Bates et al. (2015a, 2015b) to be supported by the data. Full model fixed effects along with final pruned variance components are listed in the Appendix. All reported models converged normally and were not singular fits. Variance inflation factors of all reported model terms were ≤ 4.63 .

We aimed for a sample size that was at least equivalent to that of the Heikkilä et al. (2024) study, which reported

a power of 80% based on simulations of several individual difference measures on a publicly available dataset (Gong & Shuai, 2023; for details, see Heikkilä et al., 2024). The number of observations in our sample exceed those in the Finnish study in all dependent eye-movement variables except for word-skipping probability, where the number of observations is slightly smaller (33,729 compared to 38,353 in the Finnish study).

Results

Number of observations, means, and standard deviations of the word-level eye-movement measures derived from the reading data are presented in Table 2. Means and standard deviations of foveal and parafoveal performance accuracy in lexical decision (d') for the different presentation conditions and mean accuracies are presented in Table 3.

A 2×2 repeated measures analysis of variance was conducted to examine the effects of word length (number of characters) and exposure duration in the foveal lexical decision accuracy scores (%). According to the analysis, there was a statistically significant main effect of word length ($F(1, 79) = 228.71, p < .001, \eta^2_G = .356$) and a significant main effect of exposure duration ($F(1, 79) = 14.74, p < .001, \eta^2_G = .0085$). The interaction of word length and exposure duration was non-significant. Recognition accuracy was lower for the single-character words than for the two-character words and lower for the shorter exposure duration of 40 ms than for the 60-ms duration.

For the parafovea measures, a $2 \times 2 \times 2$ repeated-measures analysis of variance was conducted to examine the effects of word length (number of characters), stimulus eccentricity, and stimulus orientation (left vs. right) on parafoveal lexical decision accuracy scores (%). According to the analysis, there were significant main effects of word length ($F(1, 79) = 50.57, p < .001, \eta^2_G = .046$), stimulus orientation ($F(1, 79) = 25.11, p < .001, \eta^2_G = .015$), and stimulus eccentricity ($F(1, 79) = 54.78, p < .001, \eta^2_G = .031$). All

Table 2 Number of observations (*N*), means (*M*), and standard deviations (*SDs*) of the word-level eye-movement measures for text reading

	<i>N</i>	<i>M</i>	<i>SD</i>
Gaze duration (ms)	50,941	231	36
First fixation duration (ms)	51,084	213	27
Probability of refixation	50,941	.10	.06
Word skipping probability	33,729	.61	.17
Between-word saccade length (characters)	47,669	3.44	1.02

Means and standard deviations have been calculated from the subject means

Table 3 Means (*M*) and standard deviations (*SDs*) of foveal and parafoveal performance accuracy in lexical decision (*d'*) for the different presentation conditions along with the relative false alarm (FA) rates in pseudoword detection. Mean accuracies are given in parentheses

Task	<i>N</i> of characters	Exposure time (ms)	Stimulus position	Eccentricity (degrees)	Accuracy <i>M</i>	Accuracy <i>SD</i>	FA-rate
Fovea	1	60	Center	-	1.37 (69%)	.94	54%
Fovea	1	40	Center	-	1.02 (67%)	.79	47%
Fovea	2	60	Center	-	2.67 (88%)	.99	19%
Fovea	2	40	Center	-	2.43 (86%)	1.10	20%
Parafovea	1	150	Right	1.35	2.50 (88%)	.74	17%
Parafovea	1	150	Left	1.35	2.15 (84%)	.72	26%
Parafovea	1	150	Right	2.27	2.02 (83%)	.72	20%
Parafovea	1	150	Left	2.27	1.92 (81%)	.86	25%
Parafovea	2	150	Right	1.35	2.92 (92%)	.75	13%
Parafovea	2	150	Left	1.35	2.59 (90%)	.80	14%
Parafovea	2	150	Right	2.27	2.49 (88%)	.87	15%
Parafovea	2	150	Left	2.27	2.28 (86%)	.92	17%

interactions were nonsignificant. Recognition accuracy was lower for the single-character words than for the two-character words and lower for the left side than for the right side, and also lower for the farther stimulus eccentricity.

To assess the stability of the individual-difference measures across tasks, we examined correlations between the *d*-prime values of foveal and parafoveal efficiency measures for both one- and two-character lexical decision tasks. These correlations are presented in the Appendix, Table 5. The tasks showed strong internal reliability. Foveal measures correlated at $r = .66$ for one-character words and $r = .81$ for two-character words. Parafoveal measures also showed high within-task consistency (one-character words: $r = .49-.73$; two-character words: $r = .63-.73$). Thus, participants' performance was consistent across foveal and parafoveal tasks.

With respect to foveal-parafoveal correlations, for one-character words, foveal-parafoveal correlations ranged from $r = .31$ to $.56$, whereas for two-character items these correlations were substantially stronger ($r = .63-.75$), indicating more unified processing demands for two-character words. Finally, performance generalized across stimulus types. Foveal performance on one- and two-character tasks was moderately correlated ($r = .50-.54$), and parafoveal performance showed strong cross-stimulus consistency ($r = .58-.70$). Together, these results demonstrate robust cross-task reliability and suggest that individual differences in lexical discrimination ability are stable across foveal and parafoveal processing and across word lengths.

First-fixation duration

The Random Forests models for first-fixation duration (Fig. 1) indicate that based on the variable importance scores the best predictor variables were foveal two-character lexical

decision at 40 ms, parafoveal two-character lexical decision at 1.35 degrees to the left, and foveal two-character lexical decision at 60 ms. These measures were selected for further LMM analyses. The results of the LMM analyses are presented in the Appendix, Table 6. Based on the number of comparisons, the alpha level for determining statistical significance in the following LMM analyses was Bonferroni corrected from .05 to .017, which corresponds to $t > 2.39$.

Two of the three LMM models showed a significant main effect of the selected measure of processing efficiency (foveal two-character lexical decision at 40 ms, $t = -3.151$, $B = -.046$, $SE = .014$; and foveal two-character lexical decision at 60 ms, $t = -3.009$, $B = -.050$, $SE = .017$). These effects are depicted in Fig. 2, where it appears that higher scores in foveal two-character lexical decision at 40 ms and 60 ms were associated with shorter first-fixation durations than lower scores of these variables. The main effect of parafoveal two-character lexical decision at 1.35 degrees to the left was clearly non-significant ($t = -.749$). In addition, all models showed significant main effects for word length and number of strokes, where single-character words and words with more strokes had longer first-fixation durations than two-character words and words with less strokes, respectively.

Gaze duration

The Random Forests models for gaze duration (Fig. 3) indicate that based on the variable importance scores the best predictor variables were foveal two-character lexical decision at 40 ms, and foveal two-character lexical decision at 60 ms. These measures were selected for further LMM analyses. The results of the LMM analyses are presented in the Appendix, Table 7. Based on the number of comparisons,

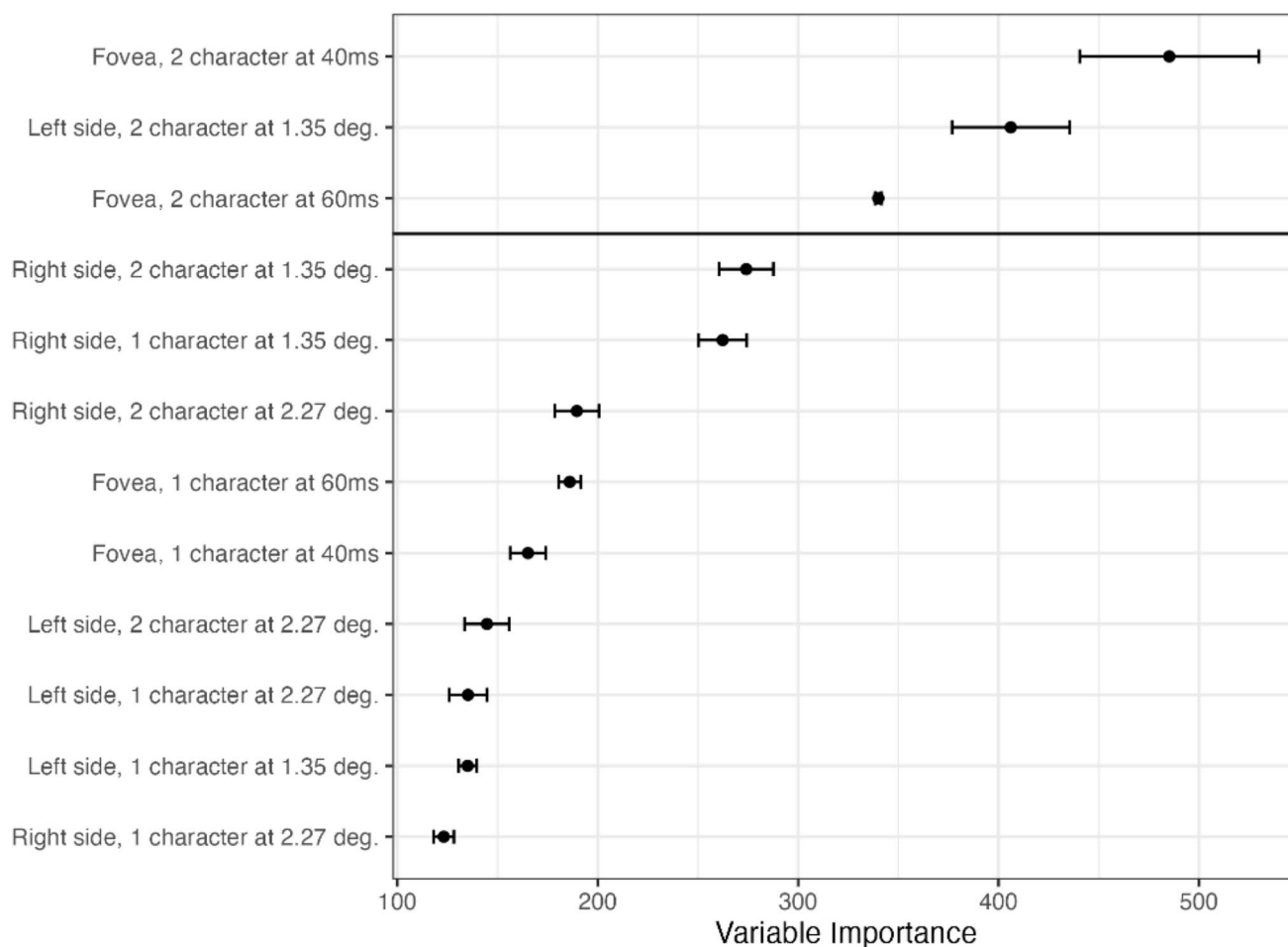


Fig. 1 Relative variable importance estimated by Random Forests models of first-fixation duration on the word. The horizontal line indicates the threshold determined by visual inspection for the best predictor variables chosen for further analysis. The error bars indicate

the standard error of variable importance calculated from the scores extracted from each run of the Random Forest with different mtry parameter

the alpha level for determining statistical significance in the following LMM analyses was Bonferroni corrected from .05 to .025 which corresponds to $t > 2.24$.

