

This is a self-archived version of an original article. This version may differ from the original in pagination and typographic details.

Author(s): Loberg, Otto; Hautala, Jarkko; Hämäläinen, Jarmo A.; Leppänen, Paavo H.T.

Title: Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading

Year: 2019

Version: Accepted version (Final draft)

Copyright: © 2019 Elsevier Ltd.

Rights: CC BY-NC-ND 4.0

Rights url: <https://creativecommons.org/licenses/by-nc-nd/4.0/>

Please cite the original version:

Loberg, O., Hautala, J., Hämäläinen, J. A., & Leppänen, P. H. (2019). Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading. *Vision Research*, 165, 109-122. <https://doi.org/10.1016/j.visres.2019.07.008>

Influence of reading skill and word length on fixation-related brain activity in school-aged children during natural reading

Otto Loberg, Jarkko Hautala, Jarmo A. Hämäläinen & Paavo H.T. Leppänen

Abstract

Word length is one of the main determinants of eye movements during reading and has been shown to influence slow readers more strongly than typical readers. The influence of word length on reading in individuals with different reading skill levels has been shown in separate eye-tracking and electroencephalography studies. However, the influence of reading difficulty on cortical correlates of word length effect during natural reading is unknown. To investigate how reading skill is related to brain activity during natural reading, we performed an exploratory analysis on our data set from a previous study, where slow reading (N=27) and typically reading (N=65) 12-to-13.5-year-old children read sentences while co-registered ET-EEG was recorded. We extracted fixation-related potentials (FRPs) from the sentences using the linear deconvolution approach. We examined standard eye-movement variables and deconvoluted FRP estimates: intercept of the response, categorical effect of first fixation versus additional fixation and continuous effect of word length. We replicated the pattern of stronger word length effect in eye movements for slow readers. We found a difference between typical readers and slow readers in the FRP intercept, which contains activity that is common to all fixations, within a fixation time-window of 50–300 ms. For both groups, the word length effect was present in brain activity during additional fixations; however, this effect was not different between groups. This suggests that stronger word length effect in the eye movements of slow readers might be mainly due re-fixations, which are more probable due to the lower efficiency of visual processing.

Keywords

FRP; reading; word length; eye-tracking; EEG; reading fluency

1 Introduction

Reading difficulties remain a significant societal problem. Previous research has shown that slow readers exhibit a wide range of differences in several cognitive test measures and in gaze behaviour as well as in brain activity during reading (Norton, Beach, & Gabrieli, 2015; Seymour, Aro, & Erskine, 2003; Ziegler & Goswami, 2005). In particular, word length effect, which is seen in the increase in duration measures of eye movements as a function of the number of letters, appears to be much larger in individuals with reading difficulties than in those with typical reading skills (Hawelka, Gagl, & Wimmer, 2010). While the influence of word length and reading difficulties on eye movements are rather well documented, individual differences in terms of how the brain operates during free reading as a function of reading ability are largely unexamined. We ran a study where participants freely read sentences while eye-movements and concurrent brain activity were recorded with co-registered eye-tracking and electroencephalography (EEG). Here, we analyse part of our experimental data, restricted to sentence beginning, which has not been previously examined. The aim is to provide exploratory results on the potential causes of enlarged word length effect associated with slow reading.

Word length effect in eye movements during reading has been widely documented and well replicated (see Barton, Hanif, Björnström, & Hills, 2014). Here, we summarize some relevant findings, acknowledging that the following account might not be an exhaustive record. The general pattern associated with word length is an increase in gaze duration (Hautala, Hyönä, & Aro, 2011; Hautala & Loberg, 2015; Kliegl, Grabner, Rolfs, & Engbert, 2004) and refixation probability measures (Hautala et al., 2011; Hautala & Loberg, 2015; Kliegl et al., 2004; Rayner, 1997). For the first fixation duration, there is a more mixed pattern of results, as a standard word length effect of a longer duration has been found (Hautala et al., 2011); however, in a few studies (Hyönä & Olson, 1995; Kliegl et al., 2004), the first fixation durations were not modulated by word length in adult readers. In Finnish, a highly inflectional language with frequent compound words, a reverse word length effect—where first fixation duration decreases as a function of word length—has been reported in first fixation duration (Bertram & Hyönä, 2003) for compound words. The direction of word length has been shown to be different between adults (standard) and children (reverse) (Tiffin-Richards & Schroeder, 2015). The reverse word length effect on first fixation duration appears to be related to situations where re-fixation probability/amount of fixation increases due to acuity or processing limitations (Kliegl, Olson, & Davidson, 1983). These

observations suggest that when the underlying cognitive system manages to estimate a need of additional fixation, it begins to decrease the duration of the first fixation in which a mere increase of duration would not be sufficient.

Reading development is strongly reflected in eye movements during reading. Generally speaking, reading skill that is still in the development stage is reflected in short saccades (Blythe & Joseph, 2012) and long fixations (Blythe, Häikiö, Bertam, Liversedge, & Hyönä, 2011; Blythe & Joseph, 2012; Tiffin-Richards & Schroeder, 2015). As reading experience accumulates and (supposedly) the brain network involved in reading becomes more specialized, the saccades become longer (Blythe & Joseph, 2012) and fixations shorter (Blythe et al., 2011; Blythe & Joseph, 2012; Tiffin-Richards & Schroeder, 2015). Stronger word length effects for children are believed to be related to processing of text that is reliant on sub-lexical decoding (Tiffin-Richards & Schroeder, 2015). This is likely related to more strict limitations of parallel letter processing—how the letter-identity span increases on account of development (Häikiö, Bertram, Hyönä, & Niemi, 2009). Letter identity span refers to the extent of an individual's ability to utilize parafoveal information to identify letters while reading; as such, it is a more specialized form of perceptual span (Rayner, 1986, 2014) and it has been found to expand with an improvement in reading ability (Häikiö et al., 2009). As mentioned earlier, the direction of the word length effect in first fixation duration has been shown to be modulated by age (Tiffin-Richards & Schroeder, 2015). It is possible that due to a more limited letter identity span, the cognitive system prepares for multiple fixations to long words more readily among children than among adults.

Reading fluency has a clear association with gaze behaviour during reading. Generally, children with reading difficulties exhibit overall longer fixation durations, gaze durations, perform many more and shorter saccades, and consequently have higher fixation count per word (De Luca, Borrelli, Judica, Spinelli, & Zoccolotti, 2002; De Luca, Di Pace, Judica, Spinelli, & Zoccolotti, 1999; Dürrwächter, Sokolov, Reinhard, Klosinski, & Trauzettel-Klosinski, 2010; Hawelka et al., 2010). However, the pattern of word length effect findings is not uniform, as there also exist studies that do not find differences in eye movements between dyslexic readers and typical readers (Hyönä & Olson, 1995), thereby illustrating that not all findings can be generalized across languages and age groups. Eye-movement reflection of word length effect has also been shown to be increased in dyslexics in terms of several variables: number of saccades (De Luca et al., 2002; Dürrwächter et al., 2010), fixation durations (Dürrwächter et al., 2010), number of fixations (De Luca et al., 1999; Hawelka et

al., 2010) and gaze durations (Hawelka et al., 2010). From these findings, it is feasible to arrive to a generalization that developmental dyslexia manifests as an inability to transition from initial reading behaviour to fluid eye movements at the same rate as that for typically developing children.

When reflecting the findings on eye movement related to dyslexia and with respect to assumptions of the eye-mind link (Reichle & Reingold, 2013), the brain responses that fit within typical fixation duration—visual P1 and visual N1—are of key interest¹. Generally, in psycholinguistic ERP experiment designs, word length is considered to be a confounding factor. Consequently, studies examining ERP signatures of word length are relatively rare, but a few studies do exist. In early studies, an increase in word length (step increase in three-letter words vs four- and five-letter words vs six- and seven-letter words vs eight-letter and longer words) led to stronger negativity around 200 ms in the occipital and parietal regions (Van Petten & Kutas, 1990). Another study found a difference between three- and four-letter words in the slope between N1 and P2, with three-letter words being more negative, although the effect was trend-like (Dehaene, 1996). An early magnetoencephalography (MEG) study found a temporally widespread effect of word length as early as 90 ms and spanning up to 800 ms, located in magnetometer sensors in proximity to the temporal and parietal cortices (Assadollahi & Pulvermüller, 2001)². In another MEG experiment, manipulating the number of letters in the stimulus (4 vs. 8) and lexicality revealed that occipital activity in time window of P1 wave was stronger for long stimuli irrespectively of lexical status (Wydell, Vuorinen, Helenius, & Salmelin, 2003). A widespread effect of word length was also found in a subsequent ERP study in which long words (mean length 6.2 letters) had more positive P1 (80–125 ms) than short words (mean length 4.1 letters) in parietal-occipital electrodes and

¹ At this point, it must be noted that the term N1 indicates the first occipital negativity occurring after P1 response, typically between 120 ms to as late as 300 ms. From this definition, it follows that in the context of this article, we do not differentiate between the plurality of occipital negative deflection labels (N130, N150, N155, N170, N200, etc.) that have been used in the literature. Similarly, it must be noted that when we are citing MEG/ERF findings as if they were EEG/ERP findings, we interpret the presented magnetic field with time-course and postulate which ERP response the ERF deflection would manifest as.

² Directions of the effects are not explained here, as this report is an early MEG study with rather limited illustrations of effect locations; thus, we cannot confidently provide insightful estimates of how the magnetic field would translate from magnetometers to an electrical field on EEG electrodes. The emphasis here is on the timescale and the fact that the word length effect is found early.

more negative response to short words than long words around the N1 time window (150–190 ms) (Hauk & Pulvermüller, 2004). In summary, even if the spatial location, timing and direction of the effect are not systematic, the broad pattern reflects that word length influences cortical processing with timing that does not preclude direct influence on eye movements.

There are certain studies that adopt a different approach to ERP estimation. In a study that analysed event-related EEG, with word length as a continuous predictor rather than a categorical average, word length was found to influence early parts of the ERP during single-word presentation (Hauk, Davis, Ford, Pulvermüller, & Marslen-Wilson, 2006). In 90–110 ms, word length had a positive dependency on the voltages in the posterior right hemisphere and negative dependency on voltages in the frontal left hemisphere of the scalp (Hauk et al., 2006). In 150–160 ms, there was positive dependency in the central location, and in 180–240 ms, there was positive dependency in the left posterior scalp and negative dependency in the central scalp (Hauk et al., 2006). However, operationalization of word length here (Hauk et al., 2006) is slightly problematic, as it is drawn from a selection of psycholinguistic factors with principal component analysis (PCA), where word length (positive weight), number of syllables (positive weight) and word neighbourhood density (negative weight) and their effects can be partially due to the inverse effect of the neighbourhood density. In a subsequent study, word length was found to have a positive dependency in the bilateral occipital scalp in 90–110 ms and positive dependency in the left occipital scalp and negative dependency on the central scalp in 212 ms (Hauk, Pulvermüller, Ford, Marslen-Wilson, & Davis, 2009). Overall, the scarce and relatively mixed results on the electrophysiological correlates of word length—particularly during unrestricted reading—makes it an interesting research topic and not merely a factor to control for in the experiment design.

Word-specific brain responses develop as reading skill is acquired. For example, N1 amplitude increases for words on account of reading acquisition from kindergarten to second grade (Maurer et al., 2007). The same response has been shown to become stronger when grapheme-phoneme correspondence is trained in children attending kindergarten (Brem et al., 2010). In another study, N150 was found to be more right lateralized in children than in adults (Spironelli & Angrilli, 2009). As early occipital negative responses have been shown to be associated with development, they have been implicated in poor reading skill. MEG studies implicate the left inferior occipital cortex to show differing activity in adult dyslexics around 150–200 ms (Helenius, Tarkiainen, Cornelissen, Hansen, & Salmelin, 1999; Salmelin,

Service, Kiesila, & Uutela, 1996), thereby suggesting that dyslexics have poor capability of treating words as unified percepts. Similar differences have been shown in ERP studies in children with dyslexia. In a study using a phonological lexical decision task, reduced occipital N1 was found in dyslexic second graders as compared to normal participants (Hasko, Groth, Bruder, Bartling, & Schulte-Körne, 2013). In a study utilizing 1-back³ Rapid Serial Visual Presentation (RSVP) to research the development of word recognition, dyslexic second graders displayed weaker N1 to words than normal readers did (Maurer et al., 2007). In another study using similar visual 1-back design, dyslexic third graders were found to have stronger N1 to words than normal participants did (Fraga González et al., 2014). Moreover, in the same population of individuals that received successful remedial intervention, the difference in N1 was reduced; however, this was not so in the case for individuals who did not benefit from the intervention (Fraga González et al., 2016). The differences in the results may very well stem from differences in experimental design. In (Maurer et al., 2007), there were three conditions (1 × 3 design: words vs. pseudowords vs. symbols) and the maximum width of the stimuli was 3.6 degrees of the eye. In contrast, (Fraga González et al., 2014) had four conditions (2 × 2 design: words vs. symbols × short vs. long) and the mean width of the stimuli was 6.4 degrees of the eye. It remains uncertain how the overall different combination of stimuli influences the relationship between two stimuli; however, stimuli with differences in visual span stimulate the visual receptive fields differently, possibly emphasizing the role of different neural generators. The susceptibility of results to be swayed by relatively small differences in experimental design erodes the basis of how well traditional RSVP approaches generalize to the actual reading process and how this process is affected by reading difficulties. Contradictory evidence from experiments emphasizes the importance of examining, in ecologically valid settings, cortical electrophysiological events associated with slow reading.

How is word length particularly relevant to dyslexia or, more generally, to slow reading? To put it simply, word length places stress on several cognitive mechanisms that have been proposed to be functioning abnormally in developmental dyslexia. For example, the visual attention span has been proposed to be smaller in dyslexics (Bosse, Tainturier, & Valdois, 2007; Valdois, Bosse, & Tainturier, 2004), thereby effectively limiting the span letters that

³ We classify the paradigms of Maurer et al. 2007 and Fraga González et al. 2014 and 2016 as 1-back paradigms as participants were required to respond when stimulus was immediately repeated.

can be processed in parallel. Thus, words receive more fixations with higher probability, as not as many words can be processed in a single fixation. On the other hand, psycholinguistic grain size theory considers the main problem of dyslexia to lie in the decoding of grapheme-phoneme correspondence (Ziegler & Goswami, 2005); the correspondence rules of phonemes to graphemes in the language that the person with dyslexia is learning to read determines the phenotype of behavioural symptoms. In languages with transparent orthography, the key problem lies in single grapheme-phoneme correspondence; thus, word length more strongly affects dyslexics. Re-fixations are then used to optimize the foveal position in relation to processed visual features (Kagan & Hafed, 2013). Re-fixations are more common in individuals with dyslexia (De Luca et al., 1999; Hawelka et al., 2010; Tiffin-Richards & Schroeder, 2015), which emphasizes the fact that not all critical issues related to reading difficulties are well known. However, previous research has employed methods that have limited explanatory power. This is evident in how ERP alone is incapable of accounting for abnormal behaviour associated with reading difficulty, since unrestricted behaviour is a confounding factor in traditional ERP experiments. Conversely, ET alone is limited in that it does not have direct observations of the neural correlates of abnormal cognition. However, the shortcomings of the two methodologies can effectively be bridged with the combination of the two—the fixation related potential (FRP) method. We used linear deconvolution methodology (Cornelissen, Sassenhagen, & Vö, 2018; Ehinger & Dimigen, 2018; Kristensen, Rivet, & Guérin-Dugué, 2017), which is an extension of the general linear model estimation of ERPs (Smith & Kutas, 2015a, 2015b), for estimation of the FRPs. We have included a short theoretical introduction to these methods in supplementary material 1.

Given that our approach is exploratory, our statistical procedure cannot be driven by specific hypotheses related to time point and channel to identify the word length effect or the difference between groups. In any ERP/FRP contrast with multiple available channels and time points, the lack of a well-defined *a priori* hypothesis raises the requirement that specific cautions must be taken against inflating type 1 error probability. It is not to say that we are hypothesis-free—we expect the effects of word length and reading fluency to be pronounced clear effects in the FRP estimate within fixation duration. This is mainly because our participant population comprises children and the transparent orthography of the utilized language, both of which are not widely represented in previous literature, makes inferences from previous literature tedious at best. Further, previous literature is not uniform in terms of the temporal and spatial characteristics of the cortical word length effect. In this study, we

adopt a conservative approach by utilizing cluster-based permutation statistics that provide strong control for the familywise error rate (Maris & Oostenveld, 2007).

In this study, we examined word length effect on deconvoluted fixation-related potentials in a large sample of school-aged children. We also examined whether first and additional fixations to a word differ from each other in terms of brain activity. The impact of level of reading ability on these factors is also examined. Broadly defined, we expected to observe the following aspects:

- 1) Word length effect difference in eye movements between typical and slow readers.
- 2) Word length effect in FRP within fixation duration from fixation onset.
- 3) Difference in FRP between first and additional fixation within fixation duration from fixation onset.
- 4) Difference in FRP word length effect between typical and slow readers within fixation duration from fixation onset.

2 Material and methods

2.1 Participants and their characteristics.

Data was collected as a part of eSeek!–Internet and Learning Difficulties: A Multidisciplinary Approach for Understanding Reading in New Media–project⁴. EEG measurements were taken from a sub-sample. Participants of the sub-sample were selected to include all individuals exhibiting difficulties in reading skill. We invited every third individual, without exhibited reading difficulties, in ascending order of their estimated internet reading skill (Kiili et al., 2018). Thus, for this particular study, the sampling procedure can be classified as convenience sampling. The sample reported here includes 92 12–13.5-year-old children (41 females, 83 right-handed). The study was approved by the Ethical Committee of the University of Jyväskylä and conducted in accordance with the Declaration of Helsinki. The parents of participants signed written consent and the children had the possibility of terminating the measurement at any given time. The participants reported normal or corrected to normal vision and had no history of neurological abnormalities. The participants were allocated to two groups based on reading fluency scores of the entire eSeek sample (542 children of the same age group): slow readers (N = 27) and typical readers (N = 65). Reading fluency was estimated with three tests: the *Word Identification Test*—a subtest of standardized Finnish reading test ALLU (Lindeman, 1998), the *Word Chain Test* (Nevala & Lyytinen, 2000), and the *Oral Pseudoword Text Reading Test* (Eklund, Torppa, Aro, Leppanen, & Lyytinen, 2015). Then, the scores of these tests were subjected to principal axis factoring with PROMAX rotation using the IBM SPSS 24 statistics program (IBM Inc.) to extract a single factor that explained 69.21% percent of the variance. Tests loaded to the fluency factor in the following manner: Word Identification Test (0.683), Word Chain Test (0.872) and Oral Pseudoword text Reading Test (0.653). The participants belonging to the weakest 15% on this factor was assigned to the slow readers group and the remainder to the typical readers group. Performance in the tests that constituted the group defining the fluency factor are presented in Table 1.

⁴ <https://www.jyu.fi/edupsy/fi/laitokset/psykologia/en/research/research-areas/neuroscience/groups/neuro/projects/eSeek>

Table 1: Descriptive information on participant performance in reading fluency tests.

	Word Identification Test		Word Chain Test	
	Mean	SE	Mean	SE
Slow readers	35.48	1.69	23.07	1.23
Typical readers	50.28	1.05	44.63	1.56

	Pseudoword Text Reading Test			
	Time		Correct	
	Mean	SE	Mean	SE
Slow readers	81.50	3.96	30.44	0.95
Typical readers	50.27	1.44	35.43	0.33

Note: In the Word Identification Test, the response variable is the number of correct answers (max. 80). In the Word Chain Test, the response variable is the number of correct answers (max. 100). In the Pseudoword Text Reading Test, the variables are reading time in seconds and number of correct spellings (max. 38).

2.2 Materials.

A total of 200 sentences with a median length of 6 words (from 5 to 9) were used as the stimuli. The sentences were presented in the Times New Roman (New) font with a font size of 20. Each letter subtended at a visual angle of 0.4 degrees on average on the screen when the distance of the participant was 60 cm. There were 3 categories of sentences: 100 plausible sentences, 50 sentences where the target word was anomalous to preceding context, and 50 sentences where the target word was anomalous yet an orthographic word neighbour of a plausible word. The target word was always the last word in the sentence. For the analysis, we excluded all eye movement events that occurred after and including the first fixation on the target word; this was done to avoid the influence of experimental manipulation on the results and focus on the non-manipulated portion of the sentence⁵. From the material, we were able to extract 912 words with a word length ranging from 5 to 13 (word frequency available from 878 of the words: mean = 75.03, standard deviation = 253.72, correlation to

⁵ For the effects of the manipulation on target words in typically reading children, see (Loberg, Hautala, Hämäläinen, & Leppänen, 2018).

word length: $r = -0.1989$, $r^2 = 0.0396$, $p < 0.00001$). Although significant, the correlation between word frequency and word length was not particularly strong and, as such, we did not consider that it would confound the results. As Times New Roman is a proportional font, we analysed the dependency between word length letters and word width in pixels; we found an almost singular relation between the two in the words ($r = 0.9498$, $r^2 = 0.9021$, $p < 0.00001$). Thus, here, the spatial width of the words is practically indistinguishable from word length. An example of the sentence and an illustration of the kept data is provided in Figure 1.



Pihan pyykkilinjalla roikkui märkä paita

Figure 1. Example sentence, translated: ‘A wet shirt was hanging on the drying rack in the yard’. Green highlights the area from where the fixations were accepted for analysis. From and including the first fixation on the last word of the sentence (highlighted in red) and until the end of the trial, fixations were not included in the analysis.

2.3 Measurement.

The recording was completed in a dimly lit soundproof room at the brain research facilities of the Department of Psychology, University of Jyväskylä. EEG was recorded using the NEURONE (Bittium Inc.) amplifier with a 1000hz sampling rate and an online high pass filter of 0.16 hz and lowpass filter of 250 hz, using a 128-channel electrode net (Electrical Geodesics Inc.) using Ag/AgCl electrodes. The quality of the EEG was maintained throughout the measurement. Eye movements were recorded with EYELINK 1000 with a 2000 hz upgrade (SR research) with a 1000 hz sampling rate. The sentences were presented on a Dell Precision T5500 workstation with an Asus VG-236 monitor (1920 × 1080, 120 Hz, 52 × 29 cm) at a viewing distance of 60 cm. The synchrony between two measures was ensured

with a mixture of Transistor to Transistor Logic (TTL)-pulses (to EEG) and Ethernet messages (to eye tracking) at the beginning and end of each trial.

2.4 Measurement procedure.

Participants held their head in a chinrest during the measurements, excluding breaks. The experiment program followed a simple loop. A 13-point calibration routine was run before each block. Before each trial, the quality of the calibration was checked by showing a black dot and running a drift-check routine on that position. If the fixation diverged from the calibration by more than 1 degree, the experimenter was alerted and the calibration was redone. Once the fixation was accepted by the experimenter, the dot disappeared and the trial was initialized. The sentence appeared on the screen and the participant read the sentence until he/she gave a manual response to the question ‘Is the sentence sensible or not?’ with a button press. Participants were instructed to read the sentences as fast as possible. The experiment was divided into four blocks, and the quality of the EEG and calibration was maintained during breaks when necessary.

2.5 Pre-processing.

Co-registered ET-EEG was pre-processed in MATLAB using EEGLAB (v14.1.2) toolbox with an EYE-EEG (0.85) add-on. The data was high-pass filtered at 0.5 hz and low-pass filtered at 30 hz. Raw gaze position data was synchronized with EEG using shared messages in both data streams at the beginning and end of each trial. Gaze positions that indicated the location outside the screen were classified as bad data; this included gaze position zero from blinks and between-trial gaps in the recording. The 100 ms before and after such a value was also considered as bad data. For the remaining gaze position data, we ran the binocular median velocity algorithm for detecting fixations and saccades (six standard deviations from median velocity was the threshold for saccades) (Engbert & Mergenthaler, 2006). The data was subjected to the OPTICAT (Dimigen, 2018) procedure, in which we sampled and copied saccade events and related data segments to the end of the data set for overweighing the dataset with the saccade data pattern. Prior to independent component analysis (ICA) training, we down-sampled the overweighed data set to 500 hz in order to make the computation more feasible. ICA was completed with extended Infomax (Lee, Girolami, & Sejnowski, 1999) and we applied the resulting weights to the original data set. In the original data set, the saccade-related independent component was selected on the basis of temporal dependency to saccade events (Plöchl, Ossandón, & König, 2012).

2.6 Deconvolution modelling of FRPs.

For estimation of the FRPs, we used UNFOLD toolbox (Ehinger & Dimigen, 2018), which uses GLM for response estimation and correction of response overlap and generalized additive model for nonlinear predictors. The following Wilkinson notation of the used formula was employed:

$$y \sim 1 + word_length *cat(first_fix)+ spl(previous_saccade_amplitude,8)'$$

Thus, we estimated the intercept which reflects the activity shared by all the fixations, continuous linear effect of word length, categorical effect of first vs. additional fixation, interaction between word length and additional fixations and nonlinear effect of previous saccade length modelled with eight splines. The saccade length was included as a predictor to model out the influence of saccade length from interesting effects, which is known to influence the latency and amplitude of early visual activity (Nikolaev, Meghanathan, & van Leeuwen, 2016). In previous studies, the influence of saccade length on brain activity is shown to be non-linear (Dandekar, Privitera, Carney, & Klein, 2012; Ries, Slayback, & Touryan, 2018); in UNFOLD toolbox, non-linear influences are modelled with splines subjected to General Additive Modelling (Ehinger & Dimigen, 2018). Thus, in the saccade amplitude portion of the equation, we arrange saccade amplitudes to partially overlapping categories and model the non-linear influence step by step. The modelled response ranged from -700 ms to 500 ms from fixation onset. The time window for modelling the response was selected so it would include preparatory activity for the incoming saccade and provide sufficient time for the next saccade-influencing activity to occur. Time points corresponding to blinks and outside screen eye movements as well as segments with large fluctuations (modified C.R.A.P. algorithm; (Ehinger & Dimigen, 2018; Lopez-Calderon & Luck, 2014) were set to zero in the regression matrix and, thus, were removed from response estimation. Fixations at the original target word (at the end of the sentence) and during re-readings were excluded from the FRP estimation by setting the timepoints of the regression matrix to zero from the onset of the first fixation on the last word until the end of the trial.