Both models for gaze duration showed a significant main effect of the selected measure of processing efficiency (foveal two-character lexical decision at 40 ms, $t = -3.732$, $B = -.055$, $SE = .015$; and foveal two-character lexical decision at 60 ms, $t = -3.541$, $B = -.060$, $SE = .017$). The nature of these main effects was such that higher scores in the foveal two-character lexical decision at 40 ms and 60 ms were associated with shorter gaze durations than lower scores of these variables. In addition, both models showed a two-way interaction of the selected measure of processing efficiency with number of characters (foveal two-character lexical decision at 40 ms, $t = -4.442$, $B = -.026$, $SE = .006$; and foveal two-character lexical decision at 60 ms, $t = -4.803$, $B = -.031$, $SE = .006$). These interactions are depicted in Fig. 4. As is apparent in the figure, individuals

with lower scores in foveal two-character lexical decision at 40 ms and 60 ms had increasingly longer gaze durations on longer words than those with higher scores. Finally, both models showed significant main effects of number of strokes and a significant two-way interaction between number of characters and word frequency. These main effects were such that words with more strokes were associated with longer gaze durations than words with less strokes, while the nature of the interaction was such that gaze durations on less common two-character words were increasingly longer than gaze durations on less common single-character words.

Probability of refixation

The Random Forests models for probability of refixation (Fig. 5) indicate that based on the variable importance scores the best predictor variables were foveal single-character lexical decision at 40 ms and foveal two-character lexical

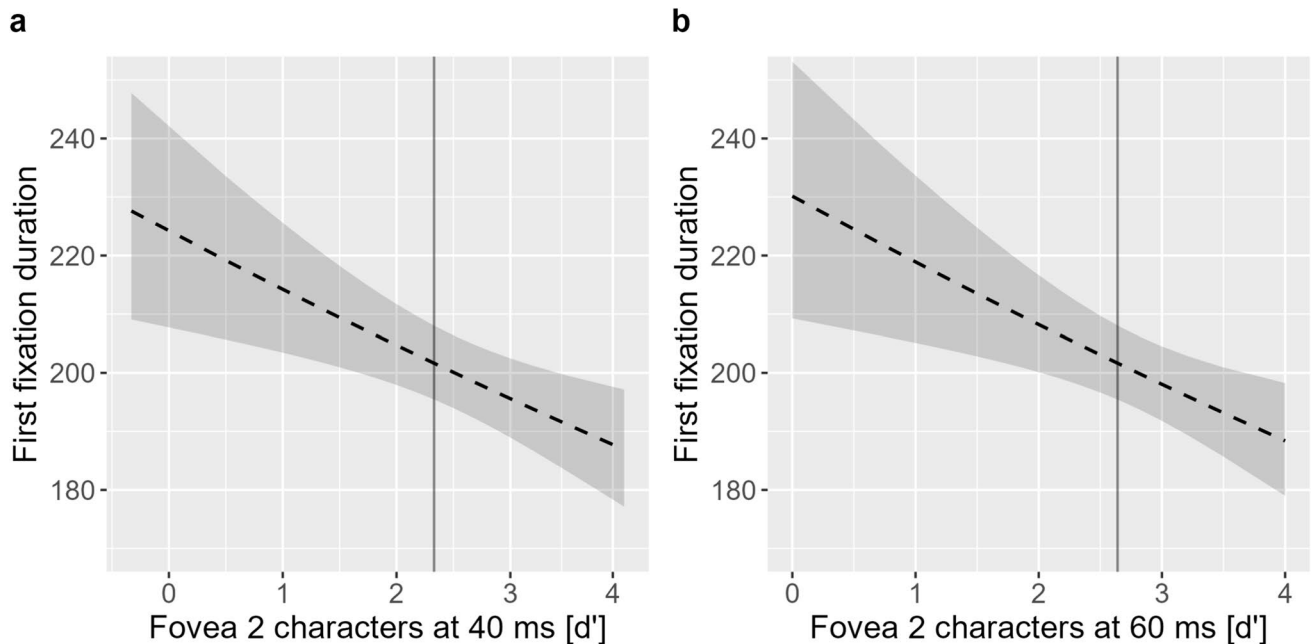


Fig. 2 Estimated marginal linear mixed-effects model (LMM) means for two different models of first-fixation duration as a function of **(a)** foveal two-character lexical decision at 40 ms; and **(b)** foveal two-character lexical decision at 60 ms. Shaded bands represent 95% confidence intervals based on standard error. The vertical lines show the

point where the main effects of centered variables in the x-axis were estimated by the LMMs (i.e., at $d' = 2.33$ for foveal two-character lexical decision at 40 ms and at $d' = 2.64$ for foveal two-character lexical decision at 60 ms)

decision at 40 ms. These measures were selected for further GLMM analyses. The results of the GLMM analyses are presented in the Appendix, Table 8. Based on the number of comparisons, the alpha level for determining statistical significance in the following GLMM analyses was Bonferroni corrected from .05 to .025, which corresponds to $z > 2.24$.

Both GLMM models showed a significant main effect of the selected measure of processing efficiency (foveal single-character lexical decision at 40 ms, $z = -3.570$, $B = .684$, $SE = .073$; foveal two-character lexical decision at 40 ms, $z = -3.502$, $B = .060$, $SE = .756$). These effects are depicted in Fig. 6, where it appears that higher scores in foveal single-character lexical decision at 40 ms and foveal two-character lexical decision at 40 ms were associated with a lower probability of refixating a word than lower scores of these variables. In addition, both models showed a main effect of number of characters. The effect was such that longer words were associated with higher probabilities of being refixated than shorter words.

Word skipping probability

The Random Forests models for probability of word skipping probability (Fig. 7) indicate that based on the variable importance scores the best predictor variables were foveal single-character lexical decision at 40 ms and foveal two-character lexical decision at 40 ms. These measures were

selected for further GLMM analyses. The results of the GLMM analyses are presented in the Appendix, Table 9. Based on the number of comparisons, the alpha level for determining statistical significance in the following GLMM analyses was Bonferroni corrected from .05 to .025, which corresponds to $z > 2.24$.

Both GLMM models showed a similar significant main effect of processing efficiency (foveal single-character lexical decision at 40 ms, $z = 3.907$, $B = 1.649$, $SE = .211$; foveal two-character lexical decision at 40 ms, $z = 5.309$, $B = 1.625$, $SE = .149$), where higher scores in processing efficiency were associated with higher probabilities in word skipping than lower scores. In addition, both models showed a similar two-way interaction of processing efficiency by number of characters (foveal single-character lexical decision at 40 ms $\times$ number of characters, $z = 2.502$, $B = 1.192$, $SE = .084$; foveal two-character lexical decision at 40 ms $\times$ number of characters, $z = 4.729$, $B = 1.273$, $SE = .065$). These interactions are depicted in Fig. 8 (panels a and b). In both interactions, individuals with lower processing efficiency scores displayed a larger word-length effect than individuals with higher processing efficiency scores. In addition, the model for foveal two-character lexical decision at 40 ms showed a two-way interaction of the processing efficiency measure and saccade launch site ($z = 4.729$, $B = 1.273$, $SE = .065$). This interaction is presented in Fig. 8 (panel c), where it can be seen that higher scores in foveal

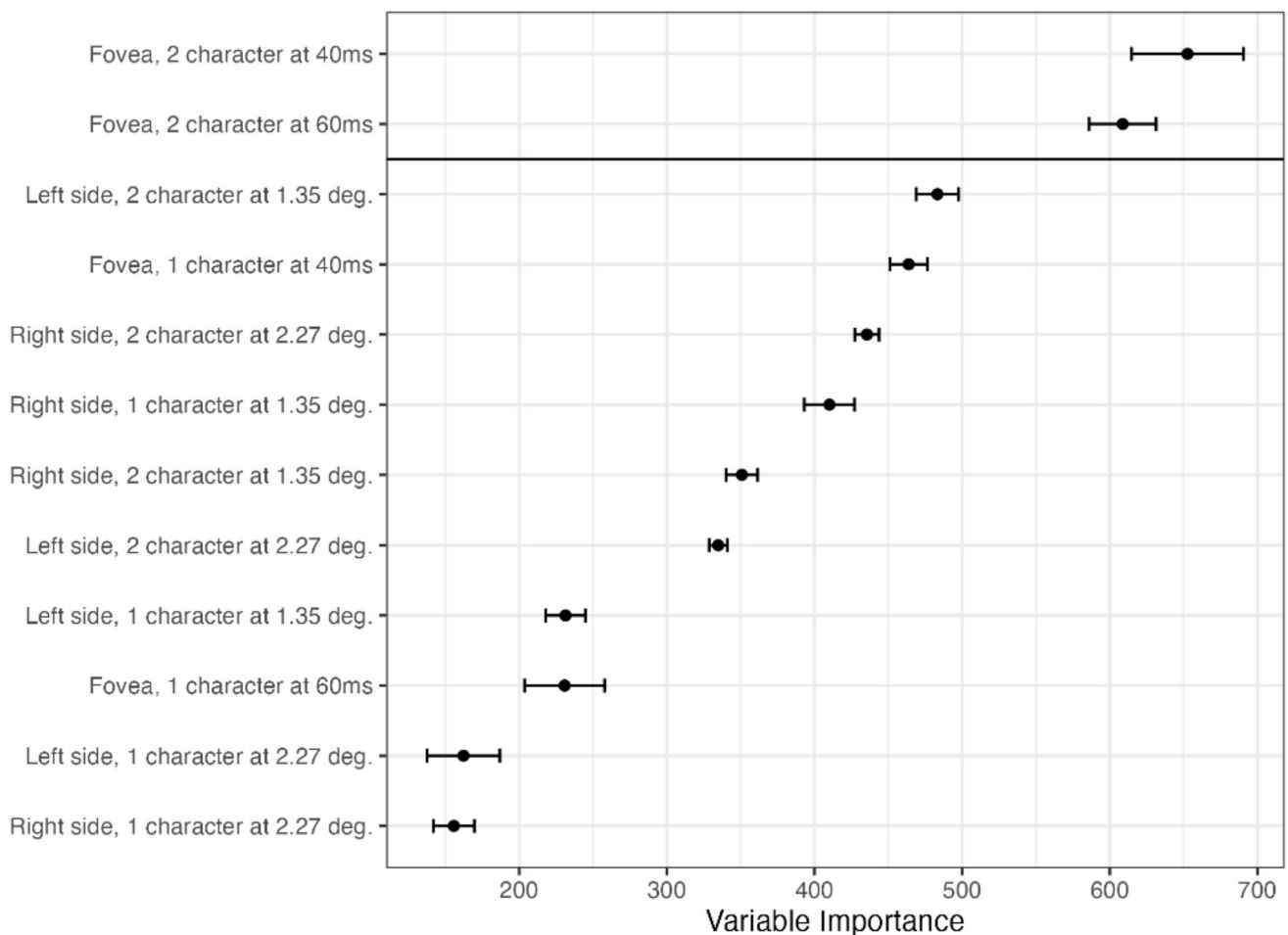


Fig. 3 Relative variable importance estimated by Random Forests models for gaze duration. The horizontal line indicates the threshold determined by visual inspection for the best predictor variables cho-

sen for further analysis. The error bars indicate the standard error of variable importance calculated from the scores extracted from each run of the Random Forest with different mtry parameter

two-character lexical decision at 40 ms were associated with high skipping probabilities with little effect of saccade launch site, while lower scores were associated with relatively low skipping probabilities especially with far saccade launch sites.

In addition, both models showed similar main effects of launch site, number of characters, word frequency, and number of strokes. These effects were such that shorter launch sites, shorter words, higher word frequencies, and words with less strokes were associated with higher skipping probabilities than longer launch sites, longer words, less common words, and words with more strokes. Also, a significant two-way interaction of launch site by number of characters was present in both models. The interaction was such that for longer launch sites both short and long words were equally unlikely to be skipped, while for shorter launch sites shorter words were more likely to be skipped than longer words. Finally, a significant two-way interaction of launch site by word frequency was present in both models. The interaction

reflects the finding that for short launch sites all words were associated with high skipping probabilities independent of word frequency, while at longer launch sites all words were less likely to be skipped overall and particularly with less common words.