2.7 Statistical analysis of eye movements.

Eye movements were analysed using linear mixed effects (LME) models within the R environment for statistical computing (R core team, 2015). In order to estimate effects in

continuous variables (first fixation durations, gaze durations), we used the lmer function of the lme4 package (Bates, Mächler, Bolker, & Walker, 2015). For estimation of the effects in the dichotomic response variable re-fixation, we used the glmer function (with family = ‘binomial’ setting) of the lme4 package (Bates et al., 2015). For the estimation of the p-values, we used the lmerTest package with Satterthwaites’s method for degrees of freedom and t-statistics (Kuznetsova, Brockhoff, & Christensen, 2017). The p-values are reported at a precision of five decimals. We fitted the simplest random structure with random intercepts for participants and items. However, we did not include random slopes for the participants, as there is no conceptual match for such a model in the permutation test framework as used for FRP analysis.

First, we ran the LME model with the formula

$$'resp_var \sim word_length * group + (1|Sub) + (1|Item)'$$

to verify that the eye movement behaviour is replicated as in previous studies. We ran this model on first fixation duration, gaze duration, and (as a binomial variant) on re-fixation probability. However, from the viewpoint of FRPs, sum variables like gaze duration are problematic as they ignore the fact that there is often additional fixation with accompanied transient brain activity—that is, even if gaze duration at the behavioural level might seem like an index of a uniform process, it most certainly is not so at the neural level, as the re-fixations that drive the variance in gaze duration bring additional visual inputs. Whether this visual information is actually utilized in word recognition is unclear based on prior research, but this can be examined at a behavioural level and also in neural signatures with modelling approaches that distinguish between the first and additional fixations. To better conceptually match the analysis of the FRP and gaze behaviour, we also estimated the following linear mixed effects model, which also uses the type of fixation as a predictor:

$$'fixation_duration \sim word_length * group * fixation_type + (1|Sub) + (1|Item)'$$

All models converged successfully.

2.8 Statistical analysis of deconvolution model parameters.

In order to examine whether the word length effect, the additional fixation effect or the interaction between these would be different between slow and typical readers, we ran nonparametric cluster-based permutation statistics (Maris & Oostenveld, 2007) on the beta waveforms in FieldTrip toolbox. The beta waveforms were baselined from 700 ms before fixation onset to 500 ms before fixation onset. The baseline was defined to avoid parafoveal processing prior to the examined fixation being present on the baseline. For intercept comparison between groups, we added the effect of word length 8 and saccade amplitude 1.8798 (mean previous saccade amplitude in the data set) to the intercepts. This was done to ensure that the exact features of the splines from differences in mean saccade amplitude among participants used to model the saccade amplitude effect would not introduce any confounding effects to the comparison. Then, we tested the between-group differences with a permutation test with an independent samples t-test as a base test. Further, in order to establish the presence of a within-group effect, we ran a one-sample t-test on the additional fixation beta and word length beta with expectation value 0 as a base test for the permutation test within both groups: slow and typical readers. Then, we compared responses from the groups with independent samples t-test as the base test for the permutation test. The neighbourhood structure was defined with the triangulation method, using the `ft_prepare_neighbours` function. The setting for the cluster statistic was 'maxsum' and the minimum number of neighbours for cluster formation was two. The time window for running the test was from 50 ms to 300 ms after fixation onset. The starting point of the window was defined at 50 ms to give time for the visual information to propagate to the visual cortex, and the end point was defined to be rounded closest 100 ms from 250 ms typical fixation duration with typical saccade duration (~10 ms) added to it. Thus, the span of the window included the observations that are likely to influence the next eye movement. The number of permutations that we used was 20000.

3 Results

3.1 Linear mixed effect models of eye movements.

The linear mixed effect (LME) analysis of standard variables showed several significant effects and effect interactions. These are presented in Figure 2 and Table 2. The LME model on the first fixation duration (FFD) showed a significant main effect of word length, where first fixation decreased as a function of word length and main effect of group, where slow readers had systematically longer fixations than typical readers. These two effects were in interaction where the word length effect was stronger for slow readers than for typical readers. Moreover, with gaze duration (GD), the main effect of word length (where GD increased as a function of word length) and main effect of group (where slow readers had systematically longer GD) were significant. These effects were in interaction where the slow readers had stronger word length effect. The generalized LME model on re-fixation probability (REFIX) showed a significant main effect of word length, where REFIX increased as a function of word length and main effect of group, where slow readers had systematically higher REFIX. Again, the main effects were in interaction, and based on the right panel of Figure 2, interaction resulted from slow readers having sharper initial increase of REFIX, which was weaker for the last two steps (11-13). On the other hand, typical readers exhibited relatively linear growth of REFIX as a function of word length.

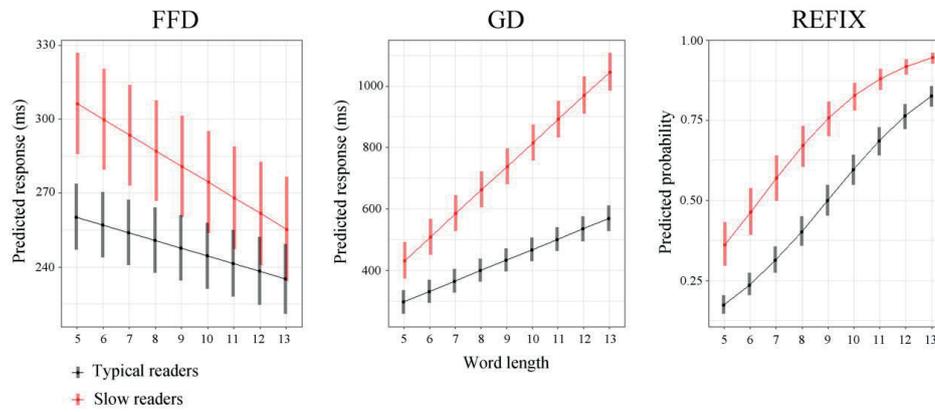


Figure 2. Predicted responses from linear mixed effect models exploring first fixation duration (FFD), gaze duration (GD) and generalized linear mixed effect model exploring Re-fixation probability (REFIX). Red lines represent slow readers ($N = 27$) and black lines represent typical readers ($N = 65$). Error bars signify 95% confidence intervals.

Table 2: Linear mixed effects model parameters on standard eye movement variables

Predicted variable	Parameter	Estimate	Standard error	t/z - value	p-value
FFD	Intercept	260.29	6.75	38.569	<0.00001
	Word length	-3.15	0.64	-6.765	<0.00001
	Group	45.9	12.2292	3.753	0.00031
	Group * Word length Interaction	-3.21	0.4558	-7.050	<0.00001
GD	Intercept	297.040	19.543	15.20	<0.00001
	Word length	33.971	1.677	20.26	<0.00001
	Group	134.575	34.769	3.87	0.00021
	Group * Word length Interaction	42.822	1.032	41.51	<0.00001
REFIX	Intercept	-1.57046	0.10176	-15.434	<0.00001
	Word length	0.39157	0.01118	35.029	<0.00001
	Group	0.99839	0.17740	5.628	<0.00001
	Group * Word length Interaction	0.03671	0.01008	3.642	0.00027

Note. FFD = first fixation duration, GD = gaze duration, REFIX = re-fixation probability

In the LME model, examining all fixations (Figure 3 and Table 3) as separate entities, we observed significant main effects of word length, fixation type and group. As observed from Figure 3, main effects of word length and fixation type were qualified by interactions. However, on the other hand, the main effect of group was systematic, and slow readers had overall longer fixations than typical readers. All combinations of two-way interactions were significant; however, the interaction between word length and group and between-fixation type and group were qualified by three-way interaction between factors. On the other hand, interaction between word length and fixation type is interpretable from Figure 3, which quite clearly shows a pattern where the word length influences mainly first fixations. In the slow readers group, the three-way interaction was generated by the stronger word length effect on first fixations.

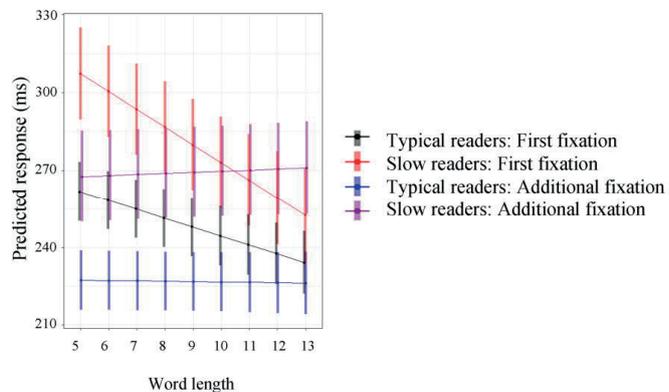


Figure 3. Predicted responses from the linear mixed effects model examining the effect of word length, group and fixation type on fixation duration. Black lines indicate predicted durations of first fixations for typical readers ($N = 65$). Blue lines indicate predicted durations of additional fixations. Red lines indicate the predicted durations of first fixations of slow readers ($N = 27$). Purple lines indicate the predicted durations of the additional fixations for slow readers. Error bars signify 95% confidence intervals.

Table 3. Linear mixed effects model parameters on fixation durations

Response variable	Parameter	Estimate	Standard error	t-value	p-value
Fixation duration	Intercept	261.8	5.773	45.354	<0.00001
	Word length	-3.46	0.411	-8.419	<0.00001
	Group	45.45	10.49	4.333	0.00004
	Fixation type	-34.48	1.388	-24.840	<0.00001
	Word length * Group interaction	-3.389	0.5001	-6.776	<0.00001
	Word length * Fixation type interaction	3.324	0.3611	9.206	<0.00001
	Group * Fixation type interaction	-5.17	2.335	-2.214	0.02680
	Word length * Group * Fixation type interaction	3.937	0.6281	6.269	< 0.00001

3.2 Deconvoluted FRP Results

Between groups, we found significant differences between the intercepts (Figure 4) of the two groups; this was visible in the positive central cluster that emerged around 140 ms and moved to occipital regions after 250 ms ($p = 0.0001$). However, there was no significant difference between groups in the beta waveforms of additional fixation (smallest $p = 0.19059$), word length (smallest $p = 0.0853$) or in the interaction between word length and additional fixation (smallest $p = 0.70097$).

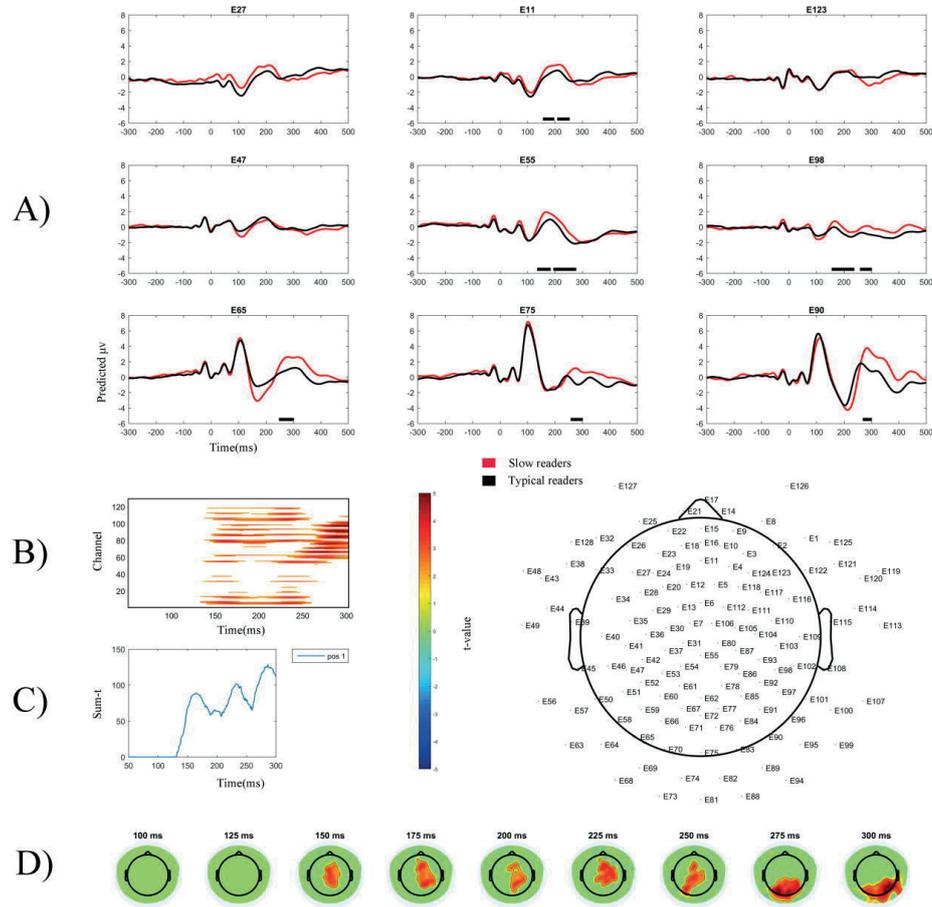


Figure 4: Between-group nonparametric cluster-based permutation test results for the intercept waveform that contains the shared activity of all responses. A) Grand average waveforms of the intercept at select channels. The red line represents the intercept response for slow readers ($N = 27$). The black line represents the intercept response for typical readers ($N = 65$). B) Raster plot of the significant cluster. C) Behaviour of sum-t parameter over time. D) Topographic representation of the significant cluster at selected time points.

Within-groups, cluster-based permutation one-sample t-tests revealed the presence of several effects, which are presented in Figure 5. For typical readers, additional fixation had a significant effect in the FRP. This was demonstrated in the positive cluster that persisted on the central scalp for the duration of the time window ($p = 0.00010$) and negative cluster that was present on the right frontal fringe channels from approximately 50 ms to 240 ms ($p = 0.03850$). The prediction of how this effect manifests on the channels is presented in Figure 6. Further, there was no significant main effect of word length (smallest $p = 0.09210$) among typical readers. The interaction between additional fixation and word length was significant. This was present in two clusters: One frontal positive cluster that emerged around 160 ms and dissipated around 270 ms ($p = 0.00800$), and the occipital negative cluster that emerged around 120 ms and dissipated around 270 ms ($p = 0.00260$). Among slow readers, there was no significant main effect of word length (smallest $p = 0.14389$) or effect of additional fixation (smallest $p = 0.07540$); however, the interaction between additional fixation and word length was significant. This was supported by three clusters: positive cluster emerging in frontal scalp around 220 ms and dissipating around 290 ms ($p = 0.02650$) and two negative clusters—one emerging around 140 ms at the occipital scalp and dissipating around 290 ms ($p = 0.00640$) and another that was present at the beginning of the time window on occipital channels and dissipated around 130 ms ($p = 0.04710$).

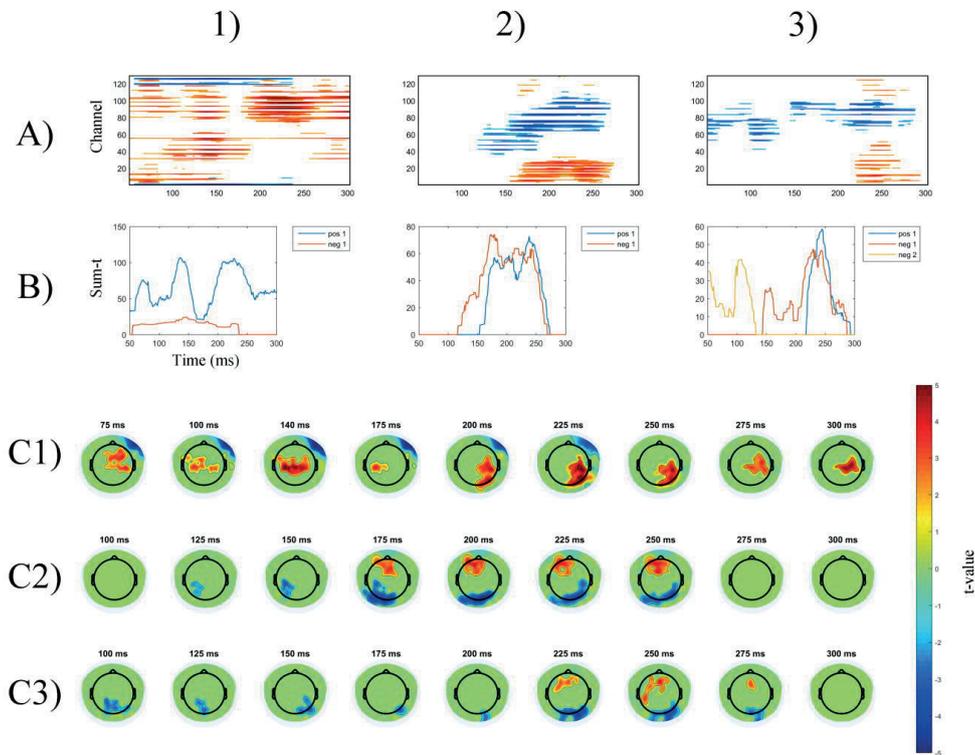


Figure 5. Within-group one-sample non-parametric cluster-based permutation test results for the factor beta waveforms, which reflect the modulation of the FRP waveform as a function of the modelled factors. A1) is the raster plot of additional fixation effect among typical readers ($N = 65$). A2) is the raster plot of interaction of word length and additional fixation effects among typical readers. A3) is the raster plot of the interaction of word length and additional fixation effects among slow readers ($N = 27$). B1): Time behaviour of the sum-t parameter from significant clusters of the additional fixation effect among typical readers. B2): Time behaviour of the sum-t parameter from significant clusters of interaction between word length and additional fixation effects among typical readers. B3) Time behaviour of the sum-t parameter from significant clusters of interaction between word length and additional fixation effects among slow readers. C1) Selected time point topographical representation of additional fixation effect among typical readers. C2) Selected time point topographical representation of interaction between word length and additional fixation among typical

readers. C3) Selected time point topographical representation of interaction between word length and additional fixation among slow readers.

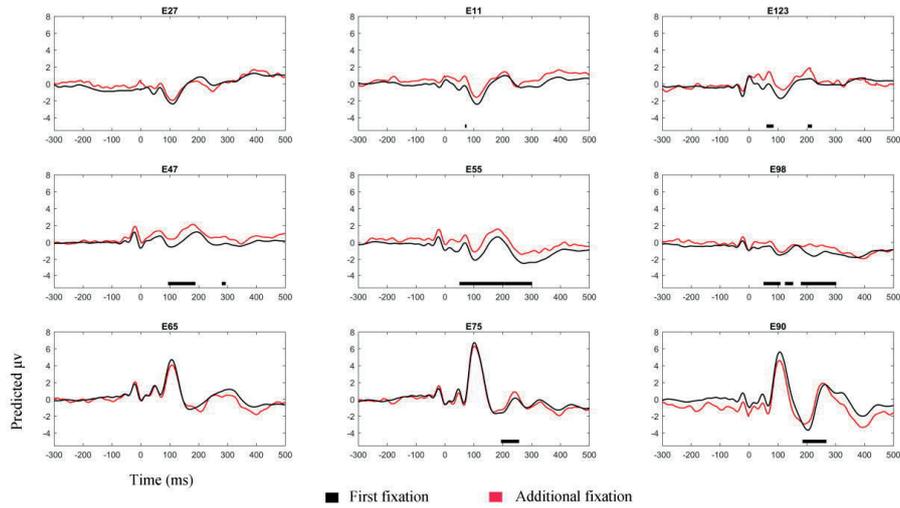


Figure 6: The predicted effect of additional fixation among typical readers ($N = 65$). The black line is the intercept which represents the first fixation in the full deconvolution model. The red line is the intercept with the effect of the additional fixation added to it. The black rectangles under the waveform signify time points that belong to a significant cluster in the one-sample permutation t-test on the additional fixation beta parameter.

3.3 Post hoc deconvolution analysis.

As evident from the permutation results, the word length is found to be in interaction with additional fixation categorical effect in both groups. To open up this interaction, we re-ran the deconvolution model where we separated the first fixation and additional fixation to separate equations, both following the Wilkinson notation:

$$y \sim 1 + word_length + spl(previous_saccade_amplitude,8)'$$

Thus, we received separate intercepts and word length betas for both fixation types. Then, we examined whether the word length betas differed from zero in both conditions and whether the word length betas were different between groups with the same statistical procedures as earlier.

3.4 Post hoc results.

Within-group tests of word length did not show any significant effect in the first fixation, neither among typical readers (smallest $p = .16159$) nor among slow readers (smallest $p = 0.21359$). However, in the additional fixation condition, word length effect was significant for both groups. In the typical readers group, two clusters were observed: a positive cluster ($p = 0.00770$) at frontal sites that emerged around 130 ms and lasted thorough the epoch, and a negative cluster ($p = 0.00050$) in occipital sites that emerged around 130 ms and persisted thorough the modelled epoch. In the slow readers group, a single negative cluster ($p = 0.01410$) was observed in right occipital sites that emerged around 170 ms and dissipated around 280 ms. There was no significant difference between the groups (smallest $p > 0.99999$). Significant results of the post hoc deconvolution analysis are presented in Figure 7.

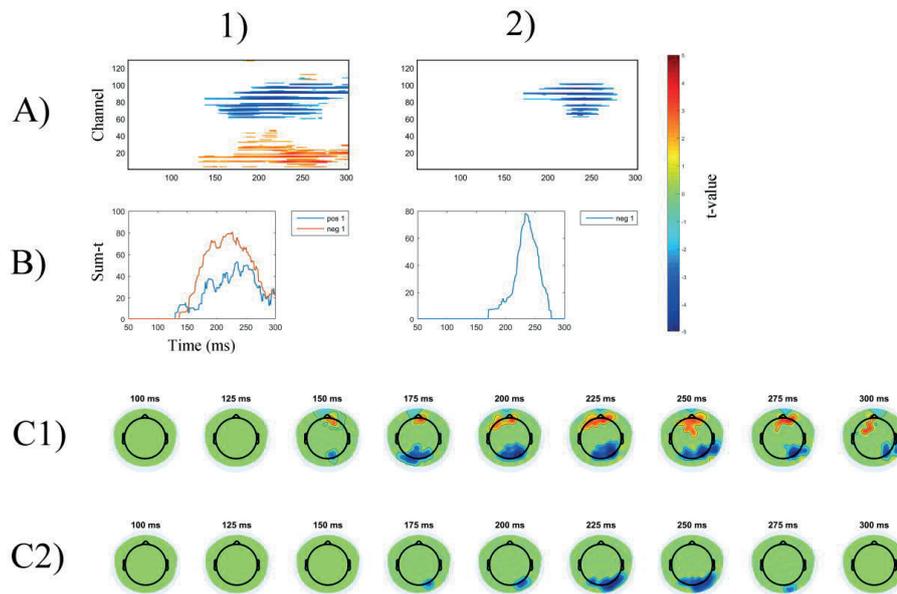


Figure 7: Post hoc analysis results of one-sample permutation t-test on word length effect in additional fixation. A1) Raster plot of word length effect in additional fixation for typical readers. A2) Raster plot of word length effect in additional fixation for slow readers. B1) Time behaviour of the sum-t parameter from significant clusters of word length effect in additional fixation for typical readers. B2) Time behaviour of the sum-t parameter from significant clusters of word length effect in additional fixation for slow readers. C1) Selected time point topographical representation of word length effect in additional fixation for typical readers. C2): Selected time point topographical representation of word length effect in additional fixation for slow readers.

In order to illustrate the effect of word length on additional fixation, we generated a prediction on 5-, 9- and 13-lettered words and added the intercept to these predictions. Predictions of word length effect on additional fixation are presented in Figure 8.