Between-word saccade length

The Random Forests models for between-word saccade length (Fig. 9) indicate that based on the variable importance scores the best predictor variables were foveal two-character lexical decision at 40 ms and foveal single-character lexical decision at 40 ms. These measures were selected for further LMM analyses. The results of the LMM analyses are presented in the Appendix, Table 10. Based on the number of comparisons, the alpha level for determining statistical significance in the following GLMM analyses was Bonferroni corrected from .05 to .025, which corresponds to $t > 2.24$.

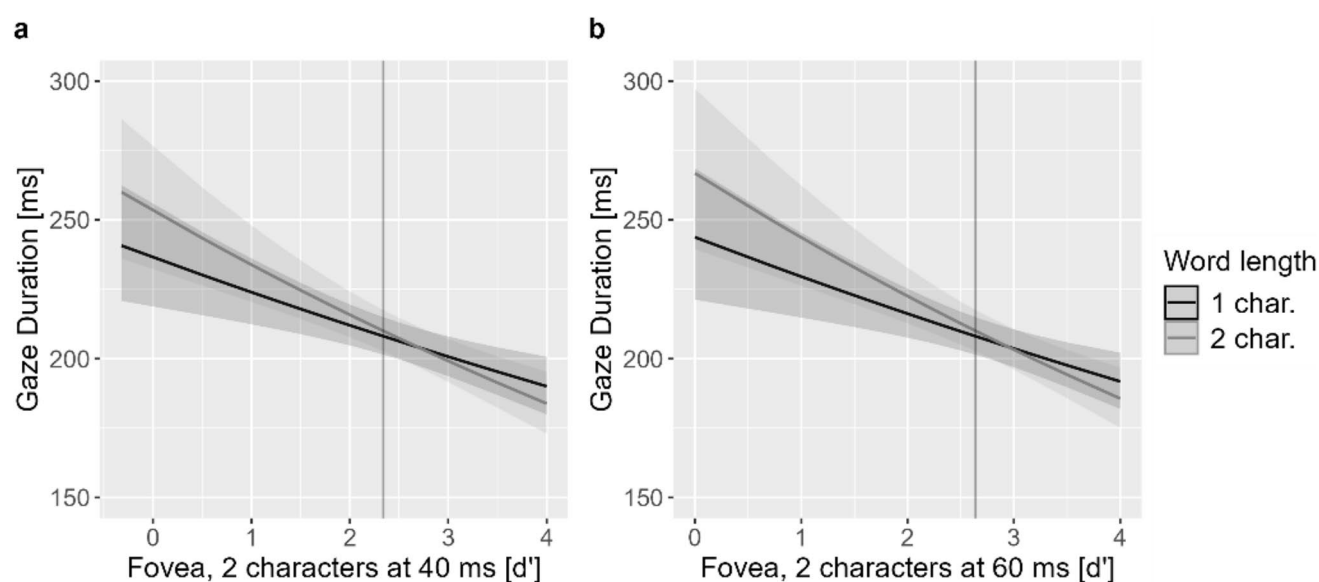


Fig. 4 Estimated marginal linear mixed-effects model (LMM) means for two different models of gaze duration as a function of (a) foveal two-character lexical decision at 40 ms and word length; and (b) foveal two-character lexical decision at 60 ms and word length. Shaded bands represent 95% confidence intervals based on standard

error. The vertical lines show the point where the main effects of centered variables in the x-axis were estimated by the LMMs (i.e., at $d' = 2.33$ for foveal two-character lexical decision at 40 ms and at $d' = 2.64$ for foveal two-character lexical decision at 60 ms)

Both LMM models showed a significant main effect of the selected measure of processing efficiency (foveal two-character lexical decision at 40 ms, $t = 4.415$, $B = .142$, $SE = .032$; foveal single-character lexical decision at 40 ms, $t = 3.520$, $B = .151$, $SE = .043$). These effects are depicted in Fig. 10, where it appears that higher scores in the foveal two-character lexical decision at 40 ms (panel a) and foveal single-character lexical decision at 40 ms (panel b) were associated with longer between-word saccade lengths than lower scores of these variables. In addition, both models showed similar main effects of word length, word frequency, and number of strokes. These effects were such that longer words, higher frequency words, and words with fewer strokes were associated with longer between-word saccade amplitudes than shorter words, less frequent words, and words with more strokes.

Summary of results

A summary of all significant effects for the word-processing efficiency measures selected for further (G)LMM analysis are presented in Table 4.

Discussion

The present study examined if text-reading fluency among adult readers of Chinese could be explained by foveal or parafoveal word-processing efficiency measured by two

different lexical decision tasks. In these tasks, single-character words versus pseudo-characters and two-character words versus two-character nonwords were presented with very short exposure durations in fovea (40–60 ms) and with variable eccentricities in parafovea (1.35 and 2.27 degrees of visual angle with 150-ms exposure duration). The single-character lexical decision task taps into processing of orthographic information, as readers need to process low-level visual properties (e.g., strokes) in order to determine if they recognize the character. The two-character lexical decision task, on the other hand, taps into semantic word processing, as the word meaning is derived by integrating the meaning of the component characters. Comparing the performance in the two task types allowed for the examination of predictive power between basic meaning unit recognition (single characters) and character combination ability (similar to word segmentation in text reading) in relation to fluency-related eye-movement measures registered during reading of short expository texts in Chinese. The experimental setup also allowed for comparison between the relative importance of foveal and parafoveal word-recognition abilities in predicting reading fluency.

The lexical decision accuracies were used to predict eye-movement patterns collected during reading of expository texts of varying topics. Word-level eye-movement measures of first-fixation duration, gaze duration, probability of refixation, word-skipping probability, and saccade length between words were calculated from the text-reading data as measures of reading fluency. The first three of these measures

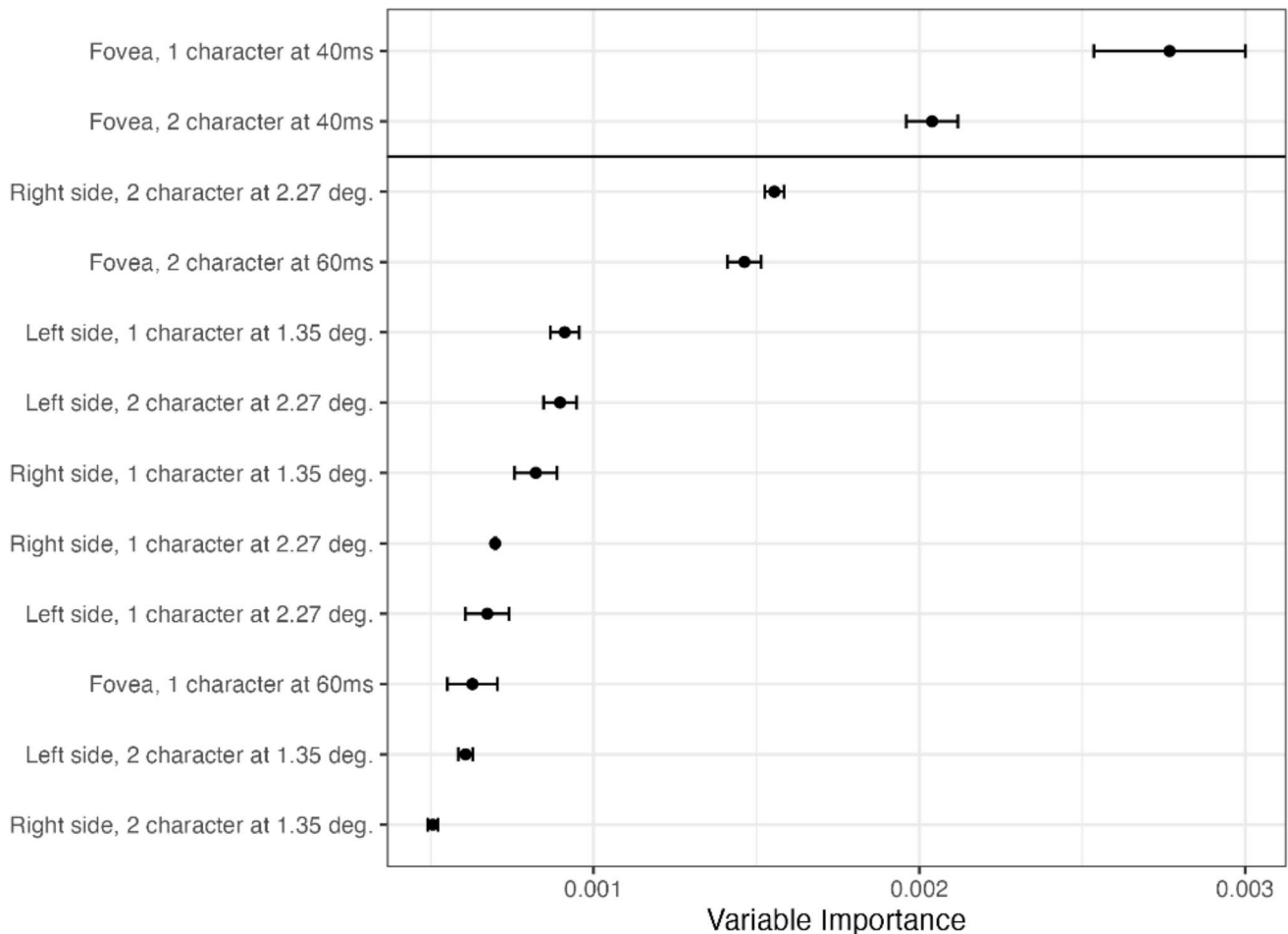


Fig. 5 Relative variable importance estimated by Random Forests models of probability of refixation. The horizontal line indicates the threshold determined by visual inspection for the best predictor vari-

ables chosen for further analysis. The error bars indicate the standard error of variable importance calculated from the scores extracted from each run of the Random Forest with different mtry parameter

were used to index foveal processing during reading and the last two to index parafoveal processing during reading. The non-parametric Random Forests statistical technique was used to estimate the relative importance of all predictor variables obtained from the lexical decision tasks in predicting the word-level eye-movement reading measures. The predictive power of the best predictor variables was further confirmed with (G)LMMs.

Foveal word-processing efficiency best predicts readers' eye movements

According to our results, measures of foveal-processing efficiency were the best predictors of both foveal and parafoveal processing during reading. Individuals able to recognize words in the fovea with the shortest exposure durations also spent the least time reading individual words and were less likely to refixate words. These results are in line with the results obtained in Finnish by Heikkilä et al. (2024), who

also found foveal measures (albeit the shortest exposure duration in lexical decision was the best predictor only for gaze duration) to be the best predictors for the same three eye-movement measures. Measures of foveal-processing efficiency were also the best predictors of parafoveal processing during reading, as the measures of word-skipping probability and saccade length between words were best predicted by the foveal lexical decision task with the shortest exposure durations.

The results on parafoveal processing in reading are also in line with the results obtained by Heikkilä et al. (2024) in Finnish, where foveal processing measures were generally better predictors than parafoveal processing measures of parafoveal processing during reading. Similar to Finnish, the results in Chinese reading can be explained by more efficient parafoveal processing resulting from efficient foveal processing, as depicted in the E-Z Reader model (Reichle et al., 1998, 2009) and in its Chinese script-specific variant, the CEZR model (Liu et al., 2025). These models state that

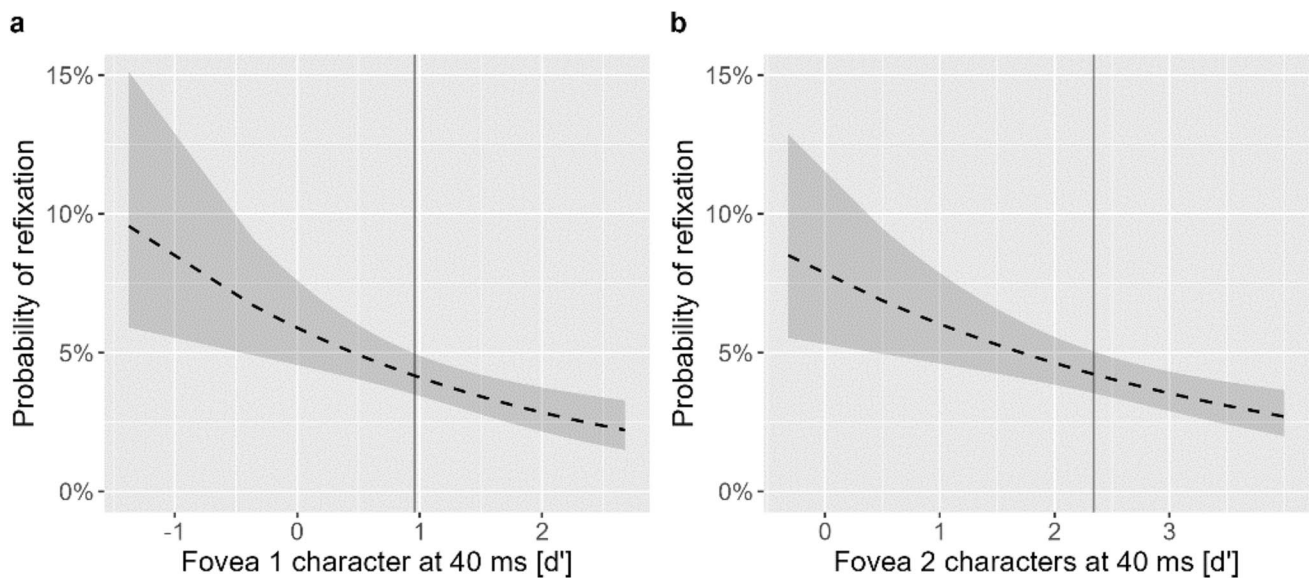


Fig. 6 Estimated marginal generalized linear mixed-effects model (GLMM) means for two different models of probability of refixation as a function of (a) foveal single-character lexical decision at 40 ms; and (b) foveal two-character lexical decision at 40 ms. Shaded bands represent 95% confidence intervals based on standard error. The verti-

cal lines show the point where the main effects of centered variables in the x-axis were estimated by the LMMs (i.e., at $d' = 0.96$ for foveal single-character lexical decision at 40 ms and at $d' = 2.34$ for foveal two-character lexical decision at 40 ms)

the extent of parafoveal processing is dependent on relative ease in processing the currently fixated word. Individuals equipped with efficient foveal processing are able to swiftly recognize the currently fixated word, which allows for their attention to swiftly shift onto the next word for parafoveal processing prior to fixating it. If such parafoveal processing has enough time to enable the parafoveal word to be recognized during the labile stage of saccade programming, the next saccade can be programmed to skip over the parafoveally recognized word. This will then lead to higher word-skipping probability and longer saccades for readers with higher foveal-processing efficiency than for those with poorer foveal processing abilities.

As stated above, individual differences in foveal-processing efficiency were also the best predictors of parafoveal processing during reading, as indexed by word-skipping probability and between-word saccade length. The single-character and two-character foveal-processing efficiency measures with the shortest 40-ms exposure duration were the best predictors. Individuals with higher scores in both the single-character and two-character task were more likely to skip over words and execute longer saccades. In addition, word length modified the pattern of results. All participants were more likely to skip one-character than two-character words. Individuals with high foveal-processing efficiency also skipped two-character words with close to the same probability to one-character words, whereas those with low foveal processing scores skipped two-character words with a very low probability compared to one-character words.

Moreover, performance in the two-character task was also associated with launch site effect. With farther launch sites, individuals with low foveal-processing efficiency hardly skipped words at all and even with near launch sites their skipping rate was still quite modest, whereas readers with high foveal-processing efficiency skipped over words with high probability and with less dependence on the launch site. This pattern of results suggests that readers with better foveal orthographic and semantic word-processing skills show heightened parafoveal-processing capabilities than those with poorer skills. As mentioned above, these results are in line with the findings of Heikkilä et al. (2024), who also found foveal-processing efficiency to be generally better determinants of parafoveal processing in reading Finnish, as indexed by the same eye-movement measures of word-skipping probability and between-word saccade length.

It is important to note that we do not argue that the parafoveal-processing efficiency measures have no predictive power on word processing during reading. The results merely suggest that the individual differences in foveal word-processing efficiency exceed the parafoveal-processing efficiency measures as predictors. Previous studies have shown that individual differences in different components of reading ability can predict the size of the perceptual span (e.g., Choi et al., 2015; Veldre & Andrews, 2014; Veldre et al., 2021). The results of the current study do not refute these findings. In addition, it is noteworthy that we are not the first to demonstrate the superiority of foveal processing skills in predicting parafoveal processing: Veldre and

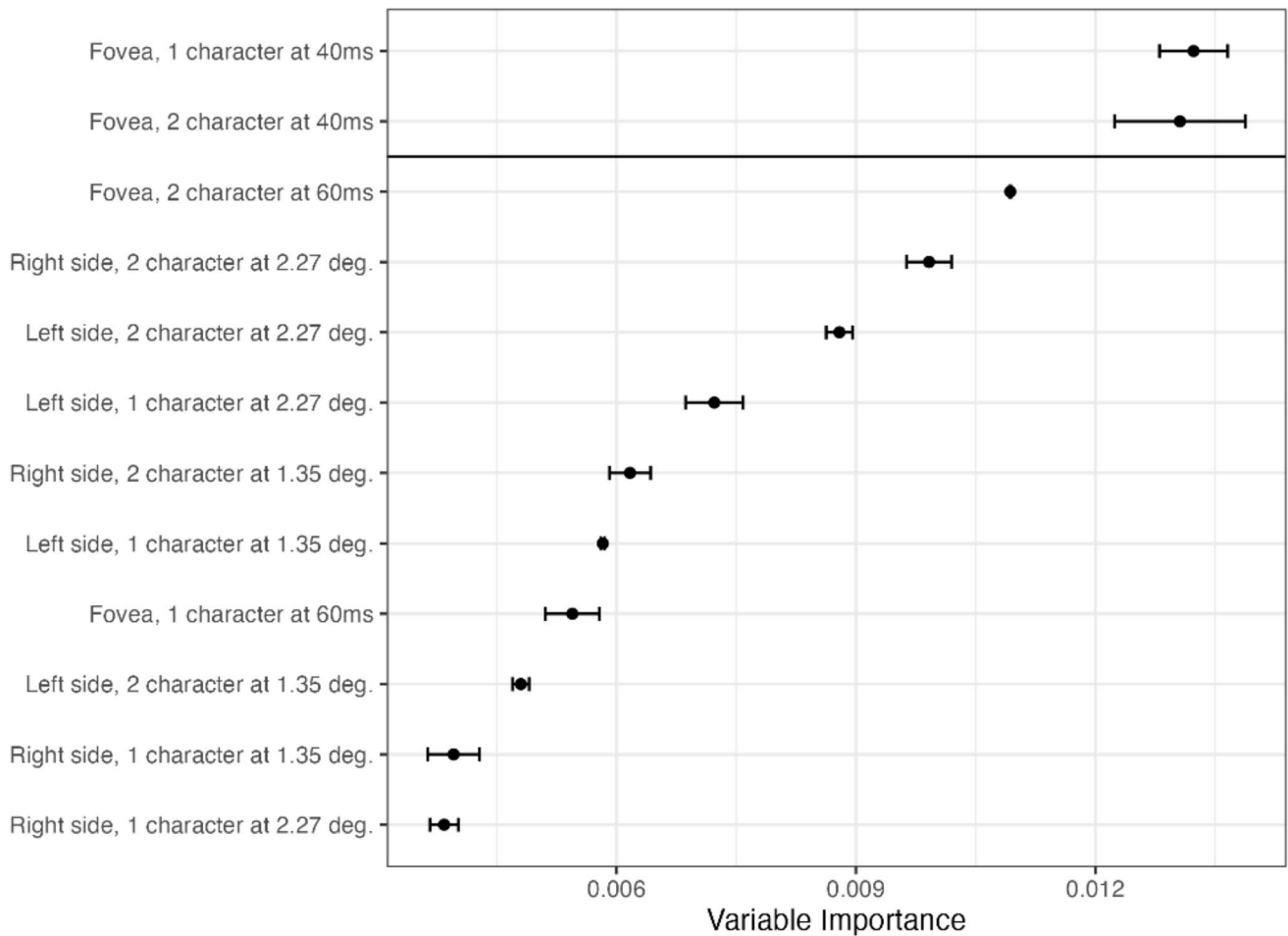


Fig. 7 Relative variable importance estimated by Random Forests models of word skipping probability. The horizontal line indicates the threshold determined by visual inspection for the best predictor vari-

ables chosen for further analysis. The error bars indicate the standard error of variable importance calculated from the scores extracted from each run of the Random Forest with different mtry parameter

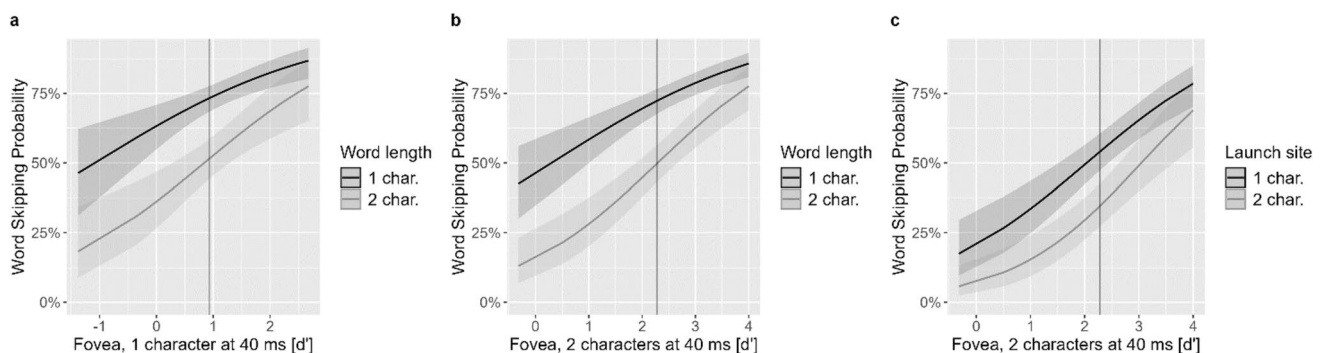


Fig. 8 Estimated marginal generalized linear mixed-effects model (GLMM) means for two models of word-skipping probability as a function of (a) foveal single-character lexical decision at 40 ms and word length; (b) foveal two-character lexical decision at 40 ms and word length; and (c) foveal two-character lexical decision at 40 ms and launch site. Shaded bands represent 95% confidence intervals

based on standard errors. The vertical lines show the point where the main effect of centered variables in the x-axis was estimated by the GLMMs (i.e., at $d' = .93$ for foveal single-character lexical decision at 40 ms and at $d' = 2.28$ for foveal two-character lexical decision at 40 ms)