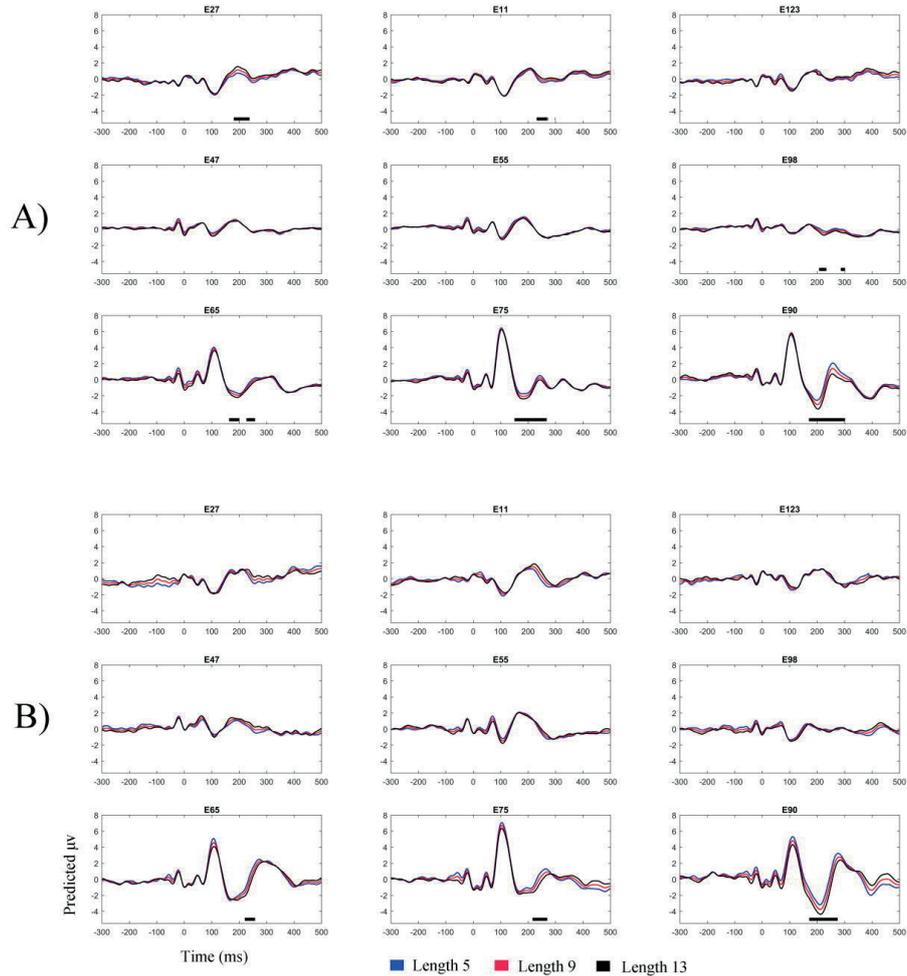


Figure 8: Predicted word length effects from additional fixations on specific word lengths added to the additional fixation intercept. A) Typical readers ($N = 65$). B) Slow readers ($N = 27$). The blue line indicates the word length effect prediction on five-lettered words added to the intercept. The red line indicates the word length effect prediction on nine-lettered words added to the intercept. The black line indicates the word length effect prediction on 13-lettered words added to the intercept. Boxes under the waveform represent the time points where the channel belongs to a significant cluster in the one-sample test of the word length continuous beta parameter from additional fixation.

4 Discussion

In this study, we examined the influence of word length and reading fluency on eye movements and fixation-related brain potentials in school-aged children. Overall, we found four noteworthy patterns of data. First, there was a difference between slow and typical readers in brain activity, overall fixation duration, gaze duration, first fixation duration and re-fixation probability, all indicating a fundamental difference in cognitive processing. Second, in eye movements, for both groups, the durations of first fixations were modulated by word length but durations of additional fixations were not, thereby suggesting that cognitive processing is different during first and additional fixations for both groups. Among typical readers, this was complemented by different brain activity for additional fixations, and among slow readers, the cortical effect was not observed (in terms of alpha level 0.05). Third, we observed stronger word length effects in the eye movements of slow readers than in those of typical readers and, thus, conceptually replicated earlier findings (Hawelka et al., 2010) of stronger word length effect in individuals with reading difficulty. Fourth, we observed that word length modulated cortical activity only during additional fixation for both groups, which has a clear discrepancy from the finding on eye movements that word length essentially modulated only first fixations and not additional fixations. Next, we discuss these patterns separately and suggest experiment designs for further examination.

The statistical analysis of the deconvoluted FRPs showed a clear pattern of results. The intercept waveforms, which represent shared activity in all responses, were different between the groups from 140 ms onwards, as the responses for the slow readers were larger. Given that there is clear polarity reversal to negative voltages in the occipital sensors in contrast to the positive difference on central sensors, it is reasonable to assume that the sources at the occipital cortex or bilateral posterior temporal cortices are the main contributors for this difference⁶. Thus, we essentially observed stronger occipital N1 and stronger subsequent occipital positivity for slow readers than typical readers. Similarly, we found clear group differences in eye movements, in which fixation duration measures and gaze durations were longer and re-fixation probability was higher for slow readers. Elevated fixation durations, gaze durations and re-fixation probabilities are indicative of cumbersome cognitive processing.

⁶ We ran source analysis on the group average intercepts and found this to be the case. Source analysis protocol and the results are presented in supplementary material 2.

The between-group differences observed in the intercept waveforms and all eye movement variables are compatible with the visual attention span hypothesis of dyslexia (Bosse et al., 2007). In terms of the visual attention span hypothesis, it may well be that the difference that we observe between slow and typical readers stems from the reduced capability of slow readers to process groups of letters in parallel or as unified percepts; however, typical readers manage to distribute their processing to a broader range of visual elements either by processing more letters in parallel (Bosse et al., 2007) or managing a larger number of letters to be connected to the corresponding phonemes during single fixation. This difficulty would then be reflected in overall longer fixations, higher re-fixation probability, and stronger intercept response. The reason why we believe that the difference we observe is related to the manner in which processing is distributed is due to previous FRP research with gaze contingent manipulations, in which it has been shown that parafoveal masking increases negativity on the left occipital electrodes in the time window from 170 ms to 300 ms (Kornrumpf, Niefind, Sommer, & Dimigen, 2016). Thus, it might be that the difference we observe between groups is related to the extent that slow readers can utilize parafoveal information, possibly due to attention span or cumbersome serial decoding that focuses the processing to a small number of letters in any single fixation.

In our opinion, the optimal design to explore the difference between slow readers and typical readers related to parafoveal processing would be to combine the moving window technique and fixation related potentials. Previous research (on eye movements only) with differing levels of reading ability has hinted that limiting the span of the window where the participant can extract the correct visual information makes good readers behave in a more similar manner to bad readers (Häikiö et al., 2009; Rayner, 1986, 2014; Veldre & Andrews, 2014). The hypothesis for such an experiment is that the brain activity difference in the FRP between typical and slow readers will disappear as the moving window is made narrower.

In the eye movements, there was clear support for differences in cognitive processing during additional fixations than during first fixations. This was evident in the manner in which word length modulated only first fixations and this applied to both groups, although the influence of word length on first fixation was stronger in the slow readers group than in the typical readers group. This processing difference was also observed in the FRP of typical readers, where a large portion of the activity was different for additional fixations from first fixations. The effect was present throughout the entire tested time window, thereby suggesting that it might be a part of a response that is carried over from pre-saccadic processing. Conversely,

we did not observe such a pronounced effect among slow readers, as statistical support was not sufficient. Overall, the results on the effect of additional fixations are not particularly clear, and in the confines of this data set we are unable to provide convincing suggestions regarding what is different in the cognitive processing, but only indicate that it is likely different. Thus, we suggest that this factor must be examined in another existing ET-EEG data set, where available, before designing specific FRP experiments to probe underlying mechanisms.

Further, we observed that word length effect in gaze duration and re-fixation probability were both increased as a function of word length. On the other hand, the first fixation duration decreased as a function of word length. Overall, slow readers exhibited longer first fixation durations and gaze duration as well as higher re-fixation probability than typical readers. Further, the influence of word length was larger for the slow readers than for typical readers, which illustrates that word length places higher demand on the cognition of slow readers than that of typical readers. At a conceptual level, these findings replicate previous research (Hawelka et al., 2010) and are well in line with the interpretation that longer word length effect in dyslexic readers is due to stronger reliance on letter-by-letter decoding. Further, the results on standard eye movement variables illustrate that our results on brain activity have a clear association with existing literature on eye movement.

The impact of word length on brain activity was found to have a fairly perplexing pattern. Recall that with regard to eye movements, we observed reversed word length effect in the first fixation duration, but no effect of word length for the duration of additional fixation. In contrast, we do not observe word length effect in brain activity during the first fixation, but we observe word length effect in brain activity during additional fixation. Moreover, we observe a difference in the word length effect in the eye movements, where slow readers are affected more severely by word length than typical readers, but any difference was not observed between word length effects in additional fixations of slow readers and typical readers. This claim must be considered carefully, as tests used for eye movement variables and FRP do not have equal sensitivity in terms of statistical power. Here, LME models utilized single trial-level information, and permutation tests utilized participant-level estimates, which leads to a very high difference in the degrees of freedom between the two test types. Thus, if the difference in the effect of word length on brain activity is weak, it might end up being unobserved in the statistical method employed in this study.

The discrepancy—that we observe word length effect in first fixations for eye movements and in additional fixations for brain activity—raises the question of whether the brain activity modulation by word length is in fact word length effect per se or an effect of some other variable/factor that has a strong correlation to the word length effect in FRPs. In this data, one obvious candidate is the spatial width of the word. It has been previously established that the two aspects of word length, the purely visual spatial width and the more linguistic number of letters, have independent contributions to gaze behaviour: temporal measures are more driven by the linguistic characteristics of words and spatial measures are more driven by the spatial characteristics of words (Hautala et al., 2011; Hautala & Loberg, 2015). The idea that the word length effect that we observe in our additional fixation FRP reflects visual characteristics is compatible with recent findings on manipulating letter sparsity in FRP studies (Weiss, Knakker, & Vidnyánszky, 2016) in which the authors found that diminishing spacing between letters led to stronger negative deflection of FRP around 200 ms in the right occipital regions. This is also supported by a recent study among children where, in a 1-back RSVP task (an RSVP paradigm where a participant monitors whether a stimulus is instantly repeated), stimulus length (both words and symbol strings) increased N1 (peak ~200ms) amplitudes, mostly in the right occipital region (Fraga González et al., 2014). To the best of our knowledge, the capability to utilize large-scale visual saliency is preserved in dyslexia and we have not found reports of evidence to the contrary; thus, the suggestion that the effect presented in Figures 7 and 8 would be generated by an increase in visuo-spatial width for both groups is feasible. In order to test this, we propose an experiment in which the number of letters and spatial width is orthogonally contrasted in materials with proportional fonts, which is similar to our earlier experiments (Hautala et al., 2011; Hautala & Loberg, 2015). If the pattern we observe is due to the spatial aspect, then we should be able to observe the same pattern we observe now by keeping the amount of letters constant and manipulating the spatial extent.

In addition, word length effect on cortical processing might be lagged in relation to eye movements. Current implementations of computational models that attempt to explain what determines eye movements during the process of reading incorporate a degree of lag between the eye movement and the cognitive process (Engbert, Nuthmann, Richter, & Kliegl, 2005; Reichle, Warren, & McConnell, 2009). The literature well acknowledges that semantic comprehension is lagged in relation to single fixation duration (Dambacher & Kliegl, 2007; Kliegl, Dambacher, Dimigen, Jacobs, & Sommer, 2012; Sereno, Rayner, & Posner, 1998).

However, we are not aware of studies that would show lag to be associated with word length and our results cannot eliminate this possibility. Fortunately, the possibility for word length effect being lagged in the cortical responses could be tested in disappearing text paradigm coupled with fixation-related potentials. Disappearing text is a gaze-contingent manipulation where words of a sentence are removed after the sentence has been looked at once. In theory, it must be possible to set the time course of the disappearance so that the participant could still program and execute small saccades leading to re-fixations yet gain no visual input from them. Thus, if the processing of word length is lagged, the effect of word length would still appear on the additional fixations with no additional input. If the cortical effect observed in the additional fixations is tied to those fixations, then elimination of the visual input from additional fixations is likely to remove the effect. In previous research, it has been shown that, in adults, the probability of re-fixations to the location of the word is not affected by disappearing text (Blythe, Liversedge, Joseph, White, & Rayner, 2009); in children, this is also true if the time window between fixation onset and word disappearance is sufficiently long (Blythe et al., 2009).

The EEG signal is generated by synchronous activity in the pyramidal cell columns of the grey matter of the cerebrum. Circuits which are associated with the generation of eye movements are not limited to the cerebrum, but include regions where EEG is effectively blind (without very specific instrumentation), such as nuclei in cerebellum and brainstem (Martinez-Conde, Otero-Millan, & Macknik, 2013). Very recently, it was also found that human cerebellum houses retinotopic maps in the oculomotor vermis (van Es, Zwaag, & Knapen, 2018), a structure also implicated in the control of eye movements (Martinez-Conde et al., 2013). Contributions of such areas to eye-movement patterns during reading are currently unknown and, therefore, the EEG measures of the present study likely miss a few neural processes related to the control of eye movements and visual processing. A further limitation of the present study is related to critical modulations of cortical activity during the first fixation, which might not be such that they impact ERP/FRP modulations. Features that could reflect the critical features of the word length effect during first fixation and are not investigated here include induced responses and brain connectivity patterns, which require time-frequency decomposition and specialized analyses to uncover. For example, recently, connectivity patterns related to words have been found in the high gamma frequency range (60–90 Hz) (Liljeström, Vartiainen, Kujala, & Salmelin, 2018) at which FRP methodology in current implementations is essentially blind. Overall, these two aspects are not intended to

diminish the applicability and value of the FRP approach. FRPs are rather powerful extensions to both EEG and eye-tracking, but the limitations of the technique do need to be understood to come to valid conclusions.