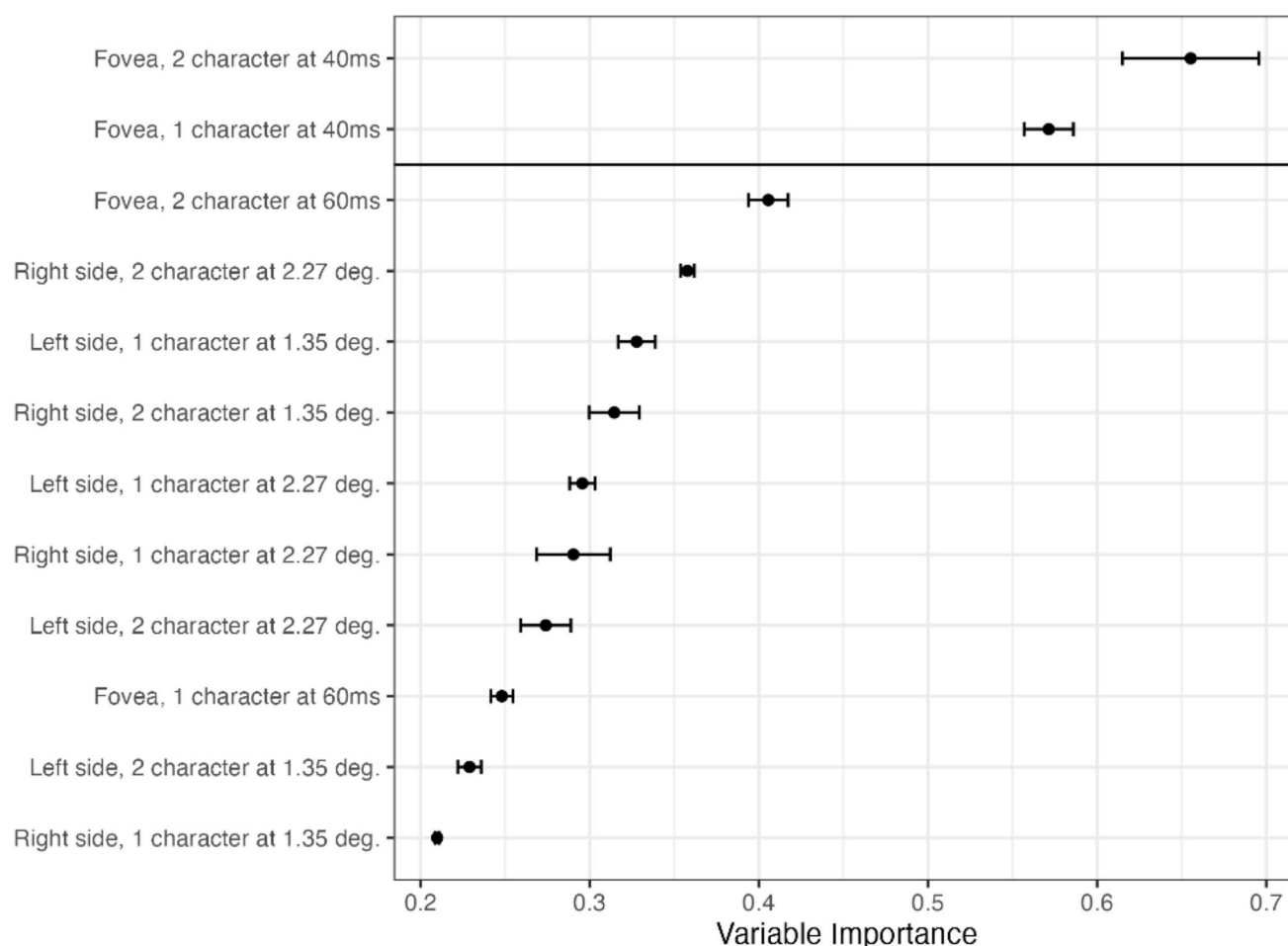


Fig. 9 Relative variable importance estimated by Random Forests models of between word saccade length. The horizontal line indicates the threshold determined by visual inspection for the best predictor variables chosen for further analysis. The error bars indicate

the standard error of variable importance calculated from the scores extracted from each run of the Random Forest with different mtry parameter

Andrews (2015) have previously shown that the parafoveal processing advantage of individuals with high reading and spelling ability can be partly explained via efficient foveal processing.

Shortest foveal exposure durations were consistently the best predictors

The measures with the lowest exposure duration of 40 ms consistently displayed the highest predictive power in each Random Forest model. This indicates that good performance in the most difficult task conditions (i.e., the shortest exposure duration) had the best predictive power for each eye-movement measure, meaning that fast activation of lexical representations in readers' minds is strongly linked to reading fluency. As discussed above, foveal word-processing skills not only predict first-pass reading of words, but also parafoveal word processing. Interestingly, this overall pattern

is similar in Chinese and Finnish reading. As the Chinese and Finnish scripts represent two vastly different scripts, the similarity in results strongly suggest that this individual-difference result is likely to be a language universal.

Efficiency in recognizing two-character words predicts better first-pass reading of words than the recognition of single characters

Another key finding of the present study is that individual differences in the two-character lexical decision task best predicted first-fixation duration and gaze duration during text reading. Also, LMM models confirmed that individuals with higher scores in the two-character lexical decision had shorter first-fixation and gaze durations than those with lower scores. This means that first-pass reading of words in Chinese text is best predicted by the efficiency in combining meanings of consecutive characters into word

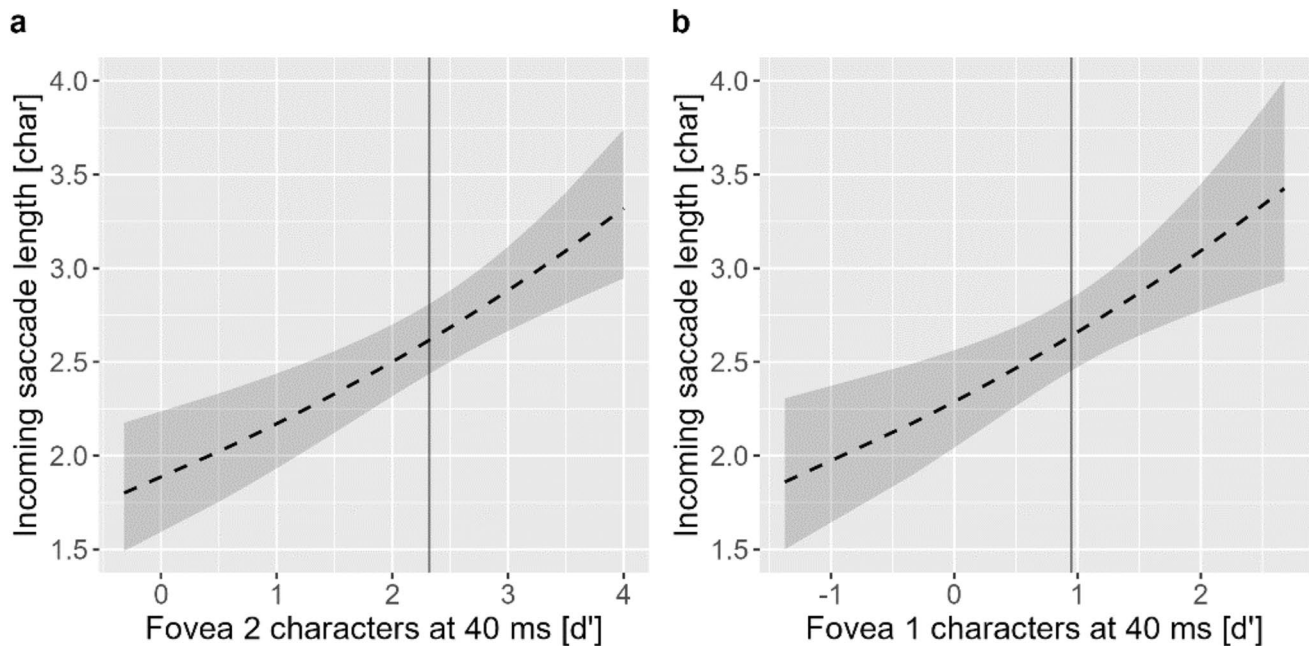


Fig. 10 Estimated marginal linear mixed-effects model (LMM) means for two different models of between-word saccade length as a function of (a) foveal two-character lexical decision at 40 ms; and (b) foveal single-character lexical decision at 40 ms. Shaded bands represent 95% confidence intervals based on standard error. The vertical

lines show the point where the main effects of centered variables in the x-axis were estimated by the LMMs (i.e., at $d' = 2.32$ for foveal two-character lexical decision at 40 ms and at $d' = 0.95$ for foveal single-character lexical decision at 40 ms)

Table 4 Summary of significant highest order effects involving processing efficiency in linear mixed-effects models (LMMs)/generalized (G) LMMs for first-fixation duration, gaze duration, probability of refixation, word skipping probability, and between-word saccade length

Eye-movement variable	Highest order significant effect in (G)LMMs involving processing efficiency		
	Single-character foveal lexical decision at 40 ms	Two-character foveal lexical decision at 40 ms	Two-character foveal lexical decision at 60 ms
First-fixation duration		Main effect	Main effect
Gaze duration		Measure $\times$ Number of characters	Measure $\times$ Number of characters
Probability of refixation	Main effect	Main effect	
Word-skipping probability	Measure $\times$ Number of characters	Measure $\times$ Number of characters, Measure $\times$ Launch site	
Between-word saccade length	Main effect	Main effect	

meaning during foveal processing. This character combination ability is an important component in word segmentation required during reading of Chinese. When reading continuous text, Chinese readers need to constantly parse characters into words by deciding which characters form a bigger meaning unit and where word boundaries are located. It is important to note that this ability to combine characters into bigger meaning units had more predictive power on first-pass duration measures than the recognition of basic meaning units (i.e., recognizing single characters). The finding that individual variability in the ability to efficiently recognize two-character words is an important

component of Chinese reading fits well with the fact that about 70% of Chinese words are two-character words.

These results fit well with the Chinese Reading Model (Li & Pollatsek, 2020). The model describes word recognition in Chinese as a process where all characters within the perceptual span are activated and compete for a winner combination to form a word to be recognized. Individuals who are good at it are swift in finding character combinations within the perceptual span, and according to our results, read words in text with shorter first fixations and gaze durations than those with poorer ability. Also, individuals with good ability in this skill (based on the LMM models) read

single-character and two-character words with similar gaze durations, whereas those with poorer ability need relatively more time to process two-character compared to single-character words, as measured by gaze durations during reading.

The probability of refixating a word during text reading, on the other hand, was best predicted by both single-character and two-character foveal recognition tasks; individuals with higher scores were less likely to refixate words than those with lower scores. The single-character lexical decision task is taken to measure the ability to process lower-level visual character properties (e.g., the combination of strokes and radicals) in order to determine if a character is an existing one or not. Thus, the result indicates that both orthographic and semantic word-processing abilities work as predictors for determining if words need to be refixated during first-pass reading.

A noticeable finding is that performance in one-character word recognition was able to predict word skipping and refixation probabilities, but had a weaker predictive power on word-reading times. As performance in single-character task reflects processing at the character level, this result may be explained via the mechanism of eye-movement target selection described in CRM (Li & Pollatsek, 2020). According to the model, an integral part of successful word recognition is the segmentation of characters comprising the word. After that, a saccade is targeted to land beyond the segmented character string, where new information can be obtained. It may be speculated that the rightmost character of the segmented word is used as an anchor in saccade programming. If so, single-character recognition becomes relevant in this process, which could explain why single-character recognition is a good predictor of skipping and making refixations, both of which entail saccade programming. Future studies are needed to test the validity of this speculation.

Two-character words are easier to recognize than one-character words

An interesting and somewhat surprising result emerged from the scores obtained from the lexical decision tasks. Both in the foveal and the parafoveal lexical decision task, the two-character words were recognized with better accuracy than one-character words. This is surprising in that two-character words contain more information than one-character words, which would suggest an opposite pattern of results. The result suggests that word representations consisting of two characters are stronger than those containing only one character. In other words, the lexical activation of two-character words is stronger, as it is jointly boosted by two rather than just one character.

A similar finding between single-character and two-character recognition accuracies has been reported by Tsang

et al. (2018) in a large dataset containing lexical decision data on 12,578 Chinese words of different length (between one and four characters) obtained from 504 native Chinese speakers. In the study, single-character words were contrasted with pseudowords constructed by adding or substituting strokes from existing characters or by modifying the character radicals. Two-character words were contrasted with pseudowords constructed by random combination of constituent characters. The results related to single-character and two-character lexical decision tasks showed that two-character words elicited both shorter reaction times (750 ms vs. 790 ms) and a lower error rate (4.20% vs. 7.24%) than single-character words. By adding their results on three- and four-character words, a U-shaped relationship was established, where the longest words were more difficult to recognize than the two-character words. It is important to note that in comparison to our study, the lexical decision tasks in Tsang et al. (2018) did not include any stimulus exposure duration manipulations and items were shown only in the central field of view.

According to Tsang et al. (2018), this difference in Chinese word recognition may be due to a similar mechanism suggested previously by Ferrand et al. (2010) based on results in French lexical decision data. Ferrand et al. suggest that difficulty in recognizing short words arises from their orthographic similarity to many other short words that causes interference in recognition even if short words are perceptually simpler than longer words. According to Ferrand et al., longer words, on the other hand, are more orthographically distinctive, which may negate the effect of their perceptual complexity compared to shorter words.

There are two points to be addressed related to the current results on lexical decision accuracy. First, the nature of our pseudocharacters in the single-character task are not visually illegal; rather, they are very rarely used but orthographically legal, real characters. Being extremely rare, participants do not recognize them but may feel a vague familiarity with them, which may trigger additional effort and uncertainty in an attempt to recognize them. This did indeed seem to be the case. There appears a tendency to categorize single-character pseudocharacters as real characters. This becomes apparent in examining the false alarm rates reported in Table 3. The false alarm rate for pseudocharacters in the foveal task is high, as in about 50% of the trials pseudocharacters are considered single-character words. For two-character words, the alarm rate was much lower (about 20%). Thus, as pseudocharacters are orthographically legal, with short exposure durations they sometimes boost sufficient lexical activation in the reader's mind to trigger an erroneous response, that is, considering a pseudocharacter a real word. This seldom happens with two-character combinations.

It has been suggested (Liu et al., 2010) that Chinese readers process single-character words differently than they do

two-character strings. According to this view, Chinese readers process two-character words holistically and one-character words analytically. Analytical processing of one-character words means that readers pay attention to the phonetic and semantic radicals inherent in the characters. Therefore, with stringent exposure times, this analytic process based on radical recognition may result in judging pseudocharacters as real words. On the other hand, in the presumed holistic processing of two-character words, readers pay more attention to the meaning of the character combination. Thus, the non-words may be relatively easily separated from real words. Moreover, in the recognition of two-character words, individual characters boost the activation of the word's lexical representation. Following the above line of reasoning, it may be argued that with short exposure times analytic processing of single characters is more difficult than holistic processing of two-character strings.

Second, why is the recognition of single-character words poorer in the foveal than parafoveal lexical decision? One possible reason is the exposure duration that was longer in the parafoveal (150 ms) than the foveal task (40–60 ms). Another possibility is that single characters are processed differently in the fovea and parafovea. When presented in the fovea, readers may adopt the analytic process described above by focusing on the radicals, that is, character subcomponents. On the other hand, parafoveal presentation does not allow the processing of visual details; instead, readers may base their decisions on the familiarity of the overall visual shape of the characters. As a consequence, pseudocharacters may no longer trigger sufficient activation to be erroneously recognized as real characters. If so, readers are less likely to make false alarms with the parafoveal stimulus. In fact, as may be seen in Table 3, the false alarm rate for pseudocharacters is half the size for parafoveal than for foveal stimuli (about 25% vs. 50%).

In contrast to visually illegal pseudocharacters that can often be rejected rapidly based on orthographic and visual cues, as argued above, or orthographically legal but extremely rare pseudocharacters probably do not elicit an early rejection. However, it should be noted that in other experiments (e.g., Tsang et al., 2018) pseudocharacters have been generated. From this perspective, there appears to be no substantial difference between these “constructed” pseudocharacters and those extremely rare, unknown real characters.

The use of legal pseudocharacters in single-character lexical decision may be questioned. However, we argue that their use is justifiable. First, the study of Tsang et al. (2018) reviewed above observed poorer recognition of single- than two-character words by generating pseudocharacters by combining radicals from different characters to create new forms with legal orthographic structures. Second, and more importantly, in a yet-to-be published study (Fu et al., 2026),

a direct comparison of preview effects has been done by using identities, matched but unrelated characters, pseudocharacters (created by Windows true font software), and very rarely used unknown characters. The results showed no differences between the latter two conditions across all measures in the pretarget, target, and post-target regions. This result confirms that very rarely used unknown characters can equally well serve as excellent baselines as “constructed” pseudocharacters. There are also other studies (e.g., Zang et al., 2020, 2023; Zang, Fu, et al., 2024a, 2024b; Zang, Wang, et al., 2024a, 2024b) that have successfully used these extremely rare, unknown characters as pseudocharacter previews.

Comparison of lexical decision accuracies in Chinese and Finnish

As the present study closely mimics that of Heikkilä et al. (2024) conducted in Finnish, it is possible to compare lexical decision accuracies across the two languages. In Finnish foveal lexical decision, the words were six- to eight-letter words, while in the parafoveal task they were five-letter words. The following observations may be made. For both exposure durations (40 ms and 60 ms), the foveal recognition of two-character Chinese words was at the high level (86% and 88%). In Finnish, the recognition accuracy reached that level (87%) with the 60-ms exposure time but was lower (73%) with the shorter exposure time. The foveal recognition accuracy was lowest for one-character Chinese words (67% vs. 69%). In sum, two-character Chinese words seem to be particularly easy to recognize even with short exposure times. Above, we have sketched a possible explanation for this result.

In parafoveal recognition accuracy, pushing the word further into the parafovea had a bigger effect in Finnish (85% vs. 75% for the near and far condition, respectively) than in Chinese (86% vs. 82% for one-character words; 91% vs. 87% for two-character words). It is also noticeable that two-character Chinese words are easier to recognize parafoveally than five-letter Finnish words, particularly in the far condition. This may reflect Chinese readers' improved parafoveal processing skills. They are likely to be more important in Chinese reading, because readers of Chinese need to constantly segment the text into words in the situation where the words are not separated from each other by spaces, as is the case in alphabetic scripts. In other words, in finding word boundaries, Chinese readers constantly need to extend their visual attention to the parafovea.

The pseudowords in the Heikkilä et al. (2024) study were constructed by changing one or two letters from five-letter Finnish words. Veldre et al. (2023b) reported data on a more robust manipulation where they changed all the vowels of English words with similar-looking consonants to form

nonwords. In this condition, discrimination can be based on orthographic legality. In the present study, the two-character nonwords resemble these orthographically illegal nonwords. When comparing the lexical decision accuracies between the two studies, it appears that the parafoveal performance in the two-character task in Chinese is slightly better than in the English study of Veldre et al. (2023b) with illegal non-words. The overall mean accuracy of the two-character words presented in the parafovea is 88% compared to 83% obtained by Veldre et al. It may reflect Chinese readers' improved parafoveal processing.

Limitations

This study contains limitations related to the participant sample, the items, and the task. They are briefly discussed below. The participants were chosen from different backgrounds to increase the heterogeneity of the sample. Differences in the age distribution in the background groups present a possible confound since all the older participants were members of the non-academic group. Hence, it is not possible to distinguish whether some of the individual variation in the data is due to age, background related reading skill, or their combination.

The items in the lexical decision sets were matched on key lexical and structural properties between the foveal and parafoveal versions to achieve item comparability between the two tasks. In hindsight, counterbalanced lists could have been used to more fully control for potential item-level differences between the tasks.

Finally, the foveal and parafoveal speeded lexical decision tasks employed in the present study do not capture two

important processes inherent in natural Chinese reading. First, as single-character and two-character words were presented in isolation for lexical decision, the task does not capture the segmentation demands inherent when processing long character strings in running text. Second, in isolated word recognition, visual crowding caused by neighboring characters is absent, which is also true for the Finnish study of Heikkilä et al. (2024). Hence, the lexical decision tasks employed in the present study may not fully capture all the processing demands of natural Chinese reading. However, as reviewed above, they do capture important individual differences among adult Chinese readers.

Conclusions

The results of the current study show that individual differences in foveal word-processing efficiency are better predictors of foveal and parafoveal processing during reading Chinese than individual differences in parafoveal word-processing efficiency. This result is analogous to that obtained in Finnish reading (Heikkilä et al., 2024), which speaks for the universality of the pattern of results, as the two scripts are at the opposite ends of the script continuum (logographic vs. perfectly regular alphabetic script). Efficiency in combining individual characters into words seems to be an important determinant in Chinese reading fluency, as constant word segmentation is a fundamental requirement of the language. This result fits well with the Chinese Reading Model of Li and Pollatsek (2020). Efficiency in recognizing basic meaning units (single characters) also plays a role in fluent Chinese reading.

Appendix

Table 5 Correlations among the single and two-character lexical decision accuracies (d') for each foveal exposure duration and each parafoveal eccentricity in both right and left parafoveal field

	1	2	3	4	5	6	7	8	9	10	11	12
1. Fovea 1CH 60 ms	-											
2. Fovea 1CH 40 ms	.66	-										
3. Right 1CH 1.35°	.49	.31	-									
4. Left 1CH 1.35°	.40	.40	.62	-								
5. Right 1CH 2.27°	.35	.41	.56	.55	-							
6. Left 1CH 2.27°	.56	.41	.60	.73	.49	-						
7. Fovea 2CH 60 ms	.50	.45	.66	.72	.50	.63	-					
8. Fovea 2CH 40 ms	.59	.54	.63	.60	.49	.61	.81	-				
9. Right 2CH 1.35°	.54	.41	.59	.52	.60	.57	.65	.67	-			
10. Left 2CH 1.35°	.49	.45	.49	.58	.51	.51	.70	.68	.63	-		
11. Right 2CH 2.27°	.54	.48	.57	.61	.66	.65	.73	.75	.71	.65	-	
12. Left 2CH 2.27°	.54	.44	.59	.59	.43	.70	.67	.68	.66	.65	.73	-

1CH = single-character lexical decision, 2CH = two-character lexical decision, Right = right parafoveal field, Left = left parafoveal field

Table 6 LMM fixed effects for three first fixation duration models: (A) number of characters $\times$ word frequency $\times$ foveal two-character lexical decision at 40 ms + number of strokes; (B) number of characters $\times$ word frequency $\times$ parafoveal two-character lexical decision at1.35 degrees to the left + number of strokes; and (C) number of characters $\times$ word frequency $\times$ foveal two-character lexical decision at 60 ms + number of strokes. Statistically significant effects ($t > 2.39$) are shown in bold