5 Conclusions

In this study, we examined cortical and eye-movement correlates of word length and reading ability during naturalistic reading in school-aged children. We found that eye movements follow previously established patterns, where reading ability and word length both modulate eye movements and that there is an interaction between these effects. In corresponding brain activity, we observe that the intercept brain response—which reflects activity common to all FRPs—is different (approximately 150 ms to 300 ms) between groups but that there is no statistical support for claiming that word length effect would manifest differently in cortical activity in children with different reading skills. Thus, we suggest that foveal difficulty in processing of multiple visual units in parallel functions as a bottleneck for word recognition among slow readers and, thus, leads to increased fixation durations and higher re-fixation probability for parsing the entire word. Thus, the apparently much stronger effect in gaze duration for slow readers is, in great likelihood, the result of each fixation contributing less towards the recognition process for slow readers and accumulates to rather long gaze durations. We suggested a set of FRP experiments to examine these conclusions.

Acknowledgements

This research was funded by grant 274022 from the Academy of Finland ‘Internet and learning difficulties: multidisciplinary approach for understanding information seeking in new media (eSeek)’ to Paavo H.T. Leppänen. The authors would like to thank all their research colleagues, research assistants and students involved in the research project.

References

- Assadollahi, R., & Pulvermüller, F. (2001). Neuromagnetic evidence for early access to cognitive representations. *Neuroreport*, *12*(2), 207–213. <https://doi.org/11209922>
- Barton, J. J. S., Hanif, H. M., Björnström, L. E., & Hills, C. (2014). *Cognitive Neuropsychology The word-length effect in reading: A review*. <https://doi.org/10.1080/02643294.2014.895314>
- 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>
- Bertram, R., & Hyönä, J. (2003). The length of a complex word modifies the role of morphological structure: Evidence from eye movements when short and long Finnish compounds. *Journal of Memory and Language*, Vol. 48, pp. 615–634. [https://doi.org/10.1016/S0749-596X\(02\)00539-9](https://doi.org/10.1016/S0749-596X(02)00539-9)
- Blythe, H. I., Häikiö, T., Bertam, R., Liversedge, S. P., & Hyönä, J. (2011). Reading disappearing text: Why do children refixate words? *Vision Research*, *51*(1), 84–92. <https://doi.org/10.1016/J.VISRES.2010.10.003>
- Blythe, H. I., & Joseph, H. S. S. L. (2012). Children's eye movements during reading. In *The Oxford Handbook of Eye Movements*. <https://doi.org/10.1093/oxfordhb/9780199539789.013.0036>
- Blythe, H. I., Liversedge, S. P., Joseph, H. S. S. L., White, S. J., & Rayner, K. (2009). Visual information capture during fixations in reading for children and adults. *Vision Research*, *49*(12), 1583–1591. <https://doi.org/10.1016/J.VISRES.2009.03.015>
- Bosse, M. L., Tainturier, M. J., & Valdois, S. (2007). Developmental dyslexia: The visual attention span deficit hypothesis. *Cognition*, *104*(2), 198–230. <https://doi.org/10.1016/j.cognition.2006.05.009>
- Brem, S., Bach, S., Kucian, K., Kujala, J. V., Guttorm, T. K., Martin, E., ... Richardson, U. (2010). Brain sensitivity to print emerges when children learn letter–speech sound correspondences. *Proceedings of the National Academy of Sciences*, *107*(17), 7939–7944. <https://doi.org/10.1073/PNAS.0904402107>
- Cornelissen, T., Sassenhagen, J., & Vö, M. L.-H. (2018). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods*. <https://doi.org/10.1016/J.JNEUMETH.2018.12.010>
- Dambacher, M., & Kliegl, R. (2007). Synchronizing timelines: Relations between fixation durations and N400 amplitudes during sentence reading. *Brain Research*, *1155*, 147–162. <https://doi.org/10.1016/j.brainres.2007.04.027>
- Dandekar, S., Privitera, C., Carney, T., & Klein, S. A. (2012). Innovative Methodology Neural saccadic response estimation during natural viewing. *J Neurophysiol*, *107*, 1776–1790. <https://doi.org/10.1152/jn.00237.2011>
- De Luca, M., Borrelli, M., Judica, A., Spinelli, D., & Zoccolotti, P. (2002). Reading words and pseudowords: An eye movement study of developmental dyslexia. *Brain and Language*, *80*(3), 617–626. <https://doi.org/10.1006/brln.2001.2637>

- De Luca, M., Di Pace, E., Judica, A., Spinelli, D., & Zoccolotti, P. (1999). Eye movement patterns in linguistic and non-linguistic tasks in developmental surface dyslexia. *Neuropsychologia*, *37*(12), 1407–1420. [https://doi.org/10.1016/S0028-3932\(99\)00038-X](https://doi.org/10.1016/S0028-3932(99)00038-X)
- Dehaene, S. (1996). The organization of brain activations in number comparison: Event-related potentials and the additive-factors method. *Journal of Cognitive Neuroscience*, *8*(1), 47–68. <https://doi.org/10.1162/jocn.1996.8.1.47>
- Dimigen, O. (2018). Optimized ICA-based removal of ocular EEG artifacts from free viewing experiments. *BioRxiv*, 446955. <https://doi.org/10.1101/446955>
- Dürwächter, U., Sokolov, A. N., Reinhard, J., Klosinski, G., & Trauzettel-Klosinski, S. (2010). Word length and word frequency affect eye movements in dyslexic children reading in a regular (German) orthography. *Annals of Dyslexia*, *60*(1), 86–101. <https://doi.org/10.1007/s11881-010-0034-9>
- Ehinger, B. V., & Dimigen, O. (2018). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *BioRxiv*, 360156. <https://doi.org/10.1101/360156>
- Eklund, K., Torppa, M., Aro, M., Leppanen, P. H. T., & Lyytinen, H. (2015). Literacy skill development of children with familial risk for dyslexia through grades 2, 3, and 8. *Journal of Educational Psychology*, *107*(1), 126–140. <https://doi.org/10.1037/a0037121>
- Engbert, R., & Mergenthaler, K. (2006). Microsaccades are triggered by low retinal image slip. *Proceedings of the National Academy of Sciences*, *103*(18), 7192–7197. <https://doi.org/10.1073/pnas.0509557103>
- Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: A Dynamical Model of Saccade Generation During Reading. *Psychological Review*, *112*(4), 777–813. <https://doi.org/10.1037/0033-295X.112.4.777>
- Fraga González, G., Žarić, G., Tijms, J., Bonte, M., Blomert, L., Leppänen, P., & van der Molen, M. W. (2016). Responsivity to dyslexia training indexed by the N170 amplitude of the brain potential elicited by word reading. *Brain and Cognition*, *106*, 42–54. <https://doi.org/10.1016/j.BANDC.2016.05.001>
- Fraga González, G., Žarić, G., Tijms, J., Bonte, M., Blomert, L., & van der Molen, M. W. (2014). Brain-potential analysis of visual word recognition in dyslexics and typically reading children. *Frontiers in Human Neuroscience*, *8*, 474. <https://doi.org/10.3389/fnhum.2014.00474>
- Häikiö, T., Bertram, R., Hyönä, J., & Niemi, P. (2009). Development of the letter identity span in reading: Evidence from the eye movement moving window paradigm. *Journal of Experimental Child Psychology*, *102*(2), 167–181. <https://doi.org/10.1016/j.jecp.2008.04.002>
- Hasko, S., Groth, K., Bruder, J., Bartling, J., & Schulte-Körne, G. (2013). The time course of reading processes in children with and without dyslexia: an ERP study. *Frontiers in Human Neuroscience*, *7*, 570. <https://doi.org/10.3389/fnhum.2013.00570>
- Hauk, O., Davis, M. H., Ford, M., Pulvermüller, F., & Marslen-Wilson, W. D. (2006). The time course of visual word recognition as revealed by linear regression analysis of ERP

- data. *NeuroImage*, 30(4), 1383–1400.
<https://doi.org/10.1016/J.NEUROIMAGE.2005.11.048>
- Hauk, O., & Pulvermüller, F. (2004). Effects of word length and frequency on the human event-related potential. *Clinical Neurophysiology*, 115(5), 1090–1103.
<https://doi.org/10.1016/j.clinph.2003.12.020>
- Hauk, O., Pulvermüller, F., Ford, M., Marslen-Wilson, W. D., & Davis, M. H. (2009). Can I have a quick word? Early electrophysiological manifestations of psycholinguistic processes revealed by event-related regression analysis of the EEG. *Biological Psychology*, 80(1), 64–74. <https://doi.org/10.1016/J.BIOPSYCHO.2008.04.015>
- Hautala, J., Hyönä, J., & Aro, M. (2011). Dissociating spatial and letter-based word length effects observed in readers' eye movement patterns. *Vision Research*, 51(15), 1719–1727. <https://doi.org/10.1016/j.visres.2011.05.015>
- Hautala, J., & Loberg, O. (2015). Breaking down the word length effect on readers' eye movements. *Language, Cognition and Neuroscience*, 30(8), 993–1007.
<https://doi.org/10.1080/23273798.2015.1049187>
- Hawelka, S., Gagl, B., & Wimmer, H. (2010). A dual-route perspective on eye movements of dyslexic readers. *Cognition*, 115(3), 367–379.
<https://doi.org/10.1016/j.cognition.2009.11.004>
- Helenius, P., Tarkiainen, A., Cornelissen, P. L., Hansen, P. C., & Salmelin, R. (1999). Dissociation of Normal Feature Analysis and Deficient Processing of Letter-strings in Dyslexic Adults. *Cerebral Cortex*, 9(5), 476–483. <https://doi.org/10.1093/cercor/9.5.476>
- Hyöna, J., & Olson, R. K. (1995). Eye fixation patterns among dyslexic and normal readers: Effects of word length and word frequency. *Journal of Experimental Psychology. Learning, Memory & Cognition*, Vol. 21, p. 1430. <https://doi.org/10.1037/0278-7393.21.6.1430>
- Kagan, I., & Hafed, Z. M. (2013). Active vision: Microsaccades direct the eye to where it matters most. *Current Biology*, 23(17). <https://doi.org/10.1016/j.cub.2013.07.038>
- Kiili, C., Leu, D. J., Utraiainen, J., Coiro, J., Kannianen, L., Tolvanen, A., ... Leppänen, P. H. T. (2018). Reading to Learn From Online Information: Modeling the Factor Structure. *Journal of Literacy Research*, 50(3), 304–334.
<https://doi.org/10.1177/1086296X18784640>
- Kliegl, R., Dambacher, M., Dimigen, O., Jacobs, A. M., & Sommer, W. (2012). Eye movements and brain electric potentials during reading. *Psychological Research*, 76(2), 145–158. <https://doi.org/10.1007/s00426-011-0376-x>
- Kliegl, R., Grabner, E., Rolfs, M., & Engbert, R. (2004). Length, frequency, and predictability effects of words on eye movements in reading. *European Journal of Cognitive Psychology*, 16, 1–2. <https://doi.org/10.1080/09541440340000213>
- Kliegl, R., Olson, R. K., & Davidson, B. J. (1983). On problems of unconfounding perceptual and language processes. *Eye Movements in Reading: Perceptual and Language Processes*, 333–343.
- Kornrumpf, B., Niefind, F., Sommer, W., & Dimigen, O. (2016). Neural correlates of word