Model	Predictor	Estimate	SE	<i>t</i>
A	Intercept	5.307	.016	333.348
	Fovea at 40 ms (2CH)	-.046	.014	-3.151
	Word frequency	-.001	.003	-.414
	Number of characters	-.037	.010	-3.827
	Number of strokes	.003	.001	3.363
	Fovea at 40 ms (2CH) $\times$ Word frequency	-.003	.001	-2.155
	Fovea at 40 ms (2CH) $\times$ Number of characters	-.003	.004	-.812
	Word frequency $\times$ Number of Characters	-.006	.003	-1.633
	Fovea at 40 ms (2CH) $\times$ Word frequency $\times$ Number of characters	.001	.002	.882
	Variance components			
	σ^2	.139		
	τ_{00} Word	.005		
	τ_{00} Subject	.014		
B	Intercept	5.302	.017	315.418
	Parafovea left at 1.35° (2CH)	-.016	.021	-.749
	Word frequency	-.001	.003	-.395
	Number of characters	-.037	.010	-3.824
	Number of strokes	.003	.001	3.368
	Parafovea left at 1.35° (2CH) $\times$ Word frequency	-.004	.002	-2.387
	Parafovea left at 1.35° (2CH) $\times$ Number of characters	-.006	.005	-1.202
	Word frequency $\times$ Number of Characters	-.006	.003	-1.648
	Parafovea left at 1.35° (2CH) $\times$ Word frequency $\times$ Number of characters	.003	.002	1.290
	Variance components			
	σ^2	.139		
	τ_{00} Word	.005		
	τ_{00} Subject	.016		
C	Intercept	5.307	.016	332.555
	Fovea Lexdec at 60 ms (2CH)	-.050	.017	-3.009
	Word frequency	-.001	.003	-.406
	Number of characters	-.037	.010	-3.829
	Number of strokes	.003	.001	3.366
	Fovea Lexdec at 60 ms (2CH) $\times$ Word frequency	-.003	.001	-2.168
	Fovea Lexdec at 60 ms (2CH) $\times$ Number of characters	-.006	.004	-1.399
	Word frequency $\times$ Number of Characters	-.006	.003	-1.643
	Fovea Lexdec at 60 ms (2CH) $\times$ Word frequency $\times$ Number of characters	.002	.002	1.036
	Variance components			
	σ^2	.139		
	τ_{00} Word	.005		
	τ_{00} Subject	.014		

First fixation duration was log-transformed. SE = standard error; 1CH = one-character word; 2CH = two-character word. N of subjects = 70, N of words = 674, N of observations = 51084

Table 7 LMM fixed effects for two gaze duration models: (A) number of characters $\times$ word frequency $\times$ foveal two-character lexical decision at 40 ms + number of strokes; and (B) number of charac-ters $\times$ word frequency $\times$ foveal two-character lexical decision at 60 ms + number of strokes. Statistically significant effects ($t > 2.24$) are shown in bold

Model	Predictor	Estimate	SE	<i>t</i>
A	Intercept	5.338	.017	323.278
	Fovea Lexical decision at 40 ms (2CH)	-.055	.015	-3.732
	Word frequency	-.002	.003	-.706
	Number of characters	.009	.012	.755
	Number of strokes	.003	.001	2.909
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency	-.002	.001	-1.819
	Fovea Lexdec at 40 ms (2CH) $\times$ Number of characters	-.026	.006	-4.442
	Word frequency $\times$ Number of Characters	-.009	.004	-2.302
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency $\times$ Number of characters	.000	.002	.286
	Variance components			
	σ^2	.173		
	τ_{00} Word	.007		
	τ_{00} Subject	.014		
	τ_{11} Subject.Number of Characters ²	.001		
	τ_{11} Subject.Number of Strokes	.000		
	ρ_{01} Subject.Number of Characters ²	.514		
	ρ_{01} Subject.Number of Strokes	-.090		
	ρ_{01} Subject. Number of Characters ² :Number of Strokes	-.203		
B	Intercept	5.338	.017	322.066
	Fovea Lexical decision at 60 ms (2CH)	-.06	.017	-3.541
	Word frequency	-.002	.003	-.701
	Number of characters	.009	.012	.764
	Number of strokes	.003	.001	2.918
	Fovea Lexdec at 60 ms (2CH) $\times$ Word frequency	-.003	.001	-1.751
	Fovea Lexdec at 60 ms (2CH) $\times$ Number of characters	-.031	.006	-4.803
	Word frequency $\times$ Number of Characters	-.01	.004	-2.306
	Fovea Lexdec at 60 ms (2CH) $\times$ Word frequency $\times$ Number of characters	.001	.002	.277
	Variance components			
	σ^2	.173		
	τ_{00} Word	.007		
	τ_{00} Subject	.014		
	τ_{11} Subject.Number of Characters ²	.001		
	τ_{11} Subject.Number of Strokes	.000		
	ρ_{01} Subject.Number of Characters ²	.515		
	ρ_{01} Subject.Number of Strokes	-.110		
	ρ_{01} Subject. Number of Characters ² :Number of Strokes	-.239		

Gaze duration was log-transformed. *SE* = standard error; 2CH = two-character word. *N* of subjects = 70, *N* of words = 674, *N* of observations = 50,941

Table 8 GLMM fixed effects for two probability of refixation models: (A) number of characters $\times$ word frequency $\times$ foveal single-character lexical decision at 40 ms + number of strokes; and (B) number of characters $\times$ word frequency $\times$ foveal two-character lexical decision at 40 ms + number of strokes. Statistically significant effects ($z > 2.24$) are shown in bold

Model	Predictor	Estimate	SE	z
A	Intercept	.044	.004	-33.238
	Number of characters	2.481	.210	10.75
	Fovea Lexical decision at 40 ms (1CH)	.684	.073	-3.57
	Word frequency	.957	.022	-1.881
	Number of strokes	1.007	.007	.995
	Fovea Lexdec at 40 ms (1CH) $\times$ Number of characters	.886	.063	-1.718
	Word frequency $\times$ Number of Characters	.985	.028	-.548
	Fovea Lexdec at 40 ms (1CH) $\times$ Word frequency	1.001	.020	.075
	Fovea Lexdec at 40 ms (1CH) $\times$ Word frequency $\times$ Number of characters	.976	.022	-1.052
	Variance components			
	σ^2	3.290		
	τ_{00} Word	0.211		
	τ_{00} Subject	0.354		
	τ_{11} Subject.Number of Characters2	0.052		
	ρ_{01} Subject	-0.052		
B	Intercept	.044	.004	-33.011
	Number of characters	2.477	.208	10.806
	Fovea Lexical decision at 40 ms (2CH)	.756	.060	-3.502
	Word frequency	.955	.022	-1.969
	Number of strokes	1.007	.007	1.001
	Fovea Lexdec at 40 ms (2CH) $\times$ Number of characters	.911	.044	-1.917
	Word frequency $\times$ Number of Characters	.986	.028	-.502
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency	.995	.013	-.421
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency $\times$ Number of characters	.987	.015	-.911
	Variance components			
	σ^2	3.290		
	τ_{00} Word	.211		
	τ_{00} Subject	.356		
	τ_{11} Subject.Number of Characters2	.048		
	ρ_{01} Subject	-.032		

SE = standard error; 1CH = one-character word; 2CH = two-character word. N of subjects = 70, N of words = 674, N of observations = 50,941

Table 9 GLMM fixed effects for two word skipping probability models: (A) launch site × word length × foveal single-character lexical decision at 40 ms + launch site × word frequency × foveal single-character lexical decision at 40 ms + number of strokes; and (B)launch site × word length × foveal two-character lexical decision at 40 ms + launch site × word frequency × foveal two-character lexical decision at 40 ms + number of strokes. Statistically significant effects ($z > 2.24$) are shown in bold

Model	Predictor	Estimate	SE	z
A	Intercept	2.738	.329	8.374
	Launch site	.460	.026	-13.615
	Number of characters	.388	.045	-8.234
	Fovea 40 ms (1CH)	1.649	.211	3.907
	Word frequency	1.089	.022	4.294
	Number of strokes	.971	.010	-3.031
	Launch site × Number of characters	1.200	.060	3.657
	Launch site × Fovea 40 ms (1CH)	1.139	.081	1.827
	Number of characters × Fovea 40 ms (1CH)	1.192	.084	2.502
	Launch site × Word frequency	1.036	.011	3.436
	Fovea 40 ms (1CH) × Word frequency	.989	.009	-1.250
	Launch site × Number of characters × Fovea 40 ms (1CH)	.878	.053	-2.155
	Launch site × Fovea 40 ms (1CH) × Word frequency	.985	.012	-1.220
	Variance components			
	σ^2	3.290		
	τ_{00} Word	.711		
	τ_{00} Subject	.668		
	τ_{11} Subject.Launch Site	.140		
	τ_{11} Subject.Number of Characters2	.133		
	ρ_{01} Subject.Launch Site	.473		
	ρ_{01} Subject.Number of Characters2	.655		
	ρ_{01} Subject.Number of Characters2:Launch Site	.965		
B	Intercept	2.611	.298	8.411
	Launch site	.450	.025	-14.517
	Number of characters	.380	.043	-8.526
	Fovea 40 ms (2CH)	1.625	.149	5.309
	Word frequency	1.088	.021	4.280
	Number of strokes	.971	.009	-3.055
	Launch site × Number of characters	1.195	.060	3.535
	Launch site × Fovea 40 ms (2CH)	1.191	.062	3.339
	Number of characters × Fovea 40 ms (2CH)	1.273	.065	4.729
	Launch site × Word frequency	1.034	.011	3.249
	Fovea 40 ms (2CH) × Word frequency	1.012	.006	1.871
	Launch site × Number of characters × Fovea 40 ms (2CH)	.952	.045	-1.044
	Launch site × Fovea 40 ms (2CH) × Word frequency	.986	.009	-1.54
	Variance components			
	σ^2	3.290		
	τ_{00} Word	.705		
	τ_{00} Subject	.560		
	τ_{11} Subject.Launch Site	.121		
	τ_{11} Subject.Number of Characters2	.112		
	ρ_{01} Subject.Launch Site	.331		
	ρ_{01} Subject.Number of Characters2	.598		
	ρ_{01} Subject.Number of Characters2:Launch Site	.936		

SE = standard error; 1CH = one-character word; 2CH = two-character word. N of subjects = 70, N of words = 674, N of observations = 33,729

Table 10 LMM fixed effects for two between-word saccade length models: (A) number of characters $\times$ word frequency $\times$ foveal two-character lexical decision at 40 ms + number of strokes; and (B)number of characters $\times$ word frequency $\times$ foveal single-character lexical decision at 40 ms + number of strokes. Statistically significant effects ($t > 2.24$) are shown in bold

Model	Predictor	Estimate	SE	<i>t</i>
A	Intercept	.962	.036	26.506
	Fovea Lexdec at 40 ms (2CH)	.142	.032	4.415
	Word frequency	.017	.007	2.569
	Number of characters	.131	.024	5.506
	Number of strokes	-.006	.002	-2.798
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency	.003	.002	1.732
	Fovea Lexdec at 40 ms (2CH) $\times$ Number of characters	.003	.006	.501
	Word frequency $\times$ Number of Characters	.002	.009	.269
	Fovea Lexdec at 40 ms (2CH) $\times$ Word frequency $\times$ Number of characters	-.001	.002	-.529
	Variance components			
	σ^2	.253		
	τ_{00} Word	.038		
	τ_{00} Subject	.072		
	τ_{11} Subject.Word Frequency	.000		
	τ_{11} Subject.Number of Characters2	.001		
	ρ_{01} Subject.Word Frequency	.041		
	ρ_{01} Subject.Number of Characters2	-.552		
	ρ_{01} Subject.Number of Characters2:Word Frequency	.413		
B	Intercept	.971	.037	25.970
	Fovea Lexdec at 40 ms (1CH)	.151	.042	3.592
	Word frequency	.018	.007	2.639
	Number of characters	.129	.024	5.416
	Number of strokes	-.006	.002	-2.673
	Fovea Lexdec at 40 ms (1CH) $\times$ Word frequency	.001	.003	.516
	Fovea Lexdec at 40 ms (1CH) $\times$ Number of characters	-.001	.009	-.069
	Word frequency $\times$ Number of Characters	.002	.009	.267
	Fovea Lexdec at 40 ms (1CH) $\times$ Word frequency $\times$ Number of characters	-.002	.003	-.61
	Variance components			
	σ^2	.253		
	τ_{00} Word	.038		
	τ_{00} Subject	.078		
	τ_{11} Subject.Word Frequency	.000		
	τ_{11} Subject.Number of Characters2	.001		
	ρ_{01} Subject.Word Frequency	.210		
	ρ_{01} Subject.Number of Characters2	-.440		
	ρ_{01} Subject.Number of Characters2:Word Frequency	.439		

Saccade length was log-transformed. SE = standard error; 1CH = one-character word; 2CH = two-character word. N of subjects = 70, N of words = 674, N of observations = 47,669

Author note Preliminary results of this study were presented at the 9th China International Conference on Eye Movements held in September 2023 at Tianjin Normal University, China and at the 22nd European Conference on Eye Movements held in August 2024 at Maynooth University, Ireland.

Authors' contributions Timo T. Heikkilä: Writing – original draft, writing – review and editing, visualization, software, methodology, formal analysis, data curation, conceptualization, project administration. Yongsheng Wang: Investigation, resources. Chuanli Zang:

Writing – review and editing, resources. Xin Li: Resources. Jukka Hyönä: Writing – review and editing, supervision, project administration, funding acquisition, conceptualization.

Funding Open Access funding provided by University of Turku (including Turku University Central Hospital). This work was supported by the Academy of Finland (grant # 315963) given to the fifth author.

Data availability The data are available at <https://osf.io/wnx7d>. For the materials used in the reading task, please consult the articles of Liversedge et al. (2016, 2024).

Code availability The scripts for statistical analyses are available at <https://osf.io/wnx7d>.

Declarations

Ethics approval This study is a part of the LogoAlpha -project, which was approved by the Ethics committee of the University of Turku (#29/2019).

Consent to participate Informed consent was obtained from all individual participants included in the study.

Consent for publication Informed consent was obtained from all individual participants included in the study, including consent for publication.

Conflicts of interest The authors have no conflicts of interest to declare.

Open practices statement The data and R-scripts for all the statistical analyses are available in the Open Science Framework repository at: <https://osf.io/wnx7d>. For the full original materials used in the reading task consult the original articles of Liversedge et al. (2016, 2024). The study design and analysis were not pre-registered.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. arXiv:1506.04967.
- Cai, Q., & Brysbaert, M. (2010). SUBTLEX-CH: Chinese word and character frequencies based on film subtitles. *PLoS One*, 5(6), Article e10729. <https://doi.org/10.1371/journal.pone.0010729>
- Choi, W., Lowder, M. W., Ferreira, F., & Henderson, J. M. (2015). Individual differences in the perceptual span during reading: Evidence from the moving window technique. *Attention, Perception & Psychophysics*, 77(7), 2463–2475. <https://doi.org/10.3758/s13414-015-0942->
- Ferrand, L., New, B., Brysbaert, M., Keuleers, E., Bonin, P., Méot, A., Augustinova, M., & Pallier, C. (2010). The French lexicon project: Lexical decision data for 38,840 French words and 38,840 pseudowords. *Behavior Research Methods*, 42, 488–496. <https://doi.org/10.3758/BRM.42.2.488>
- Fu, Y., Liversedge, S. P., Bai, X., & Zang, C. (2026). The selection of parafoveal preview baseline during Chinese reading. Manuscript in preparation.
- Gong, T., & Shuai, L. (2023). Segmented relations between online reading behaviors, text properties, and reader–text interactions: An eye-movement experiment. *Frontiers in Psychology*, 13, 1006662. <https://doi.org/10.3389/fpsyg.2022.1006662>
- Heikkilä, T. T., Soralinna, N., & Hyönä, J. (2024). Relating foveal and parafoveal-processing efficiency with word-level parameters in text reading. *Journal of Memory and Language*, 137, Article 104516. <https://doi.org/10.1016/j.jml.2024.104516>
- Hothorn, T., Hornik, K., & Zeileis, A. (2006). Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*, 15(3), 651–674. <https://doi.org/10.1198/106186006X133933>
- Hyönä, J. (2011). Foveal and parafoveal processing during reading. In S. Liversedge, I. Gilchrist, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 819–838). Oxford University Press.
- Institute of Language Teaching and Research. (1986). *Modern Chinese frequency dictionary*. Beijing Language Institute Press.
- Kuperman, V., Matsuki, K., & Van Dyke, J. A. (2018). Contributions of reader- and text level characteristics to eye-movement patterns during passage reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(11), 1687–1713. <https://doi.org/10.1037/xlm0000547>
- Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review*, 127(6), 1139–1162. <https://doi.org/10.1037/rev0000248>
- Li, X., Huang, L., Yao, P., & Hyönä, J. (2022). Universal and specific reading mechanisms across different writing systems. *Nature Reviews Psychology*, 1, 133–144. <https://doi.org/10.1038/s44159-022-00022-6>
- Liu, P. D., Chung, K. K. H., McBride-Chang, C., & Tong, X. (2010). Holistic versus analytic processing: Evidence for a different approach to processing of Chinese at the word and character levels in Chinese children. *Journal of Experimental Child Psychology*, 107(4), 466–478. <https://doi.org/10.1016/j.jecp.2010.06.006>
- Liu, Y., Yu, L., & Reichle, E. D. (2025). Towards a model of eye-movement control in Chinese reading. *Psychonomic Bulletin & Review*, 32, 493–527. <https://doi.org/10.3758/s13423-024-02570-9>
- Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, 147, 1–20. <https://doi.org/10.1016/j.cognition.2015.10.013>
- Liversedge, S. P., Olkonien, H., Zang, C., Li, X., Yan, G., Bai, X., & Hyönä, J. (2024). Universality in eye movements and reading: A replication with increased power. *Cognition*, 242, Article 105636. <https://doi.org/10.1016/j.cognition.2023.105636>
- Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, 3(22), Article 470. <https://doi.org/10.21105/joss.00470>
- Matsuki, K., Kuperman, V., & Van Dyke, J. A. (2016). The random forests statistical technique: An examination of its value for the study of reading. *Scientific Studies of Reading*, 20(1), 20–33. <https://doi.org/10.1080/10888438.2015.1107073>
- Rayner, K., Fischer, M. H., & Pollatsek, A. (1998). Unspaced text interferes with both word identification and eye movement control. *Vision Research*, 38(8), 1129–44. [https://doi.org/10.1016/s0042-6989\(97\)00274-5](https://doi.org/10.1016/s0042-6989(97)00274-5)
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105(1), 125–157. <https://doi.org/10.1037/0033-295X.105.1.125>
- Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z reader to model the effects of higher-level language processing

- on eye movements during reading. *Psychonomic Bulletin & Review*, 16, 1–21. <https://doi.org/10.3758/PBR.16.1.1>
- Schmidtke, D., Van Dyke, J. A., & Kuperman, V. (2018). Individual variability in the semantic processing of English compound words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(3), 421–439. <https://doi.org/10.1037/xlm0000442>
- Schotter, E. R., Angele, B., & Rayner, K. (2012). Parafoveal processing in reading. *Attention, Perception & Psychophysics*, 74, 5–35. <https://doi.org/10.3758/s13414-011-0219-2>
- Strobl, C., Boulesteix, A. L., Kneib, T., Augustin, T., & Zeileis, A. (2008). Conditional variable importance for random forests. *BMC Bioinformatics*, 9(1), Article 307.
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinformatics*, 8, 25. <https://doi.org/10.1186/1471-2105-8-25>
- Taylor, J. N., & Perfetti, C. A. (2016). Eye movements reveal readers' lexical quality and reading experience. *Reading and Writing*, 29(6), 1069–1103. <https://doi.org/10.1007/s11145-015-9616-6>
- Tsang, Y.-K., Huang, J., Lui, M., Xue, M., Chan, Y. W. F., Wang, S., & Chen, H.-C. (2018). MELD-SCH: A megastudy of lexical decision in simplified Chinese. *Behavior Research Methods*, 50, 1763–1777. <https://doi.org/10.3758/s13428-017-0944-0>
- Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *The Quarterly Journal of Experimental Psychology*, 67(4), 703–727. <https://doi.org/10.1080/17470218.2013.826258>
- Veldre, A., & Andrews, S. (2015). Parafoveal preview benefit is modulated by the precision of skilled readers' lexical representations. *Journal of Experimental Psychology: Human Perception and Performance*, 41(1), 219–232. <https://doi.org/10.1037/xhp0000017>
- Veldre, A., Wong, R., & Andrews, S. (2021). Reading proficiency predicts the extent of the right, but not left, perceptual span in older readers. *Attention, Perception, & Psychophysics*, 83(1), 18–26. <https://doi.org/10.3758/s13414-020-02185-x>
- Veldre, A., Reichle, E. D., Yu, L., & Andrews, S. (2023a). Understanding the visual constraints on lexical processing: New empirical and simulation results. *Journal of Experimental Psychology: General*, 152(3), 693–722. <https://doi.org/10.1037/xge000129>
- Veldre, A., Reichle, E. D., Yu, L., & Andrews, S. (2023b). Lexical processing across the visual field. *Journal of Experimental Psychology: Human Perception, and Performance*, 49(5), 649–671. <https://doi.org/10.1037/xhp0001109>
- Xing, H. (邢红兵). (2007). 现代汉字特征分析与计算研究 [The feature analysis of Chinese character and computation research]. 商务印书馆 (The Commercial Press).
- Zang, C., Du, H., Bai, X., Yan, G., & Livversedge, S. P. (2020). Word skipping in Chinese reading: The role of high-frequency preview and syntactic felicity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 46(4), 603–620.
- Zang, C., Fu, Y., Du, H., Bai, X., Yan, G., & Livversedge, S. P. (2024). Processing multi-constituent units: Preview effects during reading of Chinese words, idioms and phrases. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(1), 169–188. <https://doi.org/10.1037/xlm0001234>
- Zang, C., Livversedge, S. P., Bai, X., & Yan, G. (2011). Eye movements during Chinese reading. In S. P. Livversedge, I. D. Gilchrist, & S. Everling (Eds.), *The Oxford handbook on eye movements* (pp. 961–978). Oxford University Press.
- Zang, C., Wang, S., Bai, X., Yan, G., & Livversedge, S. P. (2024). Parafoveal processing of Chinese four-character idioms and phrases in reading: Evidence for multi-constituent unit hypothesis. *Journal of Memory and Language*, 136, Article 104508. <https://doi.org/10.1037/xlm0001234>
- Zang, C., Zhang, Z., Zhang, M., Degno, F., & Livversedge, S. P. (2023). Examining semantic parafoveal-on-foveal effects using a Stroop boundary paradigm. *Journal of Memory and Language*, 128, Article 104387. <https://doi.org/10.1016/j.jml.2022.104387>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.