- recognition: A systematic comparison of natural reading and RSVP. *Journal of Cognitive Neuroscience*, 1–35.
- Kristensen, E., Rivet, B., & Guérin-Dugué, A. (2017). Estimation of overlapped Eye Fixation Related Potentials: The General Linear Model, a more flexible framework than the ADJAR algorithm. *Journal of Eye Movement Research*, 10(1), 1–27. <https://doi.org/10.16910/JEMR.10.1.7>
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, 82(13), 1–26. <https://doi.org/10.18637/jss.v082.i13>
- Lee, T.-W., Girolami, M., & Sejnowski, T. J. (1999). Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub - Gaussian and Super - Gaussian Sources. *Neural Computation*, 11(2), 409–433. <https://doi.org/10.1162/089976699300016719>
- Liljeström, M., Vartiainen, J., Kujala, J., & Salmelin, R. (2018). Large-scale functional networks connect differently for processing words and symbol strings. *PLOS ONE*, 13(5), e0196773. <https://doi.org/10.1371/journal.pone.0196773>
- Lindeman, J. (1998). *ALLU-Ala-asteen lukutesti [ALLU-Reading Test for Primary School]* (Vol. 90). Vol. 90. [https://doi.org/10.1016/S0022-3913\(03\)00368-8](https://doi.org/10.1016/S0022-3913(03)00368-8)
- Loberg, O., Hautala, J., Hämäläinen, J. A., & Leppänen, P. H. T. (2018). Semantic anomaly detection in school-aged children during natural sentence reading – A study of fixation-related brain potentials. *PLOS ONE*, 13(12), e0209741. <https://doi.org/10.1371/journal.pone.0209741>
- Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Frontiers in Human Neuroscience*, 8, 213. <https://doi.org/10.3389/fnhum.2014.00213>
- Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190. <https://doi.org/10.1016/j.jneumeth.2007.03.024>
- Martinez-Conde, S., Otero-Millan, J., & Macknik, S. L. (2013). The impact of microsaccades on vision: towards a unified theory of saccadic function. *Nature Reviews. Neuroscience*, 14(2), 83–96. <https://doi.org/10.1038/nrn3405>
- Maurer, U., Brem, S., Bucher, K., Kranz, F., Benz, R., Steinhausen, H.-C., & Brandeis, D. (2007). Impaired tuning of a fast occipito-temporal response for print in dyslexic children learning to read. *Brain*, 130(12), 3200–3210. <https://doi.org/10.1093/brain/awm193>
- Nevala, J., & Lyytinen, H. (2000). *Sanaketjutesti [Word chain test]*. Jyväskylä: Niilo Mäki Institute.
- Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, 107, 55–83. <https://doi.org/10.1016/J.BANDC.2016.06.004>
- Norton, E. S., Beach, S. D., & Gabrieli, J. DE. (2015). Neurobiology of dyslexia. *Current*

- Opinion in Neurobiology*, 30, 73–78. <https://doi.org/10.1016/j.conb.2014.09.007>
- Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, 6, 278. <https://doi.org/10.3389/fnhum.2012.00278>
- Rayner, K. (1986). Eye movements and the perceptual span in beginning and skilled readers. *Journal of Experimental Child Psychology*, 41, 211–236. [https://doi.org/10.1016/0022-0965\(86\)90037-8](https://doi.org/10.1016/0022-0965(86)90037-8)
- Rayner, K. (1997). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Rayner, K. (2014). *Visual Cognition The gaze-contingent moving window in reading: Development and review*. <https://doi.org/10.1080/13506285.2013.879084>
- Reichle, E. D., & Reingold, E. M. (2013). Neurophysiological constraints on the eye-mind link. *Frontiers in Human Neuroscience*, 7, 361. <https://doi.org/10.3389/fnhum.2013.00361>
- 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), 1–21. <https://doi.org/10.3758/PBR.16.1.1>
- Ries, A. J., Slayback, D., & Touryan, J. (2018). The fixation-related lambda response: Effects of saccade magnitude, spatial frequency, and ocular artifact removal. *International Journal of Psychophysiology*, 134, 1–8. <https://doi.org/10.1016/j.ijpsycho.2018.09.004>
- Salmelin, R., Service, E., Kiesila, P., & Uutela, K. (1996). *Impaired Visual Word Processing in Dyslexia Revealed with Magnetoencephalography*.
- Sereno, S. C., Rayner, K., & Posner, M. I. (1998). Establishing a time-line of word recognition: Evidence from eye movements and event-related potentials. *Cognitive Neuroscience*, 9(10), 2195–2200. <https://doi.org/10.1097/00001756-199807130-00009>
- Seymour, P. H. K., Aro, M., & Erskine, J. M. (2003). Foundation literacy acquisition in European orthographies. *British Journal of Psychology (London, England : 1953)*, 94(Pt 2), 143–174. <https://doi.org/10.1348/000712603321661859>
- Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, 52(2), 157–168. <https://doi.org/10.1111/psyp.12317>
- Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, 52(2), 169–181. <https://doi.org/10.1111/psyp.12320>
- Spironelli, C., & Angrilli, A. (2009). Developmental aspects of automatic word processing: Language lateralization of early ERP components in children, young adults and middle-aged subjects. *Biological Psychology*, 80(1), 35–45. <https://doi.org/10.1016/J.BIOPSYCHO.2008.01.012>

- Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research, 113*, 33–43. <https://doi.org/10.1016/J.VISRES.2015.05.008>
- Valdois, S., Bosse, M. L., & Tainturier, M. J. (2004). The cognitive deficits responsible for developmental dyslexia: Review of evidence for a selective visual attentional disorder. *Dyslexia, 10*(4), 339–363. <https://doi.org/10.1002/dys.284>
- van Es, D. M., Zwaag, W. van der, & Knapen, T. (2018). Retinotopic maps of visual space in the human cerebellum. *BioRxiv*, 455170. <https://doi.org/10.1101/455170>
- Van Petten, C., & Kutas, M. (1990). Interactions between sentence context and word frequency in event-related brain potentials. *Memory & Cognition, 18*(4), 380–393. <https://doi.org/10.3758/BF03197127>
- Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *Quarterly Journal of Experimental Psychology, 67*(4), 703–727. <https://doi.org/10.1080/17470218.2013.826258>
- Weiss, B., Knakker, B., & Vidnyánszky, Z. (2016). Visual processing during natural reading. *Scientific Reports, 6*(May), 26902. <https://doi.org/10.1038/srep26902>
- Wydell, T. N., Vuorinen, T., Helenius, P., & Salmelin, R. (2003). Neural Correlates of Letter-String Length and Lexicality during Reading in a Regular Orthography. *Journal of Cognitive Neuroscience, 15*(7), 1052–1062. <https://doi.org/10.1162/089892903770007434>
- Ziegler, J. C., & Goswami, U. (2005). Reading acquisition, developmental dyslexia, and skilled reading across languages: a psycholinguistic grain size theory. *Psychological Bulletin, 131*(1), 3–29. <https://doi.org/10.1037/0033-2909.131.1.3>

Supplementary material 1

Fixation-related potential methodology

In ERP research, the traditional approach to study cortical correlates of reading has been to present single words in succession. Research utilizing the RSVP paradigm has formed the basis of the knowledge on brain activity during reading; however, there are severe limitations in the ecological validity of the paradigm. Typical implementations of RSVP exclude the possibility for a participant to modulate reading behaviour in relation to text features. Thus, the participant cannot adjust visual information uptake periods or return to a specific point in the material. However, there are no such restrictions during natural reading. Recently, a more ecologically valid alternative was introduced, which capitalizes on the fact that fixation onsets cause highly similar event-related perturbations on EEG as visual stimulus onsets do (Dimigen, Sommer, Hohlfeld, Jacobs, & Kliegl, 2011; Yagi, 1979). Thus, by co-registering eye movements with an eye-tracker in synchrony with EEG, there is a possibility of extracting FRPs from naturalistic reading situations (Baccino & Manunta, 2005; Degno et al., 2018; Dimigen et al., 2011; Kretzschmar, Bornkessel-Schlesewsky, & Schlewsky, 2009; Loberg, Hautala, Hämäläinen, & Leppänen, 2018; Metzner, von der Malsburg, Vasishth, & Rösler, 2015, 2016; Niefind & Dimigen, 2016). However, the FRP methodology has a set of problems that require significant attention. The better recognized problem associated with recording electrophysiological brain activity during free reading is related to eye-movement artefacts, that are in time-lock with the signal of interest; yet, ultimately, this problem has proven to be a rather trivial one that can be taken care of with blind source separation methods like Independent Component Analysis (ICA) (Kretzschmar, Schlewsky, & Staub, 2015; Meyberg, Sommer, & Dimigen, 2017; Plöchl, Ossandón, & König, 2012), which are commonly used in EEG pre-processing.

Further, FRPs have two more severe problems. First, there is the issue that eye-movement characteristics directly modulate brain activity (Cornelissen, Sassenhagen, & Vö, 2018; Dimigen, Valsecchi, Sommer, & Kliegl, 2009; Nikolaev, Meghanathan, & van Leeuwen, 2016; Ries, Slayback, & Touryan, 2018) and task requirements modulate eye-movement characteristics, which introduces confounding factors to the interpretation of observed effects in brain activity. Second, there is a temporal overlap (Cornelissen et al., 2018; Dimigen et al., 2011; Nikolaev et al., 2016) from adjacent fixations that is modulated by temporal characteristics of eye

movements, which can indirectly result as a confounding factor in the effects observed in brain activity. Both these issues can be solved by abandoning averaging as a response estimation procedure.

A typical method for estimating event-related activity from EEG is the averaging of stimulus time-locked segments by time point and channel. Averaging works well in simple experiment designs with few individual stimuli, such as the oddball paradigm; however, problems arise in experiment designs that use multiple, but different, instances of categorical stimuli—for example, printed words. Such stimuli have multiple sources of feature variance and all of these are generally not of interest, and the general idea is to control for unwanted influences by minimizing their influence or balancing them out in various categories. Often, the required level of control on naturalistic stimuli is unachievable due to the plurality of confounding factors; this can lead to sporadic and results that are not replicable in relation to the interesting factor. Essentially, this is a weakness of the averaging procedure, as it relies on the noise being random in relation to the forced category in the interesting factor.

However, new approaches are available, as it was recently illustrated that averaging is essentially least squares regression (Smith & Kutas, 2015a). The core difference in this from traditional averaging is the treatment of noise. In averaging, noise is incorporated in to the average; in contrast, when the response is estimated with a categorical regression model, the noise is allocated to a separate noise term. Thus, in estimating responses from naturalistic stimuli with multiple confounding factors, the GLM framework is more suited to deal with ‘noise’ from uninteresting stimulus features than averaging is. Strong collinearities between (an) interesting and uninteresting factor(s) remains an issue in this regard as well (Cornelissen et al., 2018). Further, it must be noted that where averaging is dependent on defining categories to group the observations, the GLM approach does not have such a limitation but can in fact handle continuous variables directly. For FRPs, this is particularly beneficial, as this enables the separation of uninteresting direct influences on brain activity that are partially correlated to interesting influences on brain activity. In practice, this implies that we can isolate the effect that word length has on brain activity from the word length modulated saccade amplitude effect on brain activity (Cornelissen et al., 2018). After estimation of regression, the ERP/FRP estimates are treated in the same manner as averaged response is. For a discussion on the conceptual

relationship between regression estimates and averaged estimates of electrophysiological brain activity, see (Smith & Kutas, 2015a).

A temporal overlap of components can be considered to be the most severe known issue associated with FRPs. A problematic amount of overlap manifests in fixation-related potentials because fixations during reading occur from 200 ms to 300 ms, on average, from each other (Rayner, 1997) and known reading-related components can occur up to 900ms from stimulus onset (Friederici & Weissenborn, 2007). Further, critical eye-movement features are modulated by experimental conditions and, consequently, this modulates an overlap between conditions; in turn, this complicates the interpretation of the effects as overlap difference may well manifest as a scalp perturbation difference that can be misidentified as a correlate of cognitive activity (Frey et al., 2013; Nikolaev et al., 2016). To the best of our knowledge, three solutions to this issue are known: eye-movement feature matching (Dimigen et al., 2011; Nikolaev et al., 2016), where characteristics of eye movements are matched between conditions, adjacent response estimation algorithms (ADJAR) for EEG (Kristensen, Rivet, & Guérin-Dugué, 2017; Woldorff, 1993) and linear deconvolution (Cornelissen et al., 2018; Ehinger & Dimigen, 2018; Kristensen et al., 2017; Smith & Kutas, 2015b), which is a method that is originally from the fMRI analysis framework. Of these approaches, the eye-movement feature matching procedure (Dimigen et al., 2011; Nikolaev et al., 2016) has an obvious drawback of potentially reducing the ecological validity of the analysis, as modulations of eye movements are associated with the same cognitive mechanisms that we are interested in in the cortical activity. Of the latter two, linear deconvolution has recently been shown to outperform ADJAR (Kristensen et al., 2017).

Recently, a GLM-based deconvolution approach was implemented in an easy-to-access toolbox (Ehinger & Dimigen, 2018); in this framework, each response is modelled in relation not only to the time-locking event but also to surrounding events that produce overlapping perturbations. The temporal variance in relative time distances between events is utilized to identify those parts of voltage variance that belong to each event. Effectively, this implies that brain activity related to events that have high temporal collinearity—for example, saccade onset and fixation onset—cannot be deconvolved from each other (Cornelissen et al., 2018).

References:

- Baccino, T., & Manunta, Y. (2005). Eye-fixation-related potentials: Insight into parafoveal processing. *Journal of Psychophysiology*, *19*(3), 204–215. <https://doi.org/10.1027/0269-8803.19.3.204>
- Cornelissen, T., Sassenhagen, J., & Võ, M. L.-H. (2018). Improving free-viewing fixation-related EEG potentials with continuous-time regression. *Journal of Neuroscience Methods*. <https://doi.org/10.1016/J.JNEUMETH.2018.12.010>
- Degno, F., Loberg, O., Zang, C., Zhang, M., Donnelly, N., & Liversedge, S. P. (2018). *Parafoveal Previews and Lexical Frequency in Natural Reading: Evidence From Eye Movements and Fixation-Related Potentials*. <https://doi.org/10.1037/xge0000494>
- Dimigen, O., Sommer, W., Hohlfeld, A., Jacobs, A. M., & Kliegl, R. (2011). Coregistration of eye movements and EEG in natural reading: Analyses and review. *Journal of Experimental Psychology: General*, *140*(4), 552–572. <https://doi.org/10.1037/a0023885>
- Dimigen, O., Valsecchi, M., Sommer, W., & Kliegl, R. (2009). Human Microsaccade-Related Visual Brain Responses. *Journal of Neuroscience*, *29*(39), 12321–12331. <https://doi.org/10.1523/JNEUROSCI.0911-09.2009>
- Ehinger, B. V., & Dimigen, O. (2018). Unfold: An integrated toolbox for overlap correction, non-linear modeling, and regression-based EEG analysis. *BioRxiv*, 360156. <https://doi.org/10.1101/360156>
- Frey, A., Ionescu, G., Lemaire, B., López-Orozco, F., Baccino, T., & Guérin-Dugué, A. (2013). Decision-making in information seeking on texts: an eye-fixation-related potentials investigation. *Frontiers in Systems Neuroscience*, *7*(39), 1–22. <https://doi.org/10.3389/fnsys.2013.00039>
- Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research*, Vol. 1146, pp. 50–58. <https://doi.org/10.1016/j.brainres.2006.08.038>
- Kretzschmar, F., Bornkessel-Schlesewsky, I., & Schlewsky, M. (2009). Parafoveal versus foveal N400s dissociate spreading activation from contextual fit. *Neuroreport*, *20*(18), 1613–1618. <https://doi.org/10.1097/WNR.0b013e328332c4f4>
- Kretzschmar, F., Schlewsky, M., & Staub, A. (2015). Dissociating word frequency and predictability effects in reading: Evidence from coregistration of eye movements and EEG. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(6), 1648–1662. <https://doi.org/10.1037/xlm0000128>
- Kristensen, E., Rivet, B., & Guérin-Dugué, A. (2017). Estimation of overlapped Eye Fixation Related Potentials: The General Linear Model, a more flexible framework than the ADJAR algorithm. *Journal of Eye Movement Research*, *10*(1), 1–27. <https://doi.org/10.16910/JEMR.10.1.7>
- Loberg, O., Hautala, J., Hämäläinen, J. A., & Leppänen, P. H. T. (2018). Semantic anomaly detection in school-aged children during natural sentence reading – A study of fixation-related brain potentials. *PLOS ONE*, *13*(12), e0209741. <https://doi.org/10.1371/journal.pone.0209741>
- Metzner, P., von der Malsburg, T., Vasishth, S., & Rösler, F. (2015). Brain Responses to World Knowledge Violations: A Comparison of Stimulus- and Fixation-triggered Event-related Potentials and Neural Oscillations. *Journal of Cognitive Neuroscience*, *27*(5), 1017–1028. https://doi.org/10.1162/jocn_a_00731
- Metzner, P., von der Malsburg, T., Vasishth, V., & Rösler, F. (2016). *The Importance of Reading*

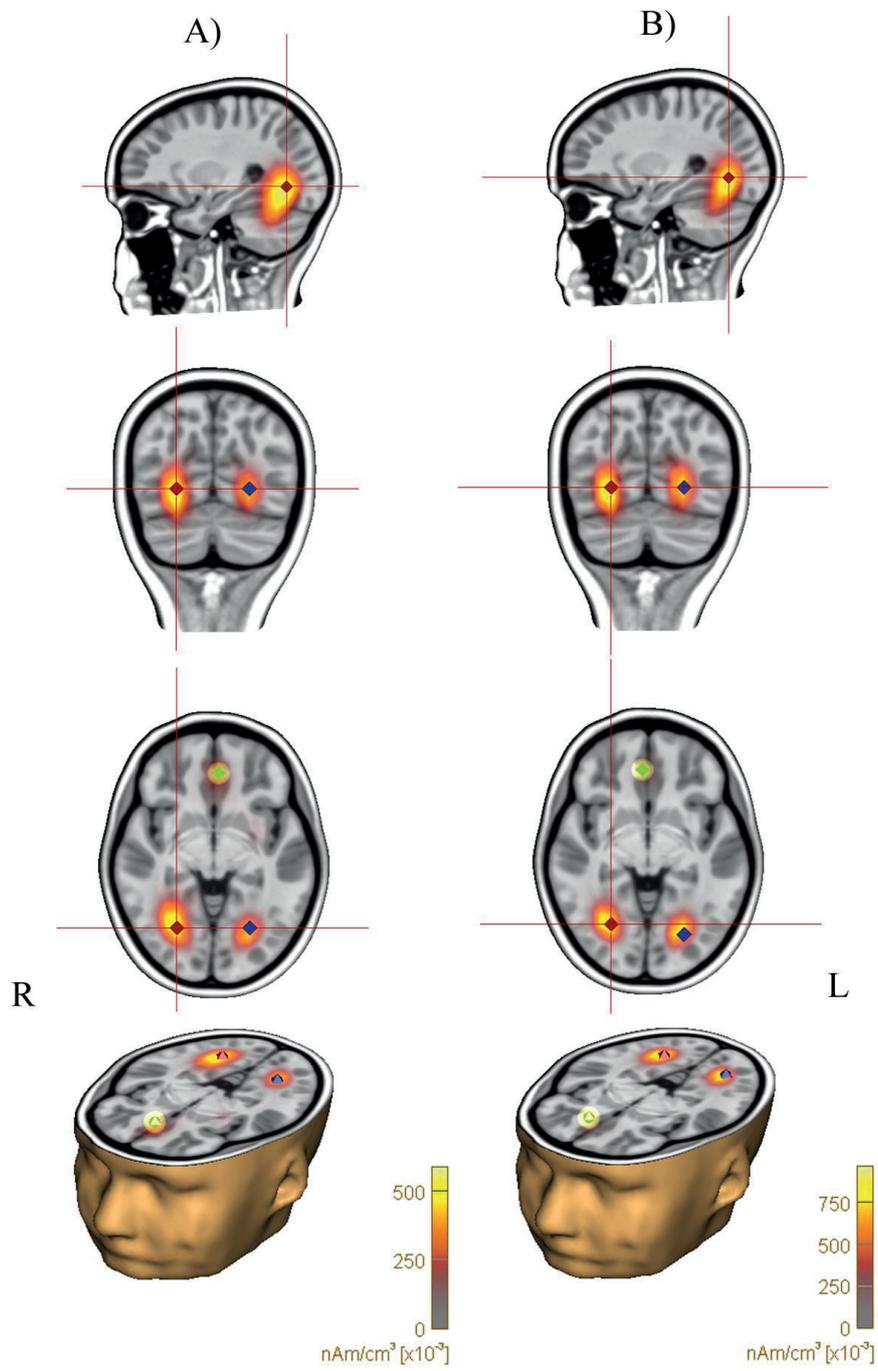
Naturally: Evidence from Combined Recordings of Eye Movements and Electric Brain Potentials. (April). <https://doi.org/10.1111/cogs.00000>

- Meyberg, S., Sommer, W., & Dimigen, O. (2017). How microsaccades relate to lateralized ERP components of spatial attention: A co-registration study. *Neuropsychologia*. <https://doi.org/10.1016/j.neuropsychologia.2017.02.023>
- Niefind, F., & Dimigen, O. (2016). Dissociating parafoveal preview benefit and parafovea-on-fovea effects during reading: A combined eye tracking and EEG study. *Psychophysiology*, *53*(12), 1784–1798. <https://doi.org/10.1111/psyp.12765>
- Nikolaev, A. R., Meghanathan, R. N., & van Leeuwen, C. (2016). Combining EEG and eye movement recording in free viewing: Pitfalls and possibilities. *Brain and Cognition*, *107*, 55–83. <https://doi.org/10.1016/J.BANDC.2016.06.004>
- Plöchl, M., Ossandón, J. P., & König, P. (2012). Combining EEG and eye tracking: identification, characterization, and correction of eye movement artifacts in electroencephalographic data. *Frontiers in Human Neuroscience*, *6*, 278. <https://doi.org/10.3389/fnhum.2012.00278>
- Rayner, K. (1997). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. <https://doi.org/10.1037/0033-2909.124.3.372>
- Ries, A. J., Slayback, D., & Touryan, J. (2018). The fixation-related lambda response: Effects of saccade magnitude, spatial frequency, and ocular artifact removal. *International Journal of Psychophysiology*, *134*, 1–8. <https://doi.org/10.1016/j.ijpsycho.2018.09.004>
- Smith, N. J., & Kutas, M. (2015a). Regression-based estimation of ERP waveforms: I. The rERP framework. *Psychophysiology*, *52*(2), 157–168. <https://doi.org/10.1111/psyp.12317>
- Smith, N. J., & Kutas, M. (2015b). Regression-based estimation of ERP waveforms: II. Nonlinear effects, overlap correction, and practical considerations. *Psychophysiology*, *52*(2), 169–181. <https://doi.org/10.1111/psyp.12320>
- Woldorff, M. G. (1993). Distortion of ERP averages due to overlap from temporally adjacent ERPs: Analysis and correction. *Psychophysiology*, *30*, 98–119. <https://doi.org/10.1111/j.1469-8986.1993.tb03209.x>
- Yagi, A. (1979). Saccade size and lambda complex in man. In *Physiological Psychology* (Vol. 7). Retrieved from <https://link.springer.com/content/pdf/10.3758%2FBF03326658.pdf>

Supplementary material 2

Post hoc source analysis. For descriptive purposes of the results, we ran source analysis on the group grand average intercept waveform. In addition, we generated a prediction of the linear word length effect on word length 13 and attempted to localize the source of the effect. We used Brain Electrical Source Analysis (BESA 6.1, Besa GmbH, Gräfelfing, Germany) software to explore the potential brain areas involved in the distributed source modelling method named Classical Loreta Recursively Applied (CLARA). In order to improve the accuracy of the source localization procedure, we applied a further high-pass filter of 1 HZ and used an age-appropriate template MRI for 12 year olds (Richards, Sanchez, Phillips-Meek, & Xie, 2016; Richards & Xie, 2015). Then, the model was estimated from a time window of 150–300 ms separately for both groups. We obtained the distributed model and begun inserting regional dipoles in order from the strongest to the weakest source maxima until the residual variance in the modelled time window was less than 5%. In both groups, this resulted in three almost identical source models. The results are presented in Supplementary Figure 1 and Supplementary Table 1.

We were unable to obtain a reasonable source solution for predicted word length effect on length 13. This was due the fact that as an effect separated from the intercept, the predicted waveform contained too much noise in relation to the size of the effect (~1 microvolt); thus, the distributed model showed widespread activity at the borders of the model volume (temporal poles and cerebellum), which is symptomatic of excessively high noise contamination for the source model to be accurate.



Supplementary figure 1: Results of the descriptive source analysis on the intercept grand averages. Left panel A: Typical readers. Right panel B: Slow readers. The red regional dipole is source 1. The blue regional dipole is source 2. The green regional dipole is source 3. Talairach coordinates of the sources are presented in Table 3. The source images are centred on source 1 to optimally display this location for both groups.

Supplementary Table 1: Locations of the sources derived from the descriptive source analysis.

	Typical readers				Slow readers			
	Talairach			C-RV%	Talairach			C-RV%
	x	y	z		x	y	z	
SC 1	24.5	-65.9	-4.3	20.365	24.5	-65.9	-4.3	21.28
SC 2	-24.5	-65.9	-4.3	10.686	-24.5	-72.9	-4.3	7.223
SC 3	-3.5	39.1	2.7	4.919	3.5	39.1	2.7	4.161

Note: Locations of the sources are provided in Talairach coordinates. C-RV% is the cumulative residual variance in order of setting the regional source ‘on’ in the model; thus, the C-RV% of SC 1 is when SC 1 is active alone, C-RV% of SC 2 is when SC 1 & 2 are active and so on.

Results of the descriptive source analysis. In both groups, source 1 was detected in the vicinity of the posterior portions of the right fusiform gyrus. This source had a contralateral correspondence in source 2, which located to the vicinity of posterior parts of the left fusiform gyrus. Moreover, source 3 was localized very similarly in the two groups in the vicinity of the anterior cingulate gyrus.

References

- Richards, J. E., Sanchez, C., Phillips-Meek, M., & Xie, W. (2016). A database of age-appropriate average MRI templates. *NeuroImage*, *124*, 1254–1259.
<https://doi.org/10.1016/j.neuroimage.2015.04.055>
- Richards, J. E., & Xie, W. (2015). Brains for All the Ages: Structural Neurodevelopment in Infants and Children from a Life-Span Perspective. *Advances in Child Development and Behavior*, *48*, 1–52. <https://doi.org/10.1016/bs.acdb.2014.11.001>